



Abschlussbericht  
der  
Operationellen Gruppe „OPTIKO“  
mit dem  
Innovationsprojekt „Optimierung des Kohlanbaus durch  
Messungen mit UAVs und Maschinellem Lernen“



Alle Fotos © OG OPTIKO

Lead Partner:

Universität Bonn

Kompetenzzentrum Gartenbau

Campus Klein-Altendorf 2

53359 Rheinbach

© 2023



# Inhalt

A Kurzdarstellung (in Alltagssprache).....	1
1 Ausgangssituation und Bedarf.....	1
2 Projektziel und konkrete Aufgabenstellung .....	1
3 Mitglieder der OG.....	1
4 Projektgebiet .....	2
5 Projektlaufzeit und – dauer .....	2
6 Budget .....	2
7 Ablauf des Vorhabens .....	2
8 Zusammenfassung der Ergebnisse .....	3
B Eingehende Darstellung .....	5
1 Verwendung der Zuwendung.....	5
2 Detaillierte Erläuterung der Situation zu Projektbeginn .....	5
2.1 Ausgangssituation .....	5
2.2 Projektaufgabenstellung .....	5
3 Ergebnisse der OG in Bezug auf.....	6
3.1 Wie wurde die Zusammenarbeit im Einzelnen gestaltet (ggf. mit Beispielen, wie die Zusammenarbeit sowohl organisatorisch als auch praktisch erfolgt ist)?.....	6
3.2 Was war der besondere Mehrwert des Formates einer OG für die Durchführung des Projekts?.....	7
3.3 Ist eine weitere Zusammenarbeit der Mitglieder der OG nach Abschluss des geförderten Projekts vorgesehen? .....	8
4 Ergebnisse des Innovationsprojektes.....	8
4.1 Zielerreichung (wurde eine Innovation im Projekt generiert?).....	8
4.2 Abweichungen zwischen Projektplan und Ergebnissen .....	8
4.3 Projektverlauf (ggf. mit Fotodokumentation).....	10
4.4 Beitrag des Ergebnisses zu förderpolitischen EIP Zielen .....	20
4.5 Nebenergebnisse.....	20
4.6 Arbeiten, die zu keiner Lösung geführt haben .....	20
5 Nutzen der Ergebnisse für die Praxis.....	20
6 (Geplante) Verwertung und Nutzung der Ergebnisse .....	21
7 Wirtschaftliche und wissenschaftliche Anschlussfähigkeit .....	21
8 Wo relevant: Nutzung Innovationsdienstleister (IDL) .....	21
9 Kommunikations- und Disseminationskonzept.....	21
10 Fazit .....	24
C ANHANG .....	27



## A Kurzdarstellung (in Alltagssprache)

### 1 Ausgangssituation und Bedarf

Blumenkohl gehört zu den im Anbau anspruchsvollsten Gemüsearten. Während der Kulturdauer von rund 8-12 Wochen werden Felder regelmäßig durch Landwirte und landwirtschaftliche Berater kontrolliert, um Ertragseinbußen durch abiotische und biotische Stresse zu minimieren. Hierzu gehören insbesondere Nährstoffmangel sowie der Befall durch Pilzkrankheiten oder Schadinsekten, die sich negativ auf Wachstum und Entwicklung der Pflanzen auswirken und zu Qualitäts- und Ertragseinbußen führen können. Derartige Feldkontrollen können jedoch nur stichprobenhaft innerhalb eines Feldes durchgeführt werden, und eine flächendeckende Einschätzung ganzer Felder ist nicht möglich. Dadurch können erste, punktuell auftretende Stressfaktoren übersehen werden, sodass notwendige Managementmaßnahmen wie z.B. der Einsatz von Pflanzenschutzmitteln erst bei flächendeckend auftretenden Stressen durchgeführt werden.

Ähnliche Schwierigkeiten ergeben sich bei der Blumenkohlernte. Blumenkohl hat nur ein kurzes Erntezeitfenster von wenigen Tagen, innerhalb derer der Kopf die notwendige Größe aufweist, jedoch noch kompakt im Wuchs ist. Da es zurzeit nur unzureichende Verfahren zur Vorhersage des Erntezeitpunktes gibt, müssen täglich stichprobenhafte Feldkontrollen durchgeführt werden. Die Ernte der jeweils erntereifen Pflanzen pro Feld erfolgt von Hand, wobei meist 3-4 Durchgänge pro Feld über einen Zeitraum von 1-2 Wochen notwendig sind.

Der Einsatz automatischer Verfahren zur frühzeitigen Erkennung von Stressfaktoren und zur Abschätzung der Erntereife bietet ein hohes Potential zur Optimierung des Blumenkohlanbaus.

### 2 Projektziel und konkrete Aufgabenstellung

Unbemannte Luftfahrzeuge (UAV) bzw. Drohnen ermöglichen den automatisierten Einsatz von hochauflösenden Multispektralkameras zur räumlichen Erfassung von landwirtschaftlichen Kulturen für die Präzisionslandwirtschaft. Jedoch kam es bisher meist nur zur wissenschaftlichen Anwendung und kaum zur Integration in die landwirtschaftliche Praxis.

Ziel des Projekts war die Entwicklung eines Verfahrens zur Auswertung von UAV-basierten Multispektralmessungen über landwirtschaftlichen Kulturen. Hierbei sollten maschinelle Lernverfahren entwickelt werden, die die erhobenen spektralen und räumlichen Daten automatisiert auswerten und Heterogenitäten in Wuchs und Pflanzengesundheit aufgrund von abiotischen und biotischen Stressfaktoren innerhalb eines Feldes erkennen sowie den Erntezeitpunkt der Pflanzen vorhersagen.

Das Verfahren sollte zunächst für Blumenkohl entwickelt werden, soll jedoch in Zukunft auf weitere Kulturen adaptiert werden. Damit das entwickelte Verfahren reibungslos in den Betriebsablauf integriert werden kann, wurde ein besonderer Fokus auf eine zügige Datenauswertung gelegt. Die frühzeitige Erkennung von Stressfaktoren trägt zu einer Optimierung notwendiger Managementmaßnahmen bei. Die Vorhersage des Erntezeitpunktes für alle Felder eines Betriebs ermöglicht eine verbesserte Steuerung des Verkaufs und somit eine Stabilisierung des Ertrags.

### 3 Mitglieder der OG

- Universität Bonn: Kompetenzzentrum Gartenbau (*KoGa*, Leadpartner) und Institut für Geodäsie und Geoinformatik (*IGG*)
- Betrieb Konrad und Markus Schwarz, Bornheim (*Betrieb Schwarz*)
- Forschungszentrum Jülich, Institut für Bio- und Geowissenschaften-II (*FZI*)

- JB Hyperspectral Devices GmbH (JB)

#### 4 Projektgebiet

Bornheim, Rheinland, NRW

#### 5 Projektlaufzeit und –dauer

01.01.2020 – 31.12.2022 (36 Monate)

#### 6 Budget

Die Zuwendung belief sich auf € 499.346,10. Davon waren € 79.239,60 für die OG und € 420.106,50 für die Arbeiten des IP.

#### 7 Ablauf des Vorhabens

Die geplanten Aktivitäten und Meilensteine sind in Abbildung A1 dargestellt. Der Zeitplan wurde weitgehend eingehalten, im Projektverlauf ergaben sich kleinere zeitliche Verzögerungen und Änderungen in den Meilensteinen von AP 4.

Ifd. Nr.	Bezeichnung des AP	Dauer (Monate)	2020				2021				2022									
			Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4						
1	Datenerhebung anhand von Überflügen mit UAVs	24				M1.1														
2	Erfassung des Expertenwissens	24				M2.1														
3	Analyse der erhobenen Daten mittels maschineller Lernverfahren	36	M3.1	M3.2					M3.3		M3.4									
4	Entwicklung eines theorie-getriebenen maschinellen Lernverfahrens zur Integration von Expertenwissen	36	M4.1	M4.2					M4.3		M4.4									
5	Koordination und Demonstrationsaktivitäten	36	M5.1								M5.2									
<b>Phasen des Projektfortschritts:</b>																				
Phase 1: Einarbeitung und Abstimmung der Projektpartner, Planung und Vorbereitung der Datenerhebungen (Monat 1-3)																				
Phase 2: Erhebung von Multispektraldaten sowie Expertenwissen während der ersten Vegetationsperiode, Entwicklung erster maschineller Lernverfahren (Monat 4-12)																				
Phase 3: Erhebung von Multispektraldaten sowie Expertenwissen während der zweiten Vegetationsperiode, Weiterentwicklung der maschinellen Lernverfahren (Monat 13-24)																				
Phase 4: Validierung der entwickelten maschinellen Lernverfahren unter Praxisbedingungen im Feld (Monat 25-33)																				
Phase 5: Projektabschlussphase: Zusammenfassen der Ergebnisse, Publikationen und Öffentlichkeitsarbeit (Monat 34-36)																				
<b>Liste der Meilensteine:</b>																				
M1.1: Trainingsdaten für die Heterogenität im Feld mithilfe von regelmäßigen Drohnenüberflügen erhoben und in Datenbank zusammengestellt (Monat 10)																				
M1.2: Daten der zweiten Vegetationsperiode anhand regelmäßiger Drohnenüberflüge erhoben und in Datenbank zusammengestellt (Monat 22)																				
M1.3: Daten ausgewählter Zeitpunkte der dritten Vegetationsperiode zur Validierung der entwickelten Modelle und Ableitung von Erntevorhersagen erhoben und in Datenbank																				
M2.1: Feldbonituren zur Komplementierung der Trainingsdaten für die Heterogenität im Feld durchgeführt und in Datenbank zusammengestellt (Monat 10)																				
M2.2: Feldbonituren zur Validierung der von maschinellen Lernverfahren (AP3) detektierten Heterogenität und Erfassung des Expertenwissens zu Erntezeitpunkten hunderter Pflanzen																				
M2.3: Regelmäßige Durchführung von Feldkontrollen zur Validierung der von maschinellen Lernverfahren (AP3) vorhergesagten Erntezeitpunkte und in Datenbank zusammengestellt																				
M3.1: Einarbeitung in Methoden maschineller Lernverfahren und Ableitung der Eignung für die angedachten Anwendungen (Monat 3)																				
M3.2: Detektion von Kohlpflanzen im Feld anhand maschineller Lernverfahren möglich (Monat 6)																				
M3.3: Schätzung von Wuchs und Ausreife zur Erkennung von Heterogenität anhand maschineller Lernverfahren möglich (Monat 18)																				
M3.4: Vorhersage des Erntezeitfensters anhand maschineller Lernverfahren möglich (Monat 24)																				
M3.5: Quantifizierung der zeitlichen und räumlichen Heterogenität im Auswuchs anhand maschineller Lernverfahren möglich (Monat 36)																				
M4.1: Planung der experimentellen Durchführung, insbesondere der Methoden für die Datenerhebung im Feld, abgeschlossen (Monat 3)																				
M4.2: Einarbeitung in die Methodik maschineller Lernverfahren und deren Eignung für die angedachten Anwendungen (Monat 6)																				
M4.3: Schätzung von Wuchs und Pflanzengesundheit der Blumenkohlpflanzen mit wissensbasierten neuronalen Netzen für einen Zeitpunkt möglich (Monat 18)																				
M4.4: Vorhersage der Erntereife der Blumenkohlpflanzen basierend auf Zeitreihenbeobachtungen mit wissensbasierten neuronalen Netzen möglich (Monat 24)																				
M4.5: Bestätigung der Transferleistung des optimierten Modells anhand von Messungen in Chinakohl (Monat 36)																				
M5.1: KICKOFF-Workshop mit allen Projektpartnern und weiteren Stakeholdern (Monat 3)																				
M5.2: Vorstellung des Projekts bei landwirtschaftlichen Veranstaltungen (z.B. Feldtag) zur Sicherstellung der Praxisnähe des Projekts und Evaluation weiterer																				
M5.3: Abschlussveranstaltung (Monat 36)																				

**Abbildung A1:** Gantt-Chart und Meilensteine des Projekts bei Projektende.

Das Projekt begann am 1.1.2020. Bis Ende des ersten Quartals 2020 waren alle Personalstellen bei den Partnern besetzt und es erfolgte eine Einarbeitung in die Thematik. Das Kick-Off-Treffen des Projekts fand am 26.2.2020 in Bonn statt. Bei diesem ersten Treffen wurden die Anforderungen an

das UAV-System diskutiert und entschieden eine Drohne mit RGB (=Farbfoto) und Multispektralkamera zu nutzen. Das System wurde von FZJ aus Eigenmitteln bereitgestellt.

Früh im ersten Versuchsjahr wurde das erste Blumenkohlfeld befliegen. Zeitgleich wurden die ersten phänologischen Daten erhoben um einen kompletten Datensatz mit Referenzdaten zu erhalten. Später im gleichen Jahr wurde ein zweites Feld befliegen und einige der Schwierigkeiten mit Feld 1 durch Anpassung des Flugplans behoben. Erste Daten konnten so generiert werden, die in einer Datenbank abgelegt wurden und es erlaubten, die Entwicklung des maschinellen Lernverfahrens zu beginnen. Die Feldarbeiten fanden trotz Corona im Zeitplan statt; einige Veranstaltungen konnten wegen des Lockdowns nicht besucht bzw. organisiert werden.

Im zweiten Versuchsjahr wurde ein drittes Blumenkohlfeld mit einer weiter angepassten Methodik befliegen und nun ein kompletter Datensatz generiert, der in der Datenbank abgelegt wurde. Dadurch konnten Zeitreihen für die Bestimmung von Pflanzenwachstum als Vorläufer der Vorhersage des Erntezeitfensters erstellt werden. Hiermit konnte das maschinelle Verfahren weiter verfeinert werden. Im Jahr 2021 wurde das Projekt auf verschiedenen wissenschaftlichen Foren vorgestellt und eine Publikation vorbereitet.

Im dritten Versuchsjahr wurden auf einem vierten und letzten Blumenkohlfeld schlussendlich Daten von sehr hoher Qualität erhoben, die dazu dienten, das maschinelle Verfahren zu testen. Dabei konnten Cluster von erntereifen Pflanzen identifiziert werden, die eine Erntevorhersage mit einem Zeitrahmen von drei Tagen im Voraus möglich machen können. In diesem Jahr fanden auch einige Veranstaltungen statt, auf denen das Projekt vorgestellt werden konnte. Auch wurde eine Publikation mit den erhobenen Daten veröffentlicht (Kierdorf et al., 2022 b), die auf sehr große positive Resonanz stieß.

## 8 Zusammenfassung der Ergebnisse

- Das Projekt zeigt das Potential des erarbeiteten Verfahrens, das zukünftig auch auf andere Kulturen angepasst werden kann und so zur Digitalisierung in der Landwirtschaft beiträgt.
- Eine Methodik zur großflächigen bildgebenden Datenerfassung von Feldern mit sehr gleichmäßigem Kronendach wurde etabliert, die sowohl für RGB- als auch Multispektralaufnahmen geeignet ist.
- Maschinelle Lernverfahren können zur Detektion von Einzelpflanzen und Abschätzung ihrer Größe genutzt werden.
- Es ist möglich, mit maschinellen Lernverfahren einzelne Blattinstanzen zu segmentieren, was eine Abschätzung des phänologischen Entwicklungsstadiums ermöglicht.
- Es ist möglich, mit einer Genauigkeit von 75 % den Anteil erntereifer Blumenkohlpflanzen 3 Tage im Voraus abzuschätzen.
- Die Datensätze sind publiziert (Kierdorf et al. 2022b), stehen dadurch für weiterführende Arbeiten zur Verfügung und stoßen bereits auf großes Interesse.



## B Eingehende Darstellung

### 1 Verwendung der Zuwendung

Der Verwendungsnachweis wird separat gestellt.

### 2 Detaillierte Erläuterung der Situation zu Projektbeginn

#### 2.1 Ausgangssituation

In NRW gibt es rund 1.500 Gemüsebaubetriebe, die auf ca. 20.000 ha frisches Gemüse produzieren. Etwa ein Sechstel der Anbaufläche entfällt dabei auf Kohlgemüse wie Blumenkohl, Weißkohl, Kohlrabi, Grünkohl, etc. Kohlanbauende Betriebe sind häufig auf 1-4 Kohlarten spezialisiert, um die Wettbewerbsfähigkeit sichern.

Blumenkohl gehört zu den im Anbau anspruchsvollsten Gemüsearten. Während der Kulturdauer von rund 8-12 Wochen werden Felder regelmäßig durch Landwirte und landwirtschaftliche Berater kontrolliert, um Ertragseinbußen durch abiotische und biotische Stresse zu minimieren. Hierzu gehören insbesondere Nährstoffmangel sowie der Befall durch Pilzkrankheiten oder Schadinsekten, die sich negativ auf Wachstum und Entwicklung der Pflanzen auswirken und zu Qualitäts- und Ertragseinbußen führen können.

Derartige Feldkontrollen können jedoch nur stichprobenhaft innerhalb eines Feldes durchgeführt werden, und eine flächendeckende Einschätzung ganzer Felder ist nicht möglich. Dadurch werden häufig erste, punktuell auftretende Stressfaktoren übersehen und notwendige Managementmaßnahmen wie z.B. der Einsatz von Pflanzenschutzmitteln werden erst bei flächendeckend auftretenden Stressen durchgeführt.

Ähnliche Schwierigkeiten ergeben sich bei der Blumenkohlernte. Blumenkohl hat nur ein kurzes Erntezeitfenster von wenigen Tagen, innerhalb derer der Kopf die notwendige Größe aufweist, jedoch noch kompakt im Wuchs ist. Da es zurzeit nur unzureichende Verfahren zur Vorhersage des Erntezeitpunktes gibt, müssen täglich stichprobenhafte Feldkontrollen durchgeführt werden. Die Ernte der jeweils erntereifen Pflanzen pro Feld erfolgt von Hand, wobei meist 3-4 Durchgänge pro Feld über einen Zeitraum von 1-2 Wochen notwendig sind.

Der Einsatz automatischer Verfahren zur frühzeitigen Erkennung von Stressfaktoren und zur Abschätzung der Erntereife bietet ein hohes Potential zur Optimierung des Blumenkohlanbaus.

#### 2.2 Projektaufgabenstellung

##### **Beitrag des Vorhabens zur Unterstützung einer umweltgerechten Entwicklung sowie Nachhaltigkeit im Gartenbau**

Das geplante Vorhaben leistet einen Beitrag zur Entwicklung der Präzisionslandwirtschaft. Durch die automatisierte Erfassung von Feldern ermöglicht es eine höhere Effizienz der Landwirte und landwirtschaftlichen Berater, da diese bereits vor Begehung der Felder auf auffällige Pflanzen bzw. Bereiche innerhalb eines Feldes hingewiesen werden können. Zudem bietet es eine Entscheidungshilfe für Landwirte, und ermöglicht den präzisen Einsatz von Managementmaßnahmen. So ist z.B. auf lange Sicht ein räumlich differenzierter Düngemittleinsatz möglich, und die frühzeitige Erkennung von biotischen Stressfaktoren ermöglicht eine spezifische Behandlung mit Pflanzenschutzmitteln und trägt so langfristig zu einer Reduzierung des Pflanzenschutzmitteleintrags in die Umwelt bei. Die Vorhersage des optimalen Erntezeitpunktes

verringert zudem Ernteauffälle. Insgesamt kann das geplante Vorhaben so die Ertragssicherheit erhöhen, wodurch sich auch die einzuplanende Überschussproduktion und der damit verbundene Ressourceneinsatz verringern werden.

### **Beitrag des Vorhabens zu einer Stärkung der Wertschöpfungsketten sowie der Wettbewerbsfähigkeit des Gartenbaus in NRW**

In NRW wird rund ein Drittel des in Deutschlands produzierten Blumenkohls angebaut. Da der Anbau von Blumenkohl mit einem hohen Zeit- und Personaleinsatz verbunden ist, bietet sich hier im besonderen Maße die Möglichkeit, durch den Einsatz digitaler Verfahren notwendige Managementmaßnahmen wie z.B. Düngung und Pflanzenschutzmitteleinsatz zu optimieren, den Arbeitsaufwand zu reduzieren und so die Ertragssicherheit zu steigern. Das zu entwickelnde Verfahren soll mit Hinblick auf die mögliche Anpassung an weitere Kulturen entwickelt werden, um langfristig auf verschiedene Kulturen und Fragestellungen übertragen werden zu können. Dabei ist perspektivisch davon auszugehen, dass die Kosten für UAVs und Multispektralkameras weiter sinken, sodass der Einsatz dieses innovativen digitalen Verfahrens ökonomisch vorteilhaft ist. Zudem bietet das entwickelte Verfahren das Potential zur Etablierung einer neuen Dienstleistung für den Acker- und Gartenbau: der Einsatz derartiger „digitaler Berater“ ermöglicht eine genauere Erfassung von Ackerschlägen ohne erhöhten Personaleinsatz bzw. reduziert langfristig den notwendigen Personaleinsatz. Die Optimierung des Anbaus trägt dazu bei, den in NRW ansässigen, regionalen Erwerbsgartenbau auch unter sich verändernden Umweltbedingungen und strenger werdenden Umweltschutzauflagen zu sichern.

### **Beitrag zum Transfer von Wissen und Know-How zum Nutzen der Praxis**

Das Projekt unterstützt die Entwicklung der wissenschaftlichen Grundlage für ein innovatives Verfahren, das für die gartenbauliche Praxis von großem Interesse ist. Um die neue Technik und Ihre Möglichkeiten vorzustellen und weitere Anwendungsgebiete zu identifizieren, wird KoGa die innerhalb des Verbundes durch die Landwirtschaftskammer NRW jährlich stattfindenden Veranstaltungen, Seminare und Feldtage nutzen. Der Wissenstransfer in die Praxis wird durch den Betrieb Schwarz gefördert. Durch den Kontakt mit vielen weiteren regionalen Landwirtschaftsbetrieben sind Demonstrationen der Technik auf dem Betrieb sowie in weiteren praxisnahen Foren möglich. Die wissenschaftlichen Erkenntnisse des Projekts werden auf praxisnahen Konferenzen präsentiert und zeitnah in Open Access-Journalen publiziert, um über die Grenzen NRWs bzw. Deutschlands hinaus der wissenschaftlichen Community zugänglich gemacht zu werden. Um die Akzeptanz der breiten Öffentlichkeit für diese innovative Methoden in der Landwirtschaft zu erhöhen, wird das Projekt zudem bei Veranstaltungen der KoGa-Verbundpartner vorgestellt, z.B. dem „Tag der offenen Tür“ der Uni Bonn mit über 1.000 Besuchern und dem „Tag der Neugier“ am FZJ mit über 20.000 Besuchern. Über eine Website wird das Projekt überregional bekannt gemacht.

## 3 Ergebnisse der OG

### 3.1 Wie wurde die Zusammenarbeit im Einzelnen gestaltet (ggf. mit Beispielen, wie die Zusammenarbeit sowohl organisatorisch als auch praktisch erfolgt ist)?

Das Kompetenzzentrum Gartenbau (KoGa) ist ein Verbund der Universität Bonn, des Dienstleistungszentrums Ländlicher Raum, Rheinland-Pfalz, des Forschungszentrums Jülich und der Landwirtschaftskammer NRW. Fokus der Zusammenarbeit ist der schnelle Austausch zwischen Forschung und Praxis unter den Themenbereichen „Adaptationsstrategien an sich ändernde



Klimabedingungen“, „Mechanisierung von Prozessen“ sowie „Optimierung der Ressourceneffizienz und umweltschonender Verfahren“. Das Thema für das vorliegende Projekt wurde aus der Praxis in der vormaligen Operationellen Gruppe „Präzisionsgartenbau“ entworfen. Weitere relevante Partner fanden sich dann aus dem weiteren Kreis der KoGa Partner zusammen.

Während der Projektlaufzeit fand eine enge Abstimmung der Partner in Bezug auf geplante Aktivitäten im Rahmen von Projekttreffen statt. Wegen der Kontaktbeschränkungen während der Corona-Pandemie wurde meistens online kommuniziert.

Insgesamt war der Austausch zwischen den Partnern konstruktiv. Allerdings war die Kommunikation mit dem Praxisbetrieb – aufgrund von dringenden anderen Prioritäten – zuweilen schwierig.

Im November 2022 fand eine Projektabschlussevaluierung unter den neun Projektmitarbeiterinnen und –mitarbeitern statt. Die Fragen lauteten:

- Warum haben Sie sich an diesem Projekt beteiligt?
- Welche Erwartungen hatten Sie an das Projekt? Haben sich diese erfüllt? Inwieweit haben sich Ihre Erwartungen im Laufe des Projektes geändert?
- Gab es irgendwelche Probleme/Schwierigkeiten? Wenn ja – Welche?
- Was würden Sie beim nächsten Mal anders machen?
- Was war besonders zufriedenstellend?
- Inwieweit sind die Ergebnisse Ihrer Meinung nach auf die Praxis anwendbar?
- Zum Schluss: Was nehmen Sie aus diesem Projekt mit? Was ist Ihr persönliches Highlight aus diesem Projekt?

Sechs Rückmeldungen liegen vor, aus denen deutlich wird, dass die Motivation an diesem Projekt mitzuarbeiten, maßgeblich an der Verbindung von Wissenschaft und Praxis lag sowie dem Erlernen neuer Erfassungsmethoden. Auch wenn Schwierigkeiten, insbesondere in der Datenerhebung und durch die Corona-Pandemie und die damit verbundenen Beschränkungen, auftraten und den Projektfortschritt zunächst behinderten, wurde die Lösung dieser Probleme im Team als sehr wertvoll erachtet. In Hinblick auf die Anwendbarkeit der Ergebnisse in die Praxis sind die Befragten einig, dass die Komplexität des Ansatzes eine Praxiseinführung zum Zeitpunkt des Projektendes unmöglich macht, dass aber ein sehr wertvoller erster Schritt mit qualitativ guten Basisdaten für eine Weiterentwicklung der Erntevorhersage mittels künstlicher Intelligenz gemacht wurde. Alle Rückmeldungen beschreiben das persönliche Lernen als größten Gewinn dieses Projekts.

### 3.2 Was war der besondere Mehrwert des Formates einer OG für die Durchführung des Projekts?

In großen Verbundprojekten ist ein Gremium mit Vertretern der Partnereinrichtungen eine Notwendigkeit. Auch bei unserem Projekt war das Gremium „OG“ in Projektplanungsmeetings aktiv. Da an unserem Projekt eine überschaubare Anzahl von Personen beteiligt war, war die OG mit dem IP nahezu deckungsgleich. Das vorliegende Format hat durch die administrative Trennung bei uns eher für eine verwaltungstechnische Mehrbelastung und für Verwirrung gesorgt (unterschiedliches Abrechnen, Zeitanrechnung OG/IP). Insgesamt hat die enge Zusammenarbeit in OG/IP aber für ein gutes Vertrauensverhältnis der Partner zueinander geführt, so dass weitere Kooperationsmöglichkeiten für die Zukunft positiv gesehen werden können.

### 3.3 Ist eine weitere Zusammenarbeit der Mitglieder der OG nach Abschluss des geförderten Projekts vorgesehen?

Eine Zusammenarbeit einiger Partner bestand bereits vor Abschluss des vorliegenden Projekts innerhalb des KoGa-Verbunds aber auch im Exzellenzcluster „PhenoRob“ der Universität Bonn. Im Verlauf des Projekts haben sich Möglichkeiten der weiteren Zusammenarbeit insbesondere der Partner FZJ und Uni Bonn IGG mit JB ergeben. Auch der Landwirt ist bereit, die Zusammenarbeit mit den Wissenschaftlern von Uni Bonn und Forschungszentrum Jülich weiteren Kollegen zu empfehlen.

## 4 Ergebnisse des Innovationsprojektes

### 4.1 Zielerreichung (wurde eine Innovation im Projekt generiert?)

Das Projekt konnte das Potential des erarbeiteten Verfahrens demonstrieren. Dieses kann zukünftig auch auf andere Kulturen angepasst werden und trägt so zur Digitalisierung in der Landwirtschaft bei. Insbesondere wurde eine Methodik zur großflächigen bildgebenden Datenerfassung von Feldern mit sehr gleichmäßigem Kronendach etabliert, die sowohl für RGB- als auch Multispektralaufnahmen geeignet ist. Es zeigte sich, dass maschinelle Lernverfahren zur Detektion von Einzelpflanzen und Abschätzung ihrer Größe genutzt werden können. Es ist möglich, mit maschinellen Lernverfahren einzelne Blattinstanzen zu segmentieren, was eine Abschätzung des phänologischen Entwicklungsstadiums ermöglicht und so eine frühzeitige Erkennung von Stress ermöglichen kann.

Die Datensätze, insbesondere der Multispektraldatensatz, stehen der Wissenschaft für weiterführende Arbeiten zur Verfügung und stoßen bereits auf großes Interesse.

### 4.2 Abweichungen zwischen Projektplan und Ergebnissen

Anhand der Entwicklung der Meilensteine (**Tabelle B1**) ist sichtbar, dass die Planung in der ersten Phase des Projekts gut eingehalten werden konnte. In AP 4 kam es zu einigen Verzögerungen und Abweichungen von den ursprünglich geplanten Meilensteinen, die jedoch die Zielerreichung insgesamt nicht beeinträchtigt haben. Verschiedene Ansätze wurden analysiert und dokumentiert, dabei wurden vielversprechende Ansätze identifiziert, die auch nach Projektende weiter bearbeitet werden sollen.

**Tabelle B1:** Gegenüberstellung der Meilensteine: Planung vis-à-vis Durchführung

Meilenstein	Geplant	Erreicht
M1.1: Trainingsdaten für die Heterogenität im Feld mithilfe von regelmäßigen Drohnenüberflügen erhoben und in Datenbank zusammengestellt	Monat 10	Erreicht
M1.2: Daten der zweiten Vegetationsperiode anhand regelmäßiger Drohnenüberflüge erhoben und in Datenbank zusammengestellt	Monat 22	Erreicht
M1.3: Daten ausgewählter Zeitpunkte der dritten Vegetationsperiode zur Validierung der entwickelten Modelle und Ableitung von Erntevorhersagen erhoben und in Datenbank zusammengestellt	Monat 34	Erreicht M35
M2.1: Feldbonituren zur Komplementierung der Trainingsdaten für die Heterogenität im Feld durchgeführt und in Datenbank zusammengestellt	Monat 10	Erreicht

M2.2: Felddonituren zur Validierung der von maschinellen Lernverfahren (AP3) detektierten Heterogenität und Erfassung des Expertenwissens zu Erntezeitpunkten hunderter Pflanzen durchgeführt und in Datenbank zusammengestellt	Monat 22	Erreicht
M2.3: Regelmäßige Durchführung von Feldkontrollen zur Validierung der von maschinellen Lernverfahren (AP3) vorhergesagten Erntezeitpunkte und in Datenbank zusammengestellt	Monat 34	Erreicht
M3.1: Einarbeitung in Methoden maschineller Lernverfahren und Ableitung der Eignung für die angedachten Anwendungen	Monat 3	Erreicht
M3.2: Detektion von Kohlpflanzen im Feld anhand maschineller Lernverfahren möglich	Monat 6	Erreicht
M3.3: Schätzung von Wuchs und Ausreifung zur Erkennung von Heterogenität anhand maschineller Lernverfahren möglich	Monat 18	Erreicht
M3.4: Vorhersage des Erntezeitfensters anhand maschineller Lernverfahren möglich	Monat 24	Erreicht M36
M3.5: Quantifizierung der zeitlichen und räumlichen Heterogenität im Aufwuchs anhand maschineller Lernverfahren möglich	Monat 36	Erreicht
M4.1: Planung der experimentellen Durchführung, insbesondere der Methoden für die Datenerhebung im Feld, abgeschlossen	Monat 3	Erreicht
M4.2: Einarbeitung in die Methodik maschineller Lernverfahren und deren Eignung für die angedachten Anwendungen	Monat 6	Erreicht M18
M4.3: Schätzung von Wuchs und Pflanzengesundheit der Blumenkohlpflanzen mit wissensbasierten neutralen Netzen für einen Zeitpunkt möglich <b>Geändert:</b> „Schätzung von Wuchs und Pflanzengesundheit der Blumenkohlpflanzen <b>basierend auf Multispektraldaten</b> “ (Monat 32)	Monat 18	Vielversprechende Ansätze identifiziert. Dies sollte weiter vertieft werden.
M4.4: Vorhersage der Erntereife der Blumenkohlpflanzen basierend auf Zeitreihenbeobachtungen mit wissensbasierten neuronalen Netzen möglich <b>Geändert:</b> „Vorhersage der Erntereife der Blumenkohlpflanzen basierend auf Zeitreihenbeobachtungen <b>der Blattentwicklung</b> “ (Monat 24)	Monat 24	Vielversprechende Ansätze identifiziert. Dies sollte weiter vertieft werden.
M4.5: Bestätigung der Transferleistung des optimierten Modells anhand von Messungen in Chinakohl <b>Geändert</b> auf Brokkoli	Monat 36	Aufgrund von fehlenden Alternativkulturen nicht erreicht.
M5.1: KICKOFF-Workshop mit allen Projektpartnern und weiteren Stakeholdern	Monat 3	Erreicht
M5.2: Vorstellung des Projekts bei landwirtschaftlichen Veranstaltungen (z.B. Feldtag) zur Sicherstellung der Praxisnähe	Monat 24	Erreicht M36, aufgrund der

des Projekts und Evaluation weiterer Anwendungsmöglichkeiten		Corona-bedingten Lockdowns aber leider nicht im angestrebten Umfang.
M5.3: Abschlussveranstaltung	Monat 36	Erreicht

#### 4.3 Projektverlauf (ggf. mit Fotodokumentation)

##### 4.3.1: Einrichten der OG, Einarbeitung und Abstimmung der Projektpartner, Planung und Vorbereitung der Datenerhebungen (Januar – März 2020)

Das Projekt begann am 1.1.2020. Am 26.2.2020 fand das Kick-off Treffen im Institut für Geodäsie und Geoinformation der Uni Bonn statt.

In den ersten Monaten wurden die Ideen des Antrags konkretisiert, die Feldversuche geplant und vorbereitet, technische Tests durchgeführt und die Einarbeitung in die Methodik des maschinellen Lernens begonnen. Die nötigen Vorbereitungen für die Befliegung der Blumenkohlfelder im Betrieb Schwarz wurden getätigt und ein passendes Kopter-System mit den entsprechenden Sensoren konfiguriert. Flugpläne wurden erarbeitet und an die gegebenen Standorte angepasst. Ground-Control-Points sowie Farbreferenzkarten für den Feldeinsatz wurden hergestellt. Weiter mussten die rechtlichen Rahmenbedingungen hergestellt werden, um dem Personal die regelmäßigen Befliegungen zu erlauben. Der für die Flüge zuständige Mitarbeiter machte einen sogenannten „Drohnenführerschein“ inklusive Prüfung; verschiedene Einverständniserklärungen von Ämtern wurden eingeholt.



Abbildung B1: Die Drohne erfasst RGB- und Hyperspektralbilder von einem Blumenkohlfeld. © OPTIKO

##### 4.3.2: Erhebung von Multispektraldaten sowie Expertenwissen während der ersten Vegetationsperiode, Entwicklung erster maschineller Lernverfahren (April – Dezember 2020)

#### **Arbeitspaket 1: Datenerhebung anhand von Überflügen mit UAVs**

Mit Beginn der Bepflanzung wurde mit den Befliegungen von Feld 1 begonnen. Es wurden Überblickflüge aus großer Höhe sowie Detailflüge aus niedriger Höhe durchgeführt. Je nach

Wetterlage wurden Flüge wöchentlich oder öfter durchgeführt. Die Bilder wurden gespeichert. Im Zeitraum bis zum 30.06.2020 wurde das Feld an 14 Befliegungsterminen erfasst, bei denen insgesamt ca. 350.000 Bilder erstellt wurden. Es zeigte sich, dass die Datenmenge und –qualität nicht zielführend waren. Nach Auswertung dieser Überfliegungen wurde daher ein verbessertes Drohnensystem entwickelt, das dann für Feld 2 an 19 Flugtagen in Einsatz kam. Nach Abschluss der Befliegungen im Oktober wurde das Befliegungssystem weiter verbessert.

Parallel zu den Arbeiten im Feld wurde der für das Projekt angeschaffte Prozessierungsrechner intensiv genutzt um aus den Einzelbildern jeder Befliegung Orthomosaik zu generieren. Von Feld 2 wurden 15 dieser mehrere Gigabyte großer Bilddateien generiert, die über einen Server der Uni Bonn (AP 3) bereitgestellt wurden. Um diese Arbeiten effizient und automatisiert zu gestalten wurde eine Stapelverarbeitung angelegt, mit deren Hilfe die manuelle Arbeit bei der Prozessierung der Bilder deutlich reduziert ist. Dies wurde durch das neue Field-Setup mit akkurat positionierten und gleichbleibenden Ground Control Points möglich.



Abbildung B2: Neben den Drohnenaufnahmen wurden regelmäßig phänologische Daten der Blumenkohlpflanzen erhoben. © OPTIKO

### **Arbeitspaket 2: Erfassung des Expertenwissens**

Auf dem Feld „Neugrabenweg 6 ha“ (Feld 1) wurden während der Drohnenüberflüge die phänologische Entwicklung nach BBCH, die Höhe, maximaler Durchmesser, Durchmesser der Blume und Auffälligkeiten, insbesondere biotische Stressfaktoren, von 180 Blumenkohlpflanzen (je 90 Pflanzen der Sorten *Clarina* und *Guideline*) erfasst. Dieselben Daten wurden auf dem Feld „Am Mühlchen“ (Feld 2) bei 240 Blumenkohlpflanzen der Sorte *Korlanu* erfasst. Zudem wurden Daten von 60 Brokkolipflanzen an Feld 1 erfasst. Neben der phänologischen Datenerfassung wurden Feldbegehungen mit dem landwirtschaftlichen Berater der LWK-NRW durchgeführt, um sein Expertenwissen ebenfalls zu nutzen.

### **Arbeitspaket 3: Analyse der erhobenen Daten mittels maschineller Lernverfahren**

Nach einer Evaluierung der zur Verfügung stehenden Techniken des neuronalen Lernens wurde beschlossen, sich für OPTIKO auf Neuronale Netze und Deep Learning Verfahren zu fokussieren



(verwendetes maschinelles Lernverfahren für die Detektion: Mask R-CNN, verwendetes maschinelles Lernverfahren für die Ableitung der Blütengröße: CycleGAN).

Es erfolgte die Einarbeitung in die dazugehörige Literatur und erste Tests an gleichgearteten, externen Pflanzen-Datensätzen wie dem ‚LSC Datensatz‘ (<https://www.plant-phenotyping.org/CVPPP2017-challenge>) und dem ‚Brassica Datensatz‘ (<https://cloudstor.aarnet.edu.au/plus/s/N0giZD5QTMB32Xj>).

In ersten Tests wurde das Deep Learning Verfahren auf die in AP 1 aufgenommenen Blumenkohldaten angewendet. Die Blumenkohlpflanzen konnten in den Aufnahmen bereits als einzelne Instanzen separiert erfasst werden.

Um ein Detektionsmodell zu lernen, das auf unsere Aufgabenstellung und unsere Daten angepasst ist, werden die für das Detektionsmodell notwendigen Annotationen der Blumenkohl-Daten erstellt. Die Annotationen bestehen pro aufgenommenem Bild aus einer Maske (siehe Abbildung B3), die auf Pixelbasis angibt, ob das Pixel dem Boden oder einem Teil einer Pflanze zugeordnet wird. Zusätzlich wird die entsprechende Pflanzeninstanz sowie Blattinstanz des jeweiligen Pixels vermerkt. So kann jedes Blatt einer Pflanze zugeordnet werden. Dies ermöglicht eine Ableitung phänotypischer Merkmale, wie die Größe der Blumenkohlpflanze und später auch Blumenkohlblüte, für die Schätzung des Erntezeitpunkts.



*Abbildung B3: Beispielhafte Annotation eines aus dem in AP1 erstellten Orthophotos extrahierten Bildes. Lila entspricht hierbei den Bodenpixeln, rot den Begrenzungen einer Pflanzeninstanz und die weiteren Farben geben die verschiedenen Blattinstanzen wieder. Alle farblich markierten Blätter innerhalb einer roten Umrandung gehören zu derselben Pflanzeninstanz. © OPTIKO*

#### **Arbeitspaket 4: Entwicklung eines Theorie-getriebenenen maschinellen Lernverfahrens zur Integration von Expertenwissen**

Mit der Einarbeitung in die Methodik maschineller Lernverfahren und deren Eignung für die angedachten Anwendungen wurde in dieser Phase begonnen. Leider wurde sie nicht im angestrebten Umfang erreicht, da der zeitliche Aufwand bei der Koordination der experimentellen Durchführung weitaus grösser als angenommen war. Außerdem schränkte die Covid-19-Pandemie den Austausch mit anderen Wissenschaftlern erheblich ein.

Dennoch fanden eine profunde Literaturrecherche, Erlernen der Grundsätze der für maschinelles Lernen verwendeten Programmiersprache Python statt.

#### 4.3.3: Erhebung von Multispektraldaten sowie Expertenwissen während der zweiten Vegetationsperiode, Weiterentwicklung der maschinellen Lernverfahren (Januar – Dezember 2021)

##### **Arbeitspaket 1: Datenerhebung anhand von Überflügen mit UAVs**

Die Planung und Erprobung in der vorigen Projektphase bewährte sich bei der Überfliegung von Feld 3 („Bannweg, kleines Feld“) im Frühsommer 2021, für das bei jedem Flug ein Orthomosaik erstellt werden konnte, das bei der Bildausrichtung im Rahmen der Datenverarbeitung perfekt ausgerichtet werden konnte. Nach mehr als einem Jahr des OPTIKO-Projekts war es aufregend, endlich einen perfekten Datensatz ohne ein einziges fehlerhaftes Orthomosaik zu erhalten. Weitere wichtige Effizienzgewinne wurden im Bereich der Bildverarbeitung nach dem Flug erzielt, wobei viele Prozesse stärker automatisiert werden konnten.

##### **Arbeitspaket 2: Erfassung des Expertenwissens**

Feldbonituren zur Validierung der von maschinellen Lernverfahren (AP3) detektierten Heterogenität und Erfassung des Expertenwissens zu Erntezeitpunkten hunderter Pflanzen wurden in einer Datenbank zusammengestellt.

##### **Arbeitspaket 3: Analyse der erhobenen Daten mittels maschineller Lernverfahren**

Das bereits für die Pflanzenextraktion gelernte Deep Learning Modell (Mask R-CNN) wurde auf die in 2021 aufgenommenen Daten angewendet, um so Koordinaten für die Einzelpflanzen zu bestimmen (Abbildung B4).

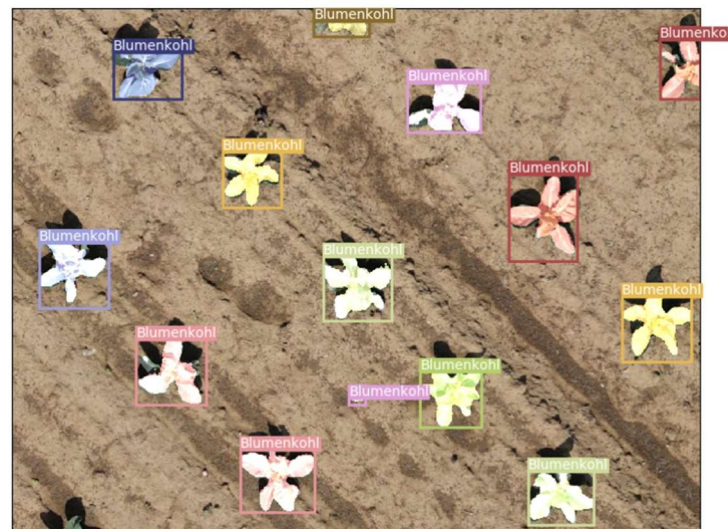


Abbildung B4: Beispielhafte Ergebnisse der Pflanzendetektion basierend auf den in AP1 aufgenommenen Daten. Farblich hervorgehoben werden die vorhergesagte Maske der Pflanzen und die Bounding Box. © OPTIKO

Die Georeferenzierung der Daten ermöglichte es, Zeitreihen für die Einzelpflanze zu extrahieren und die Berechnungen über den gesamten Wachstumszeitraum durchzuführen. Dies ermöglicht die Bestimmung des Pflanzenwachstums, was später für die Vorhersage des Erntezeitfensters genutzt werden kann. Beispiele von Zeitreihen, die unterschiedliche Entwicklungen zeigen, werden in Abbildung B5 dargestellt. Eine Zeile stellt dabei eine Pflanzenzeitreihe dar. Die Spalten zeigen die verschiedenen Aufnahmezeitpunkte. Die Zeitreihen ermöglichen Berechnungen über den gesamten Wachstumszeitraum durchzuführen. Dies ermöglicht die Bestimmung des Pflanzenwachstums, was später für die Vorhersage des Erntezeitfensters genutzt werden kann.

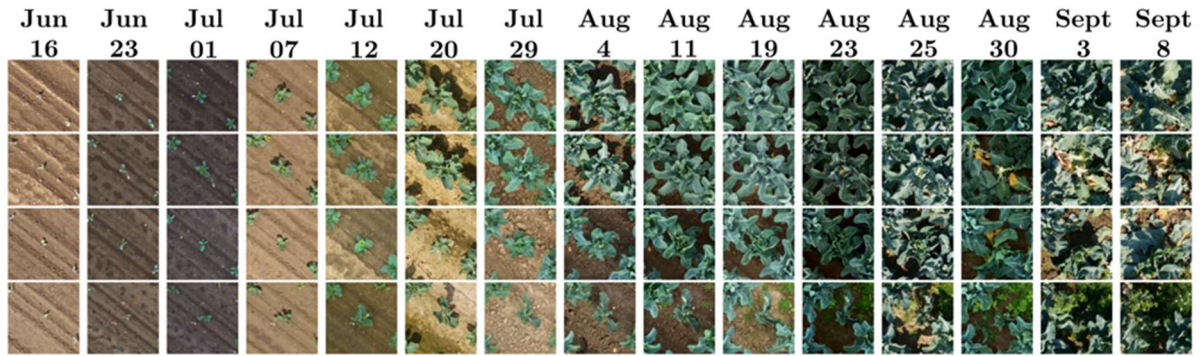


Abbildung B5: Zeitreihen für Einzelpflanzen extrahiert aus den 2021 aufgenommenen Daten. © OPTIKO

Entblätterungsversuche ermöglichen eine visuelle Analyse zwischen äußerem Erscheinungsbild der Pflanzen und der inneren Blumenkohlblüte. In dieser Projektphase wurden im Vergleich zum Jahr 2020, mit ca. 600 Pflanzen, wesentlich mehr Pflanzen systematisch entblättert. Ein Beispiel eines entblätterten Plots vor und nach der Entblätterung ist in Abbildung B6 zu finden.

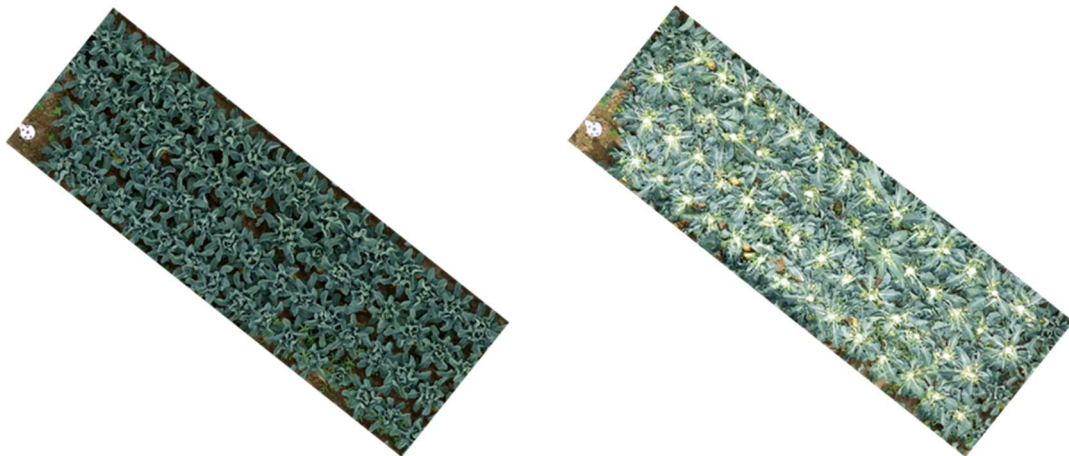


Abbildung B6: Darstellung eines Plots vor (links) und nach (rechts) der Entblätterung zur Sichtbarmachung der inneren Blüte. © OPTIKO

Maschinelle Lernmethoden, wie das Conditional Generative Adversarial Network Pix2Pix, ermöglichen es, ein Abbild zu lernen wie die verdeckte Blumenkohlblüte in Abhängigkeit der sichtbaren Informationen im Bild aussehen könnte (Abbildung B7, rechts). Die Bilder in Abbildung B7 zeigen ein beispielhaftes Ergebnis dieser Methodik angewandt auf die OPTIKO Blumenkohl-Daten aus dem Jahr 2020. Das linke Bild zeigt eine Pflanze, bei der die Blüte nicht sichtbar ist. Das mittlere Bild repräsentiert die entblätterte Blüte auf dem Feld. Die Entblätterung wurde mit Hilfe der anderen APs durchgeführt. Das rechte Bild zeigt ein Output des Modells, wobei zu beachten ist, dass das Modell aufgrund zu weniger vorhandener Daten an die vorliegenden Daten angepasst wurde. Die Beispielbilder wurden bereits vom Modell gesehen, wodurch sie nicht repräsentativ sind, sondern ausschließlich zeigen, was mit der Anwendung erreicht werden soll. Das Potential dieses Ansatzes wurde analysiert. Gegen Projektende wurde der Fokus jedoch auf vielversprechendere Ansätze gelegt.





Abbildung B7: Darstellung einer möglichen Visualisierung der in den Bildern nicht sichtbaren Blumenkohlblüte mittels eines Conditional Adversarial Networks (cGAN) (rechtes Bild). Links: Input-Bild der Pflanze mit versteckter Blüte. Mitte: Im Feld aufgenommene entblätterte Pflanze mit sichtbarer Blüte. Rechts: Ergebnis des cGAN Modells mit möglicher visueller Darstellung der inneren Blüte basierend auf dem Input-Bild mit verdeckter Blüte (links).

#### **Arbeitspaket 4: Entwicklung eines Theorie-getriebenen maschinellen Lernverfahrens zur Integration von Expertenwissen**

In diesem Arbeitspaket fanden aufgrund der Elternzeit von Frau Dr. Junker-Frohn nur geringe Aktivitäten statt. Zudem wurde klar, dass expertenwissengetriebene Modelle noch nicht etabliert sind und die Ausarbeitung eines funktionalen Modells im Rahmen der Projektlaufzeit unwahrscheinlich ist. Daher wurde nach der Rückkehr von Frau Dr. Junker-Frohn aus der Elternzeit im Januar 2022 die Entscheidung gefällt, den Fokus auf die Analyse multispektraler Daten zu legen, die bisher nicht bearbeitet wurden.

#### *4.3.4: Validierung der entwickelten maschinellen Lernverfahren unter Praxisbedingungen im Feld (Januar – September 2022)*

##### **Arbeitspaket 1: Datenerhebung anhand von Überflügen mit UAVs**

Mit der Überfliegung von Feld 4 (wieder „Neugrabenweg 6 ha“) wurden Daten erhoben, die in dieser Phase der Validierung der Lernverfahren dienen.

##### **Arbeitspaket 2: Erfassung des Expertenwissens**

Die in den Vorjahren erhobenen Daten wurden für die Veröffentlichung überarbeitet, Zusätzlich wurden im Rahmen der Datenerfassung für das Projekt Wetterstationen und verfügbare Datensätze zu Wetterbedingungen während der Versuchszeiträume recherchiert und entsprechende Daten bereitgestellt. Im Frühjahr fand ein Treffen mit dem Partnerbetrieb statt, um die bisherigen Ergebnisse zu kommunizieren und die geplanten Feldmessungen abzusprechen.

##### **Arbeitspaket 3: Analyse der erhobenen Daten mittels maschineller Lernverfahren**

Für die Vorhersage des Erntezeitpunktes wurden Verfahren des self-supervised learnings genutzt. Der Hauptgrund war hierfür die Datengrundlage, die aus wenigen gelabelten Daten (Daten mit Annotationen wie Wachstumsparameter oder Erntezeitpunkt) und vielen ungelabelten Daten (Daten mit keinen Annotationen) bestehen. Verfahren des self-supervised learnings kombinieren beide

Arten von Daten, sodass die Datengrundlage bestmöglich genutzt wird. In einem Pretraining Schritt wird zunächst ein Modell auf den ungelabelten Daten trainiert. Als Optimierungskriterium werden dabei Informationen aus den Daten verwendet, die direkt beim Training des Modells erzeugt werden oder bekannt sind, wie z.B. den Aufnahmezeitpunkt des Bildes. Das Pretraining sorgt dafür, dass das Modell die Daten einmal sieht und sich an sie anpasst. Bei dem sogenannten Downstream Task wird das Modell dann basierend auf den gelabelten Daten hinsichtlich einer gegebenen Aufgabe optimiert. Das kann z.B. das Clustering von Eigenschaften innerhalb der Daten sein oder die Klassifikation des Erntezeitpunktes. Abbildung B8 zeigt drei exemplarische Cluster.

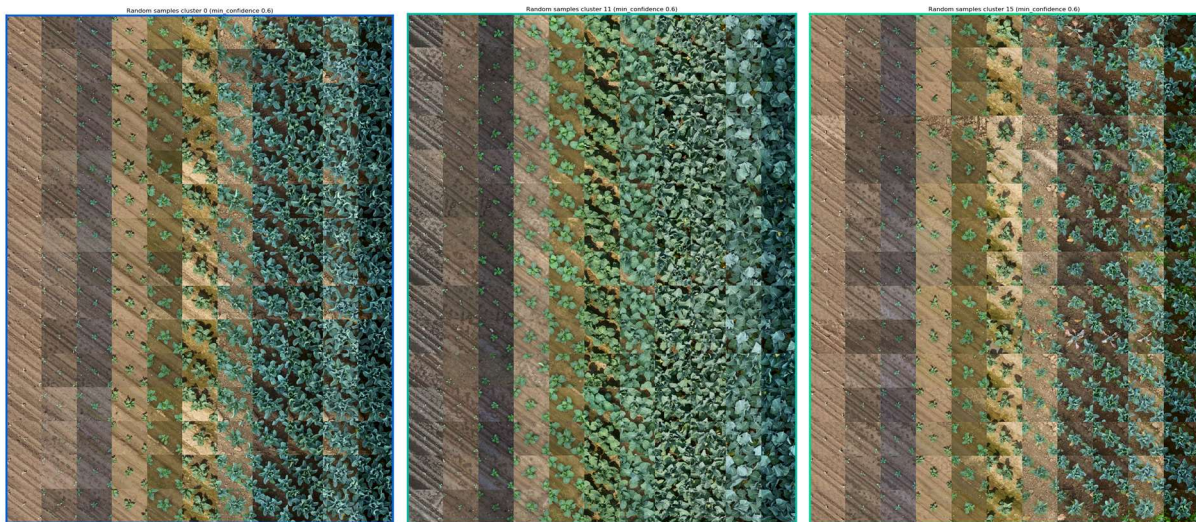


Abbildung B8: Darstellung von drei exemplarischen Clustern, welche verschiedene Entwicklungen und somit unterschiedliche Wachstumseigenschaften darstellen. Eine Reihe repräsentiert eine Zeitreihe und eine Spalte einen Aufnahmezeitpunkt innerhalb der Zeitreihe. Pflanzen innerhalb eines Clusters entwickeln sich ähnlich.  
© OPTIKO

Ein weiterer Ansatz ist die Klassifikation des Erntezeitpunktes. Dafür haben wir den zuvor beschriebenen Ansatz erweitert um anstelle von Clustern Klassen vorherzusagen, die mit den verschiedenen Erntezeitfenstern übereinstimmen. Dafür haben wir die Daten aus dem Jahre 2021 genutzt, für die es vier definierte Erntezeitpunkte (= 4 Klassen) gab, sowie eine Klasse mit noch nicht geernteten Pflanzen sowie eine Klasse, die während des Wachstums abgestorbenen Pflanzen beinhaltet. Analysen haben gezeigt, dass die ausschließliche Verwendung der vier Erntezeitpunkte als Klassen zu einer höheren Genauigkeit in der Vorhersage des Erntezeitfensters führen als die zusätzliche Verwendung der anderen beiden Klassen, siehe Abbildung B9. Außerdem wurde unser Ansatz inklusive des Pretrainings (orange) dem Standardansatz ohne datenspezifischem Pretraining (rot) gegenübergestellt. Auch hier ist zu erkennen, dass unser Ansatz inklusive datenspezifischem Pretraining zu besseren Ergebnissen führt. Weiterführend analysieren wir welche Zeitpunkte für die Vorhersage des Erntezeitpunktes einen größeren Einfluss haben.

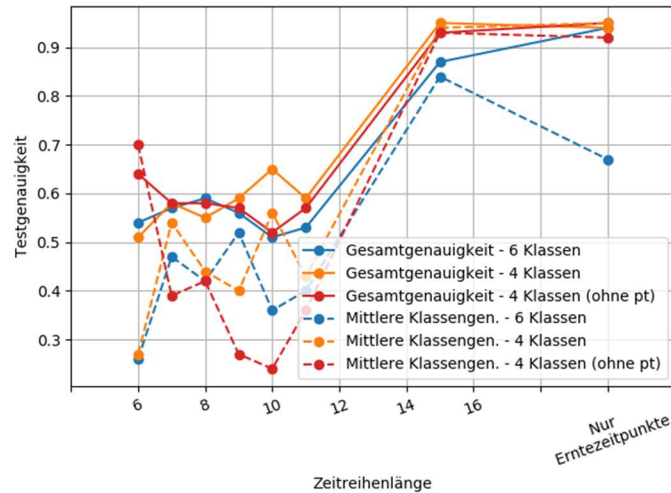


Abbildung B9: Gegenüberstellung der Klassifikationsergebnisse für (blau) 6 und (orange) 4 Klassen inklusive pretraining und (rot) 4 Klassen ohne pretraining. Die durchgezogene Linie zeigen die Gesamtgenauigkeit und die gestrichelte Linie die mittlere Klassengenauigkeit. © OPTIKO

Für die Vorhersage des Erntezeitpunktes auf Basis von Einzelzeitpunkten wurde das Verfahren der Klassifikation verwendet. Explizit wurde die Modell Architektur ResNet18 angewendet. Als Datengrundlage dienten die aufgenommenen Referenzdaten aus 2021 ab der Woche vor Beginn der Ernte bis Ende der Datenaufnahmeperiode (23.8.2021, 25.8.2021, 30.8.2021, 3.9.2021). Die Daten wurden in die folgenden Klassen aufgeteilt: 0 - Noch nicht bereit für die Ernte innerhalb der nächsten drei Tage, 1 - Bereit für die Ernte innerhalb der nächsten 3 Tage. Bilder von Pflanzen, die zu einem Zeitpunkt bereits geerntet wurden, wurden aus der Datenmenge heraus genommen um eine höhere Genauigkeit des Modells zu gewährleisten und den Bias zu verringern. Unser trainiertes Modell zeigt eine Gesamtgenauigkeit von 74.23% und mittlere Klassengenauigkeit von 75.19 % auf Testdaten. Dies bedeutet, dass wir bei 3 von 4 Pflanzen korrekt vorhersagen können, ob sie innerhalb von 3 Tagen erntereif sind.

Für die Quantifizierung der Heterogenität im Feld wird das gelernte Modell zur Klassifizierung des Erntestatus auf solche Pflanzen im Feld angewendet, für die keine im Feld aufgenommenen Referenzen bzgl. Erntezeitpunkt vorhanden sind.

Abbildung B10 zeigt den geschätzten Erntestatus der Pflanzen zu den vier verschiedenen Zeitpunkten. Die Pflanzen wurden je nach Vorhersage in grün (Klasse 1) oder rot (Klasse 0) eingefärbt. Ersichtlich wird, dass vom 23. August 2021 bis zum 3. September 2021 der Anteil erntereifer Pflanzen zunimmt (visualisiert durch zunehmende grün markierte Pflanzen). Der Erntegang der Arbeiter auf dem Feld verlief von süd-westlicher Ausrichtung des Feldes in Richtung nord-östlicher Ausrichtung des Feldes. Am 23. August ist zu erkennen, dass im süd-westlichen Streifen unterhalb der Traktorspur vermehrt grüne Markierungen auftreten. In Realität sind auch hier die Arbeiter am ersten Erntetag entlang gelaufen um die Ernte zu starten, haben die Ernte aber aufgrund zu weniger erntereifer Pflanzen abgebrochen. Aus subjektiver Sicht (nicht qualitativ bewertbar aufgrund fehlender Referenzen) zeigt der vorhergesagte Erntestatus der Pflanzen ähnliche Ergebnisse an, wie sie auch auf dem Feld aufgetreten sind.



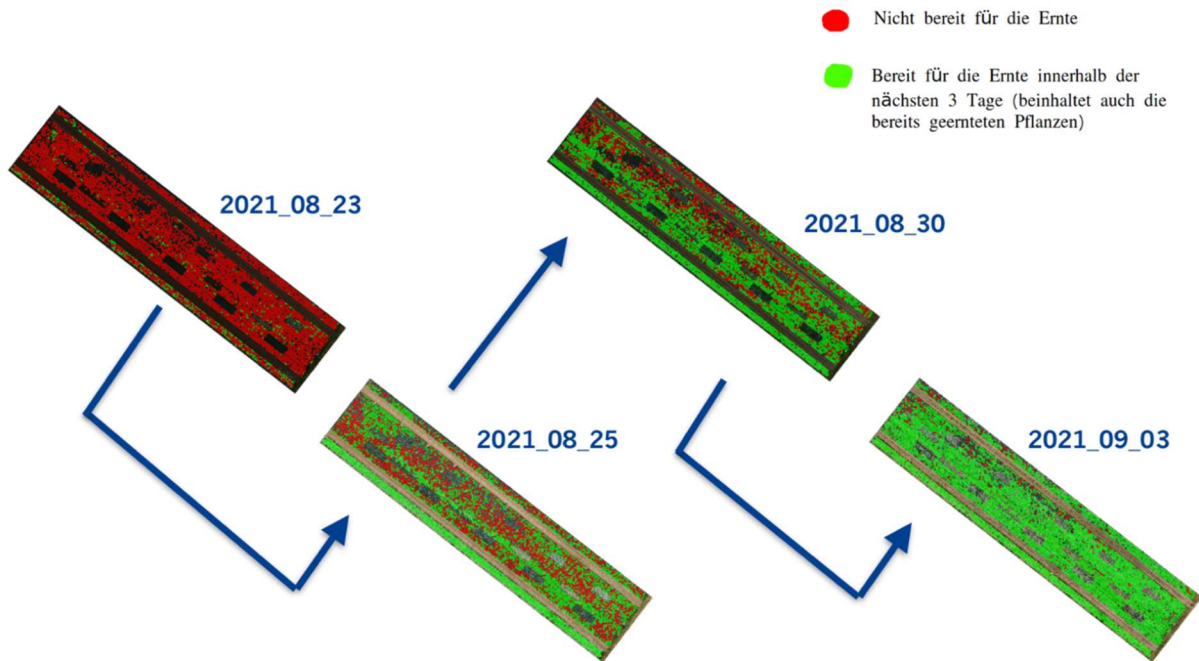


Abbildung B10: Visuelle Darstellung der Vorhersage des Erntestatus. Abgebildet sind die vorhergesagten Klassen (rot) nicht bereit für die Ernte und (grün) bereit für die Ernte innerhalb der nächsten drei Tage. © OPTIKO

Es sieht vielversprechend aus, dass auch für diese Ergebnisse eine Rate von 75% für korrekte Vorhersagen zutrifft. Zu beachten ist, dass in Abbildung B10 die verschiedenen Zeitpunkte unabhängig voneinander vorhergesagt werden, ohne eine Vorhersage derselben Pflanze zu einem früheren Zeitpunkt zu berücksichtigen. Basierend auf den Ergebnissen könnte den Landwirten eine Übersicht zur Hand gegeben werden, auf deren Grundlage sie einschätzen könnten, wie viele Arbeiter wann auf welche Bereiche im Feld zur Ernte eingeteilt werden sollten.

Für die Verbesserung und Analyse der Ergebnisse wurden Sensitivitätsanalysen durchgeführt. Anwendung fanden Methoden wie SHAP values, Occlusion Sensitivity Maps (OSM) und Grad-CAM. Sensitivitätsanalysen werden durchgeführt um zu analysieren anhand welcher Merkmale in den Daten ein Modell seine Entscheidung trifft, um Rückschlüsse darauf zu ziehen, welche Merkmale für eine bestimmte Klasse von Bedeutung sind. Nachdem uns der Landwirt das Vorwissen mitgeteilt hat, dass keine äußeren Merkmale der Pflanze dabei helfen den Erntestatus der Pflanze abzuleiten, haben unsere Versuche gezeigt, dass ein markantes Merkmal das im Zentrum der Pflanze liegende Blätterdach ist, welches um den Blumenkohlkopf herum wächst. Abbildung B11 zeigt beispielhafte Ergebnisse nach Anwendung von Grad-CAM. Rote Pixel weisen dabei auf Regionen hin, die für eine spezielle Klasse sprechen. Dunkelblaue Pixel werden im Netzwerk nicht aktiviert und haben somit keinen Einfluss auf das Klassifikationsergebnis. Unter Betrachtung der Ergebnisse in Abbildung B11 fällt auf, dass rote Hotspots genau dort auftreten wo das Zentrum der Pflanze ist. Anhand der Ergebnisse wird analysiert, ob die Ergebnisse des Modells vertrauenswürdig sind und ob man die Datenaufbereitung in eine Richtung leiten kann, die die gelernten Modelle verbessert.

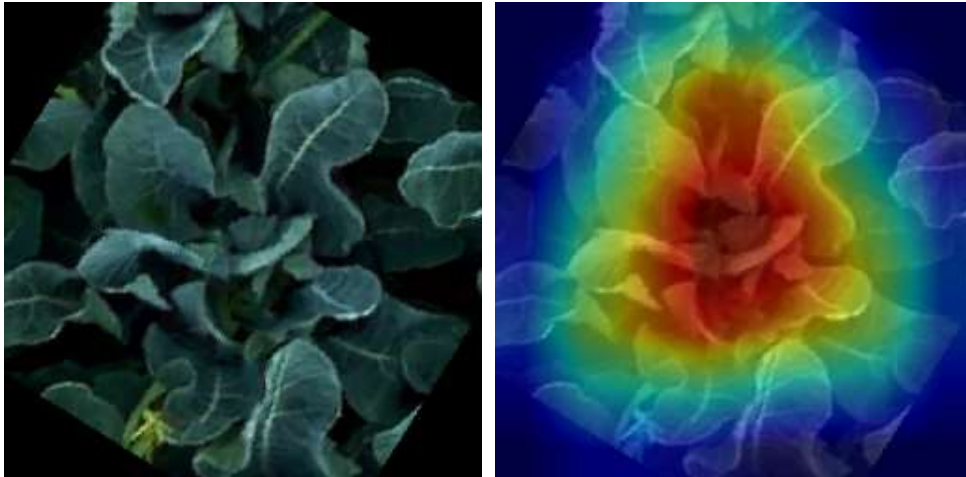


Abbildung B11: Ergebnisse der Sensitivitätsanalyse mittels Grad-CAM. Links wird das originale Bild dargestellt. Rechts eine Kombination aus Sensitivität und originalem Bild. © OPTIKO

#### **Arbeitspaket 4: Entwicklung eines Theorie-getriebenenen maschinellen Lernverfahrens zur Integration von Expertenwissen**

Nachdem Dr. Junker-Frohn aus der Elternzeit zurück war, wurde entschieden, den Fokus des Arbeitspaketes zu verändern, um bis zum Projektende möglichst verwertbare Ergebnisse zu erhalten. Aufgrund der Komplexität und Neuartigkeit Expertenwissen-getriebener Modelle im Rahmen der Projektlaufzeit erschien die Entwicklung derselben als unwahrscheinlich. Ursprünglich war angedacht gewesen, die Datenanalyse mittels maschineller Lernverfahren im Projekt auf Multispektraldaten anzuwenden. Da sich die Prozessierung der Multispektraldaten jedoch als schwierig erwies und RGB-Daten früher und in höherer Auflösung zur Verfügung standen, sind diese für die Entwicklung der maschinellen Lernverfahren genutzt worden. Da Multispektraldaten jedoch zusätzliche Informationen über den physiologischen Zustand der Pflanzen enthalten können, wurde (mit Zustimmung der bewilligten Behörde) entschieden, den Fokus in AP4 auf die korrekte Prozessierung und Analyse der Multispektraldaten zu legen. Die Projektziele „Schätzung von Wuchs und Pflanzengesundheit der Blumenkohlpflanzen“, beschrieben in M 4.3 und M 4.4 mit wissensbasierten neuronalen Netzen wurden entsprechend angepasst.

Die korrekte Prozessierung der erhobenen multispektralen Einzelaufnahmen zu Orthomosaiken erwies sich als äußerst komplex und zeitaufwändig. Die Daten von Feld 2 waren wegen zu geringer Überlappung nicht prozessierbar. Für Feld 3 und 4 war eine Prozessierung möglich, jedoch lagen aufgrund der komplexen Kalibrierung (siehe Anhang C3) die Orthomosaik von Feld 3 und Feld 4 erst im November 2022 in hoher Qualität vor. Aufgrund des resultierenden Zeitmangels konnten die multispektralen Orthomosaik im Rahmen des Projektes nicht vollumfänglich ausgewertet werden. Es ist geplant, einen entsprechenden Manuskriptentwurf sowie den vollständigen Datensatz im Anschluss an das Projekt im Rahmen einer Masterarbeit zu verwenden, um den Datensatz wissenschaftlich zu verwerten und den Erkenntnisgewinn im Sinne des Projektziels zu ermöglichen. Dr. Junker-Frohn wird auch nach Beendigung ihrer Projektstätigkeit als wissenschaftliche Betreuerin zur Verfügung stehen.

#### 4.3.5: Projektabschlussphase: Zusammenfassen der Ergebnisse, Publikationen und Öffentlichkeitsarbeit (Oktober – Dezember 2022)

in der Projektabschlussphase wurden die Ergebnisse zusammengefasst. Erste Fachpublikationen waren schon in der vorigen Phase veröffentlicht worden. Die derzeit laufenden Arbeiten an weiteren Manuskripten werden auch nach Projektabschluss weitergeführt werden. Es liegen vollständige Datensätze als Excel-Dateien zur phänologischen Entwicklung der Blumenkohlpflanzen pro erfasstem Feld vor, sowie eine Auswertung der durchschnittlichen phänologischen Entwicklung und des Wachstums. Ebenfalls sind Multispektraldaten in hoher Qualität unter <http://rs.ipb.uni-bonn.de/data/growliflower/> verfügbar. Eine weiterführende Abschlussarbeit von Uni Bonn und FZJ ist geplant um die vorhandenen Daten für weitere Auswertungen zu nutzen.

Das Projektabschlusstreffen fand am 28.11.2022 statt.

#### 4.4 Beitrag des Ergebnisses zu förderpolitischen EIP Zielen

Das Projekt OPTIKO hat einen wichtigen Beitrag vordringlich zum Ziel der „Digitalisierung der Landwirtschaft“ gemacht<sup>1</sup>. In einer zukunftsgerichteten Landwirtschaft werden automatisierte Prozesse immer wichtiger werden. Höhere Lohnkosten bei gleichzeitig weniger verfügbarem Fachpersonal mit langjähriger einschlägiger Erfahrung sowie ein weiterhin von starkem Preisdruck getriebener Lebensmittelhandel verlangen nach innovativen Verfahren um die Lebensmittelproduktion weiterhin nachhaltig gestalten zu können. Der im Projekt OPTIKO erforschte Ansatz, mithilfe eines maschinellen Lernverfahrens das Erntefenster für Blumenkohlanbau vorhersagen zu können ist ein Schritt in diese Richtung. Aufgrund der Komplexität des Ansatzes (Phänologie des Blumenkohls, Marktbedingungen, Klimawandel) ist eine Praxisreife noch nicht greifbar. Dennoch gehen wir davon aus, dass die in unserem Projekt erfassten und einer breiten Fachöffentlichkeit zur Verfügung gestellten Daten für eine weitere Entwicklung – auch bei anderen Kulturen – immens wertvoll sein werden.

#### 4.5 Nebenergebnisse

Keine

#### 4.6 Arbeiten, die zu keiner Lösung geführt haben

Die Arbeiten an der Entwicklung datengetriebener machine learning Ansätze zeigten sich arbeitsintensiver als zu Projektbeginn vorgesehen und konnten nicht in der angestrebten Tiefe durchgeführt werden.

### 5 Nutzen der Ergebnisse für die Praxis

Die Ergebnisse unseres Projekts sind derzeit für die Praxis nur eingeschränkt zu nutzen. Obwohl noch kein praxistaugliches Verfahren entwickelt werden konnte, zeigt das Projekt das Potential des erarbeiteten Verfahrens, das zukünftig auch auf andere Kulturen angepasst werden kann. Die generierten Erkenntnisse sind wichtig, um zukünftig eine zielgerichtete (Weiter-)Entwicklung robotischer Systeme voranzutreiben und dadurch die Digitalisierung in der Landwirtschaft zu fördern.

---

<sup>1</sup> In der Digitalisierung sieht das Land eine große Chance zur Lösung aktueller Herausforderungen in der Landwirtschaft. Dabei liegt ein Schwerpunkt im Farmdatenmanagement und der verständlichen Darstellung der Daten. Die Auswirkungen auf Beschäftigungszahlen und Agrarstruktur sowie die Datensicherheit muss gleichzeitig mitgedacht werden (Quelle: Wettbewerbsaufruf für den 2. EIP Wettbewerb)

## 6 (Geplante) Verwertung und Nutzung der Ergebnisse

Die wissenschaftlichen Erkenntnisse könnten in weitere Projekte zum Präzisionsgartenbau, z.B. innerhalb des PhenoRob Clusters der Universität Bonn einfließen. Dies kann durch die bestehenden Kooperationen insbesondere des FZJ und Uni Bonn IGG verstärkt werden.

Ein Highlight ist die Bereitstellung des „GrowliFlower“ Datensets. Es ist anzunehmen, dass viele Erkenntnisse darauf basierend gewonnen werden können, die insgesamt die Möglichkeiten zur Digitalisierung des Blumenkohlanbaus aufzeigen – voraussichtlich weit über den Rahmen dieses Projektes hinaus.

## 7 Wirtschaftliche und wissenschaftliche Anschlussfähigkeit

Wir sehen das Projekt auf dem TRL 3-4. Es ergeben sich daher viele Möglichkeiten zur wissenschaftlichen Weiterführung (s.o.) und zu einem späteren Zeitpunkt, vielleicht in einem weiteren Projekt, zur Erarbeitung von Prototypen.

## 8 Wo relevant: Nutzung Innovationsdienstleister (IDL)

Die Einrichtung der Position der Innovationsdienstleisterin war für unser Projekt sehr wichtig und sehr nützlich. Die Möglichkeit, Fragen, insbesondere zur Administration, schnell und kompetent beantwortet zu bekommen, hat die Durchführung des Projekts maßgeblich unterstützt. Ebenso wurde durch die Organisation von Veranstaltungen und ein sehr gutes Informationsangebot seitens der IDL das Netzwerk der laufenden EIP-Projekte in NRW maßgeblich gestaltet.

## 9 Kommunikations- und Disseminationskonzept

Die Kommunikation der Ergebnisse erfolgte über zahlreiche Workshops bzw. Besprechungen der OG, die über die gesamte Projektlaufzeit stattgefunden haben. Außerdem wurden verschiedene Veranstaltungen besucht, auf denen das Projekt mittels Poster oder Präsentationen vorgestellt wurde. Über LinkedIn / <https://www.linkedin.com/company/77699395/admin/> wurden Projektaktivitäten und –Ergebnisse mit der Öffentlichkeit geteilt (Abbildung B12).

Auch wenn durch die Covid-19 Pandemie die Kommunikation in die Öffentlichkeit und breite Praxis nicht wie gewünscht durchgeführt werden konnte, wurden die Projektergebnisse mit einem angepassten Konzept breitmöglichst gestreut.

### 9.1 Veranstaltungen (wissenschaftlich und nicht-wissenschaftlich):

#### **Veranstaltungen 2020**

- Foodhub NRW (14.09.2020)
- Shoot dynamics focus seminar (09.11.2020)

#### **Veranstaltungen 2021**

- Core Projekt 2 Meeting des Exzellenzcluster PhenoRob der Universität Bonn (23.07.2021)
- EIP Innovationstagung (7.09.2021)
- EIP OG-Workshop, online (08.2021)
- Frontiers of Geodetic Science (23.09.2021)

Posted by Annika Frohn · 12/16/2022

...



OPTIKO

205 followers

1mo ·

Our drone is equipped with a multispectral camera and takes pictures of the individual plants, which gave us good results this year. But it was a long way until we found our preferred settings from the drone. This involved changin ...see more



Ulrich Schurr and 10 others

Abbildung B12: Eintrag auf der LinkedIn-Seite des Projekts. © OPTIKO

### Veranstaltungen 2022

- DIGICROP 2022 (28.03.22)
- ISPRS Konferenz 2022, Nizza (9.06.2022)
- Tag der Neugier, FZJ (21.08.2022)
- Agrarforschungstag, Düsseldorf (1.09.2022)
- EIP OG-Workshop 2022, Hannover (7.-8.09.2022)
- Core Projekt 2 Meeting des Exzellenzcluster PhenoRob der Universität Bonn (07.10.2022)
- PhenoRob ISAB Meeting (7.11.2022)
- PhenoRob AIFARMS Delegation Meeting (14.11.2022)

### 9.2 Poster

Poster wurden auf verschiedenen Veranstaltungen vorgestellt. Hierbei kamen die folgenden drei Poster zum Einsatz (Abbildung B13a-c).



**Optimierung des Kohlanbaus durch Messungen mit UAVs und maschinellem Lernen**

J. Kierdorf<sup>1</sup>, L.V. Junker-Frohn<sup>2</sup>, R. Roscher<sup>1</sup>, U. Rascher<sup>1</sup>, O. Müller<sup>1</sup>, M. Delaney<sup>1</sup>, A. Burkart<sup>1</sup>, M. Schwarz<sup>1</sup>, H. Jaenicke<sup>1</sup>

**Hintergrund**

- Blumenkohl ist im Anbau anspruchsvoll und muss hohe Qualitätsanforderungen erfüllen
- Kontrollen der Felder durch Landwirte und ihre Berater zur frühzeitigen Erkennung von abiotischen und biotischen Stressen sind nur stichprobenhaft möglich
- Der Zeitpunkt zur Ernte muss aufgrund des kurzen, stark wetterabhängigen Erntezeitfensters gut abgeschätzt werden, um den Ertrag zu maximieren

**Projektziele und Erwartungen**

- Frühzeitige Erkennung von Stressfaktoren und Abschätzung der Erntereife durch ein automatisches Verfahren
- Abschätzung von Wuchs und Pflanzengesundheit von Blumenkohl mittels Überflügen mit UAVs und Anwendung maschineller Lernverfahren zur Satenauswertung
- Anwendung verhältnismäßig günstiger Technik, die in bestehende Arbeitsabläufe integriert werden kann
- Präzise Erkennung notwendiger Managementmaßnahmen wie Düngung und Pflanzenschutzmittelbehandlungen
- Vorhersage des idealen Erntezeitfensters

**Arbeitspakete (AP)**

AP 1: Drohnenüberflüge      AP 2: Erfassung von Expertenwissen

AP 3 + 4: Wissensbasiertes maschinelles Lernen

AP 1: Erstellung von Multispektralaufnahmen von Blumenkohlfeldern mit hoher räumlicher Auflösung durch Einsatz von Drohnen

AP 2: Erfassung der Entwicklung der Blumenkohlpflanzen sowie des Expertenwissens zu auftretenden Stressfaktoren

AP 3: Entwicklung maschineller Lernverfahren zur Abschätzung von Wuchs und Pflanzengesundheit sowie Vorhersage des Erntezeitpunkts anhand von Drohnenaufnahmen

AP 4: Integration von Expertenwissen in maschinelle Lernverfahren (AP3) zur Entwicklung eines wissensbasierten maschinellen Lernverfahrens

AP 5: Koordination und Öffentlichkeitsarbeit, Unterstützt die dargestellten vier technischen Arbeitspakete

In diesem EF-Initiativprojekt arbeiten Universität Bonn, Kompetenzzentrum Gärtenbau und Institut für Geodäsie und Geoinformation (IGG), Fernstudienzentrum Jülich GmbH (IGG-2), JAG, Hyperspectral Devices und KUGA & Markus Schwarz GbR zusammen an einer innovativen Lösung.

Dr. Hannah Jaenicke, Uni Bonn Kompetenzzentrum Gärtenbau, Campus Klein-Altendorf 2, 53129 Preibach, h.jaenicke@ig-ga.eu

Abbildung B13a: Poster 2020

**GrowliFlower: An image time-series dataset for GROWth analysis of cauLIFLOWER**

Jana Kierdorf<sup>1</sup> (CP2), Laura Verena Junker-Frohn<sup>2</sup>, Mike Delaney<sup>3</sup>, Mariela Donoso Olave<sup>1</sup>, Andreas Burkart<sup>1</sup>, Hannah Jaenicke<sup>3</sup>, Onno Müller<sup>2</sup> (CP1), Uwe Rascher<sup>2</sup> (CP1), Ribana Roscher<sup>1</sup> (CP2)

**Motivation**

- Agricultural dataset that is suitable for the development of ML approaches
- intended to address the growth analysis and development of cauliflower plants
- Derivation of phenotypic traits relevant for agricultural applications
- promote the development of automation in agriculture

**Study area**

- Monitoring of two cauliflower fields (0.39 and 0.60 ha) in two years, 2020 and 2021
- Location: West Germany (50°46'6.742" N, 6°58'20.271" O), 20 km south of Cologne
- Cauliflower cultivars: Korlanu (2020) and Guideline (2021)

**Data collection**

- RGB and multispectral UAV image data with high spatial resolution
- Georeferenced ground control points (GCP) to locate the data in space
- In-situ measurements of phenotypic traits characterizing the development state and stress factors that serve as reference observations

**Dataset**

**GrowliFlowerO / GrowliFlowerM:**

- RGB and multispectral orthophotos of two different cauliflower fields acquired over the entire growing period (from planting to harvest time)

**GrowliFlowerI / GrowliFlowerF:**

- Complete and incomplete time-series of image patches showing individual plants accompanied with in-situ reference data captured manually on the field

**GrowliFlowerD:**

- Image pairs of plants pre- and post-defoliation accompanied with a time-series of the respective plant to facilitate analysis of the correlation between the external appearance and internal head of the cauliflower plant

**GrowliFlowerL:**

- Pixel-accurate labeled data providing plant instance masks, leave instance masks and stem annotation masks

The poster includes a detailed timeline of data collection from June to September and a diagram showing the dataset structure with labeled image patches, defoliation time series, and time series of reference plants.

Abbildung B13b: Poster 2021

**Optimierung der Blumenkohlernte mit Drohnen und maschinellem Lernen**

J. Kierdorf<sup>1</sup>, L.V. Junker-Frohn<sup>2</sup>, M. Delaney<sup>3</sup>, M. Schwarz<sup>1</sup>, A. Wehr<sup>1</sup>, A. Burkart<sup>1</sup>, O. Müller<sup>1</sup>, U. Rascher<sup>1</sup>, R. Roscher<sup>1</sup>, H. Jaenicke<sup>1</sup>

**Problemstellung**

- Blumenkohl hat ein enges Erntezeitfenster
- Köpfe müssen groß sein, aber nicht überreif
- Pflanzen innerhalb eines Feldes entwickeln sich unterschiedlich schnell
- Blumenkohl muss von Hand in mehreren Erntegängen geerntet werden
- Erntezeitpunkte sind für Landwirt erst spät abschätzbar

**Ziel des Projekts**

- Abschätzung der Erntereife durch automatisches Verfahren
- Datenerfassung durch Überflüge mit Kameras und Sensoren
- Datenauswertung mittels maschineller Lernverfahren
- Verfahren kann bestehende Ernteprozesse optimieren

**Ergebnisse der Befliegungen ...**

In 2020 und 2021 wurden zwei Felder wöchentlich befliegen und fotografiert

- RGB- und Multispektralaufnahmen
- Georeferenzierten Kontrollpunkten
- Referenzmessungen zur Entwicklung von Blumenkohlpflanzen

**... und maschineller Lernverfahren**

- Computerprogramme können lernen, die Entwicklung der Pflanzen abzuschätzen
- Erstes Verfahren: Frühe Entwicklung anhand Anzahl der Blätter erkennen

Veröffentlicht als GrowliFlower Datenset:

Jun 10, Jun 23, Jul 01, Jul 07, Jul 12, Jul 20, Aug 04, Aug 11, Aug 19, Aug 23, Aug 25, Aug 30, Sept 3, Sept 8

The poster includes a grid of drone images showing plant development over time from June to September.

Abbildung B13c: Poster 2022

9.3 Projektvorstellung im KoGa-Newsletter

Das Projekt wurde regelmäßig im KoGa-Newsletter vorgestellt (Abbildung B14a, b).



Abbildung B14a: KoGa Newsletter 2020



Abbildung B14b: KoGa Newsletter 2021

### 9.4 Publikationen (s. Anhang C4)

Kierdorf, J., Weber, I., Kicherer, A., Zabawa, L., Drees, L., Roscher, R. (2022 a) Behind the Leaves: Estimation of Occluded Grapevine Berries With Conditional Generative Adversarial Networks. *Front. Artif. Intell.* 5:830026. doi: 10.3389/frai.2022.830026

Kierdorf, J., Junker-Frohn, LV., Delaney, M., Donoso Olave, M., Burkart, A., Jaenicke, H., Muller, O., Rascher, U., Roscher, R. (2022 b) GrowliFlower: An image time - series dataset for GROWth analysis of cauliflower. *J Field Robotics.* 2022; 1 – 20. DOI: 10.1002/rob.22122

### 9.5 Abschlussarbeiten

Hembach J.M. (2021): „Ableitung phänotypischer Merkmale im Blumenkohl mittels bildbasierter Instanzsegmentierung“, Bachelorarbeit, Universität Bonn.

Askar, F. (2022): „Growth Analysis of Cauliflower Plants using Machine Learning Methods“, Masterarbeit, Universität Bonn.

## 10 Fazit

Nach der in diesem Projekt generierten Erfahrung ist das Instrument EIP-Agri grundsätzlich gut geeignet, um landwirtschaftliche Innovationen im Grenzbereich Wissenschaft und Praxis zu fördern. Es bietet einen Raum, in dem der Praxis eine Stimme gegeben wird und kreative Lösungsansätze gemeinsam von Vertretern der Wissenschaft und der landwirtschaftlichen Praxis entwickelt werden können. Hier muss immer eine Balance aus wissenschaftlichem Erkenntnisgewinn und Praxisnähe gefunden werden. Dies ist nicht immer einfach.

Wie mit vielen anderen ähnlichen Programmen auch, ist es aber auch nach unserer Erfahrung schwer, Praxisbetriebe und Startups zur Erfüllung der administrativen Vorgaben zu bewegen. Die Notwendigkeit zur Vorfinanzierung über Zeiträume von z.T. über 12 Monate macht die Mitarbeit von kleineren Firmen in dem Konstrukt EIP-Agri sehr schwierig, wenn nicht unmöglich.



## C ANHANG

Anhang C1: Bericht von Arbeitspaket 1

Anhang C2: Bericht von Arbeitspaket 2 & 4

Anhang C3: Bericht von Arbeitspaket 3

Anhang C4: Publikationen



# **Optiko-Bericht**

## **Arbeitspaket 1**

Michael Delaney  
JB Hyperspectral  
24. November 2022

## Arbeitspaket 1:

Im Projekt Durchführung von UAV-Überflügen und Prozessierung der spektralen Aufnahmen.

### Einleitung:

Die Datenerfassung für das Optiko-Projekt fand im Laufe von drei Jahren (2020, 2021, 2022) statt. Im Jahr 2020 wurde die Fernerkundung des Blumenkohls in zwei separaten Wachstumszyklen versucht. In den Jahren 2021 und 2022 wurde die Fernerkundung für einen Wachstumszyklus pro Jahr abgeschlossen.

Bei der Datenerfassung wurden zwei verschiedene Arten von Bilddaten erfasst: RGB-Bilddaten und multispektrale Bilddaten. Diese Daten wurden aus der Ferne mit einer Drohne erfasst. Nach der Erfassung wurden die Daten zu Orthomosaiken verarbeitet, kalibriert, georeferenziert und dann an das Team des neuronalen Netzes übermittelt.

### Material und Methode:

Die RGB-Bilddaten wurden größtenteils mit einer Sony A7 Riii Kamera und einem Zeiss / Batis 2.0 Objektiv (Brennweite 25mm) aufgenommen. Für einige Flüge im Jahr 2020 und einen Flug im Jahr 2022 wurde stattdessen eine Sony A7 Riv-Kamera verwendet. In diesen Fällen wurde ein identisches Zeiss-Objektiv verwendet.

Die multispektralen Bilddaten wurden hauptsächlich mit einer MicaSense RedEdge 3 5-Kanal-Kamera erfasst. Für einige Flüge im Jahr 2020 und einen Flug im Jahr 2022 wurde stattdessen eine MicaSense RedEdge MX Dual 10-Kanal-Kamera verwendet.

Sowohl die RGB- als auch die Multispektralkamera wurden über einen DJI Ronin mx Gimbal an einer DJI Matrice M600 Drohne befestigt.



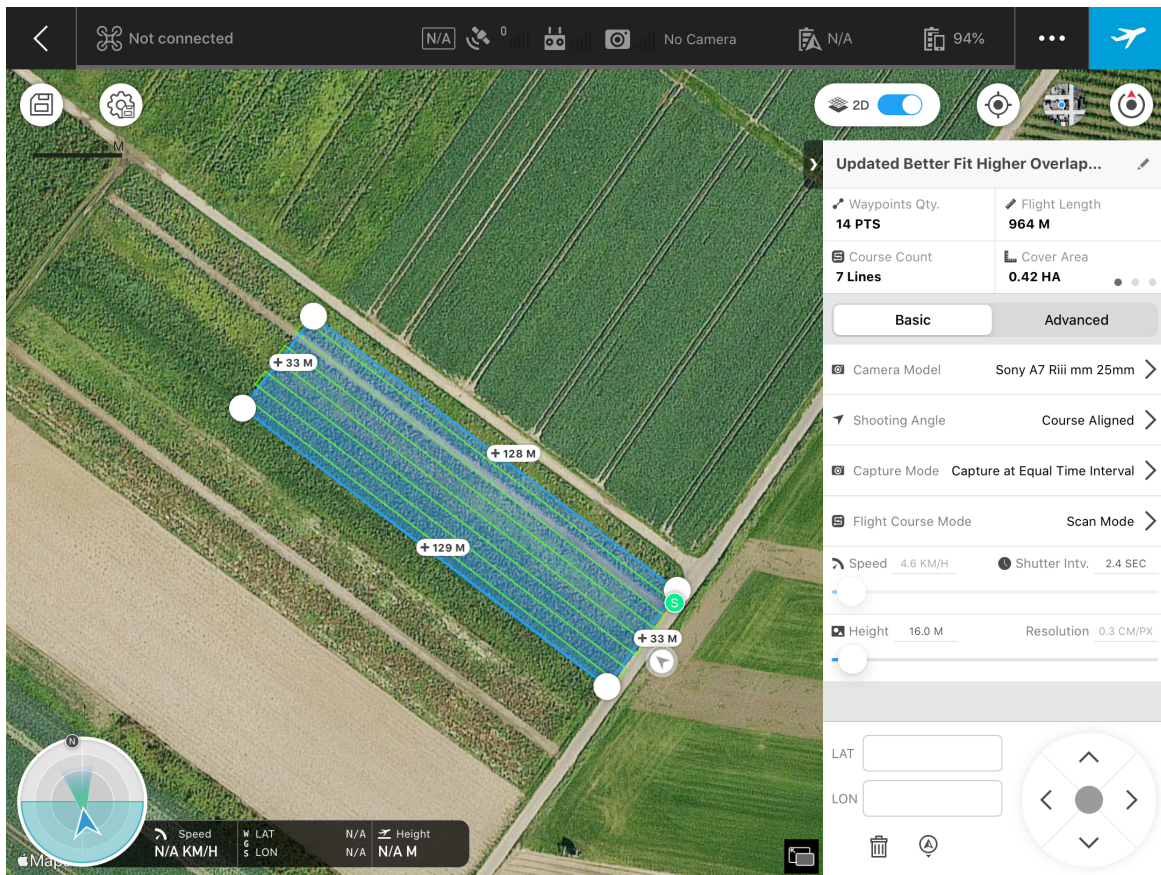


Abbildung 1: Beispielflugplan (mit Feld 3)

Die Flugplanung wurde mit der DJI Ground Station Pro App auf einem iPad durchgeführt. Neben den Bilddaten wurden auch grundlegende Metadaten wie Wetterbedingungen und Temperatur zum Zeitpunkt der Flüge erfasst. Wenn möglich, wurden die Flüge bei einer Windgeschwindigkeit von unter 6 m /s und minimaler Bewölkung durchgeführt. Wenn es die Bedingungen zuließen, wurden die Flüge so geplant, dass sie so nah wie möglich am Zenit der Sonne stattfanden.

Im ersten Jahr des Projekts (2020) wurde eine Flugfrequenz von zweimal pro Woche angestrebt. In den Jahren 2021 und 2022 wurde die Häufigkeit auf einen Flug pro Woche erhöht.

Bei jedem Flug wurden Luftbilddaten der Blumenkohlpflanzen innerhalb des vorgesehenen Gebiets (siehe Beispiel in Abb. XXXX) sowie neun 0,3 m x

0,3 m große Tafeln mit unterschiedlicher Grauintensität und Bodenkontrollpunkte erfasst.

Die vorgenannten grauen Flächen variierten in der Sättigung und waren in den Bildern enthalten, um die empirische Linienkorrektur (ELC) an den multispektralen Daten durchführen zu können. Von den neun Grauf Flächen, die bei jedem Flug aufgenommen wurden, wurden nur 7 als brauchbar erachtet. Ihre Graustufenintensitäten waren 2 %, 3 %, 4 %, 5 %, 6 %, 40 % und 63 %.

Tabelle 1: Überblick über die Flugtage in vier Wachstumszyklen.

	<b>First Flight</b>	<b>Last Flight</b>	<b>Total Flight Days</b>	<b>Defoliation flight Days</b>
<b>Field 1 (2020)</b>	27.04.2020	10.07.2020	16	0
<b>Field 2 (2020)</b>	05.08.2020	02.11.2020	16	0
<b>Field 3 (2021)</b>	16.06.2021	08.09.2021	15	6
<b>Field 4 (2022)</b>	16.06.2021	29.08.2021	14	4

Tabelle 2: Überblick über die Kameraeinstellungen in vier Wachstumszyklen

	<b>Aperture</b>	<b>Shutter</b>	<b>ISO</b>
<b>Field 1 (2020)</b>	f/8	Auto	800
<b>Field 2 (2020)</b>	Auto	1/1250	Auto
<b>Field 3 (2021)</b>	Auto	1/1250	50
<b>Field 4 (2022)</b>	Auto	1/1250	100

Nach der Erfassung wurden die Daten mit Agisoft Metashape zu Orthomosaiken verarbeitet. Für RGB-Orthomosaike betrug die Pixelgröße 3,1 mm und für multispektrale Orthomosaike 15 mm. Die Orthomosaiken wurden unter Verwendung des Koordinatenreferenzsystems WGS 84 / UTM Zone 32N (EPSG::32632) georeferenziert.

Um sicherzustellen, dass die Orthomosaiken (und die darin enthaltenen Pflanzen) so genau wie möglich georeferenziert wurden, wurden Bodenkontrollpunkte verwendet. Für die Felder 3 und 4 wurden 44 Bodenkontrollpunkte verwendet, die über die jeweiligen Felder verstreut waren. Für jeden Punkt wurden die Positionsdaten mit einem RTK-GPS-Gerät gemessen.

#### Ergebnisse:

Aus Sicht der Fernerkundung lieferten die beiden Wachstumszyklen 2020 keine verwertbaren Daten, 2021 und 2022 dagegen hervorragende Fernerkundungsdaten. Die Multispektraldaten von einem Flug im Jahr 2021 gingen verloren. Abgesehen von diesem einen Fehler waren die gesamten Datensätze für 2021 und 2022 vollständig.

#### Diskussion:

Der Grund für das Scheitern von 2020 war unter anderem eine schlechte Entscheidungsfindung, unzureichende Planung und Vorbereitung sowie die Unfähigkeit, einen Konsens zwischen den Wissenschaftlern zu erzielen.

Die Jahre 2021 und 2022 waren aufgrund eines klarer definierten Entscheidungsfindungsprozesses erfolgreich.

# 1 Erhebung und Auswertung von multispektralen Orthomosaiken von Blumenkohlfeldern

2 Laura Verena Junker-Frohn

3 Institut für Bio- und Geowissenschaften: Pflanzenwissenschaften (IBG-2), Forschungszentrum Jülich  
4 GmbH, Wilhelm-Johnen-Str., 52428 Jülich

5

## 6 Hintergrund

### 7 *Einsatz von Drohnen in der acker- und gartenbaulichen Landwirtschaft*

8 Die landwirtschaftliche Nahrungsmittelproduktion steht vor der Herausforderung, die Ernährung der  
9 wachsenden Menschheit in Zeiten der globalen Erwärmung sicherzustellen (Yang et al. 2017).  
10 Gleichzeitig müssen besonders gartenbauliche Produkte hohen Qualitätsstandards entsprechen, um  
11 vermarktungsfähig zu sein. Aus diesen Gründen ist ein präzises Feldmanagement nötig, um den Ertrag  
12 zu maximieren, und es kommt zunehmend zu einer Digitalisierung der eingesetzten Maschinen und  
13 Nutzung von Sensoren zur Erfassung der Pflanzenbestände (Rasmussen et al. 2016). Die jüngsten  
14 technologischen Fortschritte in der unmanned aerial vehicle (UAV)- und Kameratechnologie  
15 ermöglichen Aufnahmen ganzer Pflanzenbestände mit hoher räumlicher und zeitlicher Auflösung und  
16 ermöglichen die Entwicklung neuer Verfahren für die Landwirtschaft und den Gartenbau (Poncet et al.  
17 2019). Dabei ist die nötige Investitionsaufwand überschaubar und eine Integration in bestehende  
18 landwirtschaftliche Verfahren verhältnismäßig kostengünstig, zudem können die UAV-basierten  
19 Kameras für verschiedene Kulturen verwendet werden. Insbesondere für high value crops wie  
20 Blumenkohl, die einen hohen personellen Aufwand zur Beobachtung des Pflanzenwachstums und  
21 möglicher abiotischer und biotischer Stresssymptome erfordern, könnte die UAV-basierten RGB- und  
22 multispektralen Bildgebung zur Digitalisierung der Landwirtschaft im Rahmen des Präzisionsgartenbau  
23 beitragen (Yang et al. 2017). Bei der Bildverarbeitung können zum einen Vegetationsindizes, die  
24 repräsentativ für verschiedene Pflanzeigenschaften sind, auf der Grundlage von Farbkanälen aus  
25 RGB-Bildern oder der Reflexion in Spektralbändern aus multispektralen Bildern berechnet werden.  
26 Zum anderen ist es möglich, maschinelle Lernverfahren zur Abschätzung von Wuchs und  
27 Pflanzengesundheit zu entwickeln, die auch komplexe strukturelle Parameter einbeziehen.

28 Digitalkameras, die Informationen im Rot-grün-blauen Bereich erfassen (RGB-Kameras), sind die am  
29 häufigsten eingesetzten UAV-Kameras in der Pflanzenphänotypisierungsforschung. Der Sensor hat die  
30 Vorteile von geringen Kosten, geringem Gewicht, und einfacherer Datenverarbeitung. Obwohl die  
31 Belichtung an die Wetterbedingungen angepasst werden sollte, können Befliegungen prinzipiell bei  
32 Sonnenschein als auch bei bewölktem Himmel durchgeführt werden. Obwohl der Bedeckungsgrad und  
33 die Intensität der Grünfärbung der Pflanzen gut erfasst werden können, ist die Aussagekraft über  
34 physiologische Pflanzeigenschaften gering, da nur drei Lichtbänder im sichtbaren Bereich des Lichts  
35 detektiert werden können (Yang et al. 2017). Um die Entwicklung von einzelnen Pflanzen darzustellen,

36 müssen diese wiederholt fotografiert werden und die erstellten Zeitreihen analysiert werden, was eine  
37 fortgeschrittene Datenanalyse erfordert.

38 Im Vergleich zu RGB-Kameras erfassen Hyperspektralkameras mehr Bänder, sind daher  
39 datenintensiver, und die Anschaffungskosten sind höher. Jedoch ermöglicht die höhere Anzahl  
40 (üblicherweise 5 oder 10) erfasster Lichtbänder im sichtbaren und auch nahinfraroten Bereich eine  
41 Aussagekraft über den physiologischen Zustand der Pflanzen auch auf Basis einzelner Aufnahmen.  
42 Jedoch erfordert die Datenerhebung zusätzliche Kalibrierungen während der Datenaufnahmen im Feld  
43 und stabile Belichtungsbedingungen (Candiago et al., 2015). Der komplexeren Datenerhebung stehen  
44 daher einfachere Ansätze zur Datenanalyse entgegen. Es könnte daher für den Praxiseinsatz von Vorteil  
45 im Vergleich zu RGB-Kameras sein, dass Multispektralkameras ohne die Notwendigkeit komplexer  
46 Modelle für verschiedene Nutzpflanzen anwendbar sind, ohne dass Modelle gelernt werden müssen.  
47 Der Einsatz von Hyperspektralkameras würde weitere Bereiche des elektromagnetischen Spektrums  
48 kontinuierlich abdecken, jedoch mit sehr viel höheren Kosten, geringerer Drohnenreichweite und  
49 komplexerer Datenauswertung verbunden sein, sodass der zum jetzigen Zeitpunkt nicht praxisnah ist.

50 Zum jetzigen Zeitpunkt wurde der Einsatz von UAV-basierten Kameras für verschiedene Kulturen und  
51 Aufgabenbereiche entwickelt: zur Kartierung der Bedeckungsgrades von Getreide (Torres-Sánchez et  
52 al., 2014), des Chlorophyllgehalts von Getreide (Berni et al., 2009), des Unkrautbefalls in  
53 Sonnenblumenfeldern (Torres-Sánchez et al., 2013), von bakteriell übertragbaren Pflanzenkrankheiten  
54 in Zitrusplantagen (Garcia-Ruiz et al., 2013) und der Auswirkungen von Unkrautbefall und -  
55 bekämpfung auf den Ernteertrag von Gerste (Rasmussen et al., 2013; Streibig et al., 2014).

#### 56 *Technische Herausforderungen beim Einsatz von Multispektralkameras*

57 Zur Auswertung von Multispektraldaten müssen die verwendeten Sensoren gut kalibriert sein und die  
58 Belichtungseigenschaften zur Zeit der Befliegung anhand von grauen Reflektanzstandards  
59 aufgenommen werden. So kann bei der Datenanalyse eine radiometrische Kalibrierung zur Berechnung  
60 der Reflektanz durchgeführt werden kann, d.h. die Umwandlung roher digitaler Zahlen in den  
61 Reflexionsgrad unter Einbeziehung einer empirischen Linienkorrektur auf der Grundlage der  
62 Graureferenzen (Poncet et al 2019) oder auch unter Einbeziehung von nach oben gerichteten  
63 Lichtsensoren der Kameras. Effekte, die die Reflektanz beeinflussen, schließen die Auswirkungen  
64 atmosphärischer Absorption, Streueffekte und Änderungen des Spektrums des einfallenden Lichts  
65 aufgrund von Änderungen der Wolkenbedeckung, des Zenitwinkels der Sonne, der Sonnenaktivität und  
66 sogar der Position der Erde relativ zur Sonne ein (Jones et al 2010). Diese Effekte treten vor allem  
67 zwischen Flügen an verschiedenen Tagen auf, jedoch auch innerhalb eines Fluges. Um Abweichungen  
68 während des Überfluges zu minimieren, sollten Überflüge bei möglichst bei konstanten Wetter- und  
69 Lichtbedingungen durchgeführt werden (Bendig et al. 2021). Idealerweise steht die Sonne im Nadir,  
70 sodass Variationen des Winkel zwischen Sonne, Vegetation und Kameralinse minimiert werden  
71 (Rasmussen et al. 2016). Aufgrund der geringen Flughöhe treten jedoch auch innerhalb eines Bildes



72 Winkelvariationen auf, die die Reflektanz der Pflanzen sowie das Stitching-Verfahren zur Herstellung  
73 von Orthomosaiken beeinflussen. Verglichen zu den oft untersuchten Getreidekulturen ist die  
74 Morphologie von Blumenkohlpflanzen äußerst komplex, was das Stitching-Verfahren erschwert, da  
75 Pflanzen im Randbereich angrenzender Bilder aus unterschiedlichen Winkeln aufgenommen werden.  
76 Aufgrund der unterschiedlichen Repräsentation derselben Pflanze auf den Bildern können diese schwer  
77 automatisiert miteinander verknüpft werden. Eine hohe räumliche Auflösung und kleine Bildgröße  
78 führen zu hoher spektraler Varianz aufgrund der bidirektionalen Reflexionsverteilungsfunktion  
79 (BRDF), und auch empirische Korrekturfunktionen können dies oft nicht vollumfänglich ausgleichen,  
80 was zu Varianz in den Daten beiträgt (Tu et al 2018). Andererseits ist eine hohe Bildauflösung  
81 notwendig, um auch kleinere Pflanzen abzubilden. Um bei der Auswertung Einflüsse der Bodenfarbe  
82 auszuschließen, kann ein Segmentierungsschritt durchgeführt werden, bei dem Bodenpixel von der  
83 Analyse ausgeschlossen werden Rasmussen (2013).

84 Im Rahmen des Projekts OPTIKO sollte untersucht werden, ob mit Drohnen aufgenommen  
85 Multispektralaufnahmen von Blumenkohlfeldern dazu beitragen können, auftretende abiotische und  
86 biotische Stresse zu erkennen und die Erntereife frühzeitig abzuschätzen. Blumenkohl ist hierbei  
87 insofern eine vielversprechende Kultur, als dass die Pflege der Kulturen einen hohen personellen  
88 Aufwand bedeuten und hohe Qualitätsanforderungen gelten, was ein großes Potenzial für die Integration  
89 neuer Technologien bedeutet. Zudem ermöglicht die Selbstbedeckung der Blumenkohlpflanzen keine  
90 direkte Abschätzung des Kopfzustands und der Kopfgröße. Uns ist bisher nur ein Projekt bekannt, in  
91 dem der Einsatz proximaler hyperspektraler Messungen zur Interpretation des Reifegrads auf einer  
92 Erntemaschine untersucht wurde, jedoch wurden keine Ergebnisse veröffentlicht. Der Einsatz von  
93 Drohnen ermöglicht jedoch eine kostengünstigere Überwachung ganzer Flächen, ohne die  
94 Notwendigkeit für kostspieligere traktorbasierte Sensoren und die Befahrung ganzer Flächen.

95 Die experimentelle Durchführung wurde so geplant, dass anhand der erhobenen Datensätze untersucht  
96 werden konnte, ob anhand von RGB- und multispektrale Orthomosaike von Blumenkohlfeldern  
97 zusätzliche Informationen über das Pflanzenwachstum erfasst werden, die den Landwirt bei seiner  
98 Entscheidungsfindung bzgl. Pflanzenmanagement und Ernte-/ Verkaufsplanung unterstützen. Wir  
99 gehen davon aus, dass sich aus Multispektraldaten bestimmte Entwicklungsstadien der  
100 Blumenkohlpflanzen abgeleitet werden können, wie z. B. die initiale Etablierung, der Kronenschluss  
101 und die Kopfbildung, da sie mit Veränderungen der spektralen Signatur der Pflanzen einhergehen

102 Die in AP3 entwickelten maschinellen Lernverfahren (ML) wurden anhand von RGB-Aufnahmen  
103 entwickelt, da aufgrund der technischen Schwierigkeiten zunächst keine Multispektralaufnahmen  
104 vorlagen, und aufgrund der geringeren Auflösung kaum Strukturparameter erkennbar sind. Daher wurde  
105 in AP4 darauf fokussiert, korrekt kalibrierte Multispektralaufnahmen der Blumenkohlfelder zu erstellen,  
106 um untersuchen zu können, ob Mutlispektralaufnahmen relevante zusätzliche Informationen enthalten,  
107 und zukünftig in die Entwicklung von praxisnahen Verfahren einbezogen werden sollten. Daher wurde

108 hier zunächst eine klassische Auswertung der Daten angestrebt, die jedoch in Zukunft auch durch die  
109 Entwicklung entsprechender ML ausgewertet werden können oder zur Erhöhung der Aussagekraft in  
110 die bereits in AP3 entwickelten ML einbezogen werden können.

111

## 112 **Datenerhebung**

### 113 *Versuchsflächen und verwendete Pflanzen*

114 Die untersuchten Blumenkohlfelder wurden in Bornheim, Deutschland, angelegt. Dieser Standort liegt  
115 im Rheintal auf 55 m ü.d.M. und zeichnet sich durch ein mildes Klima (gemäßigte/ozeanische  
116 Klimazone nach der Köppen-Klimaklassifikation) mit einer mittleren Jahrestemperatur von 10 °C,  
117 einem Jahresniederschlag von 458 mm, flachem Gelände und fruchtbarem Boden aus. Die Felder  
118 wurden mit jungen, in Gewächshäusern gezüchteten Blumenkohlpflanzen eines lokalen Züchters in  
119 einer Dichte von 33000 Pflanzen/ha (60 cm Abstand zwischen den Reihen, 50 cm Abstand zwischen  
120 den Pflanzen) bepflanzt und konventionell angebaut. Zu den pflanzenbaulichen Maßnahmen gehörten  
121 Düngung, mechanische Unkrautentfernung durch Hacken, Pflanzenschutzmittelbehandlungen gemäß  
122 den Empfehlungen der landwirtschaftlichen Berater und bei Bedarf Bewässerung mit Sprinklern, um  
123 ein gleichmäßiges Pflanzenwachstum und eine gleichmäßige Entwicklung ohne größere abiotische oder  
124 biotische Belastungen zu ermöglichen. Innerhalb jedes Feldes wurde ein 18 m breiter Abschnitt des  
125 Feldes, der von Traktorspuren für Sprüh- und Bewässerungszwecke begrenzt wird, als  
126 Untersuchungsgebiet (region of interest, ROI) definiert, mit einer Länge von je 130 m (Abb. 1).  
127 Aufgrund der technischen Schwierigkeiten bei der Erhebung der Multispektraldaten sind nur die Daten  
128 der in 2021 und 2022 erhobenen Felder verwertbar.

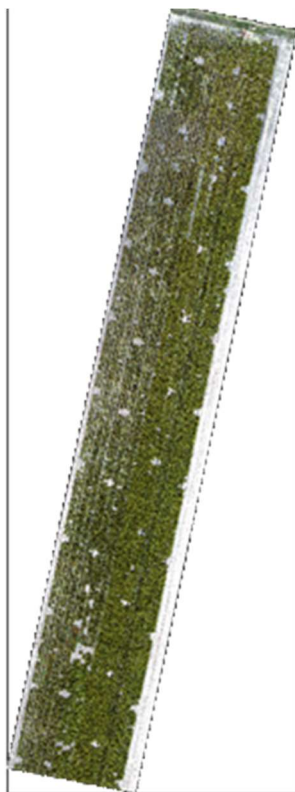
129 Feld A (61°50'47 "N 24°17'44 "E) wurde am 15.06.2021 angelegt und zwischen dem 19.08.2021 und  
130 dem 08.09.2021 geerntet, Feld B (61°50'47 "N 24°17'44 "E) wurde am 16.06.2022 angelegt und  
131 zwischen dem 21.08.2022 und dem 03.09.2022 geerntet. Beide Felder wurden ab dem Zeitpunkt der  
132 Aussaat bis zur letzten Ernte wöchentlich befliegen, wobei die Flüge während der Erntezeit häufiger  
133 durchgeführt wurden, um die Ernte der Pflanzen zu verfolgen. Auf beiden Feldern wurden im Rahmen  
134 von AP2 fünf Referenzparzellen mit auf den Drohnenbildern nicht erkennbaren Plastikpflöcken markiert  
135 und ihre Position durch RTK-GPS-Messungen aufgezeichnet. Die phänologische Entwicklung nach der  
136 BBCH-Skala (Feller et al. 1995), die maximale Höhe und der Durchmesser von fünf Pflanzen pro  
137 Parzelle wurden am Tag der Drohnenaufnahmen als Bodendaten erfasst. Darüber hinaus wurden die  
138 Anzahl der Blätter und der Durchmesser des Strunks von 100 Pflanzen pro Parzelle während der  
139 Blattentwicklung (BBCH  $\leq 19$ ) bzw. des vegetativen Wachstums (BBCH  $\geq 40$ ) gemessen und die  
140 Pflanzen auf abiotische oder biotische Stresssymptome überprüft.

141

142



143



144

145

146 **Abb. 1 RGB- und Multispektralaufnahmen (Normalfarbendarstellung) von Feld A (oben) am 12.7.2021 und**  
147 **Feld B (unten) im Maßstab von ungefähr 1:500 (links) sowie ungefähr 1:35 (rechts).**

148

149 *UAV-basierte Multispektralaufnahmen und Erstellung von Orthomosaiken*

150 Die Multispektralaufnahmen der Blumenkohlfelder wurden mit einem DJI Matrice 600 Hexacopter  
151 aufgenommen, der mit einer A7 rIII RGB Kamera (Sony Europe B.V. Surrey, UK) mit einem Batis 2.0  
152 Objektiv (Carl Zeiss AG, Oberkochen, Deutschland) und einer MicaSense RedEdge 3  
153 Multispektralkamera (AgEagle Aerial Systems Inc, Wichita, KS, US), die 5 Spektralbänder abdeckt

154 (475 nm, blau; 560 nm, grün; 668 nm, rot; 717 nm, roter Rand; 840 nm, nahes Infrarot). Die  
155 Drohnenflüge wurden in einer Höhe von 16 m für beide Felder durchgeführt, Flugrichtung entlang der  
156 Reihe, mit einer 80/80 Überlappung für die Multispektralkamera. Um sicherzustellen, dass bei der  
157 Erstellung von Orthomosaiken für die komplette ROI eine gleiche Anzahl von Bildern überlagert wird,  
158 wurden bei der Flugplanung ausreichend große Grenzen berücksichtigt, so dass ein 30 m breiter  
159 Feldabschnitt aufgenommen wurde. Die Überflüge wurden vorzugsweise an sonnigen oder bewölkten  
160 Tagen durchgeführt, um Beleuchtungsschwankungen innerhalb des erzeugten Orthomosaiks zu  
161 minimieren. Ein nach oben gerichteter Bestrahlungssensor mit denselben Spektralbändern wird zur  
162 Messung der einfallenden Bestrahlungsstärke für die radiometrische Normalisierung der  
163 aufgenommenen Bilder verwendet. Zusätzlich wurden 5 radiometrische Graukalibrierungstafeln mit 2  
164 %, 4 %, 6 %, 12 % und 24 % Reflexionsgrad zu Beginn und am Ende jedes Fluges aufgenommen.

165 Um eine genaue Georeferenzierung zu ermöglichen und das Zusammenfügen der Bilder zu  
166 Orthomosaiken zu erleichtern, wurden 44 Bodenkontrollpunkte (GCPs) im Feld platziert und ihre  
167 Position zu Beginn und am Ende der Flüge über jedes Feld gemessen. Die Orthomosaiken wurden mit  
168 der Software Agisoft Metashape Professional (Agisoft LLC, St. Petersburg, Russland) erstellt. Mit den  
169 Standardeinstellungen wurden RGB-Orthomosaiken mit einer Auflösung von 3,1 cm/ px (Feld A) und  
170 3,0 cm/ px (Feld B) und multispektrale Orthomosaiken mit einer Auflösung von 2,5 cm/ px (Feld A) und  
171 2,5 cm/ px (Feld B) erstellt, die im Koordinatensystem WGS84/ UTM 32 exportiert wurden. Insgesamt  
172 liegt für Feld A eine Zeitreihe von 15 georeferenzierten Orthomosaiken vor, für Feld B besteht die  
173 Zeitreihe aus 16 georeferenzierten Orthomosaiken. Orthomosaiken und weitere Referenzdaten aus Feld  
174 A sind Teil des Datensatzes "Feld 2", der von Kierdorf et al. (2022) veröffentlicht wurde.

175 Aus der Information der fünf Kanäle der Multispektralkamera für die georeferenzierten Orthomosaiken  
176 wurden folgende Vegetationsindizes berechnet:

177 Normalised Difference Vegetation index:  $NDVI = (R_{NIR} - R_{RED}) / (R_{NIR} + R_{RED})$

178 Enhanced Vegetation Index:  $EVI = 2.5 * (R_{NIR} - R_{RED}) / (R_{NIR} + 6 * R_{RED} - 7.5 * R_{BLUE} + 1)$

179 Optimized Soil Adjusted Vegetation Index:  $OSAVI = (1 + 0.16) * (R_{NIR} - R_{RED}) / (R_{NIR} + R_{RED} + 0.16)$

180 Excess Green:  $ExG = (2 * R_{GREEN} - R_{BLUE} - R_{RED}) / (R_{GREEN} + R_{BLUE} + R_{RED})$

181 Excess Red:  $ExR = (1.4 * R_{BLUE} - R_{GREEN}) / (R_{GREEN} + R_{BLUE} + R_{RED})$

182  $ExG\_ExR = ExG - ExR$

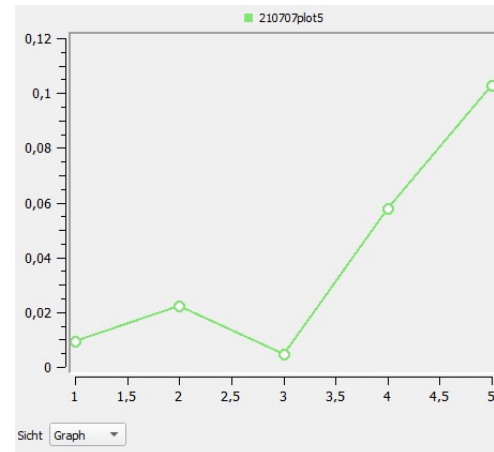
183 Zur Berechnung der Reflektanz von Einzelpflanzen wurden die Positionen der Pflanzen in Feld A aus  
184 Kierdorf et al. (2022) übernommen bzw. die Pflanzenpositionen für Feld B wurden ebenso erzeugt. In  
185 einem Umkreis von 25cm um den Mittelpunkt der Pflanze wurde eine ROI per Pflanze bestimmt. Da  
186 vor allem zu früheren Zeitpunkten noch Bodenpixel in dieser ROI liegen, wurde zudem mithilfe des

187 ExG\_ExR Indices segmentiert, um nur Pixel in die Analyse einzubeziehen, die durch eine  
188 Pflanzensignatur dominiert sind.

189

## 190 Erste Ergebnisse

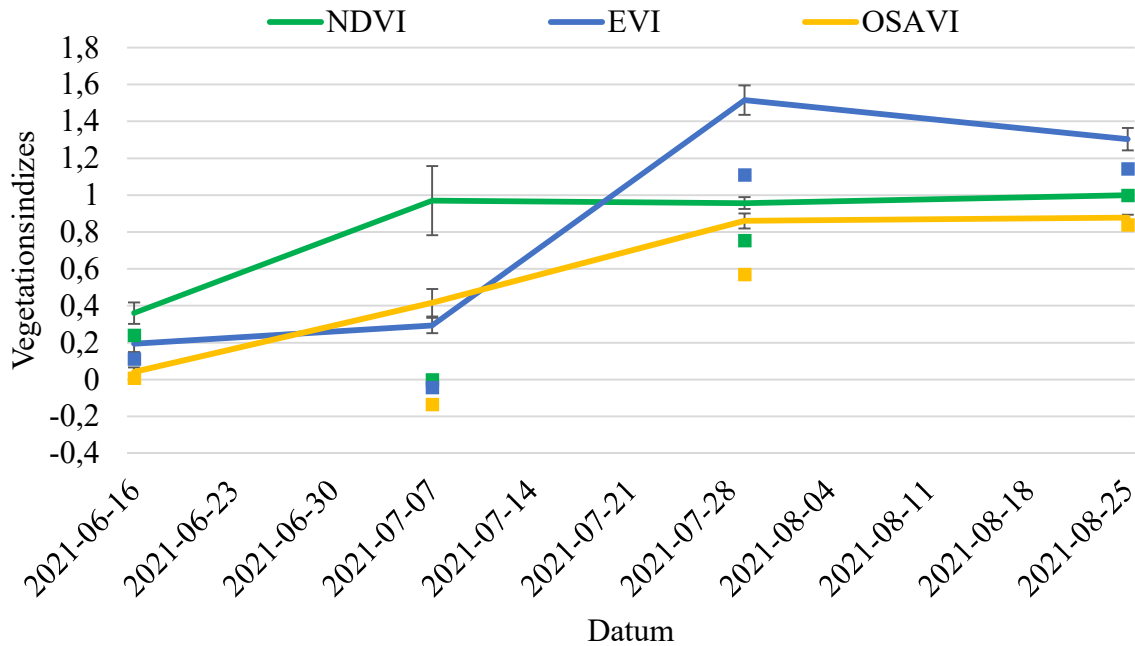
191 Die Prozessierung der Multispektraldaten erwies sich als komplexer als gedacht, ergab jedoch nach  
192 langer Arbeit Daten von hoher Qualität, wie das beispielhafte Spektrum in Fig. 2 zeigt.



193

194 **Abb. 2 Multispektralsignatur einer Blumenkohlpflanze (Feld A) zeigt das typische Reflektanzmuster von**  
195 **Pflanzen, mit einem Peak im Grünen Bereich (Kanal 3, entspricht 560 nm) und hoher Reflektanz im**  
196 **nahinfraroten Wellenlängenbereich (Kanal 5, entspricht 840 nm).**

197 Eine erst Auswertung der Reflektanz von N=97 Pflanzen an vier Zeitpunkten zeigt, dass sich die  
198 spektrale Reflektanz über die Kulturdauer verändert (Abb.3). NDVI bleibt nach einem initialen Anstieg  
199 konstant, was in Übereinstimmung mit der generell schnellen Sättigung dieses Indizes steht. Im  
200 Gegensatz dazu zeigt EVI einen späteren, aber stärkeren Anstieg, während OSAVI von Beginn an  
201 ansteigt. Interessant ist, dass die einzige Pflanze, die während der Kulturdauer eingegangen ist (zwischen  
202 dem 23.6. und 1.7. 2021), deutlich andere Reflektanzwerte zeigt, was dafür spricht, dass von Drohnen  
203 erhobene Multispektraldaten sensitiv für physiologische Veränderungen der Pflanzen sind.



204

205 **Abb. 3** Änderungen der drei Vegetationsindizes NDVI, EVI und OSAVI über die Kulturdauer der  
 206 **Blumenkohlpflanzen von Pflanzung (15.6.2021) bis Beginn der Ernte (19.08.-08.09.2021) für N=97**  
 207 **Referenzpflanzen aus Feld B, Versuchsfläche 5, die bis zum 25. August nicht geerntet wurden. Quadrate**  
 208 **repräsentieren Messungen an einer Pflanze, die Ende Juni eingegangen ist.**

209 Diese Grafik stellt nur einen Bruchteil des verfügbaren Datensatzes dar, und es lassen sich viele weitere  
 210 Fragestellungen anhand dessen untersuchen. Dazu gehören unter anderen die Untersuchung  
 211 vollständiger Zeitreihen mit wöchentlicher Auflösung, und wie sich verschiedene Indizes über die Zeit  
 212 verhalten. Gibt es einen Zusammenhang mit phänologischen Entwicklungsschritten wie z.B. der Beginn  
 213 der Blütenentwicklung? Lässt sich aus der Veränderung der Reflektanz, quantifiziert anhand von  
 214 Vegetationsindices, frühzeitig die weitere Entwicklung abschätzen? Hier wäre es z.B. möglich, dass  
 215 Pflanzen, die sich nicht etablieren können, frühzeitig eine Stresssignatur entwickeln (siehe Abb. 3), der  
 216 Zeitpunkt des Kronenschlusses im Zusammenhang mit dem Erntetermin steht oder sich die  
 217 Selbstbedeckung der Pflanzen bei der Blütenentwicklung auf die Reflektanz auswirkt und daher eine  
 218 Abschätzung der Erntereife erlaubt. Die Auswertung der Multispektraldaten zur Beantwortung könnte  
 219 hierbei auch auf die in AP3 erstellten Cluster zurückgreifen, um die Multispektralsignaturen besonders  
 220 gut/ schlecht gewachsener Blumenkohlpflanzen zu vergleichen. Die vor und nach Entblätterung der  
 221 Pflanzen aufgenommenen Bilder erlauben zudem zu untersuchen, ob es einen Zusammenhang zwischen  
 222 der multispektralen Signatur und der Größe des Kopfes besteht. Des weiteren erlaubt die Größe des  
 223 Datensatzes mit mehreren tausend erfassten Pflanzen auf zwei Feldern/ in zwei Jahren eine gute  
 224 Abschätzung der auftretenden Variation. Des weiteren können anhand eines derart großen Datensatzes  
 225 die Auswertung in Hinblick auf verschiedene Parameter optimiert werden. Zum Beispiel ist es auch  
 226 denkbar, dass die Pflanzenpositionen nicht genau bestimmt werden müssen, d.h. der Schritt der



227 Erstellung der Orthomosaik übersprungen werden kann, und die Position der Pflanzen anhand der  
228 Segmentierung in Zusammenhang mit der von der Drohne erfassten Position abgeleitet wird – dies ist  
229 für wissenschaftliche Zwecke zu ungenau, könnte jedoch einen weiteren Schritt in Richtung Praxisnähe  
230 bedeuten, da ein zeit- und kostenaufwändiger Prozessierungsschritt entfallen würde.

231

### 232 **Ausblick**

233 Erste Ergebnisse der Analyse der multispektralen Orthomosaik von zwei Blumenkohlfeldern mit  
234 tausenden erfassten Pflanzen deuten an, dass die multispektralen Reflektanzen die phänologische  
235 Entwicklung der Pflanze und ggf. auftretende Stresse abbilden. Zum jetzigen Zeitpunkt kann keine  
236 Aussage darüber getroffen werden, ob und welche Vegetationsindizes oder Auswertungsmethode die  
237 besten Ergebnisse liefert, da die Erstellung der multispektralen Orthomosaik sehr viel  
238 problembehafteter und zeitaufwändiger war als geschätzt. Zum Projektende liegt nun ein umfassender  
239 Datensatz vor, der Grundlage für weitere geplante Studien sein soll, und in Zukunft auch anderen  
240 Forschungsgruppen zur Verfügung stehen wird (die optimierten Orthomosaik werden im Rahmen der  
241 Projektpublikation von Kierdorf et al. 2022 für Forschungszwecke bereitgestellt).

242

### 243 **Mitwirkung**

244 Die Erhebung und Prozessierung der vorgestellten Datensätze wurde durch Juliane Bendig, Andreas  
245 Burkart, Mike Delaney, Mariele Donoso Olave, Jana Kierdorf, Onno Muller, Uwe Rascher, Ribana  
246 Roscher, Markus Schwarz, Anton Wehr und Jonas Westheider unterstützt.

247

### 248 **Literaturreferenzen**

249 Berni, J.A.J., Zarco-Tejada, P.J., Suárez, L., Fereres, E. (2009): Thermal and narrowbandmultispectral  
250 remote sensing for vegetation monitoring from an unmannedaerial vehicle. *IEEE Trans. Geosci. Remote*  
251 *Sens.* 47, 722–738.

252 Garcia-Ruiz, F., Sankaran, S., Maja, J.M., Lee, W.S., Rasmussen, J., Ehsani, R. (2013): Comparison of  
253 two aerial imaging platforms for identification of huanglongbing-infected citrus trees. *Comput.*  
254 *Electron. Agric.* 91, 106–115.

255 Jones, H.; Vaughan, R. (2010): Chapter 622, Preparation and Manipulation of Optical Data, Image  
256 Correction, Radiometric Correction. In *Remote Sensing of Vegetation: Principles, Techniques, and*  
257 *Applications*; Oxford University Press: Oxford, UK.

258 Kierdorf J, Junker-Frohn LV, Delaney M, Donoso Olave M, Burkart A, Jaenicke H, Muller O, Rascher  
259 U, Roscher R: GrowliFlower (2022): An image time series dataset for GROWth analysis of  
260 cauLIFLOWER. *Journal of Field Robotics*, DOI: 10.1002/rob.22122.

- 261 Poncet, A. M., T. Knappenberger, C. Brodbeck, M. Fogle, J. N. Shaw, and B. V. Ortiz (2019):  
262 Multispectral UAS Data Accuracy for Different Radiometric Calibration Methods. *Remote Sensing*  
263 11.16, 1917.
- 264 Rasmussen, J., Nielsen, J., Garcia-Ruiz, F., Christensen, S., Streibig, J.C. (2013): Potential uses of small  
265 unmanned aircraft systems (UAS) in weed research. *Weed Res.* 53, 242–248.
- 266 Rasmussen, J., Ntakos, G., Nielsen, J., Svensgaard, J., Poulsen, R., Christensen, S. (2016): Are  
267 vegetation indices derived from consumer-grade cameras mounted on UAVs sufficiently reliable for  
268 assessing experimental plots?. *European Journal of Agronomy.* 74. 75-92. 10.1016/j.eja.2015.11.026.
- 269 Streibig, J.C., Rasmussen, J., Andújar, D., et al. (2014): Sensor-based assessment of herbicide effects.  
270 *Weed Res.* 54, 223–233.
- 271 Torres-Sánchez, J., López-Granados, F., De Castro, A.I., Penã-Barragán, J.M. (2013): Configuration  
272 and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. *PLoS*  
273 *One* 8 (3).
- 274 Torres-Sanchez, J., Pena, J. M., de Castro, A. I., and Lopez-Granados, F. (2014). Multi-temporal  
275 mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Comp.*  
276 *Electron. Agric.* 103, 104–113.
- 277 Torres-Sánchez, J., Pena, J.M., de Castro, A.I., López-Granados, F. (2014): Multi-temporal mapping of  
278 the vegetation fraction in early-season wheatfields using images from UAV. *Comput. Electron. Agric.*  
279 103, 104–113.
- 280 Tu, Y., Phinn, S., Johansen, K., Robson, A. (2018): Assessing Radiometric Correction Approaches for  
281 Multi-Spectral UAS Imagery for Horticultural Applications. *Remote Sensing.* 10. 1684.  
282 10.3390/rs10111684.
- 283 Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, Xu B, Yang X, Zhu D, Zhang X, Zhang R, Feng H, Zhao  
284 X, Li Z, Li H, Yang H. (2017): Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop  
285 Phenotyping: Current Status and Perspectives. *Front Plant Sci.* 30 (8) 1111.

## **Anhang: Analyse der erhobenen Daten mittels maschineller Lernverfahren**

Jana Kierdorf

Data Science for Crop Systems, Remote Sensing Group, Institute of Geodesy and Geoinformation

### **Einarbeitung in Methoden und Ableitung der Eignung für die angedachten Anwendungen**

Um das Pflanzen Management der Landwirte zu optimieren und eine effektive Entscheidungsfindung hinsichtlich der Blumenkohlernte zu unterstützen, sind die Landwirte auf eine häufige Überwachung der Pflanzen angewiesen. Dies ist jedoch ein zeitaufwändiger Prozess, der Expertenwissen erfordert. Normalerweise überwachen Landwirte und landwirtschaftliche Berater die Felder regelmäßig durch Stichproben einzelner Pflanzen [6]. Mit Hilfe von Fernerkundungsdaten, bei uns Dronenbildern, ist es möglich diese Überwachung kontinuierlicher durchzuführen und eine großflächige Übersicht zu erlangen ohne die Kulturen zu beschädigen. Dabei hat sich gezeigt, dass maschinelles Lernen (ML) bei der Verarbeitung von großen Mengen an Daten an Bedeutung gewonnen hat [9]. Mit Hilfe von ML ist es z.B. möglich Pflanzenmerkmale aus den Bildern abzuleiten [1, 13].

Die Schätzung und Vorhersage der Erntereife von Blumenkohl ist für Landwirte von großer Bedeutung. Die Blumenkohlernte ist arbeitsintensiv, da jede Pflanze innerhalb von ein bis zwei Wochen geerntet werden muss. Dabei darf der Blumenkohlkopf nicht zu klein oder zu überreif sein, um die Marktstandards zu erfüllen. Außerdem müssen die Pflanzen aufgrund der unterschiedlichen Entwicklungen innerhalb eines Feldes mit der Hand geerntet werden. Da der Blumenkohlkopf mit Blättern bedeckt ist, müssen die Arbeiter die Pflanzen einzeln abtasten um die Gütekriterien zu überprüfen.

Die Schätzung des Entwicklungszustandes von Blumenkohl mit ML stellt insbesondere in den letzten Wochen des Wachstums gewisse Herausforderungen dar. Die Ernte einer Pflanze ist wie bereits erwähnt abhängig von der Größe des Blumenkohlkopfes. Da der Kopf bei den meisten der Pflanzen jedoch durch das Blätterdach verdeckt ist, ist er nicht von außen sichtbar. Dazu kommt, dass äußerlich visuell gleich wachsende Pflanzen nicht zwangsweise einen Kopf mit der selben Größe in sich tragen. Oder anders herum haben Pflanzen mit der gleichen Kopfgröße nicht unbedingt die identische Pflanzengröße. Ein weiterer Aspekt ist die starke Abhängigkeit vom Klima, sodass Felder die zu unterschiedlichen Zeiten gepflanzt wurden ggf. zur selben Zeit erntereif sind und so die Entwicklungsdauer eines Feldes bzw. einer Pflanze vom Pflanzzeitpunkt sowie Klima abhängt.

Das Ziel dieses Arbeitspakets ist die Schätzung phänotypischer Merkmale des Blumenkohls zur Ableitung des Entwicklungszustandes aus den erhobenen Daten von Arbeitspaket 1 (AP1) und Arbeitspaket 2 (AP2). Dazu wurden zu Beginn Grundlagen, die allgemeine Thematik sowie verwandte Arbeiten aufgearbeitet. Für die methodische Umsetzung wurde sich im speziellen auf Neuronale Netze und Deep Learning Verfahren fokussiert.

### **Detektion von Kohlpflanzen im Feld**

Um Aussagen über den Entwicklungszustand von einzelnen Blumenkohlpflanzen zu treffen, muss die Position der Pflanzen bekannt sein um u.A. eine zeitliche Entwicklung der Pflanzen analysieren zu können. Die Koordinate einer Pflanze wird mit Hilfe des maschinellen Lernverfahrens Mask R-CNN [3] bestimmt, welches der Instanz Segmentierung dient. Das Ziel des Verfahrens ist die Bestimmung bzw. Erstellung einer Maske sowie Bounding Box eines Objektes, in unserem Fall einer bzw. mehrerer Pflanzen. Die Maske sowie Bounding Box ermöglichen eine Ableitung phänotypischer Merkmale wie den Durchmesser oder die projizierte Blattfläche der Pflanze. Über den Mittelpunkt der Bounding Box kann ein ungefährer Pflanzenmittelpunkt bestimmt werden, welcher für die Lokalisierung der Pflanzen genutzt wird.

Erste Versuche der Pflanzen Detektion wurden auf externen Pflanzen-Datensätzen wie dem *LSC Datensatz* (<https://plant-phenotyping.org/CVPPP2017-challenge>) und dem *Brassica Datensatz* (<https://cloudstor.aarnet.edu.au/plus/s/N0giZD5QTMB32Xj>) durchgeführt. Die beiden Datensätze eigneten sich für erste Tests, da zu Beginn des Projekts noch keine projektinternen Daten vorhanden waren. Ein weiterer positiver Aspekt waren die vorhandenen Annotationen (Label) der Daten, die für maschinelle Lernverfahren wie dem Mask R-CNN von Bedeutung sind. Beispielhafte Ergebnisse aus dem LSC Datensatz sind in Abb. 1 dargestellt.

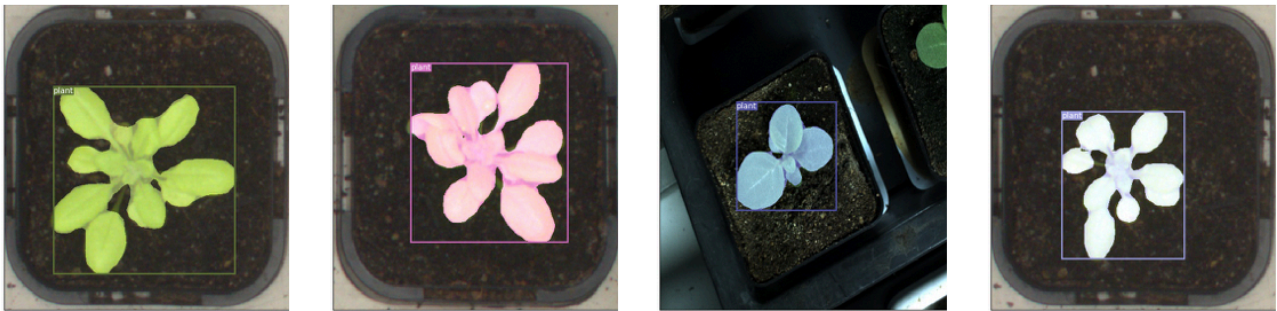


Abbildung 1: Beispielhafte Ergebnisse der Pflanzendetektion basierend auf dem LSC Datensatz. Farblich hervorgehoben wird die vorhergesagte Maske der Pflanze und die Bounding Box (Farbe hat keine Bedeutung).

Für die Anwendung des Mask R-CNN auf projektinterne Daten wurden zunächst die aufgenommenen Daten durch studentische Hilfskräfte annotiert. Die Annotationen bestehen pro aufgenommenem Bild (Abb. 2, Spalte 1) aus einer Maske, die auf Pixelbasis angibt, ob das Pixel dem Boden oder einem Teil einer Pflanze zugeordnet wird (Abb. 2, Spalte 2), dabei wird direkt in verschiedene Pflanzen Instanzen unterschieden. Zusätzlich werden die unterschiedlichen Blatinstanzen pixelgenau vermerkt und den entsprechenden Pflanzen zugeordnet (Abb. 2, Spalte 3). Zuletzt wurden noch Pflanzenteile annotiert, die zwar auf dem Bildausschnitt zu sehen sind, aber der daraus geschätzte Mittelpunkt durch den geringen Pflanzenanteil zu weit vom korrekten Pflanzenmittelpunkt abweicht.

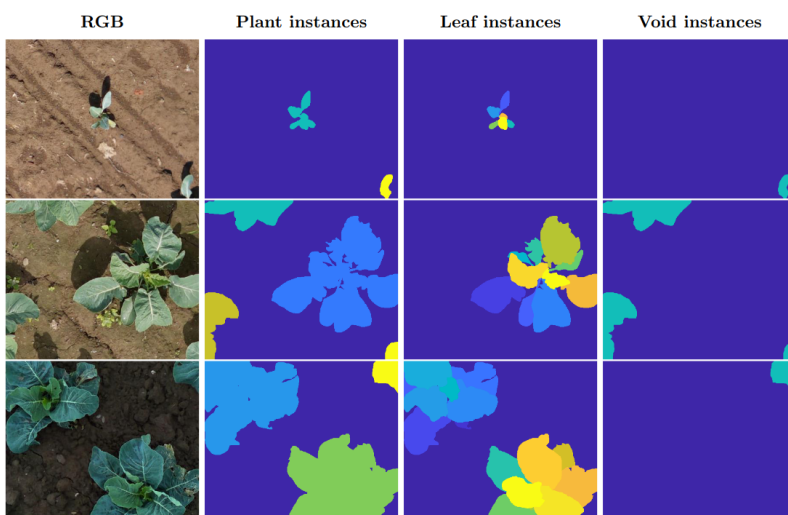


Abbildung 2: Beispielhafte Annotation eines aus dem in AP1 erstellten Orthophotos extrahierten Bildes. Die großflächige dunkelblaue Fläche entspricht hierbei den Bodenpixeln. Weitere Farben stellen Pflanzen- oder Blatinstanzen dar.

Nach erfolgreicher Datenprozessierung wurde die Methodik des Mask R-CNN auf unsere Daten angewendet und so die Koordinaten von jährlich mehr als 8000 Pflanzen bestimmt. Abb. 3 zeigt einen Ausschnitt des Feldes in dem automatisiert die einzelnen Pflanzen erkannt wurden. Es ist zu erkennen, dass es auch fehlerhafte Annotationen gibt, die in einem post-processing Schritt korrigiert wurden.



Abbildung 3: Beispielhafte Ergebnisse der Pflanzendetektion basierend auf den in AP1 aufgenommenen Daten. Farblich hervorgehoben wird die vorhergesagte Maske der Pflanzen und die Bounding Box (Farbe hat keine Bedeutung).

Die Experimente der Detektion von Pflanzen wurde ebenfalls auf Blattinstanzen durchgeführt um die Möglichkeit zu gewährleisten zusätzliche phänotypische Merkmale, wie z.B. die Anzahl Blätter pro Pflanze zu bestimmen. Ergebnisse hierzu sind in Abb. 4 zu finden. Zur visuellen Vereinfachung werden die Bounding Boxes in der Abbildung nicht dargestellt.



Abbildung 4: Ergebnisse der Blatt Instanz Segmentierung mit Mask R-CNN. Die verschiedenen Farben stellen unterschiedliche Instanzen dar.

Die Georeferenzierung der Daten aus AP1 ermöglicht es die Koordinaten der Pflanzen für jeden Aufnahmezeitpunkt zu nutzen um Bilder der Einzelpflanzen zu extrahieren. Daraus ergeben sich Bildzeitreihen der Einzelpflanzen, die für spätere Aufgaben dieses Arbeitspakets genutzt werden. Beispielsweise können Berechnungen über den gesamten Wachstumszeitraum durchgeführt werden. Dies ermöglicht die Bestimmung des Pflanzenwachstums, was für die Vorhersage des Erntezeitfensters genutzt werden kann. Beispiele für vier Zeitreihen sind in Abb. 5 dargestellt. Dabei zeigt eine Reihe eine Pflanze und jede Spalte den entsprechenden Aufnahmezeitpunkt. Es ist gut zu erkennen, dass gleichzeitig gepflanzte Pflanzen, die identischen klimatischen Bedingungen ausgesetzt waren, unterschiedliche Entwicklungen vollziehen können. Dies stellt insbesondere die Schätzung von Wuchs und Ausreifung vor Herausforderungen.



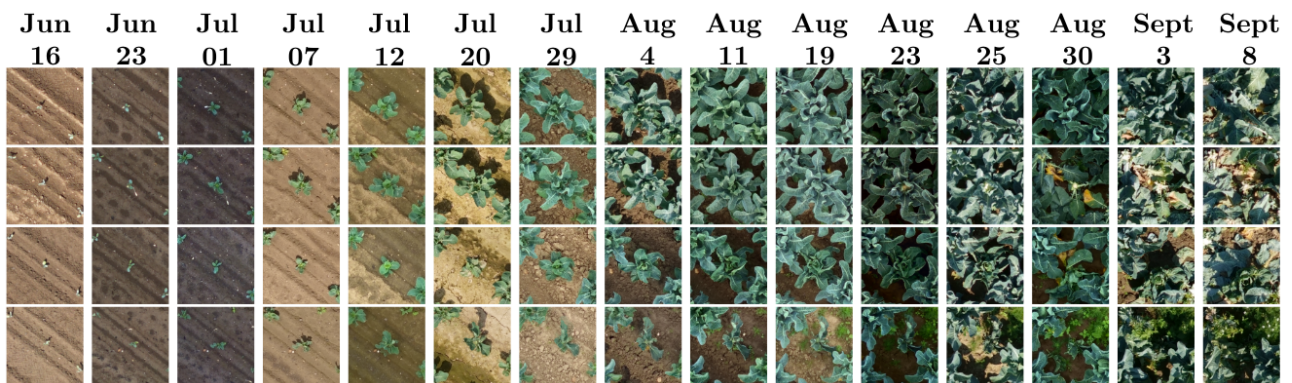


Abbildung 5: Zeitreihen für vier beispielhafte Pflanzen aus dem Jahre 2021. Eine Reihe stellt eine Pflanze dar während die verschiedenen Spalten den Aufnahmezeitpunkt darstellen.

Als unterstützende Arbeit wurden innerhalb einer Bachelorarbeit [5] die Einflüsse unterschiedlicher Bildauflösungen analysiert, um Rückschlüsse für eine optimierte Befliegung, wie zum Beispiel einer geeignete Flughöhe für AP1, zu ziehen.

### Schätzung von Wuchs und Ausreifung

Ein wichtiger Bestandteil der Schätzung des Entwicklungszustandes ist die Analyse des Wuchs und der Ausreifung der Blumenkohlpflanzen. Dazu wurden in AP2 in-situ Messungen im Feld durchgeführt. Diese wurden plotweise erhoben und die Pflanzen innerhalb der Plots werden als Referenzpflanzen bezeichnet. Der Durchmesser der Referenzpflanzen wurde jeweils entlang der breitesten Ausrichtung der Pflanze gemessen. Da eine automatisierte Analyse angestrebt wird, werden die resultierenden Masken und Bounding Boxen des Mask R-CNN für Analysen des Wuchs und der Ausreifung verwendet. Die große Halbachse der Bounding Box spiegelt dabei den geschätzten Durchmesser wieder. Für die Abschätzung der Genauigkeit der abgeleiteten Größen werden die durch die Bounding Box prädizierten Größen mit den in-situ Messungen verglichen. Abb. 6 stellt für den Aufnahmezeitpunkt am 25.08.2020 die gemessenen Werte im Feld den geschätzten Werten aus den Bildern gegenüber. Die x-Achse gibt dabei die entsprechende Pflanzen ID an. Die abgebildeten Differenzen entstehen dadurch, dass die große Halbachse der Bounding Boxen nicht exakt mit der im Feld gemessenen Achse übereinstimmt. Dennoch wird eine gute Abschätzung der Pflanzengröße mittels dieses Verfahrens erreicht.



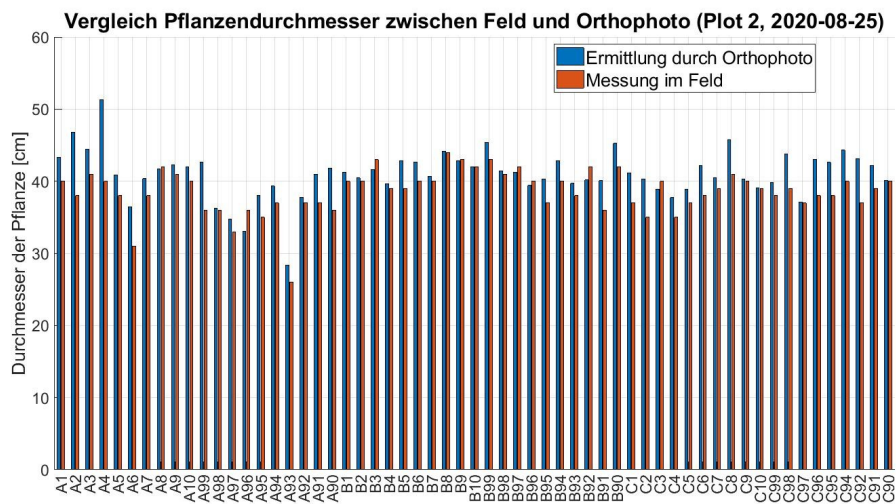


Abbildung 6: Beispielhafte Gegenüberstellung von im Feld gemessenem Pflanzendurchmesser (rot) und aus den Bildern abgeleitetem Durchmesser (blau).

In einer Masterarbeit [2] wurde sich mit der direkten Größenschätzung der Blumenkohlkopfes aus Bildern beschäftigt ohne den Zwischenschritt der zu berechneten Masken und Bounding Boxen. Dabei wurden maschinelle Lernverfahren der Klassifikation denen der Regression gegenüber gestellt. Es hat sich herausgestellt, dass es eine sehr herausfordernde Aufgabe ist, da der Blumenkohlkopf unter dem Blätterdach nicht zu sehen ist.

Um im weiteren Verlauf des Projekts eine gute Abschätzung des Erntezeitpunkts liefern zu können, ist das Wissen über die Größe der Blumenkohlblüte einer Pflanze von Bedeutung. Hierbei treten die Schwierigkeiten auf, dass beim Großteil der Pflanzen die Blüte durch Blätter verdeckt wird (siehe Abb. 8, links). Daher wurde sich mit der Verdeckung interessanter Bereiche in Bildern beschäftigt. Damit wir uns mit dieser Thematik auseinandersetzen konnten wurden nach Absprachen mit den anderen APs Entblätterungsversuche geplant und gemeinsam durchgeführt. Diese ermöglichen eine visuelle Analyse zwischen äußerem Erscheinungsbild der Pflanzen und der inneren Blumenkohlblüte. Im Laufe der Projektzeit wurde die Entblätterung der Pflanzen hinsichtlich Menge und Verteilung im Feld optimiert. Ein Beispiel eines entblätterten Plots vor und nach der Entblätterung ist in Abb. 7 zu finden. Wir haben uns für plotweises Entblättern entschieden, damit möglichst wenige benachbarte Pflanzen von der Entblätterung in ihrem Wachstum beeinflusst werden. Abb. 8 (links und mittig) zeigt einen direkten Vergleich für eine Einzelpflanze.

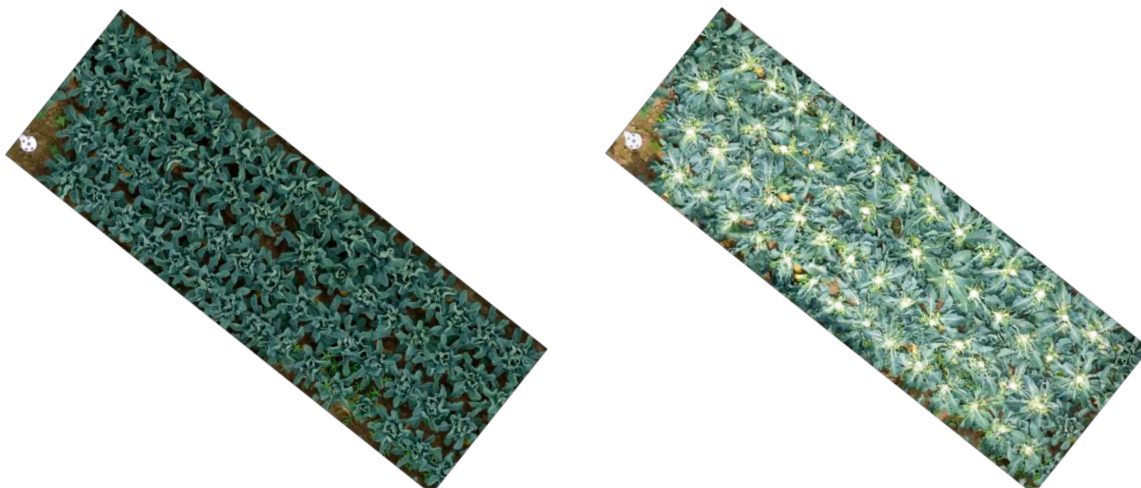


Abbildung 7: Darstellung eines Plots vor (links) und nach (rechts) der Entblätterung zur Sichtbarmachung der inneren Blüte.

Gelöst wurde die Fragestellung der Verdeckung mit machinellen Lernverfahren wie dem CycleGAN [14] oder Conditional Generative Adversarial Network Pix2Pix [6]. Es wurde sich in relevante Literatur eingearbeitet und erste Tests an einem externen Weintrauben-Datensatz durchgeführt [8] bevor die Projektdaten für Analysen bereit standen. Auch hier tritt die Problemstellung der Verdeckung der Beere, die den Ertrag widerspiegelt, auf. Die Ergebnisse sehen vielversprechend aus und die entwickelten Programme und Methodiken wurden in Folge dessen auf Blumenkohldaten angewendet.



Abbildung 8: Darstellung einer möglichen Visualisierung der in den Bildern nicht sichtbaren Blumenkohlblüte mittels eines Conditional Generative Adversarial Networks (cGAN) [14]. Links: Input-Bild der Pflanze mit verdeckter Blüte. Mitte: Im Feld aufgenommene entblätterte Pflanze mit sichtbarer Blüte. Rechts: Ergebnis des cGAN Modells mit möglicher visueller Darstellung der inneren Blüte basierend auf dem Input-Bild mit verdeckter Blüte (links).

Maschinelle Lernmethoden, wie das Pix2Pix [6], ermöglichen es ein Abbild zu lernen, wie die verdeckte Blumenkohlblüte in Abhängigkeit der sichtbaren Informationen im Bild aussehen könnte (Abb. 8, rechts). Die Bilder in Abb. 8 zeigen ein beispielhaftes Ergebnis dieser Methodik angewandt auf die OPTIKO Blumenkohl Daten aus dem Jahr 2020. Das linke Bild zeigt eine Pflanze, bei der die Blüte nicht sichtbar ist. Das mittlere Bild repräsentiert die entblätterte Blüte auf dem Feld. Das rechte Bild zeigt ein Output des Modells wobei zu beachten ist, dass das Modell aufgrund zu weniger vorhandener Daten an die vorliegenden Daten über angepasst wurde. Die Beispielbilder wurden bereits vom Modell gesehen, wodurch sie nicht repräsentativ sind sondern ausschließlich zeigen, was mit der Anwendung erreicht werden soll. Finale Ergebnisse konnten im Laufe des Projekts nicht erzeugt werden. Dies bleibt eine offene Fragestellung, die in Zukunft außerhalb des Projekts bearbeitet werden könnte.

Ein Vorteil bei der Verwendung von Fernerkundungsdaten ist die große Datenmenge, die aufgenommen wird. Für den Großteil der Pflanzen liegen uns jedoch keine in-situ Daten über das Wachstum oder die Reife des Blumenkohls vor. Dennoch ist auch für solche Pflanzen von Interesse etwas über das Wachstum herauszufinden. Mit Hilfe von halb-überwachter Lernmethoden wurde die große Anzahl nicht annotierter Daten mit der geringen Menge annotierter Daten kombiniert. Erste Tests haben gezeigt, dass Zeitreihen mit ähnlicher Entwicklung mit Hilfe des in [11] vorgestellten Siamese-Triplet-Netzwerk Modells gruppiert werden können. Abb. 9 zeigt eine Gruppe von Pflanzen, die in der frühen Entwicklungsphase abgestorben sind. Dabei geben die Reihen verschiedene Zeitreihen und die Spalten verschiedene Zeitpunkte an. Diese Gruppe liefert somit Informationen über die Anzahl an Pflanzen, die den Status der Erntereife nicht erreichen werden und so dem Landwirt nicht zum Verkauf bereit stehen.



Abbildung 9: Beispielhafte Gruppierung der Zeitreihen. Dargestellt sind Zeitreihen bei denen die Pflanzen im frühen Wachstumsstadium abgestorben sind.

Weitere Analysen wurden hinsichtlich vorhandener Stressfaktoren durchgeführt. Dabei wurde sich bei der Stressdetektion, vor allem in Hinsicht auf Düngung und Bewässerung auseinandergesetzt. Hier wurde sich ebenfalls auf ML Verfahren, speziell auf unüberwachte Verfahren, fokussiert. Da unser Projektpartner Herr Schwarz ein sehr gutes Stressmanagement hat liegt nur eine geringe Datenmenge an biotischen und abiotischen Stressen vor. Aufgrund der geringen Datengrundlage wurde beschlossen sich nicht weiter mit dem Thema Stress auseinander zu setzen und sich auf die anderen Aufgaben zu fokussieren.

### **Vorhersage des Erntezeitfensters**

Für die Vorhersage des Erntezeitfensters sind Referenzdaten auf dem Feld erforderlich. Diese wurden plotweise für die Referenzpflanzen erhoben. Dabei zeigen Abb. 10 und 11 die Menge an geernteten Referenzpflanzen pro Referenzplot für beide bisher aufgenommenen Felder. Die x-Achse gibt dabei den Tag der Ernte und die y-Achse die Anzahl der jeweils geernteten Pflanzen pro Zeitpunkt pro Plot an. Für das in 2020 aufgenommene Feld ist zu erkennen, dass entweder nicht bei jedem Erntegang unsere Plots betrachtet wurden oder die entsprechenden Pflanzen zu dem entsprechenden Zeitpunkt nicht erntereif waren. Für die in 2021 aufgenommenen Daten liegt eine bessere Verteilung unserer Plots im Feld vor, sodass wir zu jedem Zeitpunkt Daten erfassen konnten, die in unsere maschinellen Modelle integriert werden können. Diese Erkenntnisse haben wir auch in die Datenaufnahme aus 2022 einfließen lassen.

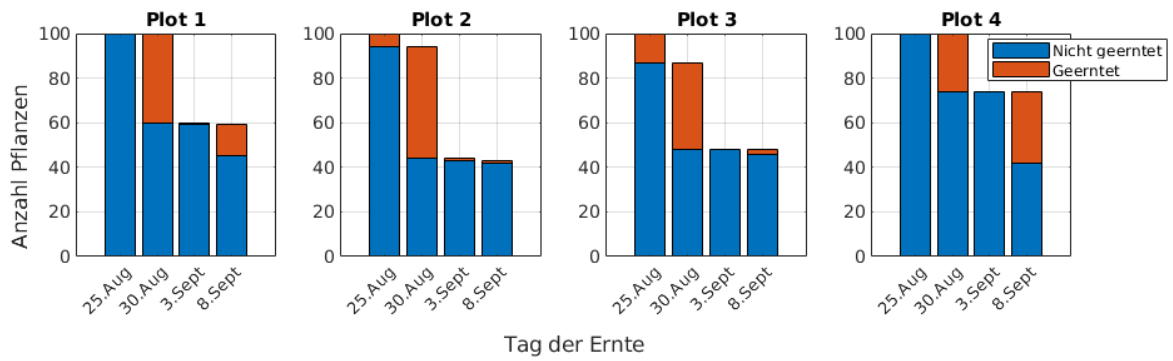


Abbildung 10: Überblick über geerntete und nicht geerntete Pflanzen pro Referenzplot pro Tag für die in 2020 aufgenommenen Daten.

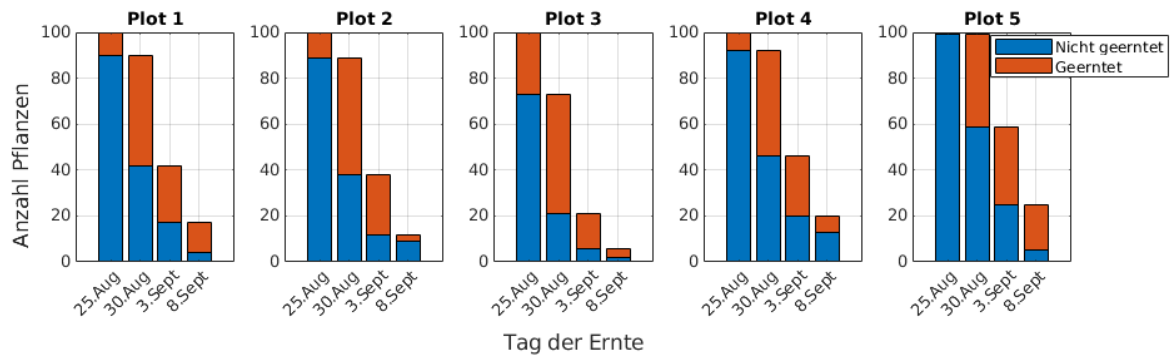


Abbildung 11: Überblick über geerntete und nicht geerntete Pflanzen pro Referenzplot pro Tag für die in 2021 aufgenommenen Daten.

Für eine präzise Vorhersage des Erntezeitfensters ist das Wachstum der Pflanzen über die Zeit von Bedeutung. Insbesondere, da die Variation im Wachstum der verschiedenen Pflanzen hoch ist. Abb. 12 zeigt die Entwicklung des mittleren Pflanzendurchmessers und Kopfdurchmessers. Alle Referenzplots zeigen eine ähnliche Entwicklung im Mittel. Variationen zwischen den Einzelpflanzen sind in den Rohdaten jedoch deutlich zu erkennen. Diese Analysen helfen uns ebenfalls dabei die Blütengröße, welche den Erntezeitpunkt bestimmt, mit der gesamten Pflanze in Relation zu setzen. Das Sinken der Kurven zu den letzten Aufnahmezeitpunkten kann damit erklärt werden, dass große Blüten bereits geerntet wurden und nur die noch nicht so weit entwickelten Pflanzen in die Messungen mit einbezogen werden.

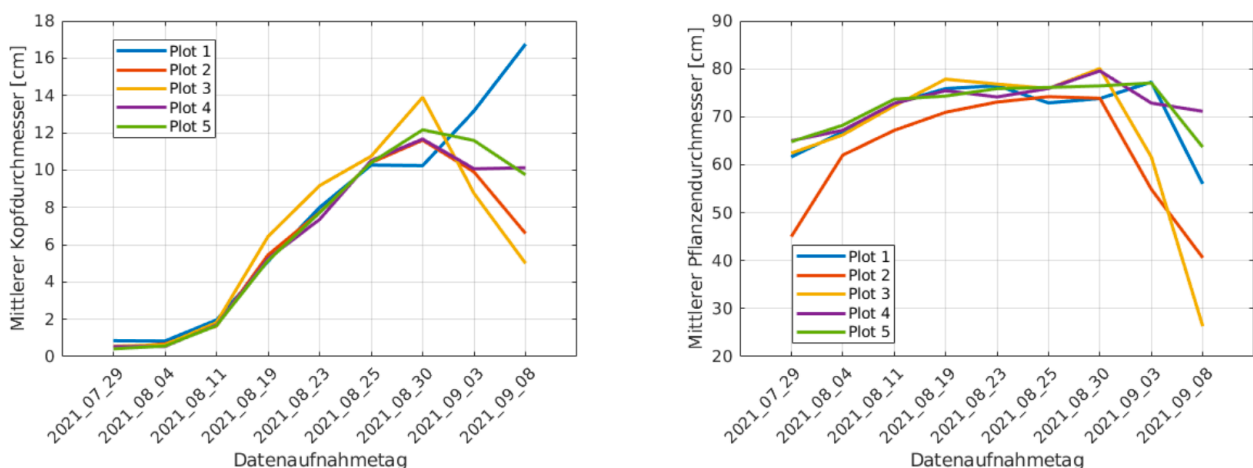


Abbildung 12: Entwicklung des mittleren Pflanzendurchmessers (rechts) und Blütendurchmessers (links) für die Referenzpflanzen des im Jahr 2021 aufgenommenen Feldes.



Für die Vorhersage des Erntezeitpunktes wurden verschiedene Ansätze verfolgt. Der erste Ansatz ist die Anwendung von Verfahren des self-supervised learnings. Der Hauptgrund ist unsere Datengrundlage, die aus wenigen gelabelten Daten (Daten mit Annotationen wie Wachstumsparameter oder Erntezeitpunkt) und vielen ungelabelten Daten (Daten mit keinen Annotationen) besteht. Verfahren des self-supervised learnings kombinieren beide Arten von Daten, sodass die Datengrundlage bestmöglich genutzt wird. In einem Pretraining Schritt wird zunächst ein Model auf den ungelabelten Daten trainiert. Als Optimierungskriterium werden dabei Informationen aus den Daten verwendet, die direkt beim Training des Models erzeugt werden oder bekannt sind, wie z.B. den Aufnahmezeitpunkt des Bildes. Das Pretraining sorgt dafür, dass das Model die Daten einmal sieht und sich an sie anpasst. Bei dem sogenannten Downstream Task wird das Model dann basierend auf den gelabelten Daten hinsichtlich einer gegebenen Aufgabe optimiert. Das kann z.B. das Clustering von Eigenschaften innerhalb der Daten sein oder die Klassifikation des Erntezeitpunktes. Mit dem self-supervised learning knüpfen wir an vorherige Analysen zu Wuchs und Ausreifung an.

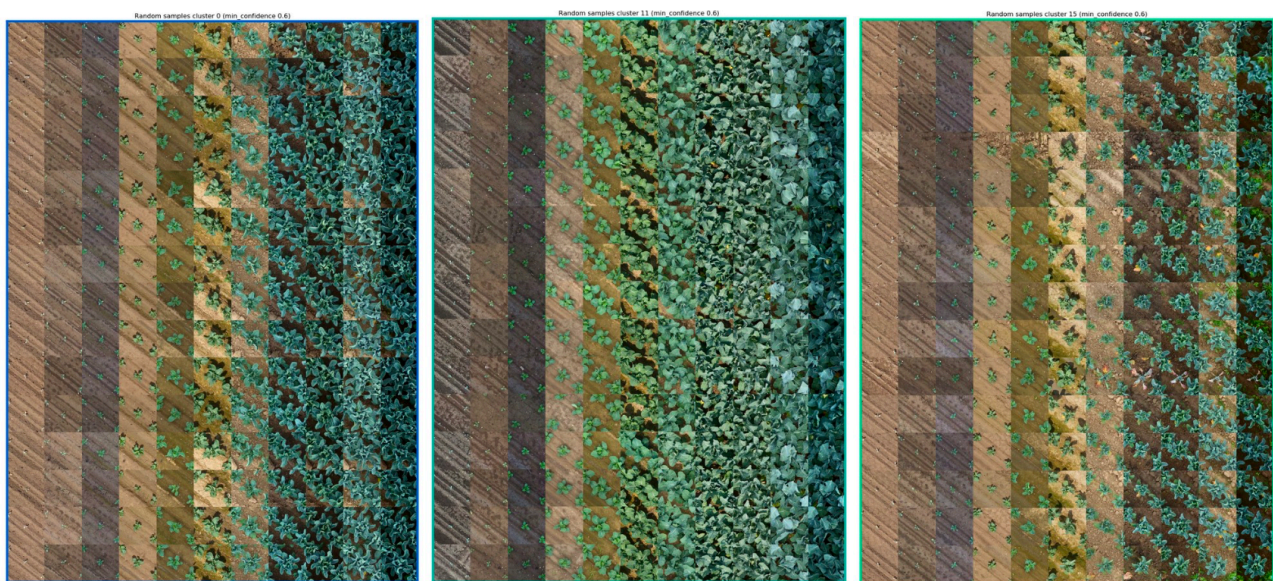


Abbildung 13: Darstellung von drei exemplarischen Clustern, welche verschiedene Entwicklungen und somit unterschiedliche Wachstumseigenschaften darstellen. Eine Reihe repräsentiert eine Zeitreihe und eine Spalte einen Aufnahmezeitpunkt innerhalb der Zeitreihe. Pflanzen innerhalb eines Clusters entwickeln sich ähnlich.

Das bereits vorgestellte Siamese-Triplet-Netzwerk Model [11] wurde um ein Positional Encoding erweitert um unterschiedliche Pflanzzeitpunkte sowie variierende Intervalle zwischen der Datenaufnahme zu integrieren. So ist es gelungen Pflanzen, die durch ihr Alter nicht so weit entwickelt, bzw. kleiner sind, von anderen Pflanzen zu separieren, die aufgrund Wachstumsschwierigkeiten eine geringere Größe aufweisen. Außerdem vereinfacht es die Anwendung des Models auf Daten, die andere Intervalle zwischen den verschiedenen Aufnahmezeitpunkten vorweisen, da so die Entwicklung der Pflanze mit dem bestimmten Alter in Verbindung gesetzt wird. Der Transfer zwischen verschiedenen Feldern wird so vereinfacht. So konnten die entwickelten Methodiken auf alle Jahre angewendet werden. Der Vorteil der Daten aus 2021 und 2022 ist die Vollständigkeit der Zeitreihen, speziell im Hinblick auf die Zeit in der die Blumenkohlköpfe anfangen zu wachsen. Diese Zeitpunkte sind in den Daten aus 2020 nicht vorhanden, in den Daten aus 2021 und 2022 jedoch schon. Wie Abb. 13 zeigt, sind eindeutig verschiedene Entwicklungen in den Zeitreihen zu erkennen. Das mittlere Cluster zeigt Pflanzen, die sich im Gegensatz zu den Pflanzen aus den anderen beiden Clustern schneller entwickeln. Analysen haben gezeigt, dass die Gruppierungen noch nicht mit Erntezeitfenstern in Korrelation gesetzt werden können. Dafür sind weitere Experimente notwendig.

Ein weiterer Ansatz ist die Klassifikation des Erntezeitpunktes. Dafür haben wir den zuvor beschriebenen Ansatz zunächst erweitert um anstelle von Clustern Klassen vorherzusagen, die mit den verschiedenen Erntezeitfenstern übereinstimmen. Dafür haben wir die Daten aus dem Jahre 2021 genutzt, da auch hier der Vorteil der Vollständigkeit gegenüber der Daten aus 2020 besteht. Für diese Daten gab es vier definierte Erntezeitpunkte (= 4 Klassen), eine Klasse mit noch nicht geernteten Pflanzen sowie eine Klasse, die während des Wachstums abgestorbenen Pflanzen beinhaltet. Analysen haben gezeigt, dass die ausschließliche Verwendung der vier Erntezeitpunkte als Klassen zu einer höheren Genauigkeit in der Vorhersage des Erntezeitfensters führen als die zusätzliche Verwendung der anderen beiden Klassen, siehe Abb. 14. Außerdem wurde unser Ansatz inklusive des Pretrainings (orange) dem Standardansatz ohne datenspezifischem Pretraining (rot) gegenübergestellt. Auch hier ist zu erkennen, dass unser Ansatz inklusive datenspezifischem Pretraining zu besseren Ergebnissen führt.

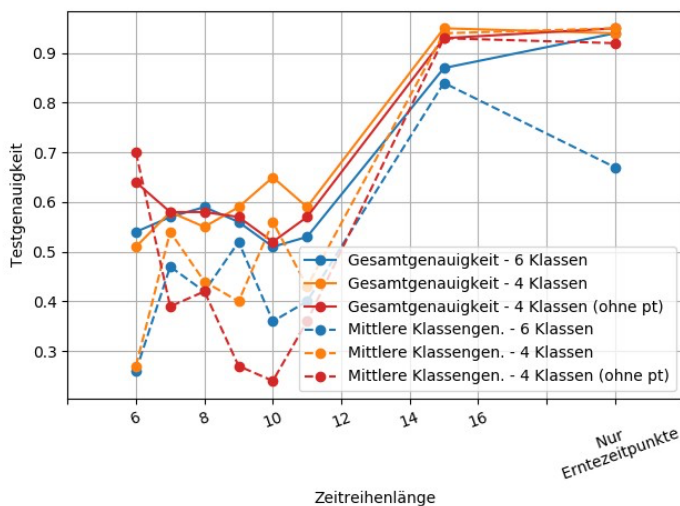


Abbildung 14: Gegenüberstellung der Klassifikationsergebnisse für (blau) 6 und (orange) 4 Klassen inklusive pretraining und (rot) 4 Klassen ohne pretraining. Die durchgezogene Linien zeigen die Gesamtgenauigkeit und die gestrichelte Linie die mittlere Klassengenauigkeit.

Um die Ergebnisse der Klassifikation zu verbessern wurde die Klassifikation der Erntezeitpunkte auch auf Einzelzeitpunkte durchgeführt und mittels Sensitivitätskarten hinsichtlich markanter Merkmale analysiert. Explizit wurde das Model ResNet18 [4] angewendet. Als Datengrundlage dienen die aufgenommenen Referenzdaten aus 2021 ab der Woche vor Beginn der Ernte bis Ende der Datenaufnahmepriode (2021\_08\_23, 2021\_08\_25, 2021\_08\_30, 2021\_09\_03). Die Daten wurden in die folgenden Klassen aufgeteilt: 0 - Noch nicht bereit für die Ernte innerhalb der nächsten drei Tage, 1 - Bereit für die Ernte innerhalb der nächsten 3 Tage. Bilder von Pflanzen, die zu einem Zeitpunkt bereits geerntet wurden, wurden aus der Datenmenge raus genommen um eine höhere Genauigkeit des Modells zu gewährleisten und den Bias zu verringern. Unser trainiertes Model zeigt eine Gesamtgenauigkeit von 74.23% und mittlere Klassengenauigkeit von 75.19% auf Testdaten. Dies bedeutet, dass wir bei 3 von 4 Pflanzen korrekt vorhersagen können, ob sie innerhalb von 3 Tagen erntereif sind.

Für die Verbesserung und Analyse der Ergebnisse wurden Sensitivitätsanalysen durchgeführt. Anwendung fanden Methoden wie SHAP values [10], Occlusion Sensitivity Maps (OSM) [12] und Grad-CAM [12]. Sensitivitätsanalysen werden durchgeführt um zu analysieren anhand welcher Merkmale in den Daten ein Model seine Entscheidung trifft um so Rückschlüsse darauf zu ziehen, welche Merkmale für eine bestimmte Klasse von Bedeutung sind. Nachdem uns der Bauer das Vorwissen mitgeteilt hat, dass keine äußere Merkmale der Pflanze dabei helfen den Erntestatus der Pflanze abzuleiten haben uns Versuche gezeigt, dass ein markantes Merkmal das Blätterdach im Zentrum der Pflanze um den Kopf herum ist. 15 zeigt beispielhafte Ergebnisse für Grad-CAM. Rote Pixel weisen dabei auf Regionen hin, die für eine spezielle Klasse sprechen. Dunkelblaue



Pixel werden im Netzwerk nicht aktiviert und haben somit keinen Einfluss auf das Klassifikationsergebnis. Unter Betrachtung der Ergebnisse in Abb. 15 fällt auf, dass rote Hotspots genau dort auftreten wo das Zentrum der Pflanze ist. Anhand der Ergebnisse wird analysiert, ob die Ergebnisse des Modells vertrauenswürdig sind und ob man die Datenaufbereitung in eine Richtung leiten kann, die die gelernten Modelle verbessert.

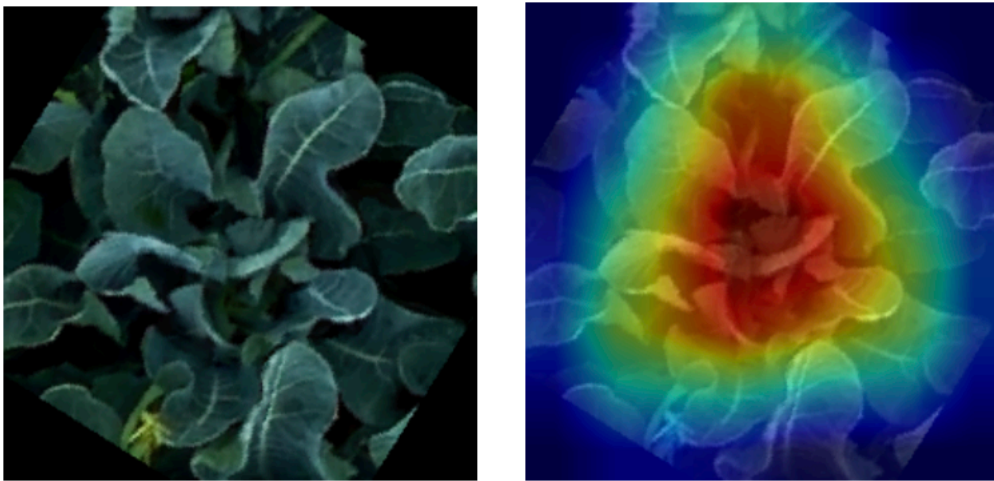


Abbildung 15: Ergebnisse der Sensitivitätsanalyse mittels Grad-CAM [12]. Links wird das originale Bild dargestellt. Rechts eine Kombination aus Sensitivität und Bild. Rote Regionen heben Regionen im Bild hervor, anhand dessen das Klassifikationsmodell seine Entscheidung trifft.

### Quantifizierung der Heterogenität für Einzelzeitpunkte und Zeitreihen

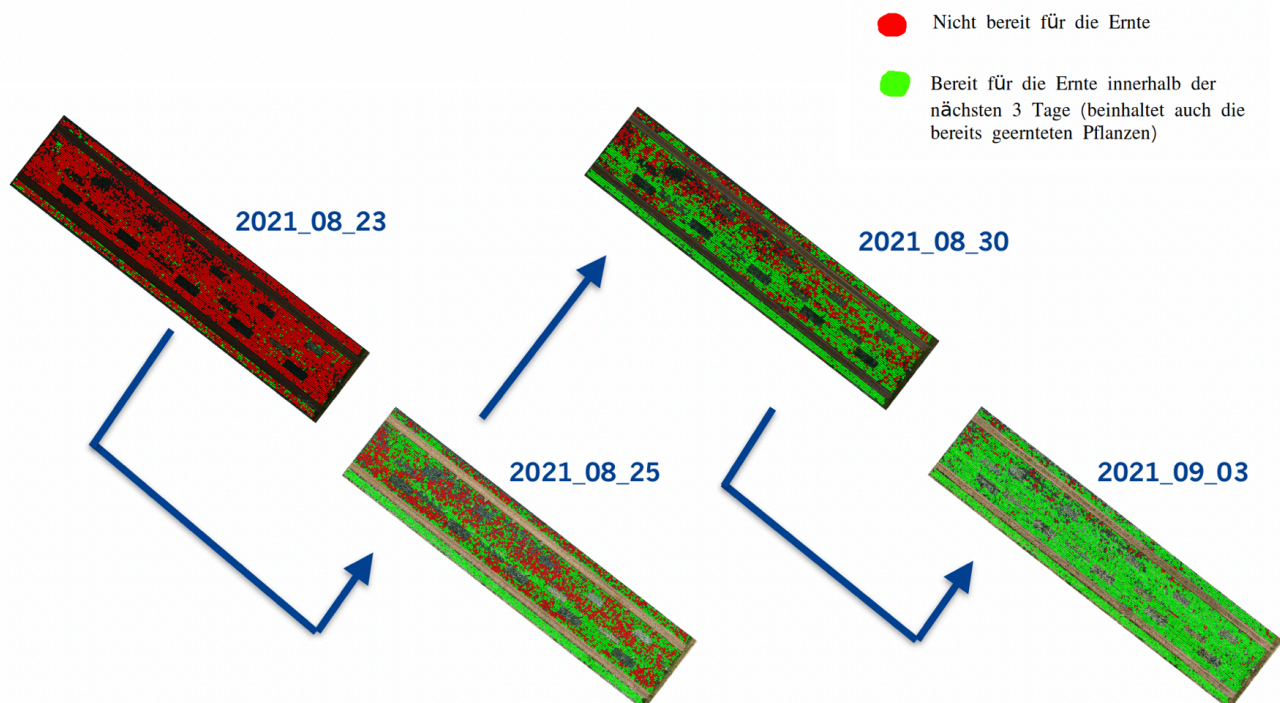


Abbildung 16: Visuelle Darstellung der Vorhersage des Erntestatus. Abgebildet werden die vorhergesagten Klassen (rot) nicht bereit für die Ernte und (grün) bereit für die Ernte innerhalb der nächsten drei Tage.

Für die geforderte Quantifizierung der Heterogenität im Feld wird das gelernte Model zur Klassifizierung des Erntestatus auf die Pflanzen im Feld angewendet, für die keine im Feld aufgenommene Referenzen bzgl. Erntezeitpunkt vorhanden sind. Abb. 16 zeigt den geschätzten Erntestatus der Pflanzen zu den vier verschiedenen Zeitpunkten. Die Pflanzen wurden je nach Vorhersage in grün (Klasse 1) oder rot (Klasse 0) eingefärbt. Evident wird, dass angefangen vom 2021\_08\_23 bis hin zum 2021\_09\_03 der Anteil erntereifer Pflanzen zunimmt (visualisiert durch zunehmende grün markierte Pflanzen). Der Erntegang der Arbeiter auf dem Feld verlief von süd-westlicher Ausrichtung des Feldes in Richtung nord-östlicher Ausrichtung des Feldes. Am 2021\_08\_23 ist zu erkennen, dass im süd-westlichen Streifen unterhalb der Traktorspur vermehrt grüne Markierungen auftreten. In Realität sind auch hier die Arbeiter am ersten Erntetag lang gegangen um die Ernte zu starten, haben die Ernte aber aufgrund zu weniger erntereifer Pflanzen abgebrochen. Aus subjektiver Sicht (nicht qualitativ bewertbar aufgrund fehlender Referenzen) zeigt der vorhergesagte Erntestatus der Pflanzen ähnliche Ergebnisse an wie sie auch auf dem Feld aufgetreten sind. Es sieht vielversprechend aus, dass auch für diese Ergebnisse eine Rate von 75% für korrekte Vorhersagen zutrifft. Zu beachten ist, dass in Abb. 16 die verschiedenen Zeitpunkte unabhängig voneinander vorhergesagt werden ohne eine Vorhersage der selben Pflanze zu einem früheren Zeitpunkt zu berücksichtigen. Basierend auf den Ergebnissen kann dem Bauern eine Übersicht zur Hand gegeben werden, anhand dessen er beurteilen kann wie viele Arbeiter er wann auf welche Bereiche im Feld zur Ernte aussendet.

## Zusätzlich Arbeiten

Unsere Daten wurden in einem Paper „GrowliFlower: An image time-series dataset for GROWth analysis of cauLIFLOWER“ [7] zusammengefasst und öffentlich zugänglich bereit gestellt (<https://phenoroam.phenorob.de/geonetwork/srv/eng/catalog.search#/metadata/cb328232-31f5-4b84-a929-8e1ee551d66a>). Dafür haben wir die Daten in 4 Kategorien unterteilt.

*GrowliFlowerR*: Zeitreihen mit in-situ Daten, die auf dem Feld für eine bestimmte Anzahl an Pflanzen aufgenommen wurde.

*GrowliFlowerD*: Bildpaare von Pflanzen vor und nach der Entblätterung des Blumenkohlkopfes mit dazugehörigen Zeitreihen.

*GrowliFlowerT*: Alle anderen Pflanzen Zeitreihen, die nicht Teil der Datensätze *GrowliFlowerR* oder *GrowliFlowerD* sind.

*GrowliFlowerL*: Annotierte Ausschnitte der Orthophotos, die für Aufgaben des maschinellen Lernens wie Klassifikation, Segmentierung, Detektion oder Instanz Segmentierung genutzt werden können.

Wir haben bereits jetzt die Erfahrung gemacht, dass die Interesse an dem Datensatz groß ist. Die Daten wurden vermehrt heruntergeladen und das Paper wurde bereits zitiert. Im Laufe des Projekts haben uns vier studentische Hilfskräfte zugearbeitet, unter Anderem bei der Datenaufnahme, -extraktion und -annotation.

## Fazit

Als Fazit zu dem Projekt kann gesagt werden, dass wir viele Einsichten in die Thematik erlangt und erste Ergebnisse für die Vorhersage des Erntestatus erreicht haben. Um Modelle zu entwickeln, die auch auf andere Felder und andere Jahreszeiten anwendbar sind fehlen jedoch noch einige Datensätze um mehr Variabilität abzudecken die für eine gute Transferbarkeit notwendig ist.

## References

- [1] Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., and Notarnicola, C. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12):16398–16421.
- [2] Askar, F. (2022) „Growth Analysis of Cauliflower Plants using Machine Learning Methods”, (Masterarbeit)
- [3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
- [4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [5] Hembach, J.M. (2021) „Ableitung phänotypischer Merkmale im Blumenkohl mittels bildbasierter Instanzsegmentierung” (Bachelorarbeit)
- [6] Isola, P.; Zhu, J.-Y.; Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition, 2017*, 1125-1134
- [7] Kierdorf, J., Junker-Frohn, L. V., Delaney, M., Olave, M. D., Burkart, A., Jaenicke, H., Muller, O., Rascher, U. & Roscher, R. (2022). GrowliFlower: An image time-series dataset for GROWth analysis of cauLIFLOWER. *Journal of Field Robotics*.
- [8] Kierdorf, J., Weber, I., Kicherer, A., Zabawa, L., Drees, L., & Roscher, R. (2022). Behind the leaves: Estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5.
- [9] Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10
- [10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- [11] Marin Zapata, P. A.; Roth, S.; Schmutzler, D.; Wolf, T.; Manesso, E. & Clevert, D.-A. Self-supervised feature extraction from image time series in plant phenotyping using triplet networks *Bioinformatics, Oxford University Press*, 2021, 37, 861-867
- [12] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).
- [13] Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., and Moreno, J. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, 40(3):589–629
- [14] Zhu, J.-Y.; Park, T.; Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE international conference on computer vision, 2017*, 2223-2232



# Behind the Leaves: Estimation of Occluded Grapevine Berries With Conditional Generative Adversarial Networks

Jana Kierdorf<sup>1\*</sup>, Immanuel Weber<sup>2</sup>, Anna Kicherer<sup>3</sup>, Laura Zabawa<sup>4</sup>, Lukas Drees<sup>1</sup> and Ribana Roscher<sup>1</sup>

<sup>1</sup> Remote Sensing Group, Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany, <sup>2</sup> Application Center for Machine Learning and Sensor Technology, University of Applied Sciences Koblenz, Koblenz, Germany, <sup>3</sup> Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany, <sup>4</sup> Geodesy Group, Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany

## OPEN ACCESS

### Edited by:

Ian Stavness,  
University of Saskatchewan, Canada

### Reviewed by:

Andrew P. French,  
University of Nottingham,  
United Kingdom  
Luigi Celona,  
University of Milano-Bicocca, Italy

### \*Correspondence:

Jana Kierdorf  
jkierdorf@uni-bonn.de

### Specialty section:

This article was submitted to  
AI in Food, Agriculture and Water,  
a section of the journal  
Frontiers in Artificial Intelligence

**Received:** 06 December 2021

**Accepted:** 28 February 2022

**Published:** 25 March 2022

### Citation:

Kierdorf J, Weber I, Kicherer A, Zabawa L, Drees L and Roscher R (2022) Behind the Leaves: Estimation of Occluded Grapevine Berries With Conditional Generative Adversarial Networks.  
*Front. Artif. Intell.* 5:830026.  
doi: 10.3389/frai.2022.830026

The need for accurate yield estimates for viticulture is becoming more important due to increasing competition in the wine market worldwide. One of the most promising methods to estimate the harvest is berry counting, as it can be approached non-destructively, and its process can be automated. In this article, we present a method that addresses the challenge of occluded berries with leaves to obtain a more accurate estimate of the number of berries that will enable a better estimate of the harvest. We use generative adversarial networks, a deep learning-based approach that generates a highly probable scenario behind the leaves exploiting learned patterns from images with non-occluded berries. Our experiments show that the estimate of the number of berries after applying our method is closer to the manually counted reference. In contrast to applying a factor to the berry count, our approach better adapts to local conditions by directly involving the appearance of the visible berries. Furthermore, we show that our approach can identify which areas in the image should be changed by adding new berries without explicitly requiring information about hidden areas.

**Keywords:** deep learning, machine learning, Generative Adversarial Networks, domain-transfer, grape generation, occlusions, yield counting

## 1. INTRODUCTION

With increasing competition on the wine market worldwide, the need for accurate yield estimations has been getting more and more important for viticulture. The variation of yield over the years is mainly caused by the berry number per vine (90%), while the remaining 10% are caused by the average berry weight (Clingleffer et al., 2001), which is generally collected manually and averaged over many years. Traditionally, yield estimations in viticulture can be done at three phenological timepoints by (1) counting the number of bunches 4–6 weeks after budburst, (2) counting the number of berries after fruit set (May, 1972), or (3) destructively sampling vines or segments of vines close to harvest. Considering that yield estimation can be more accurately and reliably determined as harvest approaches, a berry count is a promising option that can be approached non-destructively and whose process can be automated.

Several papers show that machine learning-based methods for analyzing data from imaging sensors provide an objective and fast method for counting visible berries (Diago et al., 2012; Kicherer et al., 2014; Nuske et al., 2014; Roscher et al., 2014; Aquino et al., 2017; Coviello et al., 2020; Zabawa et al., 2020), and thus for automated yield predictions in the field. One of the main challenges in deriving berry counts from image data taken in the field is occlusions, which generally causes an underestimation of the number of berries and yield (Zabawa et al.<sup>1</sup>). First, occlusions of berries by other berries make it difficult to distinguish or count individual berries. Therefore, approaches that perform a segmentation of regions of berries and regions without berries is not sufficient, and more advanced methods that recognize individual instances of berries must be applied (Zabawa et al., 2020). Second, occlusions by leaves play a major role in underestimating the number of berries. Zabawa et al. (see text footnote 1, respectively) perform leaf occlusion experiments over two years and show that the yield estimation is highly dependent on the number of visible berries. With vines defoliated (i.e., with manually removed leaves) at pea size, they report an average error of total yield estimation of 27%, whereas Nuske et al. (2014) observed average errors between 3 and 11% using images of entirely defoliated fruit zones.

In order to overcome the challenge of leaf occlusions, defoliation can be performed in the grapefruit zone, but this is immensely time-consuming and labor-intensive. Partial defoliation is carried out in viticulture, for example, for ventilation and rapid drying of the grape zone to avoid fungal infections of the grapes or yield and quality regulation (Diago et al., 2009). However, complete defoliation is not feasible on a large scale or may lead to negative effects such as increased sunburn on the berries (Feng et al., 2015) or generally have an undesirable impact on yield results. Alternatively, machine-learning-based approaches can be used to obtain a more accurate estimation of the berry number. Numerous approaches rely on information where occlusions are present, which is generally provided as a manual input (Bertalmio et al., 2003; Barnes et al., 2009; Iizuka et al., 2017; Dekel et al., 2018; Liu et al., 2018). In contrast to this, two-step approaches first detect occlusions and then fill the corresponding regions with information according to the environment (Ostyakov et al., 2018; Yan et al., 2019).

This article addresses the challenge of occlusions caused by leaves by generating images that reveal a highly probable situation behind the leaves, exploiting learned patterns from a carefully designed dataset. The generated images can then be used to count berries in a post-processing step. Our approach generates potential berries behind leaves based on RGB information obtained by visible light imaging, as this is an efficient, cheap and non-harmful approach in contrast to data from material-penetrating sensors. In order to train our machine-learning method, we use aligned image pairs showing plants with leaves and the same plants after defoliation. In detail, we

model this problem as a domain-transfer task and regard the aligned images containing occluded berries as one domain and images with revealed berries as a second domain. We resort to methods like Pix2Pix (Isola et al., 2017), that uses a conditional generative adversarial network (cGAN) (Mirza and Osindero, 2014) and can learn the described domain-transfer. In contrast to other works, we present a one-step approach that is end-to-end trainable, meaning the positions of the occlusions are identified, and patterns that need to be filled are learned simultaneously. Through the experience the model gains during training, it learns patterns such as grape instances with their appearing shapes, their environment, and where they occur in the image. This knowledge is exploited during the generation step, in which the learned domain-transfer model is applied to images of vines that have not been defoliated to obtain a high-probability and realistic impression of the scene behind the leaves. In order to obtain a berry count, the generated images are further processed with the berry counting algorithm of Zabawa et al. (2020). In this way, we provide a more accurate count of grape berries, since both visible berries and berries potentially occluded by leaves are taken into account.

A major challenge for training is that there is no large dataset of aligned natural images that includes both images with occluded berries and images with berries exposed by defoliation. In addition, in our case, the spatial alignment between the image pairs is not accurate enough since defoliation leads to a resulting movement of branches, grape bunches, and other objects in the non-occluded domain patches. As a result, the natural data is not sufficient to train a model that matches our requirements of a reliable model. Due to this, we propose the use of a synthetically generated dataset that contains paired data of both domains. Our main contributions of this article are:

- The true scenario behind the leaves without defoliation is unlikely to be identified. Therefore, our approach estimates a highly probable scenario behind the occlusions based on visible information in the image, especially of the surroundings of the occlusion, and learned patterns during the training process which include for example the berry shape and neighborhood of berries to obtain a distribution similar to the training data.
- We present a one-step approach, which can implicitly identify which image areas contain visible berries and which areas are occluded without supervision regarding occluded and non-occluded areas. This differs from approaches such as inpainting (Bertalmio et al., 2003; Barnes et al., 2009; Iizuka et al., 2017; Dekel et al., 2018; Liu et al., 2018), in which the occluded areas must be known *a priori*.
- In addition to the acquired images, we use so-called berry masks obtained by the approach presented in Zabawa et al. (2020), which uses semantic segmentation to indicate in the image which pixels belong to berry, berry-edge, and background. During training, this leads to a more stable and easier optimization process. During testing, the berry mask is only needed for the input image since our GAN-based method simultaneously generates the berry mask in which the berries are counted, in addition to the visually generated image.

**Abbreviations:** cGAN, conditional Generative Adversarial Network; SMPH, semi minimal pruned hedge; VSP, vertical shoot position.

<sup>1</sup>Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Roscher, R., and Kuhlmann, H. (2021). Image-based analysis of yield parameters in viticulture. *Biosyst. Eng.* (under review).



- Since a direct comparison of the true scenario behind the leaves and our generated scenario is not appropriate using standard evaluation methods such as a pixel-by-pixel comparison, we perform a comprehensive evaluation using alternative evaluation metrics, such as generation maps and correlation, that assesses the performance of our approach.
- We show that the application of our approach minimizes the offset compared to the manual reference berry count and the variance, which is not achieved by applying a factor.
- We create various synthetic datasets and show that our approach trained on synthetic data also works on natural data.

The article is structured as follows: After surveying related works, we start by introducing our domain-transfer framework and describe the different components, such as the conditional generative adversarial network, that are used in our approach. We explain the data acquisition and post-processing of the natural and synthetic datasets we use in our work. We explain the evaluation metrics we use and then describe our experiments in which we analyze the generation quality of different synthetic input data, compare generated results with real results in the occluded as well as the non-occluded domain and analyze the berry counting based on the generated results. Finally, we investigate the application of the synthetically learned models to natural data. We end our article with the conclusion and future directions.

## 2. RELATED WORK

### 2.1. Yield Estimation and Counting

Since an accurate yield estimation is one of the major needs in viticulture, especially on a large scale, there is a strong demand for objective, fast, and non-destructive methods for yield forecasts in the field. For many plants, including grapevines, the derivation of phenotypic traits is essential for estimating future yields. Besides 3D-reconstruction (Schöler and Steinhage, 2015; Mack et al., 2017, 2018), 2D-image processing is also a widely used method (Hacking et al., 2019) for the derivation of such traits. For vine, one plant trait that strongly correlates with yield is the number of bearing fruits, that means the amount of berries. This correlation is underlined by the study of Clingeffer et al. (2001), in which it is shown that the variation of grapevine yield over the years is mainly caused by the berry number per vine (90%).

The task of object counting can be divided into two main approaches: (1) regression (Lempitsky and Zisserman, 2010; Arteta et al., 2016; Paul Cohen et al., 2017; Xie et al., 2018) which directly quantifies the number of objects for a given input, and (2) detection and instance segmentation approaches which identify objects as an intermediate step for counting (Nuske et al., 2014; Nyarko et al., 2018). Detection approaches in viticulture are presented, for example, by Nuske et al. (2011), Roscher et al. (2014), and Nyarko et al. (2018), who define berries as geometric objects such as circles or convex surfaces and determine them by image analysis procedures such as Hough-transform. Recent state-of-the-art approaches, especially segmentation (He et al., 2017), are mostly based on neural networks. One of the earliest works combining grapevine data and neural network analysis

was Aquino et al. (2017). They detect grapes using connected components and determine key features based on them, which are fed as annotations into a three-layer neural network to estimate yield. In another work, Aquino et al. (2018) deal with counting individual berries, which are first classified into berry candidates using pixel classification and morphological operators. Afterward, a neural network classifies the results again and filters out the false positives.

The two studies by Zabawa et al. (2019, 2020) serve as the basis for this article. Zabawa et al. (2019) use a neural network which performs a semantic segmentation with the classes berry, berry-edge and background, which enables the identification of single berry instances. The masks generated in that work serve as input for the proposed approach. The article by Zabawa et al. (2020) based on Zabawa et al. (2019) extends identification to counting berries by discarding the class edge and counting the berry components with a connected component algorithm. The counting procedure applied in that work is used for the analyses of the experiments.

### 2.2. Given Prior Information About Regions to Be Transferred

A significant problem in fruit yield estimation is the overlapping of the interesting fruit regions by other objects, like in the case of this work, the leaves. Several works are already addressing the issue of data with occluded objects or gaps within the data, where actual values are missing, which is typically indicated by special values like, e.g., not-a-number. The methodologies can be divided into two areas: (1) there is prior information available about where the covered positions are, and (2) there is no prior information. In actual data gaps, where the gap positions can be easily identified a priori, data imputation approaches can be used to complete data. This imputation is especially important in machine learning, since machine learning models generally require complete numerical data. The imputation can be performed using constant values like a fixed constant, mean, median, or k-nearest neighbor imputation (Batista and Monard, 2002) or calculated using a random number like the empirical distribution of the feature under consideration (Rubin, 1996, 2004; Enders, 2001; von Hippel and Bartlett, 2012). Also, possible are multivariate imputations, which additionally measures the uncertainty of the missing values (Van Buuren and Oudshoorn, 1999; Robins and Wang, 2000; Kim and Rao, 2009). Data imputation is also possible using deep learning. Lee et al. (2019), for example, introduce CollaGAN in which they convert the image imputation problem to a multi-domain image-to-image translation task.

In case there are no data gaps, but the image areas that are occluded or need to be changed are known, inpainting is a commonly used method. The main objective is to generate visually and semantically plausible appearances for the occluded regions to fit in the image. Conventional inpainting methods (Bertalmio et al., 2003; Barnes et al., 2009) work by filling occluded pixels with patches of the image based on low level features like SIFT descriptors (Lowe, 2004). The results of these methods do not look realistic if the areas to be filled are near



foreground objects or the structure is too complex. An alternative is deep learning methods that learn a direct end-to-end mapping from masked images to filled output images. Particularly realistic results can be generated using Generative Adversarial Networks (GANs) (Iizuka et al., 2017; Dekel et al., 2018; Liu et al., 2018). For example, Yu et al. (2018) deal with generative image inpainting using contextual attention. They stack generative networks to ensure further the color and texture consistence of generated regions with surroundings. Their approach is based on rectangular masks, which do not generalize well to free-form masks. This task is solved by Yu et al. (2019) one year later by using guidance with gated convolution to complete images with free-form masks. Further work introduces mask-specific inpainting that fills in pixel values at image locations defined by masks. Xiong et al. (2019) learn a mask of the partially masked object from the unmasked region. Based on the mask, they learn the edge of the object, which they subsequently use to generate the non-occluded image in combination with the occluded input image.

### 2.3. No Prior Information About Regions to Be Transferred

Methods that do not involve any prior knowledge about gaps and occluded areas can be divided into two-step and one-step approaches. Two-step approaches first determine the occluded areas, which then are used, for example, as a mask to inpaint the occluded areas. Examples are provided by Yan et al. (2019), which visualize the occluded parts by determining a binary mask of the visible object using a segmentation model and then creating a reconstructed mask using a generator. The resulting mask is fed into coupled discriminators together with a 3D-model pool in order to decide if the generated mask is real or generated compared to the masks in the model pool. Ostyakov et al. (2018) train an adversarial architecture called SEIGAN to first segment a mask of the interesting object, then paste the segmented region into a new image and lastly fill the masked part of the original image by inpainting. Similar to the proposed approach, SeGAN introduced by Ehsani et al. (2018) uses a combination of a convolutional neural network and a cGAN (Mirza and Osindero, 2014; Isola et al., 2017) to first predict a mask of the occluded region and, based on this, generate a non-occluded output.

## 3. MATERIALS AND METHODS

### 3.1. Framework

In our work, we regard the revealing of the occluded berries as a transfer between two image domains. We first detail this and show how we model this transfer for our data. Then we will lay out the cGAN and the framework we use for this task. Finally, we show how we train this network.

#### 3.1.1. Domain-Transfer Framework

On a high level, the task of revealing the occluded berries can be described as generating a new impression of an existing image. We model this generative task as a transfer of an existing image from one domain, the source domain, to another domain, the target domain. In our work, we regard images where berries

are occluded by various objects as the source domain and call it *occluded domain*. Accordingly, our target domain contains images of defoliated plants, and we call it *non-occluded domain*. Therefore, by performing this domain-transfer, we aim to reveal hidden berries. Samples of both domains are shown at the top of **Figure 1**.

This task can typically be learned by a cGAN, like Pix2Pix in our case. We train this network using aligned pairs of images from the occluded domain and the non-occluded domain and indicate them with  $x_{occ}$  and  $x_{non}$ , respectively. The first ones are used as the network input and the latter ones, being the desired output, as the training target. Due to computational limitations, we use cropped patches from the original data and convert them to grayscale to develop an efficient approach that is independent of the berry color. In practice, we accompany the images of each domain with a corresponding semantic mask, that indicates per image pixel the content based on the classes `berry`, `berry-edge`, and `background`. This mask supports the discriminability of relevant information like the berries from the surrounding information in the image and the generation of separated berries, supporting the later counting step. After training, we use the cGAN to generate images,  $\tilde{x}_{non}$ , that we further process with a berry counting method.

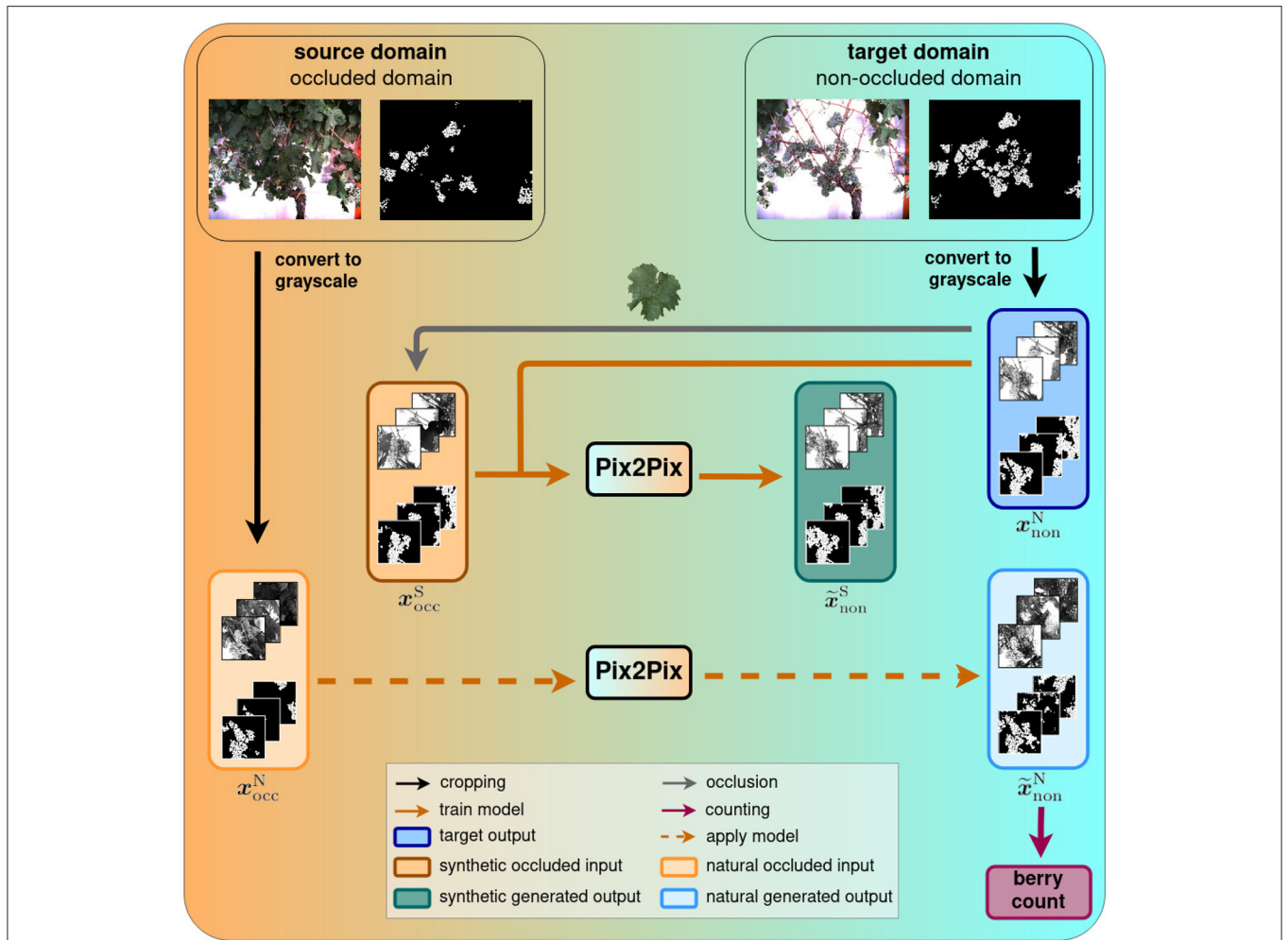
Since we only have limited amounts of data available for training and testing, we resort to a dataset consisting of synthetic images for the occluded domain and natural images for the non-occluded domain that we describe in detail in section 3.2. In addition, we test our trained model on fully natural data to analyze the generalizability of the model. For the training set, the non-occluded domain contains natural images, whereas the images from the occluded domain are derived from the former domain, where berries are artificially occluded with leaf templates. To differentiate the different datasets of images, we further qualify the natural images with index N and the synthetic images with index S, which results in the two occluded domain groups:  $x_{occ}^N$  and  $x_{occ}^S$ . The generated images are accordingly indicated by  $\tilde{x}_{non}^N$  and  $\tilde{x}_{non}^S$ . We therefore train the model with input images  $x_{occ}^S$  and use  $x_{non}^N$  as target images. Finally, we apply the model on natural images  $x_{occ}^N$  and compute the berry counts of the generated output images,  $\tilde{x}_{non}^N$ .

#### 3.1.2. Conditional Generative Adversarial Networks

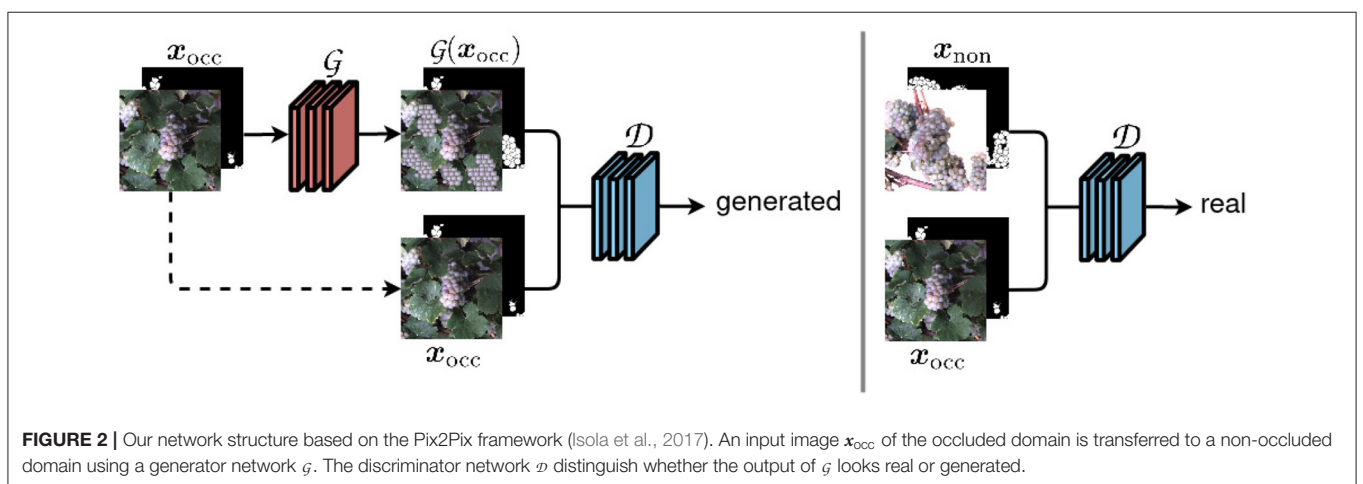
The core of our framework is the cGAN that we use to generate images with berries being revealed. Specifically, we use the Pix2Pix (Isola et al., 2017) network and training method, which is illustrated in simplified form in **Figure 2**.

The model consists of two networks, the generator, and the discriminator. The generator network  $\mathcal{G}$  takes images with occluded berries as an input and is intended to generate images with revealed berries  $\mathcal{G}(x_{occ}) = \tilde{x}_{non}$  that cannot be distinguished from real images  $x_{non}$  of the non-occluded domain. The adversarially trained discriminator network  $\mathcal{D}$ , on the other side, tries to discriminate between generated images  $\tilde{x}_{non}$  and real images  $x_{non}$ . The generator used in Pix2Pix is based on a U-Net (Ronneberger et al., 2015), the discriminator  $\mathcal{D}$  on a PatchGAN.

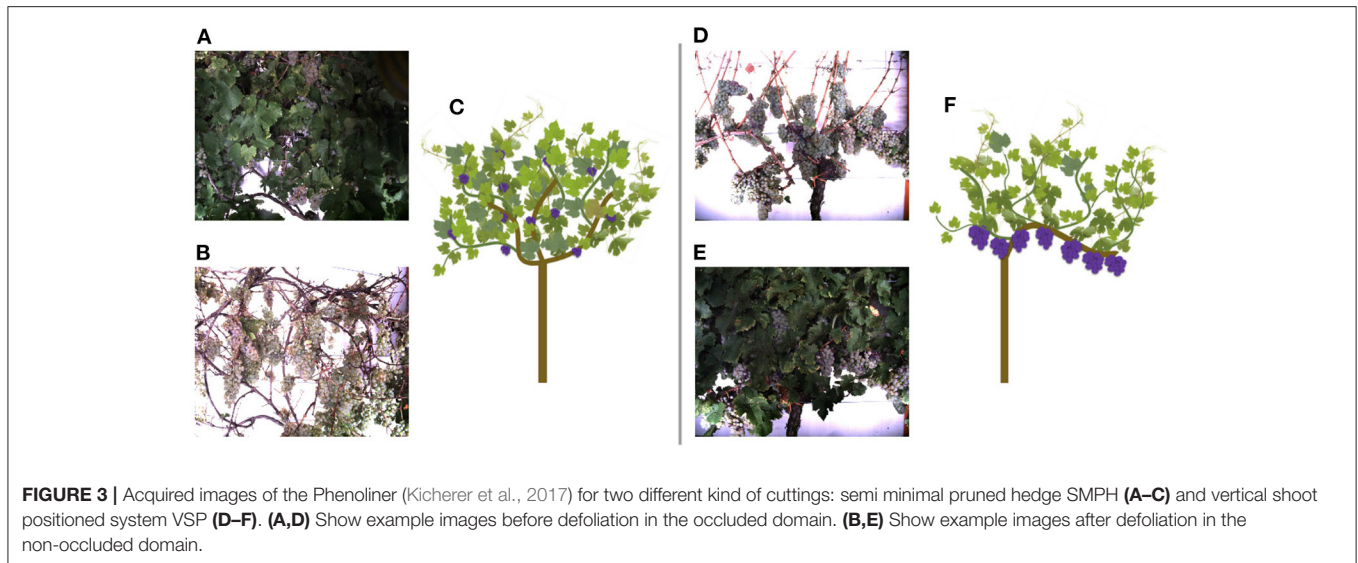
As described by Goodfellow et al. (2014), both parts of GANs are trained simultaneously using a min-max approach. The goal



**FIGURE 1 |** Domain-transfer framework. We transfer images from the source domain with occluded berries to the target domain with revealed berries using the Pix2Pix cGAN. We train and test the model on synthetic data and subsequently apply it to natural data. Finally, a berry counting is performed on the generated outputs. Further evaluation steps will be performed in our experiments.



**FIGURE 2 |** Our network structure based on the Pix2Pix framework (Isola et al., 2017). An input image  $x_{occ}$  of the occluded domain is transferred to a non-occluded domain using a generator network  $g$ . The discriminator network  $d$  distinguish whether the output of  $g$  looks real or generated.



of the discriminator during training is to be able to distinguish as good as possible between real and generated images. For this, the discriminator uses a mini-batch of input images  $x_{\text{non}}$  and computes the discriminator loss  $l_{\mathcal{D}_{\text{real}}}$ . Additionally, it uses generated images  $\tilde{x}_{\text{non}}$  obtained from the generator  $\mathcal{G}$  and computes the corresponding loss  $l_{\mathcal{D}_{\text{gen}}}$ . For both computations, the mean squared error (MSE) loss  $l_{\text{MSE}}$  is used. The overall loss  $l_{\mathcal{D}}$  of the discriminator is calculated as:

$$l_{\mathcal{D}} = \frac{1}{2} \cdot (l_{\mathcal{D}_{\text{fake}}} + l_{\mathcal{D}_{\text{real}}}) \quad (1)$$

The objective is to maximize this loss, as this means that the discriminator can distinguish between generated and real images with ease. The weights of the discriminator network are then updated with respect to this loss.

When generating new images, the generator tries to trick the discriminator at the same time, which is the adversarial part of the network. Compared to the maximization of the discriminator loss, the objective of the generator is to minimize the generator loss  $l_{\mathcal{G}}$ . This is calculated from a combination of MSE loss computed by  $\mathcal{D}[\mathcal{G}(x_{\text{occ}})]$  referred to the reference label generated and a  $l_1$  loss, which avoids blurring. The  $l_1$  loss is computed using real and generated images,  $x_{\text{non}}$  and  $\tilde{x}_{\text{non}}$ , from the non-occluded domain. The generator loss  $l_{\mathcal{G}}$  is then used to update the generator's weights.

$$l_{\mathcal{G}} = l_{\text{MSE}}(\mathcal{D}(\mathcal{G}(x_{\text{occ}}))) + \lambda \cdot l_1(x_{\text{non}}, \tilde{x}_{\text{non}}) \quad (2)$$

The weighting factor  $\lambda$  adjusts the scale of the losses to each other and is, in our case,  $\lambda = 100$ .

The minimization of the generator loss  $l_{\mathcal{G}}$  results in either a strong generator or a very weak discriminator. If the loss becomes maximal, the opposite possibilities can occur. The objective is to balance both adversarial goals at the end of the training in the best possible way by realizing both at the same time.

## 3.2. Data

### 3.2.1. Study Site

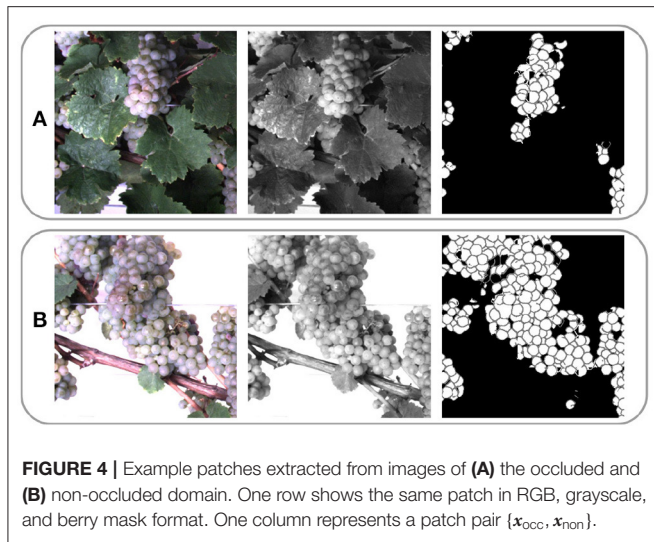
The data, we use in this work, were acquired at the experimental fields of JKI Geilweilerhof located in Siebeldingen, Germany. It was acquired using the Phenoliner (Kicherer et al., 2017), a reconstructed grape harvester that can be used as a phenotyping platform to acquire geo-referenced sensor data directly in the field. A description of the on-board camera setup can be found in Zabawa et al. (2020). The images were acquired in two different training systems of the cultivar Riesling (DEU098\_VIVC10077\_Riesling\_Weiss\_DEU098-2008-085): (1) Vertical shoot positioned (VSP) vines (Figure 3C) and (2) vines trained as semi minimal pruned hedges (SMPH) (Figure 3F) were chosen due to diverse difficulties in image analysis (Zabawa et al., 2020). The acquisition took place in September 2019 and 2020, before harvest at the plant growth stage BBCH89, and in each year the images were taken 1 day before (Figures 3A,D) and right after defoliation (Figures 3B,E). In 2019 50 cm and 2020, respectively, 100 cm of the grapevine canopy have been defoliated.

In our framework, we use three different types of inputs:

- **Natural data:** Images acquired in the vineyard before and after defoliation. For our studies, we use grayscale images. We denote this dataset with  $X^{\text{N}}$ .
- **Synthetic data:** Images acquired in the vineyard after defoliation. Images with occluded berries are synthetically generated. We denote this dataset with  $X^{\text{S}}$ .
- **Semantic segmentation masks (berry masks):** So-called berry masks obtained by a semantic segmentation approach presented in Zabawa et al. (2019). Each pixel in these images is assigned to the class `berry`, `berry-edge`, or `background`. We denote this data as  $X_{\text{B}}$ .

The use of the mentioned grayscale images is indicated by the index G and with index B we denote the use of the berry masks. Moreover, we define  $X_{\text{GB}}$  as the input where the grayscale image





and the berry mask are stacked to form a multichannel 2D input. In the following, the used data is explained in more detail.

### 3.2.2. Natural Data

We convert the acquired RGB images into grayscale images in order to develop an efficient approach that is independent of the berry color. Covering the whole variability of possible berry colors is complex and not feasible in our case. For example, in the case of green berries, the color also does not serve to differentiate them from leaves.

Since the Phenoliner platform revisits the vine row for each data collection of the two domains, the images depicting the same scene are acquired at different times and from different positions, leading to differences in translation, rotation and scale. Moreover, the defoliation of vines causes a movement of the branches and grape bunches, and additional environmental changes between the two acquisition time points can result in different scenes in the aligned patches.

However, to obtain aligned image pairs for a qualitative evaluation, we manually align images from both domains. For this, we compute a four-parameter Helmert transformation (Helmert, 1880) between the two domains, where we manually define corresponding keypoints per image pair to calculate the parameters. We apply this transformation to images from the non-occluded domain to register them to the occluded domain.

Due to computational limitations, we use a sliding window of size  $656 \text{ px} \times 656 \text{ px}$  and stride  $162 \text{ px}$  to extract patches from the grayscale images. **Figure 4** illustrates one RGB patch, the grayscale patch, and the corresponding berry mask, which is explained in the following subsection, for both domains. We denote the aligned patch pair  $x^N = \{x_{occ}^N, x_{non}^N\}$ , where  $x^N \in X^N$ .

### 3.2.3. Semantic Segmentation Mask (Berry Mask)

Besides the acquired images, we use a berry mask, obtained with a semantic segmentation approach, presented by Zabawa et al. (2019). The identification of regions containing berries and the detection of single berry instances is performed

with a convolutional neural network. The network uses a MobileNetV2 (Sandler et al., 2018) encoder and a DeepLabV3+ decoder (Chen et al., 2018). The network assigns each image pixel to one of the classes `background`, `berry-edge`, or `berry`, which corresponds to the grayscale values 0, 127, and 255. In contrast to a standard semantic segmentation without distinguishing between different instances, we use the additional class `berry-edge` to ensure the separation of single berries, which allows the counting of berries using a connected component approach.

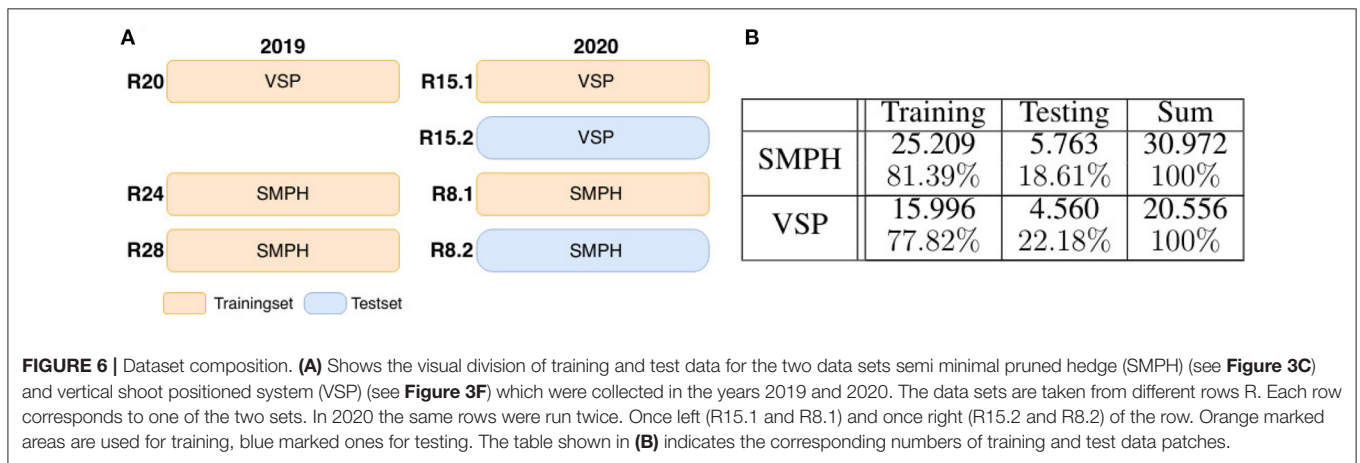
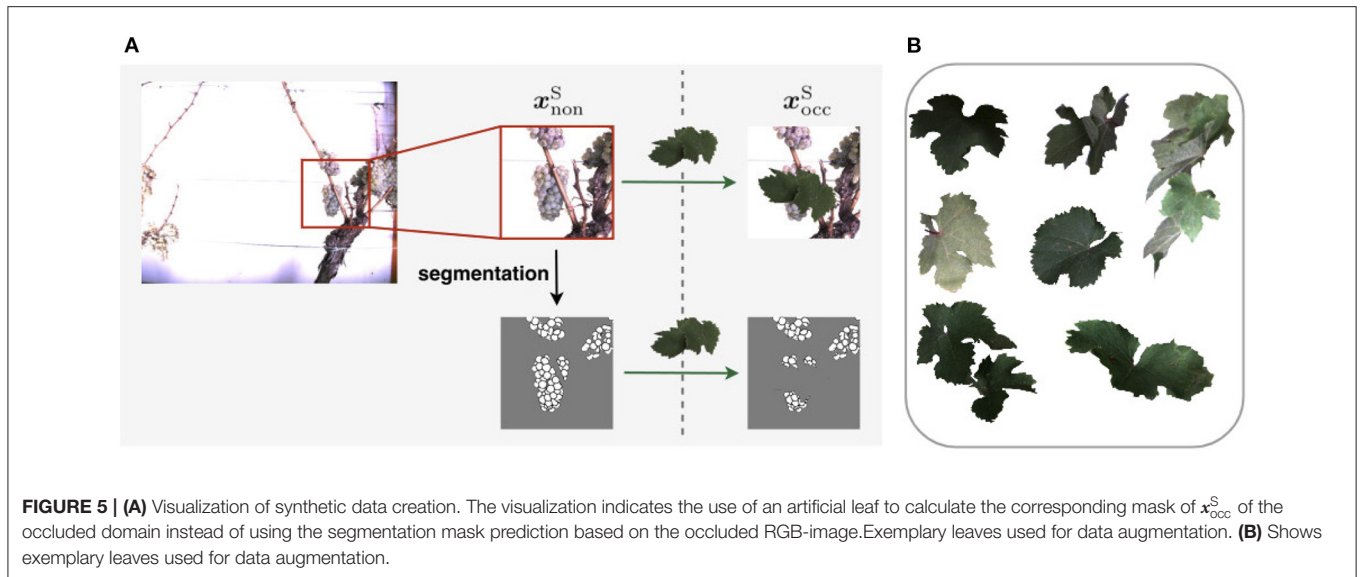
For our task of generating a highly probable scenario behind leaves, the berry mask supports the discriminability of relevant information like the berries from the surrounding information in the image, and the generation of separated berries. In addition, since the berry masks contain a masking of existing berries, it provides further knowledge about which areas in the images do not show occlusions and should be preserved in the revelation process and where potentially occlusions might appear, which are areas that are unmasked.

Since we are interested in scenes in the image that depict berries, we only integrate patch pairs in training and testing, whose berry mask of the non-occluded domain contains more than  $1/24$  background pixels and mask of the occluded domain contains at least one pixel whose class differs from the background class.

### 3.2.4. Synthetic Data

One challenge for our application is that the amount of paired data from both domains containing both, occluded and non-occluded regions of berries, is limited for training a reliable model and for evaluation. We, therefore, resort to generate artificially modified images, where berries are artificially occluded, based on natural images of defoliated plants. This allows us to generate a large dataset to ease the described lack of natural images of both domains. We denote this synthetic dataset with  $X^S$ . The natural patches  $x_{non}^N$  of the non-occluded domain serve as a basis. We create paired patches  $\{x_{non}^S, x_{occ}^S\}$  where  $x_{non}^S = x_{non}^N$ .

To generate  $x_{occ}^S$ , we apply artificial data modification on both training and test data. We artificially occlude the patches using 24 different wine leaves (**Figure 5B**) with various shapes extracted from the natural dataset and use them as occluding objects in the patches. We use 18 leaves for augmenting the training set and six leaves to augment the test set. On the basis of one image patch  $x_{non}^S$ , we create up to nine corresponding synthetically augmented versions of  $x_{occ}^S$  for the training set, resulting in nine aligned image patch pairs. During the procedure, a leaf is randomly selected from the set of leaves and rotated by a randomly chosen angle  $\alpha \in \{-50, -30, -10, 0, 10, 30, 50, 70\}$ . Converted to grayscale, it randomly overlays the grayscale patch and occludes parts of the visible berries. These steps are also performed for patches  $x_{non}^S$  of the test set. However, here only three new patch pairs are created. After applying artificial data modification, the proportion of test data amounts to  $\sim 18\text{--}23\%$  of the extracted patches depending on the type of defoliation (see **Figure 6B**). The split of the data into training and test data is illustrated visually and numerically in **Figure 6**.



The test data is taken from the dataset collected in 2020 (see **Figure 6A**).

For each synthetic grayscale image, we calculate a corresponding berry mask. However, depending on the used procedure, the appearance of the berry mask differs. In our work, we create the masks for the two domains, as illustrated in **Figure 5A**. The mask of the non-occluded patch  $x_{non}^S$  is based on the segmentation step, described in Section 3.3.2, which needs RGB images as input. We compute the mask of  $x_{occ}^S$  by overlaying the pixels of the non-occluded mask of  $x_{non}^S$  that are covered with a leaf in the RGB, or respectively grayscale patch. These pixels in the berry mask are assigned to the class background. The leaf pixels adjacent to berry pixels are changed to berry-edge pixels. In this way, the overlapped berries have a closed contour. By adding these edges, the synthetic data thus has the same characteristics as berry masks derived from the natural data. With this step, we create two corresponding masks,  $x_{occ}^S$  and  $x_{non}^S$ , which match exactly in the non-occluded pixel.

Another way to define the occluded mask is a direct computation as for  $x_{non}^S$  using the segmentation step to create a predicted mask of the patch. Since the berry mask is an estimation, the class of individual non-occluded pixels may differ between  $x_{non}^S$  and  $x_{occ}^S$ . For a simplified analysis, we have chosen the first option.

Overall, for dataset  $VSP X^S$ , we obtain 20.556 synthetic patch pairs, and for dataset  $SMPH X^S$ , we obtain 30.972 synthetic patch pairs **Figure 6B**.

### 3.2.5. Challenges

Various challenges occur in the data, which influence our training and thus our results. Since our reference masks are not manually derived but are estimations, uncertainties can occur. For example, not all visible berries are entirely shown in the images of the non-occluded domain. Therefore, it can happen that either berries are missed or only partly detected in the mask. Additionally, the estimated contour in the berry mask may not be closed and parts of the berry region may be classified as

background. Thus, these errors in the reference could be learned in the model. Furthermore, there are images in the non-occluded domain, which contain leaves despite defoliation. In an ideal case, the model learns to ignore these faults in defoliation. Other challenges are the varying sharpness of the patches. This can be caused by resizing the data, shadows, or the varying distance of the berries to the camera. Furthermore, the illumination varies within the data, e.g., due to the coverage by surrounding objects like branches or leaves or the distance of the berries to the camera. Also, worth noting are the different growth stages of the grapes in 2019 and 2020, so the grapes have different sizes due to different berry sizes.

### 3.3. Model Evaluation

#### 3.3.1. Data Post-processing

After the test phase, the generated masks do not only contain the values 0, 127, and 255. There are also mixed pixels that are not clearly assigned to one of the three classes. We use thresholding to ensure that only the values 0, 127, and 255 appear in the mask. We use the following class assignment.

- Pixel values in the interval  $[0, 50]$  are set to value 0 and assigned to class `background`.
- Pixel values in the interval  $[50, 180]$  are set to value 127 and assigned to class `berry-edge`.
- Pixel values in the interval  $[180, 255]$  are set to value 255 and assigned to class `berry`.

#### 3.3.2. Evaluation Metrics

In the following, we describe several evaluation metrics used for our experiments. The first metric we use is the *area*  $F_c$ , that we define as the number of pixels within a mask that correspond to a class  $c$  with  $c \in \{\text{background}, \text{berry-edge}, \text{berry}\}$ . With area  $F_c$  and the generated area  $\tilde{F}_c$ , which is based on the generated mask of the cGAN, we calculate the *intersection over union* IoU by dividing the area of overlap by the area of union.

$$\text{IoU}_c = \frac{F_c \cap \tilde{F}_c}{F_c \cup \tilde{F}_c} \quad (3)$$

The IoU compares the similarity between two arbitrary shapes.

The second metric we use is the *pearson product-moment correlation coefficient*. It gives a measure of the degree of linear relationship between two variables. The correlation coefficient is obtained by the correlation coefficient matrix  $Q$ , which is calculated by means of the covariance matrix  $C$ ,

$$Q_{i,j} = \frac{C_{i,j}}{\sqrt{C_{i,i} \cdot C_{j,j}}} \quad (4)$$

where  $i$  and  $j$  indicate the row and column index, respectively. The values of  $Q$  are between  $-1$  and  $1$ , inclusive. The correlation coefficient  $\rho$  between two variables can then be expressed by  $\rho = Q_{0,1}$ . A correlation coefficient  $\rho$  equals 1 indicates that both input variables are equal. We use the correlation to compare the generated images  $\tilde{x}_{\text{non}}$  from the model with the input  $x_{\text{occ}}$  as well as the target output  $x_{\text{non}}$  on pixel level.

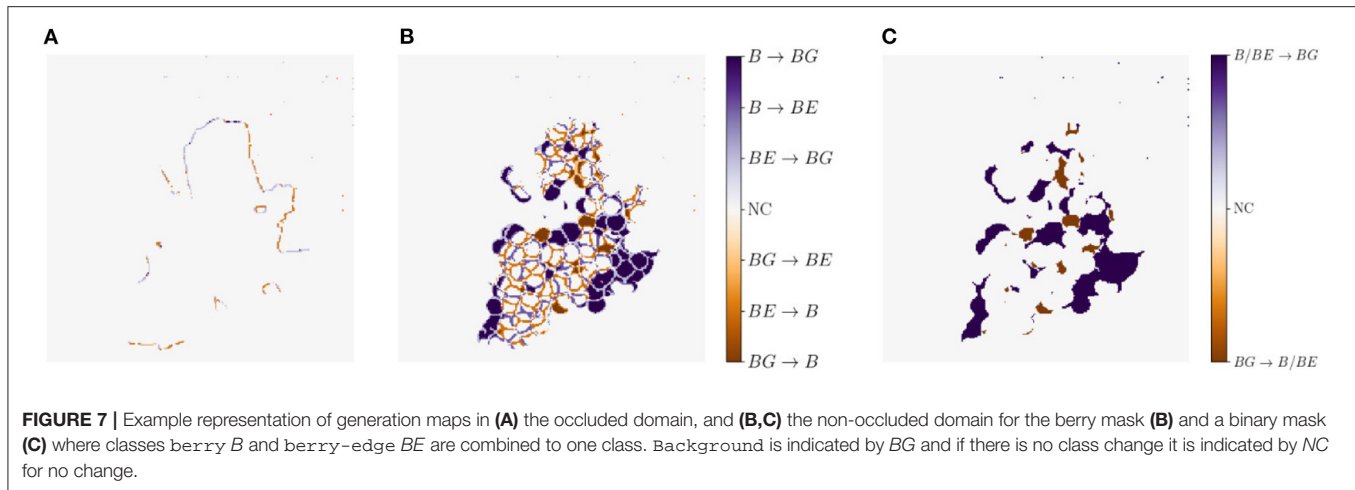
The *coefficient of determination*, also denoted by  $R^2$ , indicates the relationship between a predicted value with respect to a reference value. It provides a measure of how well-observed references are replicated by the model. In our case, we use the  $R^2$  value for the comparison between the predicted number of berries generated by the model and the reference number from the berries manually counted in the non-occluded domain. Plots, as illustrated in **Figure 10**, represent the generated distribution of the model compared to the reference. Please note, that the gray line represents the reference values. The optimal generated samples are distributed along this line, reflected in a  $R^2$  value equal to 1.

The *counting* is based on the procedure described in the work of Zabawa et al. (2020). The counting is performed based on the masks, which are predicted with the convolutional neural network presented in their work. The classes `background` and `berry-edge` are discarded, and the counting is solely performed with pixels of the class `berry`. Before counting the number of connected components of the berry mask, we introduce geometrical and qualitative filter stages to improve the count. Filtering follows the observations of Zabawa et al. (2020) (Table 3) which show that when the filter is applied, the misclassifications for VSP cutting decrease by 9% and for SMPH cutting by 11%. For the first step of filtering, we discard elements that are smaller than 25 pixels, since these artifacts are too small to represent berries. Secondly, we exploit the knowledge that berries are roughly round by removing objects with a minor-major-axis ratio below 0.3 and an insufficient area. The actual area of each component is compared to the expected area based on a radius, which is computed as the mean of the minor and major axis of the component. Lastly, we check how well each object is surrounded by an edge, since most high confidence predictions are well surrounded by an edge. For further details, we refer the reader to Zabawa et al. (2019).

Another metric we use for a visual comparison is the *generation map*. Generation mapping is used to visualize the differences between two masks. In our case the distances are calculated between (1) the input mask  $x_{\text{occ}}$  and the generated mask  $\tilde{x}_{\text{occ}}$  (**Figure 7A**), (2) the target output mask of  $x_{\text{non}}$  and the generated mask of  $\tilde{x}_{\text{non}}$  (**Figure 7B**), and lastly (3) the target output mask of  $x_{\text{non}}$  and the generated mask of  $\tilde{x}_{\text{non}}$  including only two classes, where `berry` and `berry-edge` are considered as one class (**Figure 7C**). We denote this mask as binary mask.

The different colors allow us to make a statement about the area in which, for example, berries are generated where none are present in the reference. The colors can be analyzed as follows: For **Figures 7A,B**, at pixel positions with a medium orange and medium blue discoloration, either the class `berry` is predicted to be an edge or the class `edge` is predicted to be a berry. These two cases are acceptable for our task, since we do not want to map the reference, but generate images, which provide highly probable results with a distribution that matches the input. The other pixel values are to be avoided, since at these positions for a light and dark orange discoloration the classes `berry` and `berry-edge` are generated, where in the reference `background` occurs. At the positions with a light and dark blue discoloration the class





background is generated, where in the reference the class berry or berry-edge is present. The generation map, where only two classes are included, highlights the non-acceptable pixel regions in the generated map.

## 4. RESULTS

### 4.1. Experimental Setup

Our experiments are designed to apply a domain-transfer using cGANs (section 3.1.2) to (1) learn a distribution by which we can generate a highly probable scenario of how occluded grapes could look like depending on the input, and (2) improve the counting of grapevine berries in images. To address the challenge of limited amount of natural data  $X^N$ , we perform four experiments based on a synthetic dataset  $X^S$ . In Experiment 5 (section 4.6), we show the applicability to natural data  $X^N$  based on the models and results learned in earlier experiments.

For our experiments, we define five different datasets, which are listed in Table 1. In addition to the natural data, described in section 3.3.1, we introduce a synthetic dataset in section 3.3.3. All five datasets, *Dataset 1-5*, will again be divided into the different types of defoliation SMPH and VSP. For our experiments, we also distinguish the set of input channels used. We claim that using a combination of grayscale image ( $G$ ) and berry mask ( $B$ ), denoted as  $GB$ , gives more accurate results both visually and in respect to berry counting than using the berry mask alone without grayscale information. We support this claim in Experiment 1. In the following experiments, the datasets are accordingly used with  $GB$  channels.

We resize all image patches to a uniform size of  $286 \times 286$  px with nearest neighbor interpolation. During training, we follow the procedure of Isola et al. (2017) and add small variations to the data in each epoch by randomly cropping patches of size  $256 \times 256$  px from the given patches. Additionally, patches are randomly flipped vertically, and the values within the patches are scaled and shifted to the range  $[-1, 1]$ . For testing, only scaling and shifting of the values to the range  $[-1, 1]$  is carried out. The network output is scaled back to the value range  $[0, 255]$  for visualization.

**TABLE 1** | Definitions of the used datasets.

Definition	$N$	$S$	SMPH	VSP	GB	B	Experiment				
							1	2	3	4	5
Dataset 1		X		X	X		X	X	X	X	
Dataset 2		X		X		X	X				
Dataset 3		X	X			X				X	
Dataset 4	X			X	X						X
Dataset 5	X		X		X						X

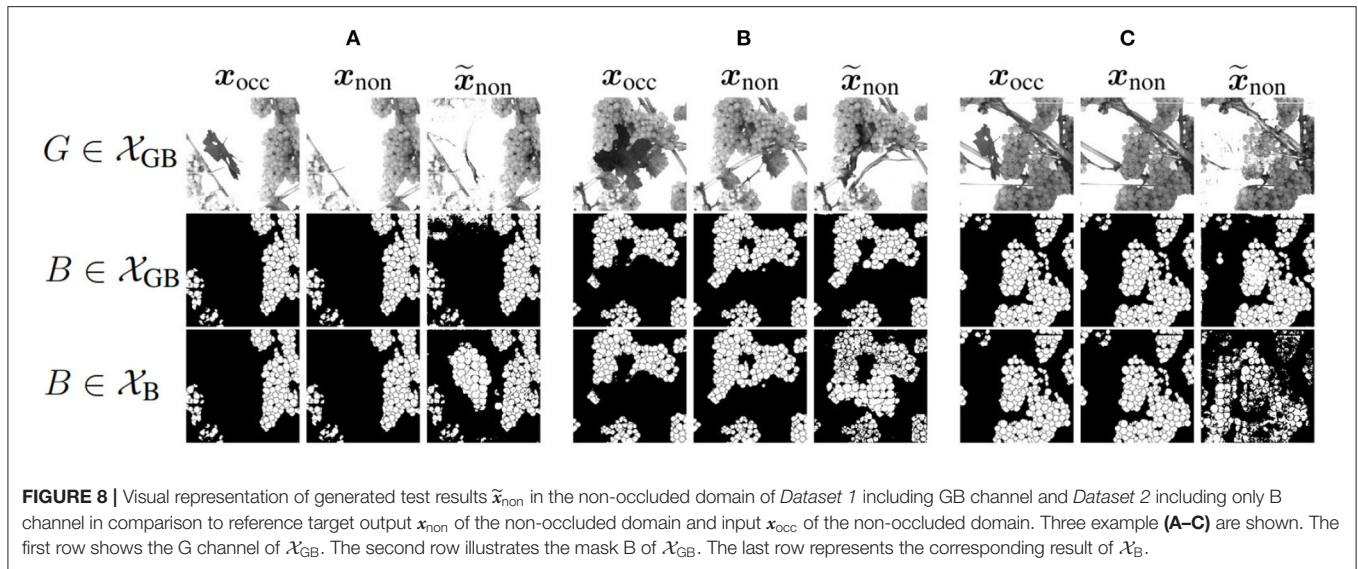
The table shows which kind of data is used for which experiment.

To train the models, we use an Intel Core i7-6850 K 3.60 GHz processor and two GeForce GTX 1080Ti with 11 GB RAM. The models are trained over 600 epochs. We use the Adam optimizer, where the learning rate is constant at 0.0004 for the first 300 epochs and is reduced linearly toward 0 for the last 300 epochs.

### 4.2. Experiment 1 – Comparison of Generation Quality Based on $GB$ and $B$ Data

With the first experiment, we analyze how the grayscale channel influences (i) the reproduction of hidden berries and (ii) the counting of berries per image. With the help of the grayscale channel  $G$ , it is possible to derive information about the presence of objects such as berries, leaves, and branches. Theoretically, this information helps to identify positions in the image where berries might be generated, for example, behind leaves or branches. In practice, however, in the non-occluded reference, a part of the berries is not present, since a proportion of berries is still occluded due to leaves or bigger branches not being cut away. This makes training more difficult, since it is generally learned that new berries should not be generated at the position of branches that have not been cut away. This implies, that we cannot expect to make new berries visible in the generated output  $\tilde{x}_{\text{non}}$  while testing, that are never present in the reference data  $x_{\text{non}}$  of the training set.

To get further insights into this, we analyze whether ignoring the  $G$  channel leads to a generation of berries in areas such as



branches. Moreover, we investigate if using channel B only is better suited on natural data, because information such as color, exposure, and lighting conditions have no influence. Thus, this experiment determines that the G channel adds value to the experiments and shows what this added value looks like.

#### 4.2.1. Used Data, Model, and Evaluation Metrics

For this experiment, we train a cGAN model on each of the training sets of *Dataset 1* and *Dataset 2*. The evaluation is based on the corresponding test sets. Since we want to determine the value of the G channel with this experiment, we limit the used data exclusively to defoliation type VSP. SMPH type shows proportionally similar outcomes to the VSP results.

To compare the two datasets  $X_B$  and  $X_{GB}$ , we use the described metrics in Section 3.4.2. We compare the correlation and the IoU in the occluded domain between the input  $x_{occ}$  and the generated input  $\tilde{x}_{occ}$ , as well as in the non-occluded domain between the target output  $x_{non}$  and the generated output  $\tilde{x}_{non}$  for both datasets. The generated input  $\tilde{x}_{occ}$  is computed by taking the generated output  $\tilde{x}_{non}$  and occlude the same pixels in the berry mask which are occluded in the input by a synthetic leaf.

#### 4.2.2. Results

**Figure 8** shows three example results to visually compare  $X_B$  and  $X_{GB}$ . The first two columns of an example show the reference of the two domains, where the third column represents the generated output  $\tilde{x}_{non}$ . The first row shows the grayscale channel of GB, the second row shows the mask channel of GB, and the bottom row shows the mask channel of B.

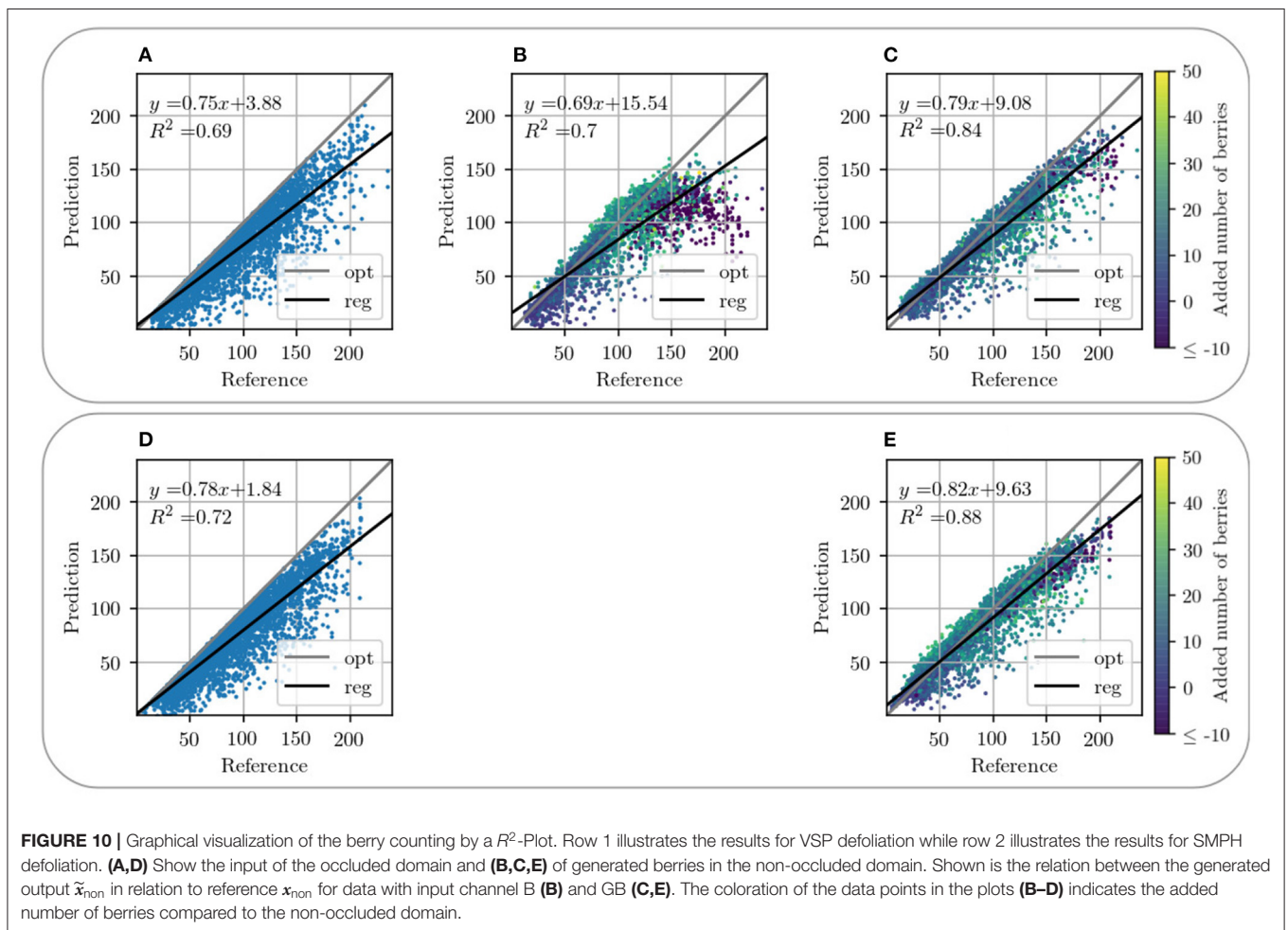
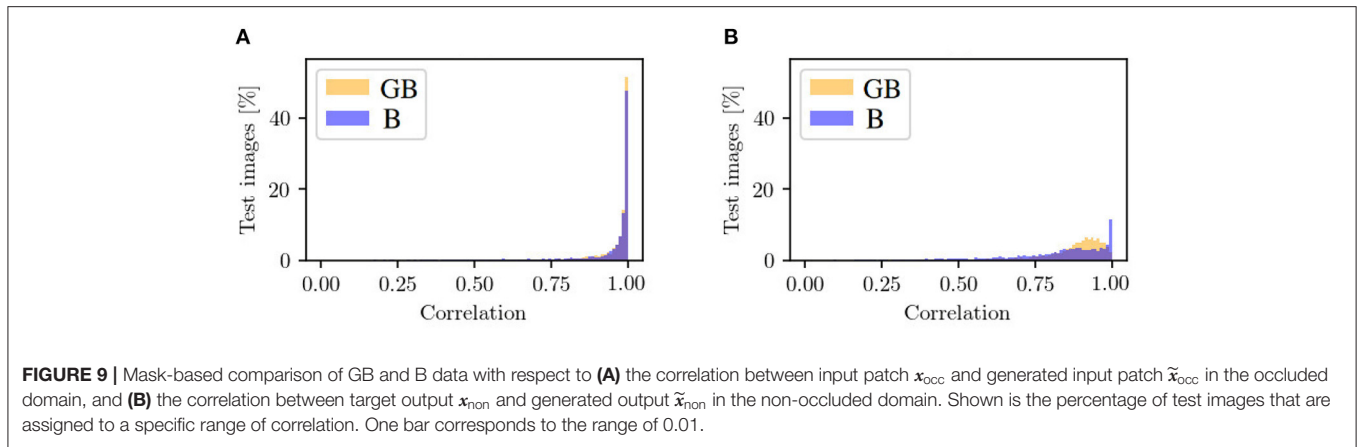
Using data without the G channel leads to higher generalizability regarding different varieties such as color, lighting conditions, and occlusions. Remarkable for the mask of B (row 3) is that for input patches containing many berries, proportionally too large and therefore too few berries are generated in  $\tilde{x}_{non}$  of the test results. This applies to the entire dataset and is demonstrated by **Figures 8A,B**. Generated berries

in  $\tilde{x}_{non}$  of  $X_{GB}$  adapt better to existing berries in mask  $x_{occ}$  than  $\tilde{x}_{non}$  of  $X_B$ . Furthermore, it turns out that the model trained on  $X_B$  has problems in generating patches with many berries. The berries are not only too big, but also in general berries are difficult to represent in their shape, as seen in **Figures 8B,C**.

Another positive aspect of  $X_{GB}$  is the already mentioned point that background information of the grayscale patch is included in the generation of new berries. The model learns to recognize where background is present in the patch and thus does not generate new berries in  $\tilde{x}_{non}$  in contrast to the model trained on  $X_B$ . This is particularly obvious in Example 1 (see **Figure 8A**), where a whole grape bunch is generated in the center of the mask. In the reference input and output of G, it is visible that on this position, background occurs.

In the following, we will take a look at the objective metrics described above. If we compare them regarding the  $X_{GB}$  and  $X_B$  input, we notice that the results for correlation between  $x_{occ}$  and  $\tilde{x}_{occ}$  are similar (see **Figure 9A**). For  $X_B$ , there are more generated patches with a correlation smaller than 0.8 and, therefore, less with a higher correlation. The correlation histogram between  $x_{non}$  and  $\tilde{x}_{non}$ , shown in **Figure 9B**, shows different distributions for the datasets. While the correlation histogram of BG, presented in orange, shows a left-skewed distribution, the amount of test patches of B increases on average with increasing correlation. At a correlation in the interval of [0.99, 1], represented by the right bar, the distribution shows a striking peak. However, there is a larger proportion of values below a correlation of 0.85. Even in the interval [0.85, 0.99], the percentages of patches for GB are higher than for B.

**Figures 9B, 10C** present a counting comparison of the different models in the non-occluded domain using a  $R^2$ -Plot. Additionally, **Figure 9A** shows the counting results without domain-transfer, i.e. no additional generated berries. Counting applied to the target  $x_{non}$  in the non-occluded domain serves



as the counting reference and is represented by the diagonal gray line. We observe that the results with input GB give the best matched results with respect to the reference. This is indicated visually as well as by the  $R^2$  value of the different models, which is the highest for our approach in the non-occluded domain with input GB. As in the visual

evaluation, the counting plot for input B in **Figure 10B** shows that the model indicates problems generating berries with a larger number of berries per patch. Also in the GB results, we observe that, especially with a reference counting number of more than 150 berries, the model does not reach the reference.

### 4.3. Experiment 2—Real vs. Generated Results in the Occluded Domain

In this experiment, we investigate whether the regions showing berries in the occluded domain stay unchanged in the transferred non-occluded domain. Furthermore, we verify that new berries are generated exclusively in the occluded area, and thus, the model detects where the appearance of berries is very likely.

#### 4.3.1. Used Data, Model, and Evaluation Metrics

For this experiment, we use synthetic *Dataset 1* of the VSP defoliation. For evaluation, we use different masks: The first mask is the so-called generated input mask  $\tilde{x}_{occ}$ , for which we take the generated output  $\tilde{x}_{non}$  of the test set and overlay it with the leaf used for data augmentation of the synthetic input  $x_{occ}$ . The other mask is the so-called baseline mask  $x_{non,leaf}$  of this experiment. For this purpose, we use the target output  $x_{non}$  and overlay it likewise with the leaf used for data augmentation of the synthetic input  $x_{occ}$ . Thus, only the non-occluded pixels of  $x_{occ}$  will remain visible in  $\tilde{x}_{occ}$  and  $x_{non,leaf}$ . The evaluation is then performed on the pairs  $\{x_{occ}, x_{occ,leaf}\}$  and  $\{x_{occ}, \tilde{x}_{occ}\}$ .

We use IoU and correlation as comparative metrics for this experiment. Additionally, we create generation maps which show the differences between the masks within each of the pairs  $\{x_{occ}, x_{occ,leaf}\}$  and  $\{x_{occ}, \tilde{x}_{occ}\}$ , as illustrated in **Figure 11**. For this experiment, the first three rows are of interest to us. The first row shows the respective grayscale patch of the generation maps. The second row shows the differences within the pair  $\{x_{occ}, x_{occ,leaf}\}$ . Row three shows the differences within the pair  $\{x_{occ}, \tilde{x}_{occ}\}$ . The columns indicate different patch examples.

#### 4.3.2. Results

The reference correlation within the mask pair  $\{x_{occ}, x_{occ,leaf}\}$  is above 0.98 for all test patches. With our method, we manage to achieve a correlation of over 0.98 within the pair  $\{x_{occ}, \tilde{x}_{occ}\}$  for about 65% of the test images (see **Figure 9A**, orange). The remaining 35% are largely distributed over a correlation within the interval [0.75, 0.98]. The correlation strongly correlates with the IoU calculation of the *berry* area. The low correlations are either due to artifacts in the generated masks or to test images with a high number of berries. In this case, the model does not transfer all non-occluded pixels one to one into the non-occluded domain. The effect of the amount of berries in the patch is shown in the generation maps in **Figure 11** {column 1, row 3} and {column 3, row 3}.

For the patch examples in columns 2, 4, and 5, the generation maps of the pairs  $\{x_{occ}, x_{occ,leaf}\}$  are almost identical to the generation maps of the pairs  $\{x_{occ}, \tilde{x}_{occ}\}$ . Such maps correspond to correlation values close to 1. It is noticeable that in all five examples, the border of the leaf used for data augmentation is highlighted in the generation maps. The coloring occurs at transitions between the leaf and the adjacent *berry-edge* and *berry* pixel.

### 4.4. Experiment 3—Real vs. Generated Results in the Non-occluded Domain

In this experiment, we investigate the similarity of our generated output  $\tilde{x}_{non}$  compared to the target output  $x_{non}$  in the non-occluded domain.

#### 4.4.1. Used Data, Model, and Evaluation Metrics

In this experiment, *Dataset 1* is used to train the model. Since we are aiming only for a highly probable result rather than the exact position and shape of specific berries, for our evaluation, we additionally create a binary mask based on the berry mask, which includes only the classes *berry* and *background*. For this, we merge the classes *berry* and *berry-edge*. We compare the mask pair  $\{x_{non}, \tilde{x}_{non}\}$  of the non-occluded domain in respect to the berry and binary mask. We evaluate the correlation and IoU within this pair. Furthermore, we create generation maps that illustrate the difference between this pair. Exclusively for the berry mask, we calculate the area and diameter of all individual berries in the entire test data set.

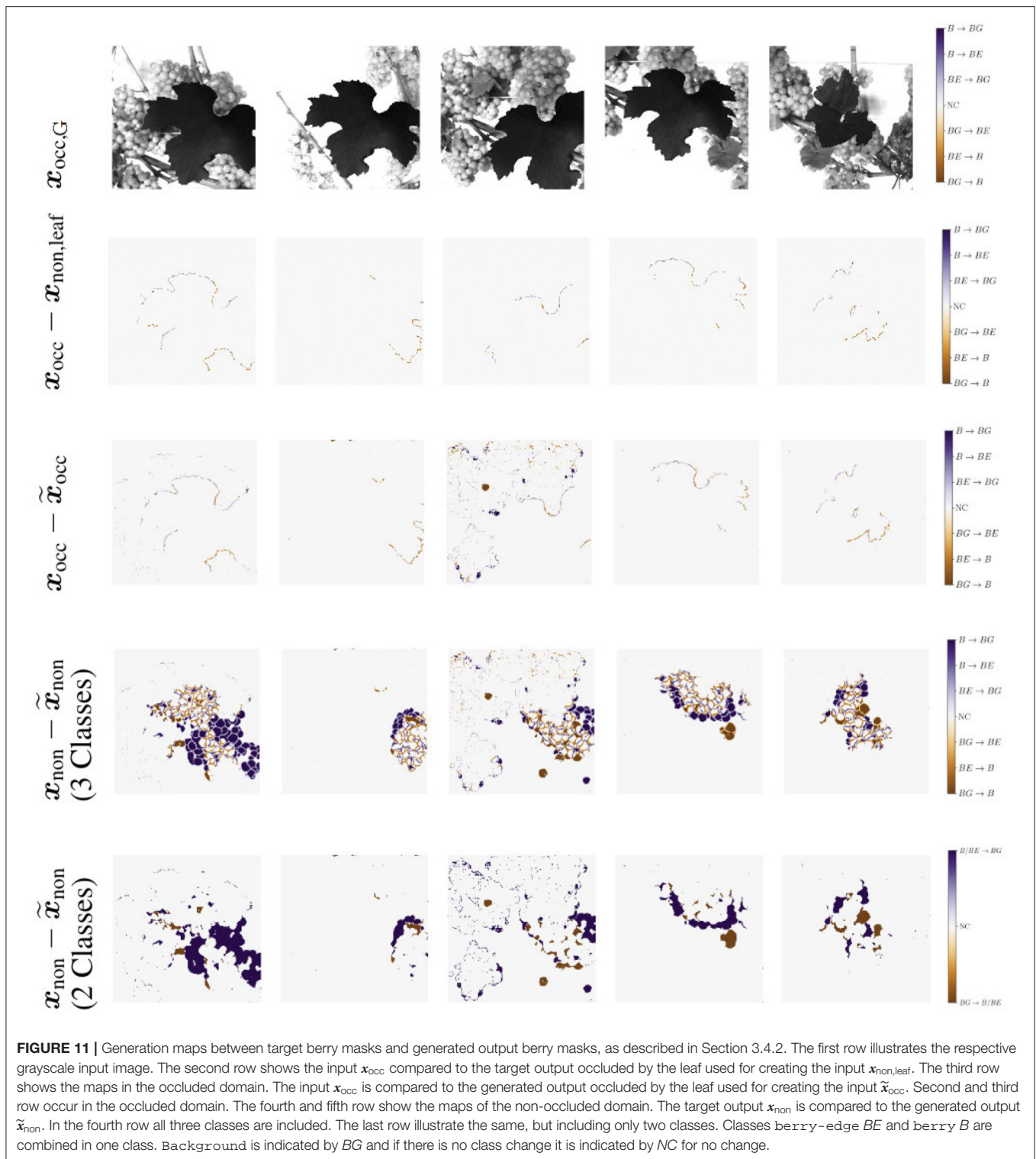
#### 4.4.2. Results

The correlation (**Figure 12A**) shows a similar left-skewed distribution for berry mask and binary mask. The majority of the test images show a correlation of above 0.8. Although our approach does not aim to generate the exact position and shape of berries, the results indicate that the similarity of the generated results and the reference are high. The IoU in **Figure 12B** also supports this finding. The IoU of the binary mask has on average higher values and is closer to the possible maximum than the berry mask. The generation maps from **Figure 11** also show this property in the fourth and fifth row. The fourth row shows example results for the berry mask, where two cases can be seen. *Case 1*: The medium orange and medium blue colors in the fourth row illustrate pixels where the classes *berry* and *berry-edge* are confused. This incorrect generation is acceptable due to the desired property of highly probable results instead of exactly matching results. *Case 2*: Dark and light blue, and dark and light orange are incorrectly generated classes that need to be avoided. In the fifth row, these pixel regions are highlighted by dark blue and dark orange. These regions either represent berries where there are no berries in the reference, or *vice versa*. Such incorrect generations shift the position and size of the grape bunches. In the example maps, however, it can be seen that *Case 1* occurs predominantly. It is obvious that berries are predicted in the right areas, but their shape and position do not correspond exactly to the reference.

At the transition from image areas with berries to background pixels, the second case occurs where too small or too large grape bunches are produced, because either too few or too many berries are generated. This is illustrated by the second and fourth column. The generation maps of the binary masks only highlight the areas that contradict the property of highly probable results.

To further check the similarity between generated and reference data, we consider the distributions for area and

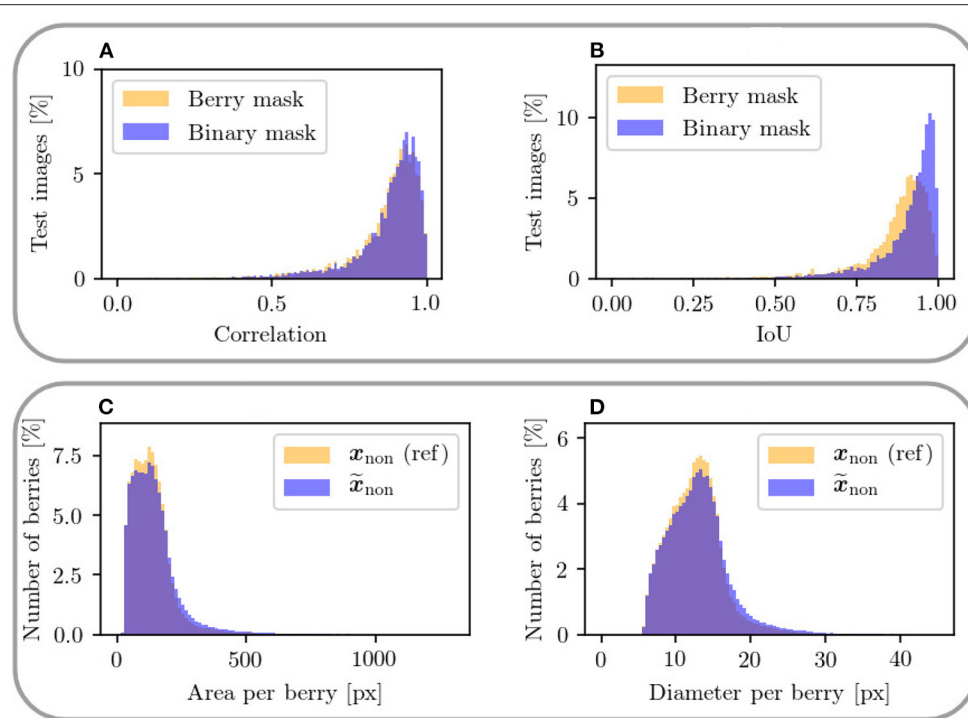




diameter within the berry masks shown in Figures 12C,D. The distributions of the metrics are highly similar between generated result and reference. For both metrics, there is a slight tendency toward an increase in area and diameter for the generated berries.

#### 4.5. Experiment 4—Counting in the Non-occluded Domain

Since the number of berries is of high importance for yield estimation, we investigate the estimation of this number in this experiment. We compare the counts based on the input



**FIGURE 12 |** The upper plots show a mask-based comparison within the non-occluded domain between berry mask and binary mask including only two classes for the metrics **(A)** correlation between  $x_{\text{non}}$  and  $\hat{x}_{\text{non}}$  and **(B)** IoU of the berry pixel in  $x_{\text{non}}$  and  $\hat{x}_{\text{non}}$ . The lower plots **(C,D)** show a comparison of area and diameter per berry between target output  $x_{\text{non}}$  and generated output  $\hat{x}_{\text{non}}$  in the non-occluded domain. Only areas up to 1,300 px and diameters up to 45 px are plotted.

patches in the occluded domain and the target patches in the non-occluded domain with the generated results of our approach.

#### 4.5.1. Used Data, Model, and Evaluation Metrics

For this experiment, we use the synthetic datasets *Dataset 1* and *Dataset 3* based on VSP and SMPH defoliation. Our model is trained on both training sets and evaluated on the corresponding test sets. During testing, we consider only the mask of the data patches. For the evaluation, we use the  $R^2$ -Plot to plot the absolute count of the input (**Figures 10A,D**) and the absolute count of the generated output of our method (**Figures 10C,E**) with the reference count from the target mask, respectively. Furthermore, we examine the distribution of the relative deviations from the reference (see **Figure 13**).

#### 4.5.2. Results

Counting in the occluded domain, presented in **Figures 10A,D**, shows that there is an underestimation of the number of berries compared to the reference. Our model shows a shift of the number of berries toward the reference for both types of defoliation. In both cases, the  $R^2$  value increases compared to the  $R^2$  value of the occluded domain, which corresponds to a better approximation of the data compared to the reference. It is important to mention that not only the sample distribution shifts, but also compresses and concentrates along the reference line.

**Figure 13** supports this observation. The plots show the relative difference of the counted berries in the occluded domain

and our method in the non-occluded domain compared to the reference counting. Our method (blue) depicts a normal distribution with a mean near zero. If the values of the occluded distribution (orange) were increased by a factor, this would lead to a shift in the distribution, but it would still be more stretched than ours. The peaks at value 0 correspond mostly to synthetic images where the synthetic leaf does not cover any berries. This is the case, for example, with images that show few berries.

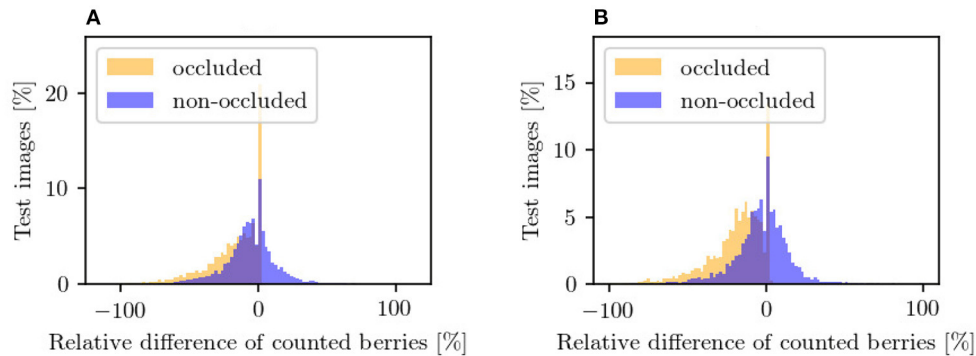
Both models exhibit problems in the generation of patches that depict more than 150 berries. This is the case for VSP (**Figure 10C**) and SMPH (**Figure 10E**). For both types of defoliation, a trend is nevertheless evident above the critical value of 150 berries. Even though an underestimation of berries tends to be counted above this value, the count fits the reference better than the count in the occluded domain.

In the occluded domain, there are data points that differ strongly from the reference. Our method reduces the amount of such points and also reduces the deviation of the highly deviating points.

## 4.6. Experiment 5—Application to Natural Data

One of the contributions of our work is to investigate the applicability of our approach to natural data. In detail, we evaluate whether our model generalizes to natural images when it is trained on synthetic data.





**FIGURE 13** | Counting in the occluded domain (orange) and after applying our approach in the non-occluded domain (blue) relative to the reference counting in the non-occluded domain. The plots illustrate the results for **(A)** VSP defoliation and **(B)** SMPH defoliation. A negative value means that fewer berries are counted than in the reference and *vice versa*. Each bar corresponds to a width of 2%.

#### 4.6.1. Used Data, Model, and Evaluation Metrics

We use the synthetic datasets *Dataset 1* and *Dataset 3* to train our model. For the test phase, we use the natural datasets *Dataset 4* and *Dataset 5*. One dataset each for VSP defoliation and one for SMPH defoliation.

The differences of the natural dataset to the synthetic dataset are the stronger coverage by a denser leaf canopy, the resulting deviating exposure ratios, and the lower contrast whereby the contours of the leaves are not easily distinguishable from berries. Other differences are found in the transformation applied to the natural dataset, since non-occluded areas are not identical in both domains, as already pointed out in the introduction. Depending on the patch position in the non-occluded domain in the original image, the transformation goes beyond the boundaries of the original image in the occluded domain. To achieve a patch size of  $656 \times 656$  px which is equivalent to the cropped patch size of the dataset, the appropriate borders of the patch are filled with black pixels.

We perform our evaluation visually, which means we compare the input from the occluded domain with the generated output of our approach in the non-occluded domain. Due to the transformation issues, direct numerical comparison and evaluation between target and generated output are not useful for the majority of patches. However, we would like to give an impression of the results by means of the visual representation.

#### 4.6.2. Results

In **Figure 14**, we provide example results of our approach applied to natural data. For each example, the first column shows the input  $x_{occ}$  of the occluded domain, the second column the reference  $x_{non}$  in the non-occluded domain, and the last column our generated output  $\tilde{x}_{non}$  in the non-occluded domain. The first row visualizes the G channel of a patch and the second row the corresponding mask. The results show that the canopy is reduced and important areas in the patch are reproduced. Generally, the observations from the previously described experiments can be repeated. Using our generative approach, *berry* and *berry-edge* pixel regions in the input

mask are also transferred to the generated output for the natural data. For input patches of the occluded domain being similar to the synthetic data (**Figures 14A–C**), the results show an expansion of the existing berry region. Our approach is also able to deal with transformation problems, as in **Figure 14A** where the transformation goes beyond the original image boundaries. There are examples, like seen in **Figure 14C**, that look similar to the target, or examples that look real compared to the input but do not reflect the target output (**Figure 14B**).

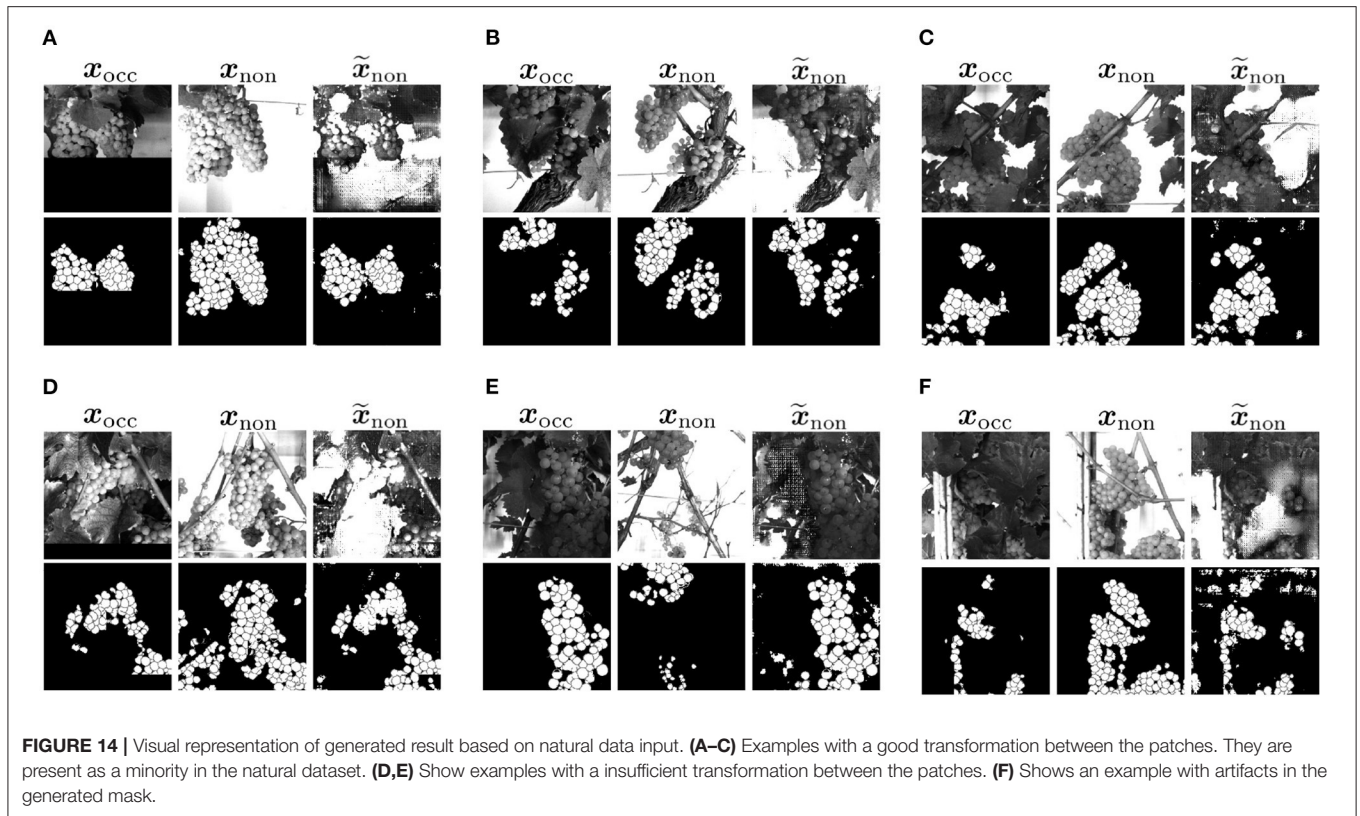
For the majority of natural data, exact transformations are not available, so this is challenging to evaluate. In examples like the one in **Figure 14D** transformation, rotation and scale fit, but due to defoliation, the orientation of the grape bunch is different in input and output target. In the input, the grape bunch is more horizontal. In the target, it is vertical. The example in **Figure 14E** shows that grape bunches are also completely different in translation due to the different weights attached to the branches. In this example, the grape bunch that is visible in the input is only partially visible at the top of the patch in the target output. The generated output adapts to the input and is also expanded, but is not comparable to the target.

Furthermore, we observe checkerboard artifacts that appear in the generated G patches (see **Figures 14A,C**). The artifacts occur more in patches that present a dense canopy.

## 5. DISCUSSION

### 5.1. Experiment 1

Our results confirm that our model trained on GB data learns where background is present. This is an important factor for realistic generated images. We found that the model trained only with berry mask B has more problems with images containing many berries than the model trained on GB data, both visually and in the counting results. The deficits in counting are explained by the fact that there are relatively few patches in the dataset with a number greater than 150 compared to the number of patches containing  $<150$  berries. This is also true for the underestimation of the count with the GB dataset. However,



**FIGURE 14** | Visual representation of generated result based on natural data input. (A–C) Examples with a good transformation between the patches. They are present as a minority in the natural dataset. (D,E) Show examples with an insufficient transformation between the patches. (F) Shows an example with artifacts in the generated mask.

by using the additional G channel, the result images can be generated more precisely. More detailed analyses of the berry counting can be found in Experiment 4. Taking into account the correlations and with the goal to generate highly probable results with a distribution that matches the input, rather than the exact image content of each image, Dataset 1 leads to better results on average as claimed in the beginning of the results section.

## 5.2. Experiment 2

We found that a high percentage of the results is correctly transferred from occluded to non-occluded domain. The occurring deviations between  $\{x_{occ}, \tilde{x}_{occ}\}$  can be traced back to the test results, which not only show the class values 0, 127, and 255 within the mask, but also pixels with values in between. This means that the model does not clearly assign the respective pixel to a class. At this point, we apply data post-processing to our generated data, as described in Section 3.4.1. Pixel in areas of class boundaries are particularly affected here, which is why the differences arise in these areas.

The deviations at the edges of the leaf are due to an additional edge with a width of about three pixels, which was added during the creation of the synthetic occluded input mask. The masks  $\tilde{x}_{non}$  and  $x_{non}$ , on which the masks  $\tilde{x}_{occ}$  and  $x_{occ,leaf}$  used in this experiment are based on, show a continuation of the depicted grape branches exactly at these transitions. This results in variations between the paired masks at this location.

The key findings from this experiment are that despite individual deviations, the visible part of the mask of the occluded

domain is safely transferred to the non-occluded domain and stays unchanged. We assume that the model will make no result-altering changes.

## 5.3. Experiment 3

Although our approach does not aim to generate the exact position and shape of berries, the results indicate that the similarity of the generated results and the references are high. The observed high IoU indicates a similar position of the grape bunches independent of the berry objects in the generated result compared to the reference. Berries are predicted in the right areas, but their shape and position do not correspond exactly to the reference. An increasing area and diameter suggest, that if the area of the total berry pixel per patch remains the same, there is a possibility that too few berries are predicted.

## 5.4. Experiment 4

In the berry counting, the underestimation of the amount of berries per patch is clearly evident in the concealed area, which can be explained by the occlusion covering part of the berries. The results indicate that we obtain better results with our approach than when we apply only a factor to the counting. We explain the deteriorating results above a berry number of 150 by the fact that the proportion of training images with a count above the critical value is relatively small in contrast to the number of images with an amount below the critical value. Our method reduces the number of outliers and additionally reduces the variance of the highly deviant points. We achieve a shift of the distribution

as well as a compression and concentration along the reference line, so that our results are more accurate than those in the occluded domain.

## 5.5. Experiment 5

Generally, our findings from the previously described experiments can be confirmed within this experiment. Although it is apparent that the model trained only on the synthetic data mentioned above is not yet strong enough to obtain similarly good results for the more complex natural data as for the synthetic data, we consider the results promising. We assume that mixing natural and synthetic data or using more complex synthetic training data can improve the results. The checkerboard artifacts that we observed could be reduced by improving the generator (Odena et al., 2016). This could also result in reduced artifacts, like they occur in the mask in **Figure 14F**. The artifacts occur more in patches that present a dense canopy.

## 5.6. Future Directions

To make the model more robust and generalizable to variations between natural and synthetic data, the synthetic data can be designed with more complex changes, for example, by increasing the synthetic occlusion through the use of more leaves per patch. In addition, brightness and contrast could be varied, for example, to reduce the dominant white background of the synthetic data and thus make it more difficult for the model to detect the occlusion. Interesting future work is the application of the model to other varieties and to see how it behaves. We assume that the model applied on varieties with a comparable or smaller grape bunch size and a similar data appearance will behave similarly to our presented results. With a larger grape bunch size and thus a larger number of berries, the model might have to be re-trained in order to achieve an accurate result for a large number of berries. Another promising future direction is to train the model from a combination of synthetic images and a limited amount of natural images. In this case, the transformation between the two required domains needs to be accurate enough and suitable data must be selected. Another possibility would involve extensive manual work on the transformation between the domains or more sophisticated techniques such as image warping. In the future, the checkerboard artifacts that occur in data could be reduced by replacing the transpose convolution layer of the decoder in the U-Net generator with bi-linear up-sampling operations, as described by Odena et al. (2016).

## 6. CONCLUSION

In this work, we have demonstrated the suitability of a conditional generative adversarial network like Pix2Pix to

generate a scenario behind occlusions in grapevine images that is highly probable based on visible information in the images. Our experiments have shown that our approach has learned patterns that characterize typical berries and clusters without occlusions so that areas where berries are added and other areas where the image remains unchanged can be identified without having to provide prior knowledge about occlusions. Compared to counting with occluded areas, we show that our approach provides a count that is closer to the manual reference count. In contrast to applying a factor, our approach directly involves the appearance of the visible berries and thus better adapts to local conditions.

We have trained our conditional adversarial network-based model on synthetic data only in order to overcome the challenge of lacking aligned image pairs. We show that the model is also applicable to natural data, given that the canopy is not too dense and the variation between natural data and synthetic data is not too high.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

JK initiated, designed, and conducted the analyses. RR helped to initiate the work and co-designed the experiments. JK, AK, and LZ contributed to the data preparation. All authors contributed to the writing of this manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This project was funded by the European Agriculture Fund for Rural Development with contribution from North-Rhine Westphalia (17-02.12.01--10/16---EP-0004617925-19-001). Furthermore, this work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070—390732324, and partially by the German Federal Ministry of Education and Research (BMBF, Bonn, Germany) in the framework of the project novisys (FKZ 031A349).

## ACKNOWLEDGMENTS

The content of the manuscript have previously appeared in a preprint (Kierdorf et al., 2021).

## REFERENCES

- Aquino, A., Diago, M. P., Millán, B., and Tardáguila, J. (2017). A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosyst. Eng.* 156, 80–95. doi: 10.1016/j.biosystemseng.2016.12.011
- Aquino, A., Millan, B., Diago, M.-P., and Tardaguila, J. (2018). Automated early yield prediction in vineyards from on-the-go image acquisition.

- Comput. Electron. Agric.* 144, 26–36. doi: 10.1016/j.compag.2017.11.026
- Arteta, C., Lempitsky, V., and Zisserman, A. (2016). “Counting in the wild,” in *European Conference on Computer Vision* (Springer), 483–498. doi: 10.1007/978-3-319-46478-7\_30
- Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2009). Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* 28, 24. doi: 10.1145/1531326.1531330
- Batista, G. E., and Monard, M. C. (2002). A study of K-nearest neighbour as an imputation method. *His* 87, 48. doi: 10.1109/METRIC.2004.1357895
- Bertalmio, M., Vese, L., Sapiro, G., and Osher, S. (2003). Simultaneous structure and texture image inpainting. *IEEE Trans. Image Process.* 12, 882–889. doi: 10.1109/TIP.2003.815261
- Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR, abs/1802.02611*. doi: 10.1007/978-3-030-01234-2\_49
- Clingeffer, P. R., Martin, S., Dunn, G., and Krstic, M. (2001). *Crop Development, Crop Estimation and Crop Control to Secure Quality and Production of Major Wine Grape Varieties: A National Approach*. Final Report. Grape and Wine Research & Development Corporation.
- Coviello, L., Cristoforetti, M., Jurman, G., and Furlanello, C. (2020). GBCNet: in-field grape berries counting for yield estimation by dilated CNNs. *Appl. Sci.* 10, 4870. doi: 10.3390/app10144870
- Dekel, T., Gan, C., Krishnan, D., Liu, C., and Freeman, W. T. (2018). “Sparse, smart contours to represent and edit images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 3511–3520. doi: 10.1109/CVPR.2018.00370
- Diago, M., Martinez De Toda, F., Poni, S., and Tardaguila, J. (2009). “Early leaf removal for optimizing yield components, grape and wine composition in tempradillo (*Vitis vinifera* L.),” in *Proceedings of the 16th International GiESCO Symposium*, ed J. A. Wolpert, Davis, CA, 113–118.
- Diago, M.-P., Correa, C., Millán, B., Barreiro, P., Valero, C., and Tardaguila, J. (2012). Grapevine yield and leaf area estimation using supervised classification methodology on rgb images taken under field conditions. *Sensors* 12, 16988–17006. doi: 10.3390/s121216988
- Ehsani, K., Mottaghi, R., and Farhadi, A. (2018). “Segan: segmenting and generating the invisible,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 6144–6153. doi: 10.1109/CVPR.2018.00643
- Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Struct. Equat. Model.* 8, 128–141. doi: 10.1207/S15328007SEM0801\_7
- Feng, H., Yuan, F., Skinkis, P. A., and Qian, M. C. (2015). Influence of cluster zone leaf removal on pinot noir grape chemical and volatile composition. *Food Chem.* 173, 414–423. doi: 10.1016/j.foodchem.2014.09.149
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.
- Hacking, C., Poon, N., Manzan, N., and Poblete-Echeverría, C. (2019). Investigating 2-D and 3-D proximal remote sensing techniques for vineyard yield estimation. *Sensors* 19, 3652. doi: 10.3390/s19173652
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 2961–2969. doi: 10.1109/ICCV.2017.322
- Helmert, F. (1880). *Die Mathematischen Physicalischen Theorien der höheren Geodäsie*. B. G. Teubner. Available online at: <https://books.google.de/books?id=g0vkwQEACAAJ>
- Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Trans. Graph.* 36, 1–14. doi: 10.1145/3072959.3073659
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 1125–1134. doi: 10.1109/CVPR.2017.632
- Kicherer, A., Herzog, K., Bendel, N., Klück, H.-C., Backhaus, A., Wieland, M., et al. (2017). Phenoliner: a new field phenotyping platform for grapevine research. *Sensors* 17, 1625. doi: 10.3390/s170471625
- Kicherer, A., Roscher, R., Herzog, K., Förstner, W., and Töpfer, R. (2014). “Image based evaluation for the detection of cluster parameters in grapevine,” in *XI International Conference on Grapevine Breeding and Genetics 1082*, Yanqing, 335–340. doi: 10.17660/ActaHortic.2015.1082.46
- Kierdorf, J., Weber, I., Kicherer, A., Zabawa, L., Drees, L., and Roscher, R. (2021). Behind the leaves-estimation of occluded grapevine berries with conditional generative adversarial networks. *arXiv preprint arXiv:2105.10325*. doi: 10.48550/arXiv.2105.10325
- Kim, J. K., and Rao, J. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* 96, 917–932. doi: 10.1093/biomet/asp041
- Lee, D., Kim, J., Moon, W.-J., and Ye, J. C. (2019). “Collagan: collaborative GAN for missing image data imputation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2487–2496. doi: 10.1109/CVPR.2019.00259
- Lempitsky, V., and Zisserman, A. (2010). Learning to count objects in images. *Adv. Neural Inform. Process. Syst.* 23, 1324–1332.
- Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. (2018). “Image inpainting for irregular holes using partial convolutions,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 85–100. doi: 10.1007/978-3-030-01252-6\_6
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94
- Mack, J., Lenz, C., Teutrine, J., and Steinhage, V. (2017). High-precision 3d detection and reconstruction of grapes from laser range data for efficient phenotyping based on supervised learning. *Comput. Electron. Agric.* 135, 300–311. doi: 10.1016/j.compag.2017.02.017
- Mack, J., Schindler, F., Rist, F., Herzog, K., Töpfer, R., and Steinhage, V. (2018). Semantic labeling and reconstruction of grape bunches from 3D range data using a new RGB-D feature descriptor. *Comput. Electron. Agric.* 155, 96–102. doi: 10.1016/j.compag.2018.10.011
- May, P. (1972). Forecasting the grape crop. *Australian Wine, Brewing and Spirit Review*. 90, 46–48.
- Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*. doi: 10.48550/arXiv.1411.1784
- Nuske, S., Achar, S., Bates, T., Narasimhan, S., and Singh, S. (2011). “Yield estimation in vineyards by visual grape detection,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, CA, 2352–2358. doi: 10.1109/IROS.2011.6095069
- Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., and Singh, S. (2014). Automated visual yield estimation in vineyards. *J. Field Robot.* 31, 837–860. doi: 10.1002/rob.21541
- Nyarko, E. K., Vidović, I., Radočaj, K., and Cupec, R. (2018). A nearest neighbor approach for fruit recognition in RGB-D images based on detection of convex surfaces. *Expert Syst. Appl.* 114, 454–466. doi: 10.1016/j.eswa.2018.07.048
- Odena, A., Dumoulin, V., and Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill* 1, e3. doi: 10.23915/distill.00003
- Ostyakov, P., Suvorov, R., Logacheva, E., Khomenko, O., and Nikolenko, S. I. (2018). Seigan: Towards compositional image generation by simultaneously learning to segment, enhance, and inpaint. *arXiv preprint arXiv:1811.07630*. doi: 10.48550/arXiv.1811.07630
- Paul Cohen, J., Boucher, G., Glastonbury, C. A., Lo, H. Z., and Bengio, Y. (2017). “Count-ception: counting by fully convolutional redundant counting,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Venice, 18–26. doi: 10.1109/ICCVW.2017.9
- Robins, J. M., and Wang, N. (2000). Inference for imputation estimators. *Biometrika* 87, 113–124. doi: 10.1093/biomet/87.1.113
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241. doi: 10.1007/978-3-319-24574-4\_28
- Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R., and Förstner, W. (2014). Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. *Comput. Electron. Agric.* 100, 148–158. doi: 10.1016/j.compag.2013.11.008
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *J. Am. Stat. Assoc.* 91, 473–489. doi: 10.1080/01621459.1996.10476908



- Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys, Vol. 81*. John Wiley & Sons.
- Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., and Chen, L. (2018). Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation. *CoRR, abs/1801.04381*. doi: 10.1109/CVPR.2018.00474
- Schöler, F., and Steinhage, V. (2015). Automated 3D reconstruction of grape cluster architecture from sensor data for efficient phenotyping. *Comput. Electron. Agric.* 114, 163–177. doi: 10.1016/j.compag.2015.04.001
- Van Buuren, S., and Oudshoorn, K. (1999). *Flexible Multivariate Imputation by MICE*. Leiden: TNO.
- von Hippel, P. T., and Bartlett, J. (2012). Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws. *Statistical Sci.* 36, 400–420. doi: 10.1214/20-STS793
- Xie, W., Noble, J. A., and Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng.* 6, 283–292. doi: 10.1080/21681163.2016.1149104
- Xiong, W., Yu, J., Lin, Z., Yang, J., Lu, X., Barnes, C., et al. (2019). “Foreground-aware image inpainting,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 5840–5848. doi: 10.1109/CVPR.2019.00599
- Yan, X., Wang, F., Liu, W., Yu, Y., He, S., and Pan, J. (2019). “Visualizing the invisible: occluded vehicle segmentation and recovery,” in *Proceedings of the IEEE International Conference on Computer Vision*, Salt Lake City, UT, 7618–7627. doi: 10.1109/ICCV.2019.00771
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. (2018). “Generative image inpainting with contextual attention,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 5505–5514. doi: 10.1109/CVPR.2018.00577
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. (2019). “Free-form image inpainting with gated convolution,” in *Proceedings of the IEEE International Conference on Computer Vision*, Long Beach, CA, 4471–4480. doi: 10.1109/ICCV.2019.00457
- Zabawa, L., Kicherer, A., Klingbeil, L., Milioto, A., Topfer, R., Kuhlmann, H., et al. (2019). “Detection of single grapevine berries in images using fully convolutional neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA. doi: 10.1109/CVPRW.2019.00313
- Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., and Roscher, R. (2020). Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS J. Photogr. Remote Sens.* 164, 73–83. doi: 10.1016/j.isprsjprs.2020.04.002



**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kierdorf, Weber, Kicherer, Zabawa, Drees and Roscher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# GrowliFlower: An image time-series dataset for GROWth analysis of cauLIFLOWER

Jana Kierdorf<sup>1</sup>  | Laura Verena Junker-Frohn<sup>2</sup> | Mike Delaney<sup>3</sup> |  
 Mariele Donoso Olave<sup>1</sup> | Andreas Burkart<sup>3</sup> | Hannah Jaenicke<sup>4</sup> | Onno Muller<sup>2</sup> |  
 Uwe Rascher<sup>2</sup> | Ribana Roscher<sup>1</sup> 

<sup>1</sup>Institute of Geodesy and Geoinformation, Remote Sensing Group, Bonn, Germany

<sup>2</sup>Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, Jülich, Germany

<sup>3</sup>JB Hyperspectral Devices GmbH, Düsseldorf, Germany

<sup>4</sup>Horticulture Competence Centre, University of Bonn, Bonn, Germany

## Correspondence

Jana Kierdorf, Institute of Geodesy and Geoinformation, Remote Sensing Group, University of Bonn, Niebuhrstraße 1A, 53113, Bonn, Germany.  
 Email: [jkierdorf@uni-bonn.de](mailto:jkierdorf@uni-bonn.de)

## Funding information

European Agriculture Fund for Rural Development with contribution from North-Rhine Westphalia (17-02.12.01–10/16–EP-0004617925-19-001); Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070—390732324

## Abstract

In this paper, we present GrowliFlower, a georeferenced, image-based unmanned aerial vehicle time-series dataset of two monitored cauliflower fields (0.39 and 0.60 ha) acquired in 2 years, 2020 and 2021. The proposed dataset contains RGB and multispectral orthophotos with coordinates of approximately 14,000 individual cauliflower plants. The coordinates enable the extraction of complete and incomplete time-series of image patches showing individual plants. The dataset contains the collected phenotypic traits of 740 plants, including the developmental stage and plant and cauliflower size. The harvestable product is completely covered by leaves, thus, plant IDs and coordinates are provided to extract image pairs of plants pre- and post-defoliation. In addition, to facilitate classification, detection, segmentation, instance segmentation, and other similar computer vision tasks, the proposed dataset contains pixel-accurate leaf and plant instance segmentations, as well as stem annotations. The proposed dataset was created to facilitate the development and evaluation of various machine-learning approaches. It focuses on the analysis of growth and development of cauliflower and the derivation of phenotypic traits to advance automation in agriculture. Two baseline results of instance segmentation tasks at the plant and leaf level based on labeled instance segmentation data are presented. The complete GrowliFlower dataset is publicly available (<http://rs.ipb.uni-bonn.de/data/growliflower/>).

## KEYWORDS

agricultural plant dataset, crop development, crop growth, instance segmentation, machine learning, plant monitoring, UAV

## 1 | INTRODUCTION

Field-grown crops are strongly affected by environmental conditions, thus, to minimize yield losses due to abiotic or biotic stresses, crop production involves careful plant management and complex

decisions. Farmers support plant growth and development through irrigation, fertilization, weeding, and pesticides applications; however, these are costly and labor-intensive processes. To optimize plant management and support effective decision-making, farmers rely on frequent crop monitoring; however, this is also time-consuming

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Field Robotics* published by Wiley Periodicals LLC.

process that requires expert knowledge. Typically, farmers and agricultural advisors monitor fields regularly via spot checks of individual plants. Here, remote sensing and analysis methods can help farmers monitor entire fields more comprehensively (Chi et al., 2016; Weiss et al., 2020), and remote sensing data can be acquired at any scale without damaging or impacting the crops. Large-scale observations from satellites or aircraft and medium-scale observations from unmanned aerial vehicles (UAVs) provide an overview of larger agricultural areas (Lillesand et al., 2015). Large-area sensor-based crop monitoring makes it possible to detect heterogeneity in the field and support the farmer's decision-making in terms of field management. With such detailed, area-wide information on biotic and abiotic stress, these factors can be counteracted more selectively to support environmentally friendly plant management. Medium-scale and close-range observations acquired from UAVs and ground robots are beneficial for collecting detailed information and can be used particularly well for phenotyping individual plants. For example, Nock et al. (2016) used optical remote sensing data to define various traits, for example, structural and phenotypical characteristics at all levels, from individual plants to large areas. Other applications using remote sensing data include yield estimation (Chaparro et al., 2018), yield forecasting (Mosleh et al., 2015), and monitoring rapid land surface changes (Verger et al., 2014).

Machine learning (ML) methods have become increasingly important (Lary et al., 2016) in processing and interpreting large amounts of remote sensing data. ML involves learning a predictive function that relates observations to the desired output, and trained models can be designed flexibly relative to the type of observations (Debolini et al., 2015; Reichstein et al., 2019). For example, using ML techniques, plant traits can be identified using remote sensing data (Ali et al., 2015; Verrelst, 2019). A main area of application is plant phenotyping, which can be made more objective and automated using advanced ML methods, for example, deep neural networks. For example, Romera-Paredes and Torr (2016), Ren and Zemel (2017), and Scharr et al. (2016) trained ML models to infer various phenotypic traits, for example, the number of leaves per plant. Similar traits can also be derived using a combination of object and leaf keypoint detection, which facilitates observation of plant growth as done by Weyler et al. (2021). Sa et al. (2016) employed deep convolutional neural networks to detect single fruits, which served as a precursor for subsequent autonomous harvesting (Arad et al., 2020). Drees et al. (2021) used time-series image data of cauliflower and broccoli to predict field growth using conditional generative adversarial networks (Isola et al., 2017). They generated an image of a plant at a later time point and employed the Mask Region-based Convolutional Neural Network (R-CNN; He et al., 2017) to calculate the projected leaf area. Another typical agricultural application is field weed control, where weeds, crops, and soil must be distinguished accurately. Using neural networks, promising results have already been achieved, where the task can be approached using classification (Lottes et al., 2017), detection (Lottes et al., 2018), or semantic segmentation (Ahmadi et al., 2021; Milioto et al., 2018) techniques.

Benchmark datasets with annotations and in-situ measurements are beneficial in terms of facilitating the development of ML methods for plant-specific tasks using remote sensing data. Various benchmark datasets already exist; however, many of these datasets are domain-specific with highly specific objects, for example, buildings (Roscher et al., 2020) and animals (Deng et al., 2009) or other semantics, for example, land cover (Cordts et al., 2015). Generally, such datasets are not suitable for plant applications. The link between ML and plant sciences is becoming increasingly important (Lary et al., 2016), as can be seen from the increasing number of related publications in recent years (Ahmadi et al., 2021; Chebrolu et al., 2017; Förster et al., 2019; Halstead et al., 2020; Kierdorf et al., 2019; Zabawa et al., 2019). Despite increased demand, to the best of our knowledge, only a few publicly available plant-specific datasets are available for ML purposes.

Among the limited number of publicly available datasets or datasets described in the literature, many were acquired in a greenhouse environment (Halstead et al., 2020; Minervini et al., 2016; Mureşan & Oltean, 2018; Scharr et al., 2014) or are based on synthetically generated data (Kierdorf et al., 2022; Ward & Moghadam, 2018), which makes it difficult to apply them to real-world scenarios. In particular, the greenhouse-grown plant *Arabidopsis thaliana* rosettes is frequently used in ML research due to its simple rosette morphology (Scharr et al., 2014). However, the morphologies of agricultural crop plants are more diverse, and their development is affected by changing environmental conditions and both abiotic and biotic stresses. Thus, agricultural datasets that represent real-world field conditions that also cover various challenges, for example, occlusion, shape variability, pose variability, the colors of plants, and plant parts, are required, such as the datasets of Kusumam et al. (2017) and Blok et al. (2021).

Modeling the temporal development of plant growth and plant traits is an active research area, and this requires datasets that monitor plants over time; however, publicly available time-series datasets of plants are rare. One such dataset is the cauliflower (*Brassica oleracea* var. *botrytis*) and broccoli (*Brassica oleracea* var. *italica*) dataset from Bender et al. (2020). The data in this dataset were acquired using a camera-equipped robot that captured close-range images at several time points. However, this dataset is limited to only a few plants and lacks semantic information and accurate georeferencing of single plants.

Cauliflower is a suitable target crop plant to develop ML algorithms because its cultivation, morphology, and economic value give rise to many potential applications in the agriculture digitization context. Cauliflower is a high-value crop that must satisfy various quality criteria. Thus, precise timing of plant management procedures is required to avoid yield losses due to abiotic or biotic stress and produce marketable cauliflowers. Cauliflower harvesting is labor-intensive because each cauliflower must be harvested within approximately 1 week period in which the heads are of sufficient size but are not yet overripe. In addition, cauliflower must be harvested by hand due to within-field variability in plant development. As the head is covered by leaves, each individual cauliflower

head must be touched to determine whether it satisfies the size criteria. After cutting and removing the surrounding leaves, product quality is assessed visually to dismiss heads with discolorations, misshapes, or stress symptoms. Note that cauliflower growth is highly dependent on climate, which makes it difficult to predict the most beneficial harvest time. Depending on the prevailing temperature, irradiance, and soil water availability, plants may develop rather heterogeneously, thus, harvesting of established fields simultaneously can take weeks. Under favorable conditions, plants in sequentially established fields may need to be harvested at the same time, which requires more workers and lowers the price per cauliflower. Early prediction of harvestable plants and harvest time would facilitate better sales planning and provide significant economic advantages to farmers.

Thus, in this paper, we present an agricultural dataset, introduced as GrowliFlower, that is suitable for the development of ML approaches. The proposed dataset is intended to address the growth analysis and development of crop plants and the derivation of phenotypic traits relevant for agricultural applications to promote the development of automation in agriculture. The proposed dataset comprises the following.

- RGB and multispectral orthophotos of two different cauliflower fields were acquired over the entire growing period (from planting to harvest time).
- Plant IDs and coordinates, which enables users to extract complete and incomplete time-series of image patches showing individual plants accompanied with insitu reference data captured manually on the field.
- The plant IDs and coordinates also allow users to extract image pairs of plants pre- and post-defoliation accompanied with a

time-series of the respective plant to facilitate analysis of the correlation between the external appearance and internal head of the cauliflower plant.

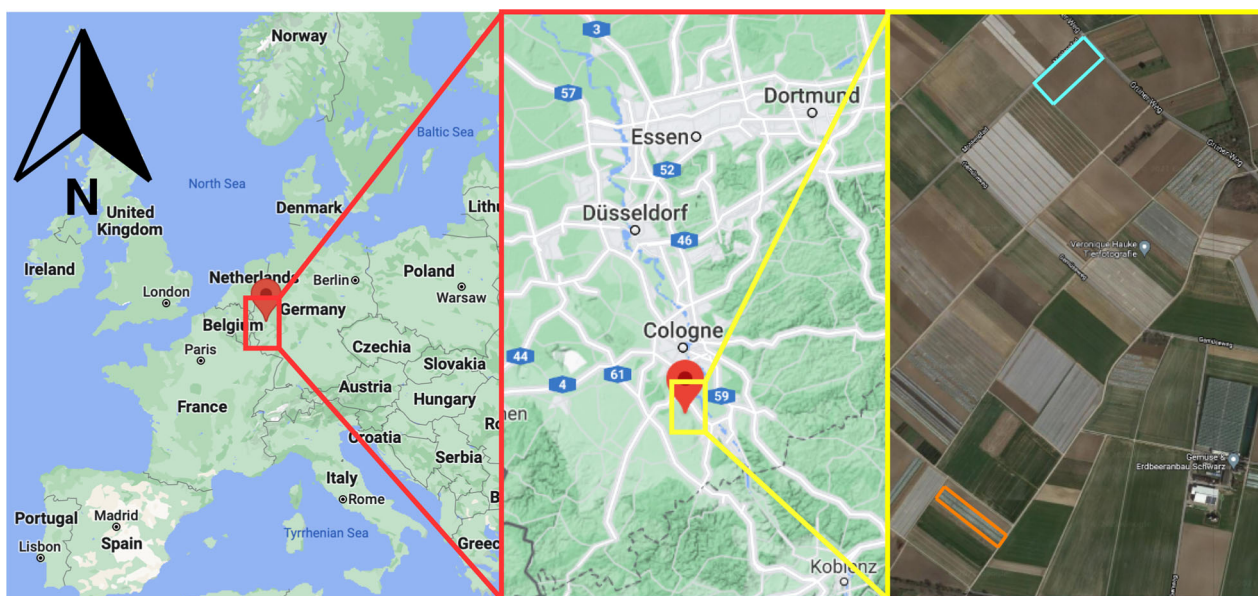
- The proposed dataset's pixel-accurate labeled data are useful for plant and leaf classification, detection, segmentation, instance segmentation, and other similar computer vision tasks.

We also present two baselines demonstrated application examples of plant and leaf instance segmentation using the proposed dataset in a Mask R-CNN (He et al., 2017) application.

## 2 | FIELD DESIGN

Here, we describe the study area. The cauliflower fields used for data acquisition in this study were located on a farm in Western Germany (50°46'6.742"N, 6°58'20.271"O) close to the city of Bornheim, which is 20 km south of Cologne (Figure 1). The mean annual temperature in Bornheim is 14°C, and the mean annual precipitation is 383 mm. This area is dry 142 days a year with an average humidity of 81%. Note that fertile loess soil is available on the farm.

We acquired data for two fields, that is, (1) the field shown in blue in Figure 1 (referred to as field 1 in this paper) in 2020, and (2) the field shown in orange (referred to as field 2 in this paper) in 2021. Note that the cauliflower plants in both fields were planted in rows in a northwest-to-southeast orientation. These fields were designed for sprayers with a working width of 18 m. Before planting, the fields were plowed to prepare the soil. Tractors with 1.8 m track width were used to plant five rows of nursery-grown young cauliflower plants simultaneously, with three rows between the tractor tracks. The distance between the rows was 0.6 m, and



**FIGURE 1** Field locations. The fields are located near Cologne, Germany. Blue: field 1 (2020); orange: field 2 (2021). Map source: Google Maps. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

the distance between the plants in a row was 0.5 m, thereby resulting in a planting density of 33,000 plants/ha. In addition, every 18 m, there was a 2-m wide lane for spraying and irrigation. The fields were subject to conventional farming practices, including hoeing cauliflower plants before canopy closure to reduce weeds and application of pesticides (including herbicides, insecticides, and fungicides). The fields were also irrigated as required using sprinklers. As a result, the abiotic and biotic stresses were rather low in both fields, and the plants developed rather uniformly.

## 2.1 | Field 1

Field 1 has a width of approximately 100 m and a length of 240 m. Thus, the total area of field 1 is approximately 2.4 ha. This field was planted with the Korlanu cultivar (Syngenta). Three-quarters of the field were planted using plants from seedling trays (Figure 2a) on July 28, 2020 from the southwest direction. The remaining northeastern part of the field was planted on July 29, 2020. Note that field 1 was generally free of weeds.

## 2.2 | Field 2

Field 2 field has a width of approximately 55 m and a length of 210 m. Thus, the area is approximately 1.32 ha. This field was planted with the Guideline cultivar (Syngenta). Here, the plants were transplanted from seedling trays on June 15, 2021. Note that field 2 contains more weeds than field 1, especially along the southwestern edge of the field due to previous rhubarb cultivation.

## 3 | DATA COLLECTION

Three types of data were acquired in the data collection process, namely:

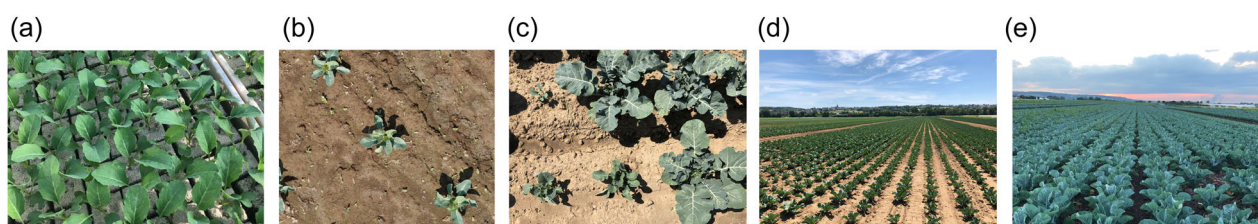
1. RGB and multispectral UAV image data with high spatial resolution, which is an indirect measurement of the phenotypic development of the plants.

2. Georeferenced ground control points (GCP) to locate the data in space, spatially arranged according to field size to ensure accurate and robust processing of the orthophotos (Persia et al., 2020).
3. In-situ measurements of phenotypic traits characterizing the development state and stress factors that serve as reference observations.

The different types of data were collected on the same day to synchronize them. However, to ensure that workers were not visible in the image data, data acquisition processes were not conducted at the same time. Data acquisition was conducted once a week during the entire growth period. During the harvest period, data were collected once between two different harvest days and once after the final harvest. Note that drone flights were only performed on sunny or overcast days to ensure stable illumination for the generation of orthophotos without shading effects due to moving clouds. As a result, the time intervals between successive flights vary. Figure 3 shows the data collection dates for fields 1 and 2. As seen in the top timeline, seven orthophotos are only partly available, which is discussed further in Section 4.1. The data collection took a few hours per day, with the in-situ measurements being the most time-intensive. In addition, data collection was adjusted to both field conditions, resulting in adaptations to camera settings, number of GCPs, and flight altitude. In the following subsections, we describe the procedure followed for fields 1 and 2.

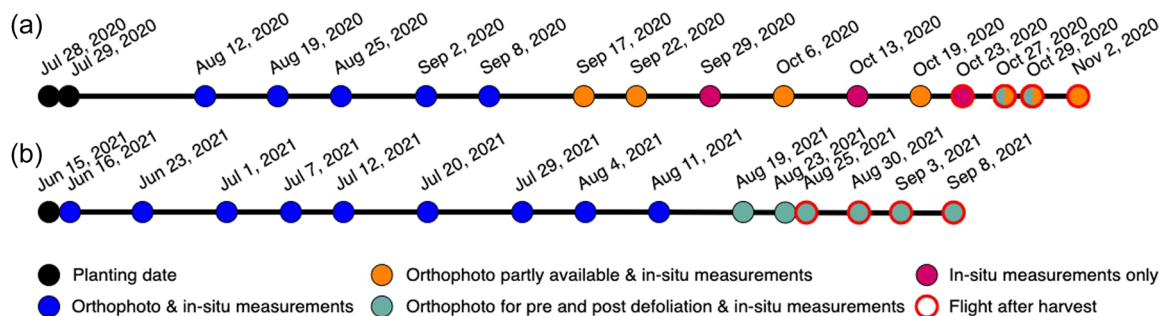
## 3.1 | RGB and multispectral imaging using UAVs

UAV images were captured using a DJI Matrice 600 hexacopter with two mounted cameras (Figure 4). The first camera was a Sony A7 rlll RGB camera with a Zeiss/Batis 2.0 lens (resolution: 47.4 MP). The focal length was 25 mm with a field of view of 71.5°. A shutter speed of 1/1250th and a *floating* aperture (highest value: 2.0) were selected. The International Organization for Standardization value was set to *automated* for field 1 and changed to 50 for field 2 to align our approach with the image-capture settings recommended by Agisoft. The second camera was a MicaSense RedEdge 3 for multispectral image data. It contains five built-in lenses (resolution: 1.2 MP per band). The wavelengths of the five acquired bands and their respective bandwidth were 475 nm (20 nm), 560 nm (20 nm),

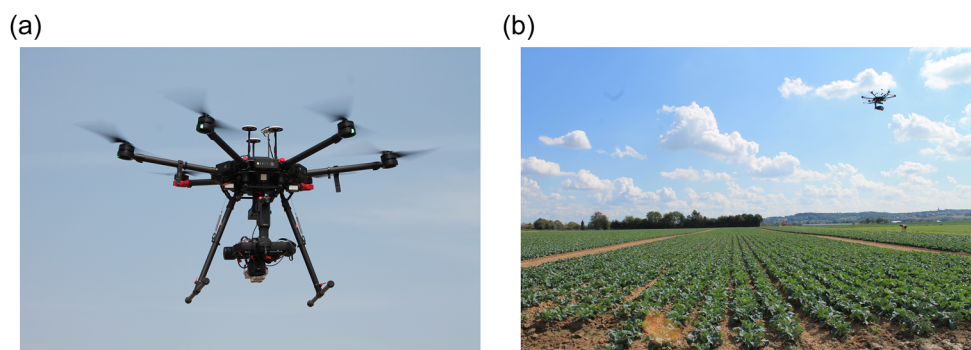


**FIGURE 2** Example field and plant images. (a) Seedling trays before planting. (b) Plants 2 weeks after planting. Images (c, d) were taken 4 weeks after planting and illustrate how different plants develop over. (e) Plants shortly before head formation. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 3** Timelines of acquired data for (a) field 1 and (b) field 2. The colors represent the data availability for images and in-situ measurements. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 4** DJI Matrice 600 hexacopter for unmanned aerial vehicle image-based measurements. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

668 nm (10 nm), 717 nm (10 nm), and 840 nm (40 nm). The focal length of the camera is 5.4 mm. For field 1, an altitude of approximately 10 m and an image overlap of 60/80 were used, and for field 2, an altitude of approximately 16 m and an image overlap of 80/80 were used to optimize the data acquisition process and subsequent image data processing. The following factors were considered in terms of the drone flights. For each flight, no irrigation was permitted in or close to the flight area, the drone was flown at temperatures and wind speeds within the device's safe operating range, and the flights only occurred during periods of no rain.

### 3.1.1 | Time-series flights

On each acquisition date, the drone was flown over a specified area of the field once, which remained the same for the entire growing period. For field 1, this area had a width of 91 m and a length of 62 m, resulting in approximately 0.60 ha. For field 2, the area had a width of 30 m and a length of 131 m, resulting in approximately 0.39 ha.

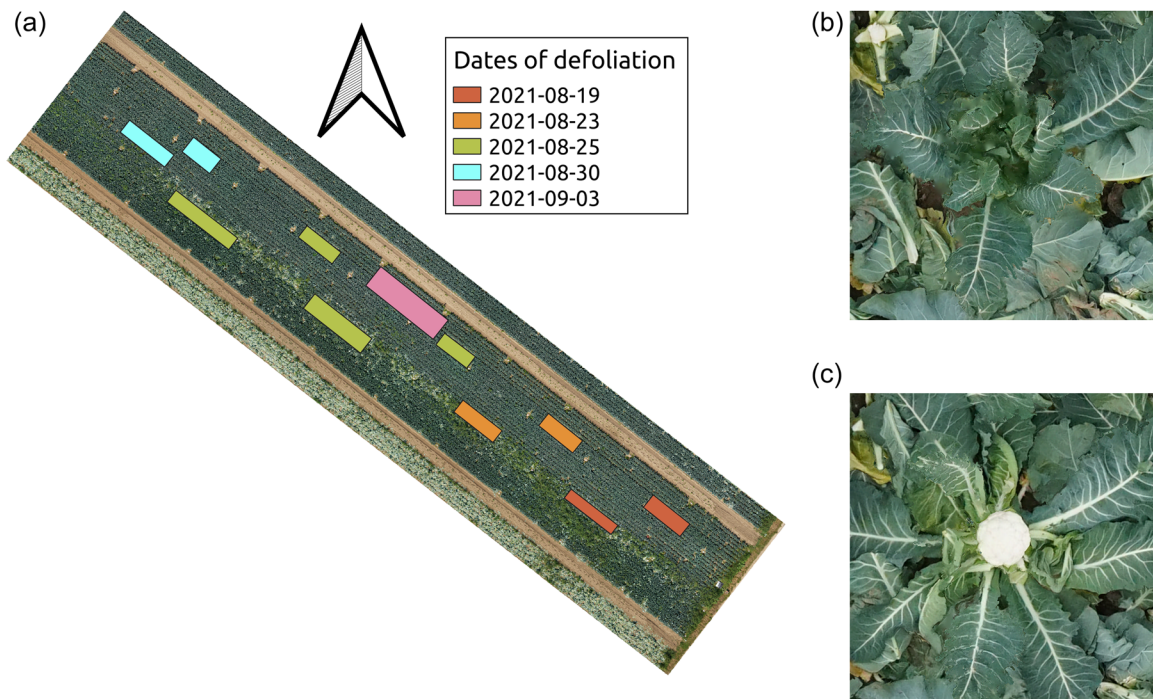
The cauliflower plant does not necessarily grow straight, thus, the center of the plant in later growing stages does not match the position of the seedling exactly (Grenzdörffer, 2019). Thus, a shift of up to  $\pm 10$  cm between the center position of the head and the stem position in the early growing stages was observed.

### 3.1.2 | Defoliation flights

In addition to the time-series flights, so-called defoliation flights were conducted. Here, the upper leaf layers covering the cauliflower head were removed manually on individual plants after the time-series flight. This step is referred to as defoliation. Note that we ensured that the defoliated leaves did not affect any neighboring plants. The defoliated plants provided information about the development of the head relative to the plant's outer appearance. By performing another UAV flight after defoliation, a dataset of images showing the time-series of the plant's outer appearance (Figure 5b) and inner head (Figure 5c) on the day of defoliation was acquired.

For field 1, the defoliation of plants was performed over 2 days, that is, October 27 and 29, after harvesting occurred. Thus, the defoliated plants represented plants whose head size did not satisfy the harvest criteria, which generally meant that the head was too small. For field 2, starting on August 19, when most of the cauliflower heads started developing, between 70 and 200 plants, were defoliated weekly. Here, all plants with developed heads were defoliated in rectangular plot regions to minimize the impact of defoliation on the biological growth of neighboring plants. Note that care was taken to not defoliate the reference plants described previously (Section 3.3). A distribution of plots for the first five defoliation time points is shown in Figure 5a. For the final flight (after the last harvest), most remaining plants that had not been harvested





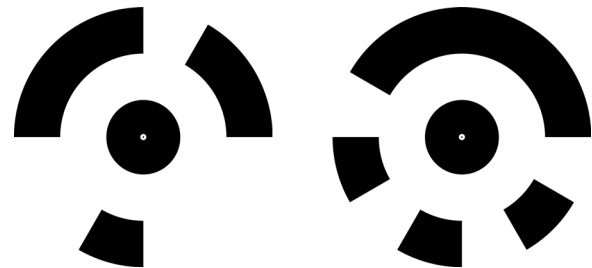
**FIGURE 5** (a) Visual overview of defoliated plant locations for the first 5 weeks of defoliation in field 2. (b, c) Images of a plant pre- and post-defoliation. The locations of random distributed defoliated plants from week six are not shown. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22122)]

were defoliated, which resulted in random distribution. Thus, this is not shown in Figure 5a.

### 3.2 | GCP points

To localize the image data globally in space, the data were georeferenced with the help of circular 12-bit GCPs with a diameter of approximately 20 cm, as shown in Figure 6. Here, the GCPs were fixed in the ground using plastic pegs, and they were distributed evenly across the field (refer to Figure A1) and positioned on tractor tracks or between plants to avoid displacement by external influences, for example, plowing. In addition, surrounding plants were removed as required to ensure the visibility of GCPs in the image data. We used 21 GCPs in field 1 (35 GCPs/ha) and 44 GCPs in field 2 (113 GCPs/ha) (refer to Figure A1), with each GCP showing a different pattern. The greater number of GCPs in field 2 was due to the fact that they facilitate subsequent image alignment by ensuring that at least three GCPs were present in each captured image, especially for growth stages with a high degree of plant overlap and dense canopies.

As a measuring device for GCP coordinates, a Trimble R4-Model 3 base station with a horizontal standard deviation of  $\pm 5 \text{ mm} + 0.5 \text{ ppm RMS}$  and vertical standard deviation of  $\pm 5 \text{ mm} + 1 \text{ ppm RMS}$  was used for both fields. The measured coordinates were acquired in the WGS84/UTM 32N coordinate system. To ensure that the markers for the GCPs were not displaced due to external influences, the GCPs

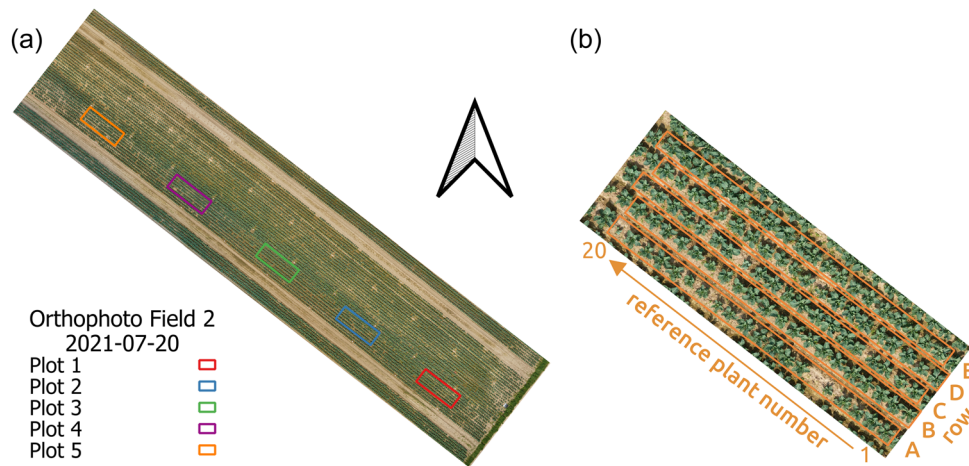


**FIGURE 6** Two ground control point patterns used for acquisition.

were measured at the beginning and end of the data acquisition period to omit displaced GCPs. A third measurement was added for field 2 in the middle of the growing period.

### 3.3 | In-situ measurements of plant development

In each field, so-called reference plots were selected to capture information from reference plants manually. For field 1, four reference plots were assigned (Figure A2a), and each plot comprised three rows with 20 plants each (Figure A2b). Thus, each plot contained 60 plants, for a total of 240 plants in all reference plots. The plots were distributed in the northwestern half of the field along the long side. Five reference plots were assigned for field 2 (Figure 7a). Here, each plot comprised five rows of 20 plants



**FIGURE 7** Visual overview of (a) reference plots for in-situ measurements in field 2 and (b) the design of reference plot 5 (including reference plants and the ordering of reference plant numbers). The plot design is valid for all reference plots in field 2. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22122)]

(Figure 7c; 100 plants/plot, 500 plants in total). The plots were distributed evenly in the southwestern half of the field along the long side. Thus, the reference data were collected along the entire field. Each reference plant was assigned a specific plant ID identifying the row (field 1: A–C; field 2: A–E) and plant number (Field 1: 1–10, 90–99; Field 2: 1–20).

The following measurements were taken for all reference plants in field 1.

1. Phenological development after BBCH-scale (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und Chemische Industrie)/Feller et al. (1995),
2. height,
3. maximum diameter,
4. other remarks, for example, stress infestation (listed in the attachment in Table A1),
5. head diameter,
6. harvesting status.

Note that the farmer followed a rigorous plant protection schedule, and very few stresses were detected in 2020; thus, information about stresses was not recorded explicitly in 2021. Due to the observed homogeneous development, focus was placed on measurements of BBCH and the height of five representative plants per plot. Here, the head diameter and harvest status were recorded for individual plants.

## 4 | DATASET

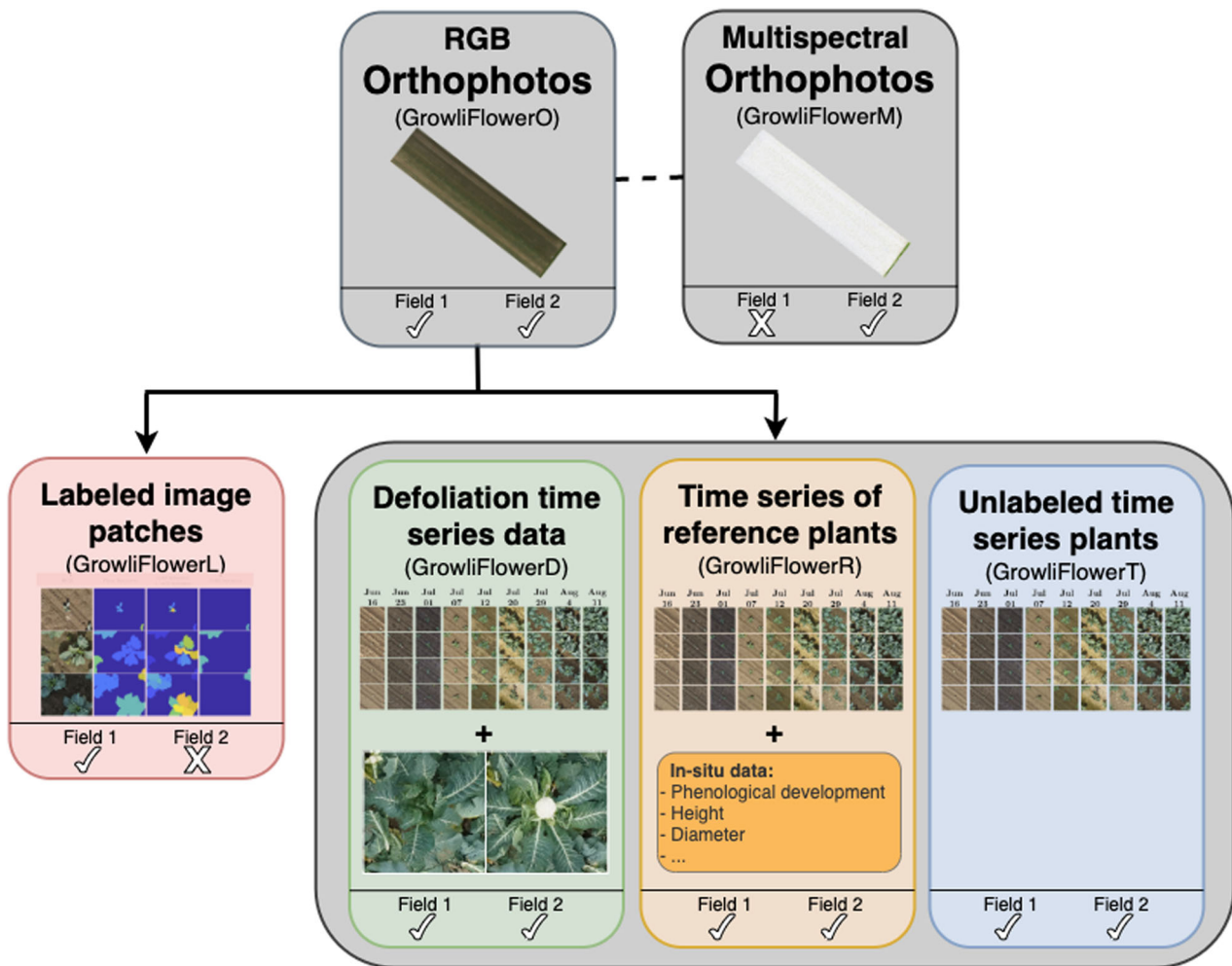
The core component of the dataset (Figure 8) comprises both RGB and multispectral orthophotos derived from the captured UAV images. In the orthophotos, single plants are identifiable by their corresponding coordinates and plant IDs. The dataset contains four subsets intended for different ML tasks. The instance

segmentation GrowliFlowerL subset contains patches extracted and processed from the RGB orthophotos, and the remaining three subsets contain time-series data of individual plants. The GrowliFlowerT subset comprises randomly selected time-series data representing a wide variety of cauliflower development. In addition to the time-series data, the GrowliFlowerD subset also contains image pairs of plants before and after defoliation. The GrowliFlowerR subset contains the insitu measurements and the time-series data. For each field, a text file containing the measured GCP coordinates at the beginning and the end of field monitoring is provided. For field 2, the coordinates measured during the growing period are also given.

### 4.1 | Orthophotos (GrowliFlowerO and GrowliFlowerM)

The acquired RGB and multispectral UAV images were aligned to orthophotos using the Agisoft Metashape Professional software to obtain a large-scale overview of the monitored fields. Here, the orthophotos were georeferenced according to the measured GCP coordinates. In addition, the individual orthophotos were exported in the WGS84/UTM 32 coordinate system.

The ground resolution for the RGB orthophotos of field 1 is 1.65 mm/px for the pixel width and height with a minimum and maximum file size of 1.64 and 6.7 GB, respectively. The ground resolution for field 2 is 3.10 mm/px for the pixel width and height with a minimum and maximum file size of 1.3 and 5.0 GB, respectively. Twelve orthophotos are available for field 1, where five are entirely processed, and seven contain data gaps for small areas where the quality of the UAV-acquired images was insufficient. For field 2, 15 orthophotos are available, as shown in Figure 3b. This set of orthophotos is provided in the GrowliFlowerO subset of the proposed dataset. In addition, the



**FIGURE 8** Overview of data in proposed GrowliFlower dataset. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

dataset contains multispectral orthophotos for field 2 with a ground resolution of 2.5 cm/px width and length, denoted as the GrowliFlowerM subset.

## 4.2 | RGB image patches

In this section, we describe the data extracted from the RGB orthophotos. Note that the ground resolution of the resulting image patches is the same as that of the respective orthophotos.

Each of the following datasets (excluding the labeled dataset described in Section 4.2.1) contains a text file with global information for each field, containing the image ID, including the plant ID, and corresponding georeferenced UTM coordinates of the plants. Note that the coordinates identify the center of the plants as observed on August 19 for field 1 and July 7 for field 2. In addition, information about the planting day and a proposed assignment as a training, validation, or testing subset are provided as a basis to compare ML methods. To minimize spatial correlation

between sets, the proposed training, validation, and testing subsets are spatially disjoint. However, certain systematic factors from a biological perspective are not excluded. The use of these sets is expected to promote the development of ML methods with high generalizability. For the reference data discussed in Section 4.2.3, the harvesting time is specified, and for the defoliation data discussed in Section 4.2.4, the defoliation date of the plants is specified. In addition, text files with local information for each acquisition date are provided, including the image ID to connect the local information with the global information, and the corresponding local pixel coordinate relative to the respective orthophoto for each data acquisition day. Also note that information about the day after planting (dap) is included.

To use image patches showing single plants, the patches must be extracted from the orthophotos using the plant IDs and coordinates. Here, an image side length and width of at least 490 px for field 1 and at least 256 px for field 2 is recommended to ensure that the entire plant is captured in the image patch regardless of the plant developmental stage.

#### 4.2.1 | Labeled image patches (GrowliFlowerL)

This subset, called GrowliFlowerL, comprises pixelwise, manually annotated images, thus, it is well-suited for classification, semantic segmentation, detection, instance segmentation, or stem detection tasks. For this subset, the image patches of four acquisition dates for field 1 are extracted using a sliding window approach. The image patches have a size of 368px × 448px. Here, the size of the patches differs from that of the proposed sizes because only plants from earlier development stages are included. In addition, in this dataset, the focus is not on individual plants but on the variability between images, thus, the plants are not located in the center of the patch.

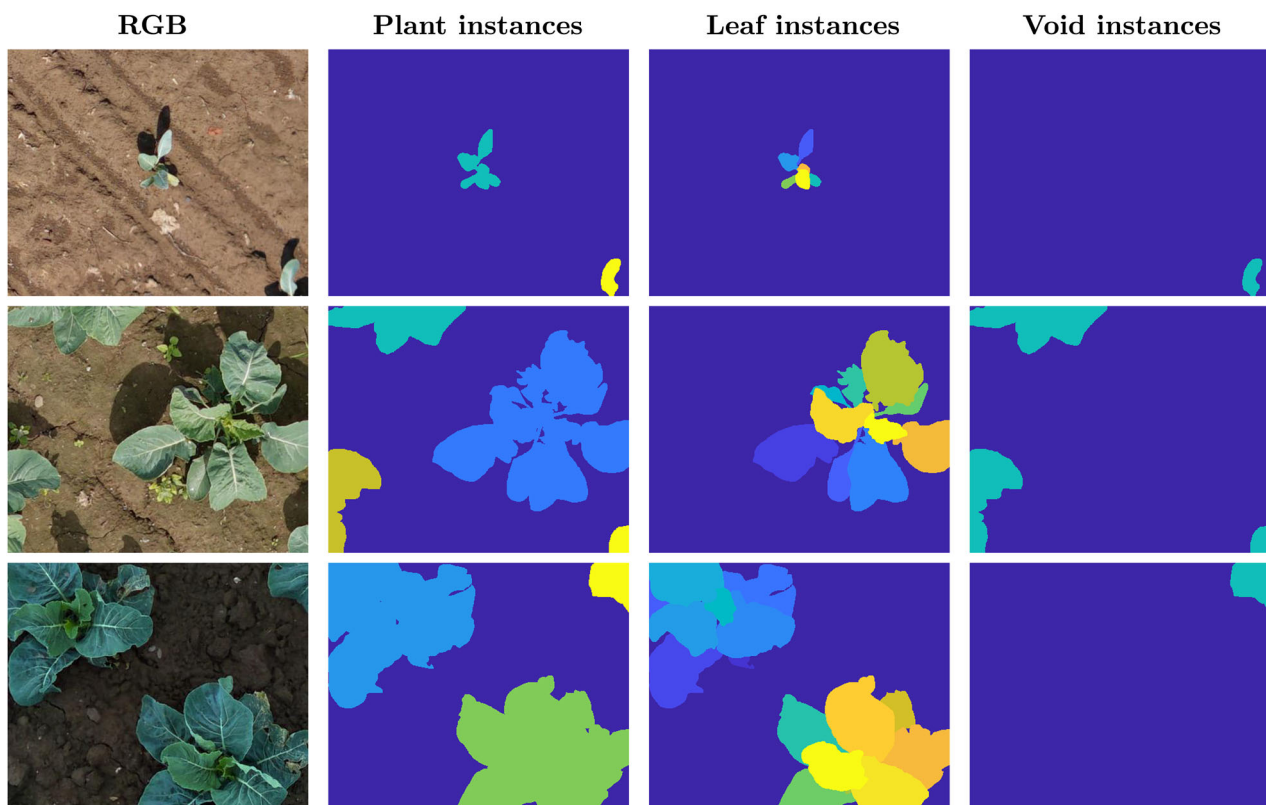
For each RGB image patch, four annotated masks are provided. These annotated masks contain segmentations of (1) plant instances, (2) leaf instances, (3) void segmentations, and (4) stem positions.

- (1) The plant instance mask segments the image in soil and plant pixels with instance information for the plants.
- (2) The leaf instance mask segments the plants into single leaves. Note that plants at image borders for which no stem or only

one-quarter of the plant is visible are annotated as void and no leaf annotation is applied.

- (3) The void segmentation mask is a binary mask where plants located at image borders where no stem is visible are segmented as void. In addition, plants with only a small amount of visible leaf material in the RGB image are also segmented as void.
- (4) The stem annotation mask represents the position of the stems of nonvoid plants.

Examples of (1) plant instance masks, (2) leaf instance masks, and (3) void segmentation masks are shown in Figure 9. Two things to note are that weed is not labeled as a plant but as a background and that stem positions are only represented by individual pixels, thus, they are difficult to recognize visually. Therefore, masks that include stem information are not shown in these examples. The annotations are provided with a defined name based on the name of the RGB image patch. Here, each patch contains a maximum of four plants, and there is a number of patches in the dataset that contain no plants (Table 1). This subset is divided into training, validation, and testing sets, and the complete labeled subset is denoted GrowliFlowerL.



**FIGURE 9** Examples of labeled images for different time points. Column 1 shows that the RGB base for columns 2–4 illustrates corresponding labeled plant instance masks, leaf instance masks, and void segmentation masks. The rows represent different points in time. Dark blue represents the background class, and the other colors represent different (leaf) instances. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/1365-3113.12122)]



## 4.2.2 | Time-series for plant data (GrowliFlowerT)

For each field, the plant coordinates are provided to allow users to extract time-series plant images. This data is denoted GrowliFlowerT. The time-series data of field 1 comprise the early plant developmental stages and the harvest dates, but lack dates, when the canopy around the cauliflower head was closed. The time-series data of field 2 comprise all growth stages.

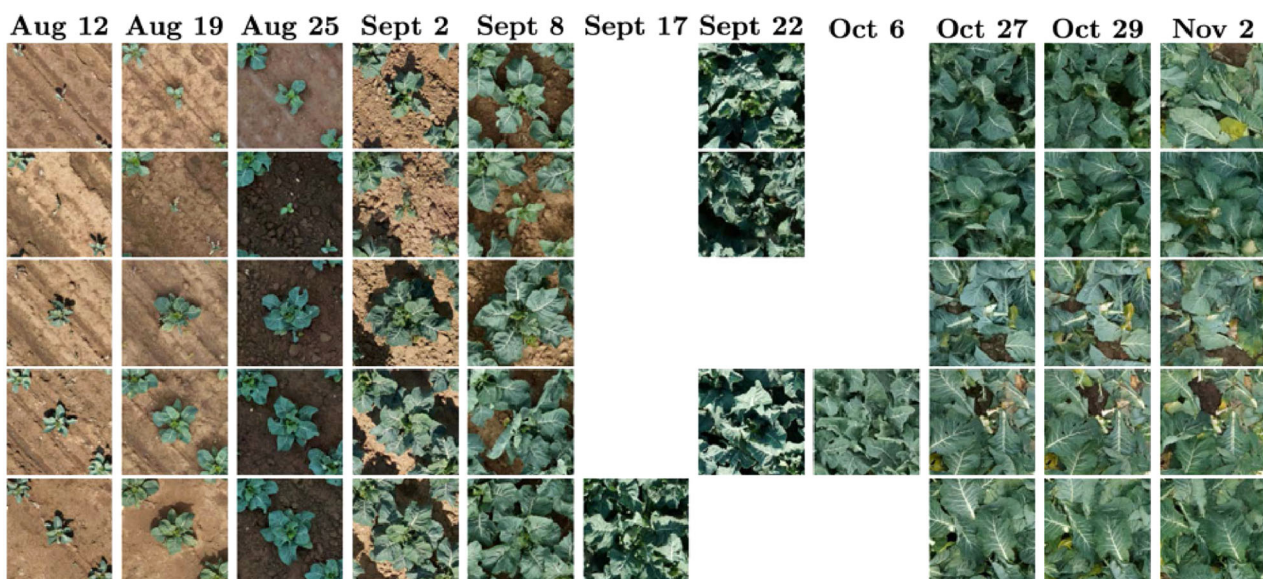
For field 1, the coordinates for approximately one-third of the plants are determined (3804 plants in total). The distribution of the location of the extracted data is visualized in Figure A3a. The selected plants are distributed along the southeastern edge of the field due to the availability of data for most time points and the ability to determine the harvest window of individual plants. The subset is divided into training, validation, and testing sets, as shown in Figure A3a. In addition, cauliflower planted on July 28 or July 29 is included in all three sets to ensure that the variability within the

**TABLE 1** Overview of distribution of labeled images acquired on different dates.

Definition	All images	Images with plants [Train/Val/Test]	Images without plants [Train/Val/Test]
2020_08_12	844	745 [521/110/112]	99 [71/15/15]
2020_08_19	892	781 [547/117/117]	111 [78/16/17]
2020_08_25	383	367 [257/55/55]	16 [12/2/2]
2020_09_08	79	79 [56/11/12]	0 [0/0/0]

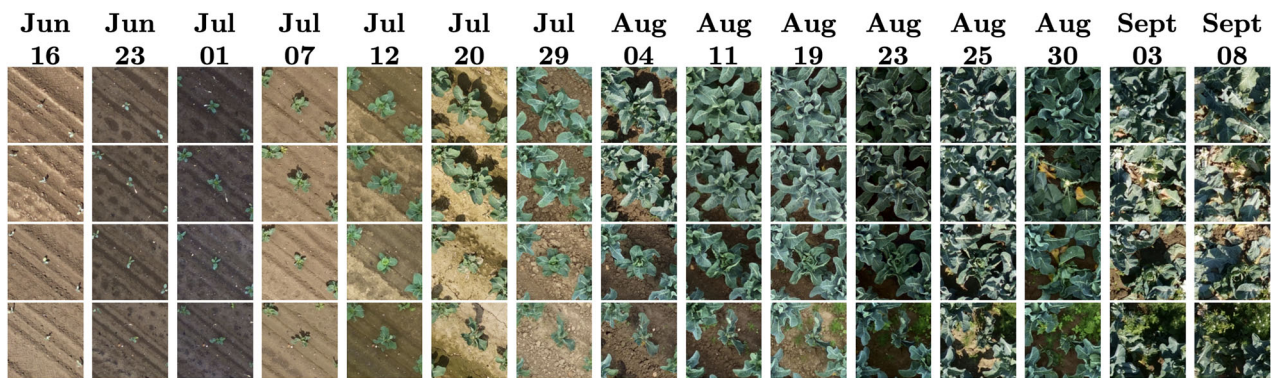
sets is guaranteed. Note that the orthophotos do not overlap entirely; thus, image data are not available for all plants at all times, which results in temporally incomplete time-series data. For field 2, 8736 plant coordinates were extracted and distributed evenly over the field. The subset is divided into training, validation, and testing sets, as shown in Figure A3b. Here, all plant coordinates are provided as georeferenced UTM coordinates.

To use individual plant images, the user must crop the patches around the local plant coordinates determined in the subset. In addition to all global plant coordinates, this subset contains the local coordinates of the patches for each acquisition date, which at a size of 490px × 490px for field 1 and 256px × 256px for field 2 lie completely within the orthophoto and are not showing spatial data gaps, as patches shown in Figure 12b. Five examples of the time-series data are shown in Figure 10 for field 1, and four examples are shown in Figure 11 for field 2. Due to the spatial data gaps, the number of coordinates per date for field 1 varies, which leads to temporal gaps in the time-series data. The largest set of time-series that includes equal time steps consists of 3611 time-series based on eight time points, including the five time points up to day after planting 42 (September 8), and all three time points from day after planting 91 (October 27). In addition to the file that contains all UTM coordinates, a text file containing the UTM coordinates for this set is also provided; thus, the user can extract the time-series data for the selected plant IDs. After removing the patches with spatial data gaps, we retained 8348 complete time-series images for field 2. Due to the heterogeneous weed occurrence in field 2, the patches contain different amounts of weed, as shown in Figure 12a. Due to the given UTM coordinates, it is possible to extract the complete time-series set of local coordinates for both fields if required.



**FIGURE 10** Time-series illustration of five different plants in GrowliFlowerT subset in field 1. All rows represent time-series of plants containing temporal data gaps due to the poor image quality of the corresponding UAV images (indicated by omitted images). The columns represent the recording days and show the five representative plants captured at the same time on that day. UAV, unmanned aerial vehicle. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 11** Four plant time-series from field 2. A row represents a time-series. The columns represent the acquisition dates. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



**FIGURE 12** Data gaps and different amounts of weed occurrences in image data during different stages of growth. (a) Different amounts of weed occurrence on the acquisition date of August 11 and (b) data gap occurrence. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

**TABLE 2** Number of reference plant image patches per acquisition date for field 1 (2020).

Date	Aug 12	Aug 19	Aug 25	Sept 2	Sept 8	Sept 17	Sept 22	Oct 06	Oct 19
# images	239	239	239	239	239	239	-	-	193
Date	Oct 27 (Post)				Oct 29 (Post)				Nov 2
# images	119				119				12

#### 4.2.3 | Time-series for reference plant data (GrowliFlowerR)

For each field, the subset includes the plant IDs and coordinates, which allows the user to extract an image time-series set of monitored reference plants that appear similar to those described in Section 4.2.2. The time-series data for field 1 comprise the early plant developmental stages and the harvest dates, but lack dates, when the canopy was closed. The time-series data for field 2 comprise all growth stages (see Figure 11). Table 2 shows the distribution of available plant IDs and the number of images of plants per time point for field 1. Note that the pre-defoliation orthophotos of October 27 and October 29 do not overlap the reference plots due to the low quality of the corresponding UAV images. Here, the reference plants were not defoliated, thus, the orthophotos of defoliation flights are used to extract images of these days to acquire a reference time-series. For field 2, all local coordinates are given for all acquisition dates, which allows the user to extract complete image time-series. Here, the data are divided into training, validation, and testing set for both fields. In addition, the plants in each plot are

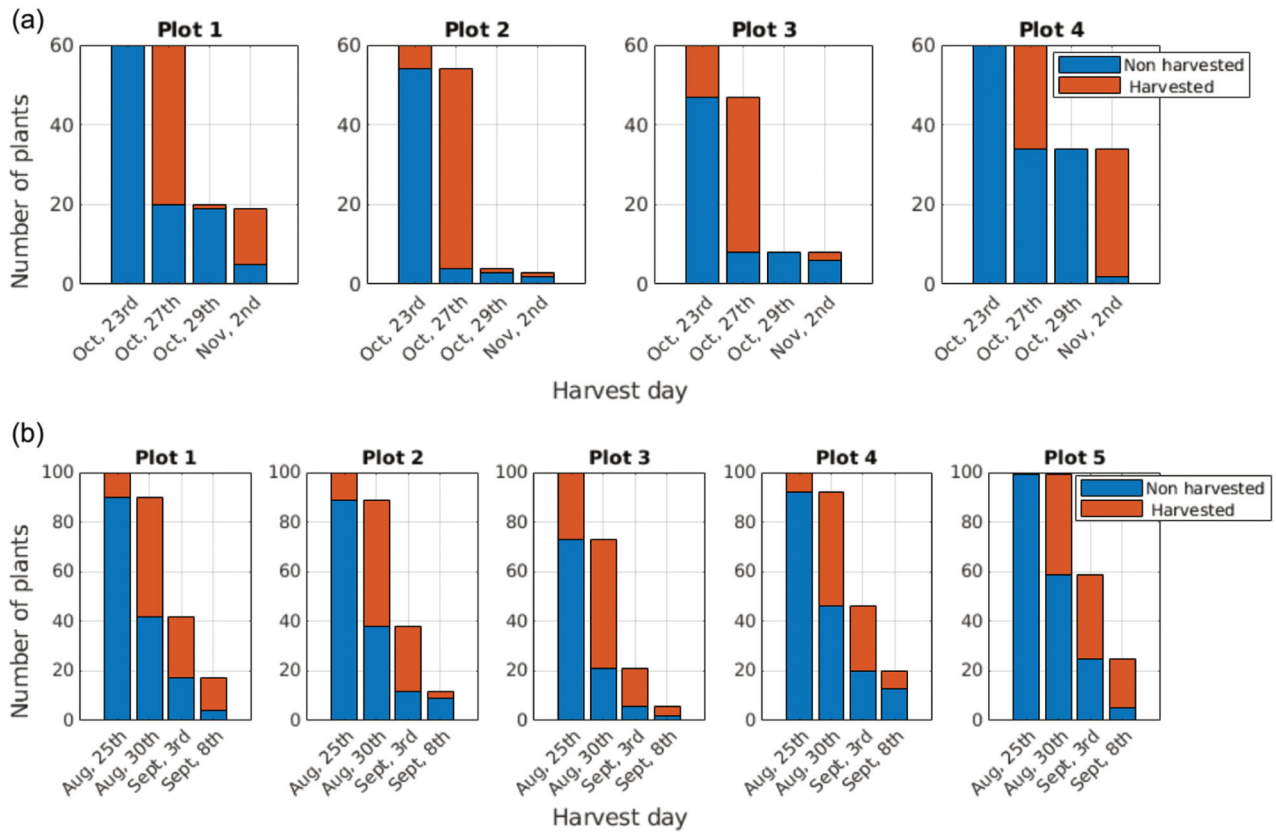
**TABLE 3** Number of defoliated plants per acquisition date for field 2 (2021).

Date	Aug 19	Aug 23	Aug 25	Aug 30	Sept 3	Sept 8
# images	110	115	251	116	71	54

presented in each set. The visual distribution for both fields is shown in Figure A4.

#### 4.2.4 | Time-series for defoliated plant data (GrowliFlowerD)

For field 1, the GrowliFlowerD subset contains a total of 130 plant IDs and coordinates for defoliated plants (30 for October 27 and 100 for October 29). For field 2, the subset contains a total of 717 plant IDs and coordinates for defoliated plants. The coordinates allow the user to extract the time-series of defoliated plants. Table 3 presents an overview of how many plants were defoliated on different acquisition days. In addition to the time-series data, pairs of pre- and post-defoliation images



**FIGURE 13** Overview of harvested and nonharvested plants per reference plot per day. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22122)]

are provided in the subset. The data are divided into training, validation, and testing sets for both fields, and each defoliation day is presented in each set. The visual distribution of both fields is shown in Figure A5.

### 4.3 | Insitu data

Two CSV files are provided, that is, one for each field, and these files contain the plant ID and the measurements described in Section 3.3 for each data acquisition day. The measured values correlate with the images in the GrowliFlowerR subset. Figure 13 shows the distribution of the number of harvested plants in the reference plots per acquisition date for fields 1 and 2.

## 5 | BASELINE FOR INSTANCE SEGMENTATION APPLICATION

### 5.1 | Experimental setup

We describe two possible applications of the proposed dataset by creating baselines using the labeled GrowliFlowerL subset and

the Mask R-CNN (He et al., 2017), which is a state-of-the-art instance segmentation method. We use the pytorch implementation available at [https://pytorch.org/tutorials/intermediate/torchvision\\_tutorial.html](https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html).

We consider plant instance and leaf instance segmentation tasks. Thus, we use the mask and bounding boxes derived from the plant instance mask as the target for the first baseline. The mask and bounding box derived from the leaf instance mask are used as the target for the second baseline. For the leaf instance segmentation baseline, the given void instances are used as the background because only leaves that do not belong to void plants are labeled. Note that the estimation of semantic masks for individual instances enables the derivation of phenotypic traits. Here, we applied a random horizontal flipping data augmentation technique with a probability of 0.5.

We trained the Mask R-CNN on a computer with Intel Core i7-6850K 3.60 GHz processor and a GeForce GTX 1080Ti GPU with 11 GB RAM. The network was pretrained on the COCO dataset (Lin et al., 2014), and training was performed over 100 epochs with a learning rate of 0.001 and batch size of 2. We used a stochastic gradient descent optimizer, and ResNet-50 was used as the backbone network.

## 5.2 | Evaluation metrics

The first evaluation metric is the Intersection over Union (IoU), which is calculated as follows:

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (1)$$

following the evaluation metrics of the COCO dataset (Lin et al., 2014), where TP represents true positives, FP represents false positives, and FN represents false negatives. We also consider two additional metrics, that is, precision  $p$  and recall  $r$ , which are defined as follows.

$$p = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad r = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (2)$$

The F1 score summarizes and balances precision and recall by calculating a harmonic mean. The F1 score is defined as follows.

$$\text{F1} = \frac{2 \cdot p \cdot r}{p + r}. \quad (3)$$

We compute precision, recall, and F1 score relative to the single object class cauliflower plant and calculate the scores for the IoU thresholds  $t_{\text{IoU}} = 0.50$  and  $t_{\text{IoU}} = 0.75$ . In addition, we determine the average precision (AP), average recall (AR), and average F1 (AF1) scores over all IoUs in the interval 0.50 – 0.95 with a step size of 0.05 as for the COCO benchmark. This is indicated by  $(\cdot)@0.5 - 0.95$ . For the leaf instance segmentation baseline, we reduce the evaluation on recall, as we do not want to penalize predictions on void pixels. The consequence of penalizing predictions on void pixels would be to penalize the model for identifying leaves that were simply not labeled as such.

## 5.3 | Results

We calculate the metrics relative to the detected bounding boxes and the segmented masks of the respective objects. The segmented masks provide information about the cumulative number of correctly classified pixels and thus, the more accurate shape of the object. The bounding box enables derivation of the detection accuracy and thus, the localization of the object.

Table 4 summarizes the results of the plant instance segmentation task for the baseline method. As can be seen, 95% at  $\text{IoU} \geq 0.5$  are predicted correctly. In addition, precision at the bounding box and pixel levels are greater than 80% for all IoU thresholds  $\leq 0.8$

(Figure 14a). At an IoU value of  $\geq 0.85$ , precision decreased rapidly. This trend is also observed for both recall (Figure 14b) and the F1 score (Figure 14c). For higher IoU values, we found that prediction at the pixel level is less accurate than at the bounding box level because slight changes in segmentation generally result in more errors in the segmentation mask than in the bounding box (Figure 15). An overview of the results is given in Table 4.

We found that many objects and masks are estimated accurately (Figure 16a). The results show all predictions with a confidence score greater than a threshold of 50%. Precise contours are estimated, and in the earlier development stages, the instances are well separated spatially. Note that the model does not predict the ground as an object in any case. In addition, smaller weeds that can be seen in some patches are also not considered objects, which is desirable, because in this way we identify that the model distinguishes cauliflower from weeds. We found that inaccuracies occur with plants that lie at the edge of the image patches. In such cases, only small parts of the plant are visible, thus, the leaves are not adjacent to each other, as shown in Figure 16b (top left and bottom left). We also found that errors occur in later developmental stages because the plants overlap (Figure 16b bottom right), which represents a more challenging scenario than well-separated plants. In particular, for overlapping plants, it is even difficult for the human eye to assign leaves to individual instances. In addition, compared with the earlier stages of development where no overlap occurs, fewer training images were available for the later stage of development. The small number of images means that less variability in the data is captured, making prediction on new unknown data more difficult.

Another distinctive feature involves plant objects from which leaves fall or plants that are impaired in their growth and thus decay. In such cases, it is difficult for the model to distinguish whether one or more plants are represented (Figure 16b top right).

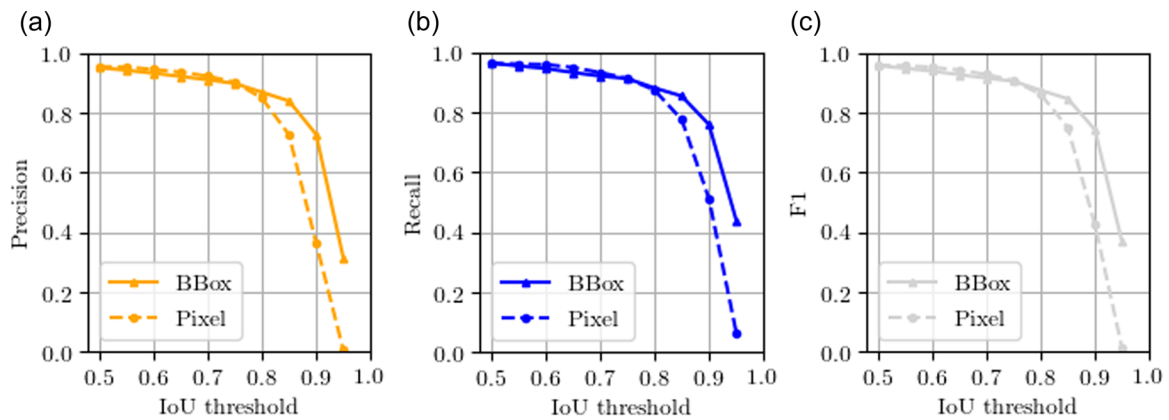
For the leaf instance segmentation task, which is a more difficult task compared with plant instance segmentation, we achieve a very good recall result of 74% at the bounding box level and 77% at the pixel level. The distinction between individual leaf instances is more complex than the distinction between plant instances. In addition, here, we assign the void-labeled objects to the `background` class for this baseline rather than the `leaf` class because individual void plants can contain several leaves, however, such leaves were not labeled individually. Note that the calculated values for recall are similar both the pixel and bounding box levels.

We can find explanations for the recall values in the visual consideration of the results, even though these results show predictions with a confidence score greater than a threshold of

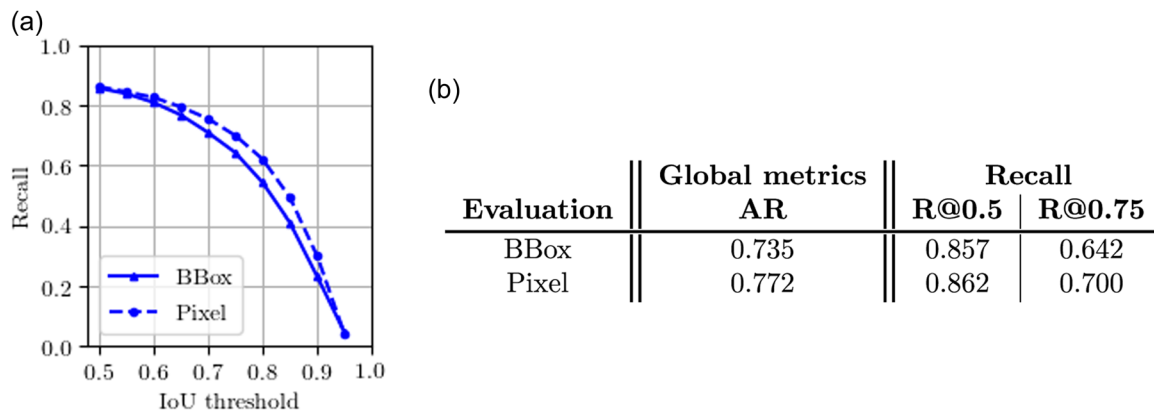
**TABLE 4** Plant instance segmentation results: precision, recall, and F1 score for predicted bounding boxes (BBox) and segmentation masks (pixel) for the class `plant`

Evaluation	Global metrics			Precision		Recall		F1	
	AP	AR	AF1	P@0.5	P@0.75	R@0.5	R@0.75	F1@0.5	F1@0.75
BBox	0.917	0.933	0.843	0.952	0.899	0.965	0.913	0.958	0.906
Pixel	0.844	0.858	0.770	0.954	0.902	0.963	0.913	0.959	0.908

Abbreviations: AP, average precision; AR, average recall; AF1, average F1.



**FIGURE 14** Representation of (a) precision, (b) recall, and (c) F1 score for class cauliflower plant. The graphs show the evaluation at different IoU thresholds on the bounding box (BBox), thus object (solid-line), and pixel (dashed-line) level. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 15** Recall results for leaf instance segmentation task. (a) Evaluation at different Intersection over Union (IoU) thresholds on the bounding box (BBox), that is, object (solid-line), and pixel (dashed-line) level. (b) shows the respective average recall (AR),  $R@0.5$ , and  $R@0.75$  values. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

50%. By defining void instances as the background, the model is challenged to predict the leaves belonging to void instances not as *leaf* objects, as shown in Figure 16d (top left and bottom left). It is difficult for the model to distinguish whether plants at the edge of the patches are void instances or leaf instances. Thus, either leaves are predicted that are not in fact present in the target (representing low precision) or no leaves are predicted even they are present in the target (representing low recall). For plants that are completely visible in the patch, the model demonstrates better prediction performance. Another source of error is the prediction of several instances on a single leaf, as shown in Figure 16d (top right and bottom right) because the model is required to learn various features, for example, leaf structure and size, because such features play a crucial role in distinguishing different leaves.

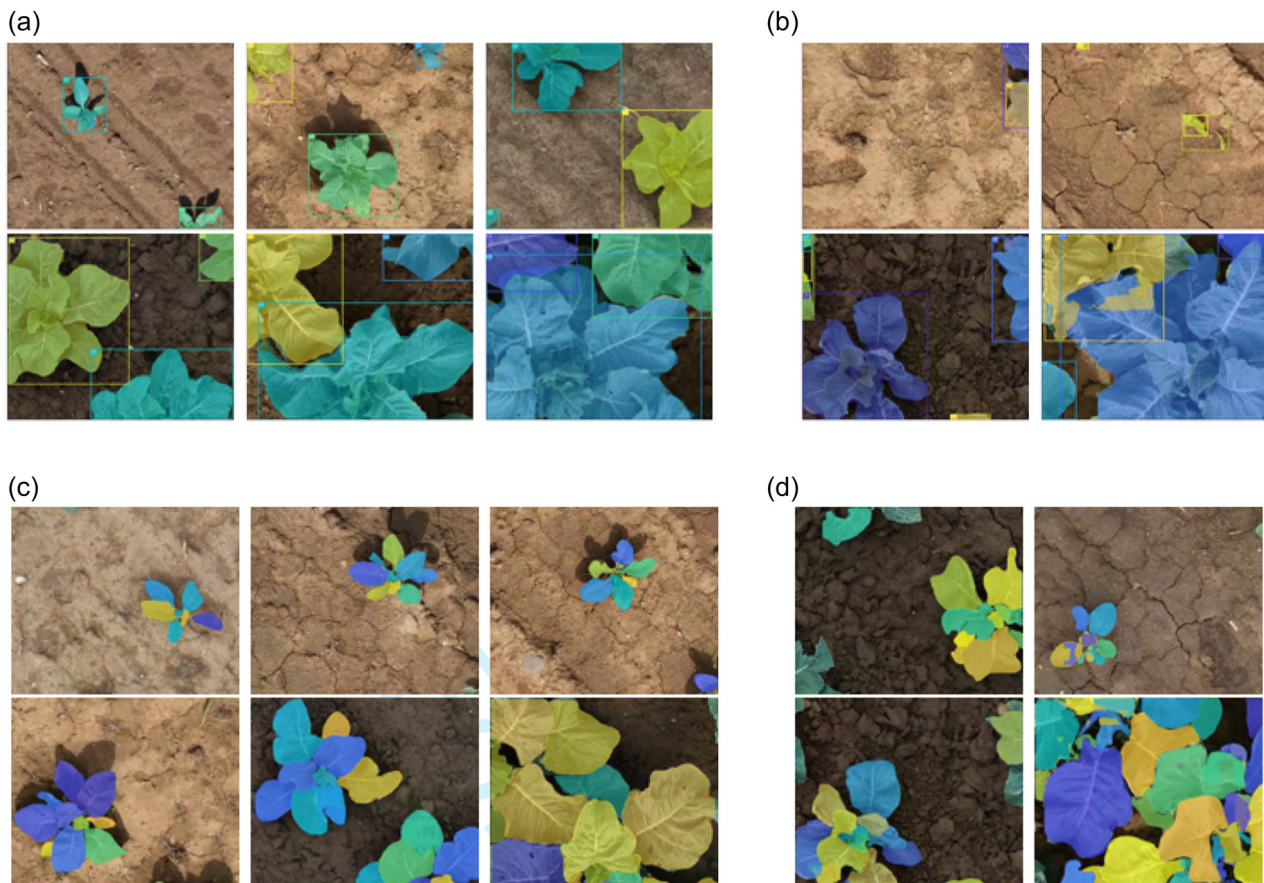
To sum up, we observe that our instance segmentations, plant instance as well as leaf instance, perform and can be used for different growth stages of the cauliflower plants.

## 6 | REFLECTION AND DISCUSSION

The results of this study provide insight into the acquisition of image time-series under field conditions. We observe that the flight altitude of the UAV must be adjusted depending on the characteristics and the height of the cultivated plants to capture images with high quality in terms of resolution and gapless spatial data. However, to obtain accurate image data, enough keypoints must be distributed in the field. Our work shows that GCPs are suitable as keypoints because they help align and georeference the orthophotos more accurately. For simplified use of the data for ML approaches, the data should have similar characteristics as exposure over time. For this purpose, data must be recorded under consistent weather conditions. However, we note that combining consistent conditions and similar interval lengths between acquisition days is a challenging task.

While previous research, such as the work of Bender et al. (2020), focuses on collecting data from many different data sources (e.g., imagery, climate, and soil data), monitoring a large number of





**FIGURE 16** Plant and leaf instance segmentation results. The different colors indicate the different instances. In the visualization of the leaf instance segmentation results, we concentrate on the visualization of the masks for clarity and omit the bounding boxes. The examples shown in (a) and (c) show accurate results and those in (b) and (d) show improvable results. (a) Accurate plant instance segmentation results, (b) improvable plant instance segmentation results, (c) accurate leaf instance segmentation results, and (d) improvable leaf instance segmentation results. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.22122)]

different instances that are specifically needed for training DL methods has not been considered sufficiently. In our work, we provide a large image number of different instances but we lack additional data, such as soil and climate data. Climate and soil characteristics are important factors to determine or predict plant growth. ML methods only learn features that are present in the training data. Thus, if external factors such as climate and soil change, the growth of plants is influenced as well. The lack of this information can cause ML models to be prone to error in their results when applied to new data. Therefore, we suggest improving the data acquisition by capturing additional soil and climate data. Another suggestion for improvement of data acquisition is the field design. To avoid systematic effects in the data, reference plots and other selected areas like defoliation plots should be distributed equally within the field, because it is difficult at the beginning of data collection to predict how much which plants (crops and weeds) will grow and how the location will affect growth and thus, the acquired data.

Regarding our baseline experiment considering instance segmentation, we observe that the application shows satisfactory results on

our data. To improve the results, future studies could integrate prior knowledge about the shape and structure of plants and leaves. Weyler et al. (2022) show an example of how to improve the results for plant and leaf segmentation by developing a combined approach of neural networks and clustering to simultaneously determine leaf and plant instances. However, to the best of our knowledge, the approach is not used for plants as large and highly overlapping as in our provided later developmental stages (see Figure 16c,d).

Another way to improve the result of instance segmentation is to vary the field of view per image. In further experiments, we observe that when applying a modified field of view, plants are segmented with only a few errors. For these experiments, we reduce the threshold by up to 20% depending on the extension of the field of view. The masks and bounding boxes of the predictions match the plants. However, the size of the objects differs from the training data due to the change in scale. This causes the model to have less confidence in its predictions, even if they are correct. Even with a changed image size, cauliflower plants can be easily distinguished from weeds and each other. As it brings greater variability to the data, adding images with a larger field of view to the training set could lead



to further improvements in the results. However, this would require labeling more data.

We recommend our dataset for further methodological developments or as evaluation dataset for existing approaches as used in the work of G nder et al. (2022).

## 7 | CONCLUSION AND FUTURE DIRECTIONS

In this *article*, we have introduced the GrowliFlower dataset, which is a georeferenced, image-based UAV time-series dataset of two monitored cauliflower fields during their entire growth period. The proposed dataset was described, and we discussed the data collection process, which may be helpful for other similar data collection procedures. The proposed dataset comprises weekly RGB and multispectral UAV orthophotos and image time-series of individual plants reflecting weekly plant growth. In subset of the proposed dataset, in-situ reference measurements, for example, plant size, are also available, and another subset provides pre- and post-defoliation images to demonstrate the relationship between the interior and exterior of the cauliflower plant. The proposed dataset also contains annotations with segmented plant and leaf instances, as well as annotations on stems. The data are available at <http://rs.ipb.uni-bonn.de/data/>. The proposed dataset is intended to promote the use and evaluation of ML methods and foster close collaboration between different disciplines, for example, agricultural sciences, remote sensing, and ML. We have also presented baseline results of two applications of the proposed dataset using the Mask R-CNN model, that is, plant instance segmentation and leaf instance segmentation tasks. In addition, we expect that the findings and descriptions presented in this paper will help realize effective data collection processes that are transferred to other areas.

### ACKNOWLEDGMENTS

This project was funded by the European Agriculture Fund for Rural Development with contribution from North-Rhine Westphalia (17-02.12.01–10/16–EP-0004617925-19-001). In addition, this study was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2070–390732324. The authors like to acknowledge and thank the farmer Markus Schwarz for providing the experimental cauliflower fields and Jonas Westheider for assistance in annotating the instance segmentation dataset. Note that this paper must be cited if data from the GrowliFlower dataset are used or aspects of the data collection process are adopted. Open Access funding enabled and organized by Projekt DEAL.

### DATA AVAILABILITY STATEMENT

The dataset presented in this study will be freely available once the paper is published. We will be happy to provide the first version to the reviewers if needed.

### ORCID

Jana Kierdorf  <http://orcid.org/0000-0003-1145-1555>

Ribana Roscher  <http://orcid.org/0000-0003-0094-6210>

### REFERENCES

- Ahmadi, A., Halstead, M. & McCool, C. (2021) Virtual temporal samples for recurrent neural networks: applied to semantic segmentation in agriculture. In: *DAGM German Conference on Pattern Recognition*. Springer, pp. 574–588.
- Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M. & Notarnicola, C. (2015) Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12), 16398–16421.
- Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellstr m, T., Hemming, J., Kurtser, P., Ringdahl, O., Tielen, T. et al. (2020) Development of a sweet pepper harvesting robot. *Journal of Field Robotics*, 37(6), 1027–1039.
- Bender, A., Whelan, B. & Sukkarieh, S. (2020) A high-resolution, multimodal data set for agricultural robotics: a ladybird's-eye view of brassica. *Journal of Field Robotics*, 37(1), 73–96.
- Blok, P.M., van Henten, E.J., van Evert, F.K. & Kootstra, G. (2021) Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosystems Engineering*, 208, 213–233.
- Chaparro, D., Piles, M., Vall-Llossera, M., Camps, A., Konings, A.G. & Entekhabi, D. (2018) L-band vegetation optical depth seasonal metrics for crop yield assessment. *Remote Sensing of Environment*, 212, 249–259.
- Chebroli, N., Lottes, P., Schaefer, A., Winterhalter, W., Burgard, W. & Stachniss, C. (2017) Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36(10), 1045–1052.
- Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J. & Zhu, Y. (2016) Big data for remote sensing: challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207–2219.
- Cordts, M., Omran, M., Ramos, S., Scharw chter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. & Schiele, B. (2015) The cityscapes dataset. In: *CVPR workshop on the future of datasets in vision*, Vol. 2.
- Debolini, M., School, J.M., Temme, A., Galli, M. & Bonari, E. (2015) Changes in agricultural land use affecting future soil redistribution patterns: a case study in Southern Tuscany (Italy). *Land Degradation & Development*, 26(6), 574–586.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009) ImageNet: a large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE, pp. 248–255.
- Drees, L., Junker-Frohn, L.V., Kierdorf, J. & Roscher, R. (2021) Temporal prediction and evaluation of brassica growth in the field using conditional generative adversarial networks. *Computers and Electronics in Agriculture*, 190, 106415.
- Feller, C., Bleiholder, H., Buhr, L., Hack, H., Hess, M., Klose, R., Meier, U., Stauss, R., Boom, T. & Weber, E. (1995) Phenological growth stages of vegetable crops. ii. Fruit vegetables and pulses. Coding and description according to the extended BBCH scale with illustrations. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, 47(9), 217–232.
- F rster, A., Behley, J., Behmann, J. & Roscher, R. (2019) Hyperspectral plant disease forecasting using generative adversarial networks. In: *2019 IEEE international geoscience and remote sensing symposium (IGARSS 2019)*. IEEE, pp. 1793–1796.
- Grenzd rffer, G. (2019) Automatic generation of geometric parameters of individual cauliflower plants for rapid phenotyping using drone images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, XLII-2/W13, 329–335.

- Günder, M., Ispizua Yamati, F.R., Kierdorf, J., Roscher, R., Mahlein, A.-K. & Bauckhage, C. (2022) Agricultural plant cataloging and establishment of a data framework from UAV-based crop images by computer vision. *GigaScience*, 11, giac054.
- Halstead, M., Denman, S., Fookes, C. & McCool, C. (2020) Fruit detection in the wild: the impact of varying conditions and cultivar. In: *2020 digital image computing: techniques and applications (DICTA)*. IEEE, pp. 1–8.
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017) Mask R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.
- Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134.
- Kierdorf, J., Weber, I., Kicherer, A., Zabawa, L., Drees, L. & Roscher, R. (2022) Behind the leaves: estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5, 830026. <https://doi.org/10.3389/frai.2022.830026>
- Kierdorf, J., Zabawa, L., Lucks, L., Klingbeil, L., Kuhlmann, H. et al. (2019) Detection and counting of wheat ears by means of ground-based image acquisition. *Bornimer Agrartechnische Berichte*, 102, 158–167.
- Kusumam, K., Krajník, T., Pearson, S., Duckett, T. & Cielniak, G. (2017) 3D-vision based detection, localization, and sizing of broccoli heads in the field. *Journal of Field Robotics*, 34(8), 1505–1518.
- Lary, D.J., Alavi, A.H., Gandomi, A.H. & Walker, A.L. (2016) Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10.
- Lillesand, T., Kiefer, R.W. & Chipman, J. (2015) *Remote sensing and image interpretation*, 7th ed. Hoboken, New Jersey: John Wiley & Sons.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C.L. (2014) Microsoft COCO: common objects in context. In: *European conference on computer vision*. Springer, pp. 740–755.
- Lottes, P., Behley, J., Milioto, A. & Stachniss, C. (2018) Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, 3(4), 2870–2877.
- Lottes, P., Hörferlin, M., Sander, S. & Stachniss, C. (2017) Effective vision-based classification for separating sugar beets and weeds for precision farming. *Journal of Field Robotics*, 34(6), 1160–1178.
- Milioto, A., Lottes, P. & Stachniss, C. (2018) Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In: *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, pp. 2229–2235.
- Minervini, M., Fischbach, A., Scharr, H., & Tsafaris, S.A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81, 80–89.
- Mosleh, M.K., Hassan, Q.K. & Chowdhury, E.H. (2015) Application of remote sensors in mapping rice area and forecasting its production: a review. *Sensors*, 15(1), 769–791.
- Muresan, H., & Oltean, M. (2018) Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 10, 26–42.
- Nock, C.A., Vogt, R.J. & Beisner, B.E. (2016) Functional traits. In: eLS. Chichester: John Wiley & Sons, Ltd.
- Persia, M., Barca, E., Greco, R., Marzulli, M.I. & Tartarino, P. (2020) Archival aerial images georeferencing: a geostatistically-based approach for improving orthophoto accuracy with minimal number of ground control points. *Remote Sensing*, 12(14), 2232.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. et al. (2019) Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743), 195–204.
- Ren, M. & Zemel, R.S. (2017) End-to-end instance segmentation with recurrent attention. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6656–6664.
- Romera-Paredes, B. & Torr, P.H.S. (2016) Recurrent instance segmentation. In: *European conference on computer vision*. Springer, pp. 312–329.
- Roscher, R., Volpi, M., Mallet, C., Drees, L. & Wegner, J.D. (2020) SemCity Toulouse: a benchmark for building instance segmentation in satellite images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-5-2020, 109–116.
- Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T. & McCool, C. (2016) DeepFruits: a fruit detection system using deep neural networks. *Sensors*, 16(8), 1222.
- Scharr, H., Minervini, M., Fischbach, A. & Tsafaris, S.A. (2014) Annotated image datasets of rosette plants. In: *European Conference on Computer Vision*. Zürich: Suisse, pp. 6–12.
- Scharr, H., Minervini, M., French, A.P., Klukas, C., Kramer, D.M., Liu, X., Luengo, I., Pape, J.-M., Polder, G., Vukadinovic, D. et al. (2016) Leaf segmentation in plant phenotyping: a collation study. *Machine Vision and Applications*, 27(4), 585–606.
- Verger, A., Baret, F. & Weiss, M. (2014) Near real-time vegetation monitoring at global scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3473–3481.
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P. & Moreno, J. (2019) Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, 40(3), 589–629.
- Ward, D. & Moghadam, P. (2018) Synthetic Arabidopsis dataset. In: *CSIRO. Data collection*.
- Weiss, M., Jacob, F. & Duveiller, G. (2020) Remote sensing for agricultural applications: a meta-review. *Remote Sensing of Environment*, 236, 111402.
- Weyler, J., Magistri, F., Seitz, P., Behley, J., & Stachniss, C. (2022). In-field phenotyping based on crop leaf and plant instance segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2725–2734.
- Weyler, J., Milioto, A., Falck, T., Behley, J. & Stachniss, C. (2021) Joint plant instance detection and leaf count estimation for in-field plant phenotyping. *IEEE Robotics and Automation Letters*, 6(2), 3599–3606.
- Zabawa, L., Kicherer, A., Klingbeil, L., Milioto, A., Topfer, R., Kuhlmann, H. & Roscher, R. (2019) Detection of single grapevine berries in images using fully convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.

**How to cite this article:** Kierdorf, J., Junker-Frohn, L.V., Delaney, M., Olave, M.D., Burkart, A., Jaenicke, H., et al. (2022) GrowliFlower: an image time-series dataset for GROWth analysis of cauLIFLOWER. *Journal of Field Robotics*, 1–20. <https://doi.org/10.1002/rob.22122>

APPENDIX A: APPENDICES

See Figure A1  
 See Figure A2  
 See Table A1

See Figure A3  
 See Figure A4  
 See Figure A5

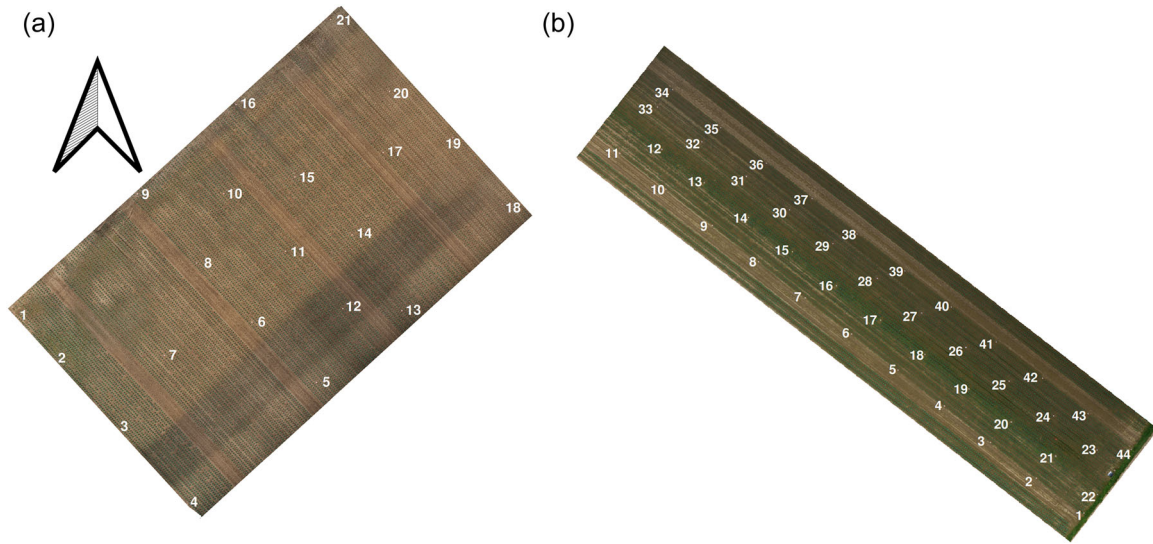


FIGURE A1 Location of ground control points in fields 1 and 2. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

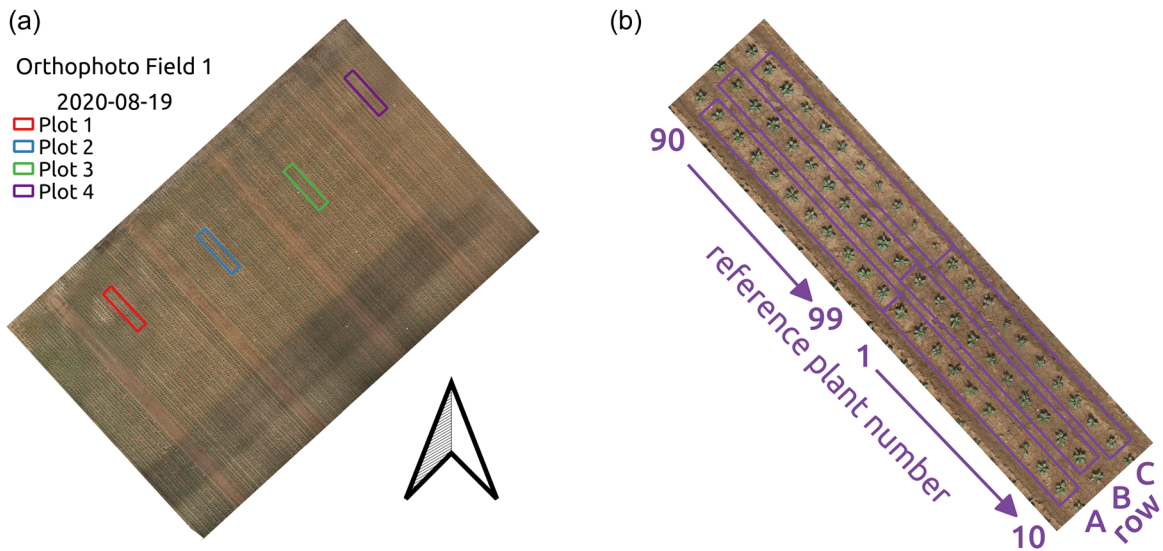
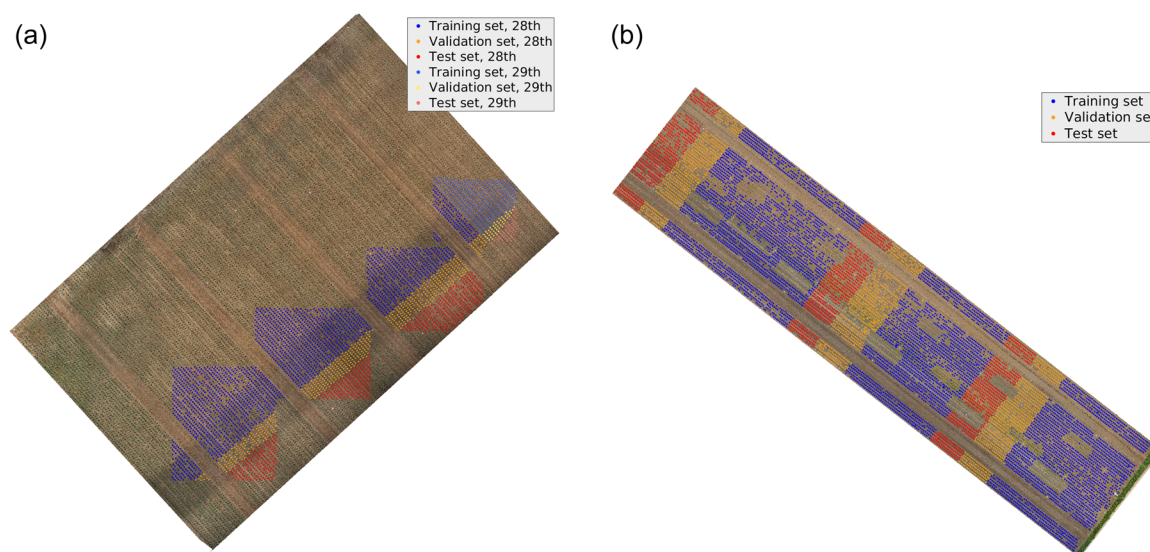


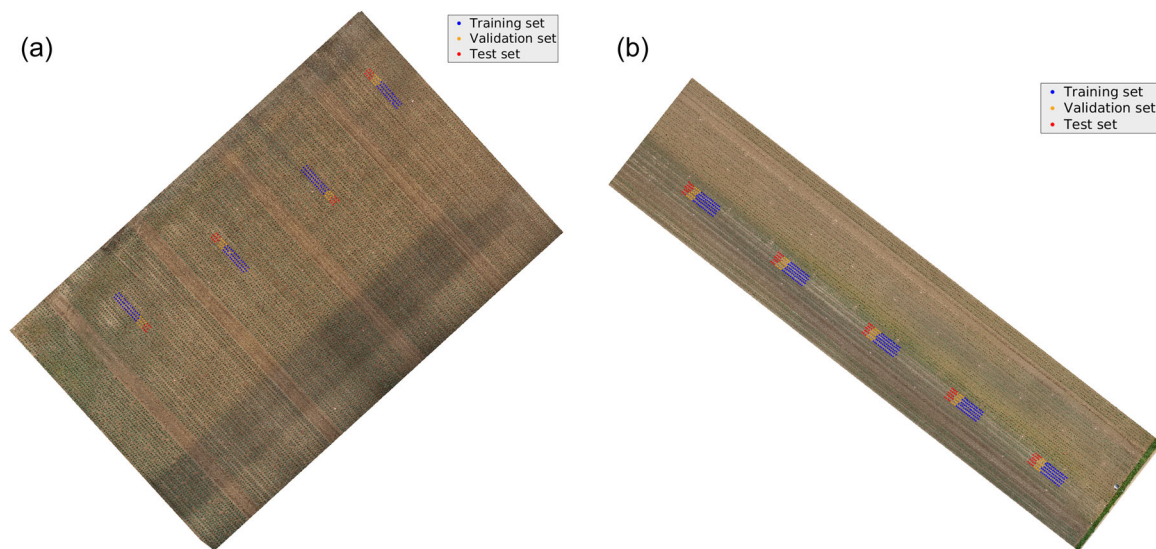
FIGURE A2 Visual overview of (a) reference plots for in-situ measurements in field 1 and (b) the respective design of reference plot 4 (including reference plants and ordering of reference plant numbers). The plot design is valid for all reference plots in field 1. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**TABLE A1** Monitored abiotic and biotic stresses.

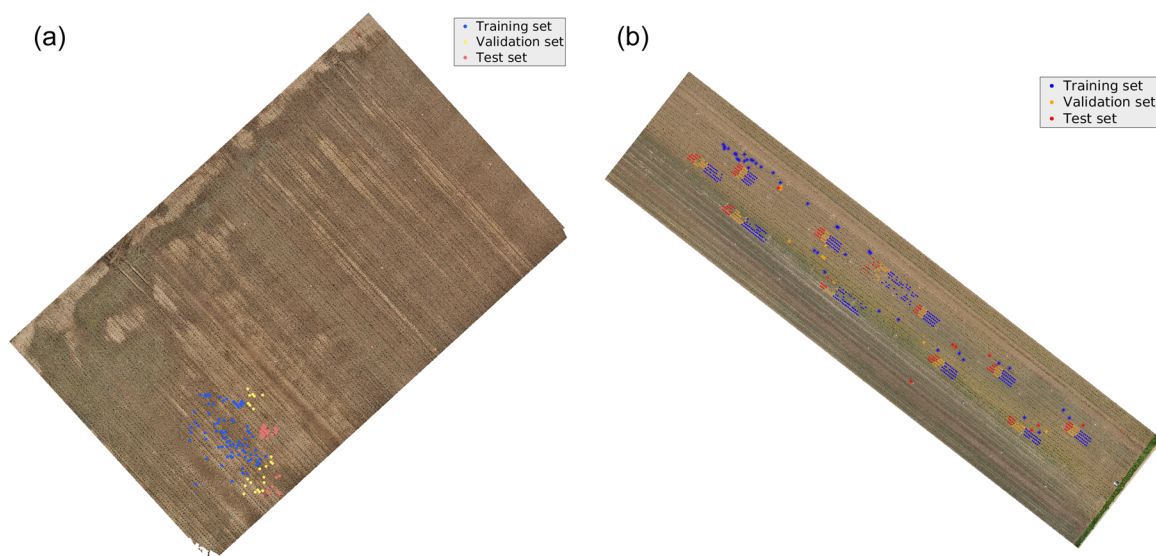
Abbreviation	Meaning	Abbreviation	Meaning
P	Plant	L	Leaf/leaves
nP	No plant	wL	Without leaves
Pl	Plant lying down	oL	Old leaves
wP	Whole plant	yL	Yellowish leaves
2P	Two plants	rL	Reddish leaves
bb	Blind bud	pgL	Pale green leaves
pd	Planted too deep	pygL	Pale yellowish green leaves
A	Aphids present	sg	Stunted growth with many shoots
C	Coal fleas present	dT	Damage to leaves caused by tractor
F	Flies present		

**FIGURE A3** Separation of plants fields 1 and 2 in GrowliFlowerT in training (blue), validation (yellow), and testing (red) sets. For field 1, the two planting days are separated using dark colors for July 28, 2020 and light colors for July 29, 2020. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]





**FIGURE A4** Separation of reference plants in fields 1 and 2 in GrowliFlowerR in training (blue), validation (yellow), and testing (red) sets. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.27122)]



**FIGURE A5** Separation of defoliated plants in fields 1 and 2 in GrowliFlowerD in training (blue), validation (yellow), and testing (red) sets. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/rob.27122)]