



# Landbedeckung in Sachsen 1961-1979 (Teil 1)

Schriftenreihe des LfULG, Heft 6/2023



Erfassung von Veränderungen  
in der Landbedeckung  
in Sachsen 1961-1979:  
Analyse anhand historischer CORONA-  
Spionageaufnahmen und  
tiefer neuronaler Netze  
(Teil 1)

Lucas Kugler, Christopher Marrs, Eric Kosczor und JProf. Dr. Matthias Forkel  
Technische Universität Dresden

Im Auftrag des Sächsischen Landesamtes für Umwelt, Landwirtschaft und Geologie

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b> .....	<b>9</b>
<b>2</b>	<b>Daten und Methoden</b> .....	<b>10</b>
2.1	Verfügbarkeit und Auswahl der Bilddaten.....	10
2.2	Georeferenzierung.....	12
2.3	Klassifikationsschema und Generierung von Referenzdaten.....	15
2.3.1	Erstellung von Referenzdaten.....	18
2.3.2	Generierung von Bildmerkmalen.....	21
2.4	Prozessierung der Eingabedaten.....	24
2.4.1	Semantische Segmentierung.....	24
2.4.2	Patch-basierte Bildklassifikation.....	25
2.4.3	Spaltung der Beispieldaten: Trainingsdaten, Testdaten, Validierungsdaten.....	26
2.4.4	Datenerweiterung.....	26
2.5	Klassifikation mit neuronalen Netzen.....	28
2.5.1	U-Net.....	29
2.5.2	U-Nets mit verschiedenen Kontraktionspfaden.....	29
2.5.2.1	ResNet.....	29
2.5.2.2	Inception.....	30
2.5.3	Spezielle Weiterentwicklungen U-förmiger Architekturen.....	30
2.5.3.1	U-Net2+.....	30
2.5.3.2	U-Net3+.....	31
2.5.3.3	Attention U-Net.....	31
2.5.3.4	Transformer U-Net (Trans U-Net).....	32
2.5.3.5	Swin U-Net.....	32
2.5.4	CNN.....	32
2.5.5	ResNets & InceptionNets.....	33
2.5.6	Semi Supervised Generative Adversarial Network (sGAN).....	33
2.6	Genauigkeitsanalyse.....	33
2.7	Training der Architekturen.....	35
2.8	Interferenz.....	35
2.9	Nachprozessierung.....	37
2.10	Konsistenz.....	37
2.11	Implementierung und Software.....	40
<b>3</b>	<b>Ergebnisse</b> .....	<b>41</b>
3.1	Georeferenzierung.....	41
3.2	Ergebnisse der Datenerweiterung.....	41
3.3	Training der semantischen Segmentierungs-Methoden.....	43
3.4	Testergebnisse der semantischen Segmentierung.....	44
3.5	Training der Patch-basierten Bildklassifikationsmethoden.....	46
3.6	Klassifikationsergebnisse der Patch-basierten Bildklassifikation.....	47
3.7	Anwendung der Architekturen.....	48
3.7.1	Datenprozessierung.....	48
3.7.2	Methodische Entscheidung zur Interferenz.....	48

3.8	Interferenz .....	50
3.9	Nachprozessierung.....	52
3.9.1	Kombination der Interferenzergebnisse .....	52
3.9.2	Prozessierung der Zielauflösung der Landbedeckung .....	52
3.9.3	Anwendung des Regelsatzes .....	54
3.10	Analyse der Landbedeckung .....	56
<b>4</b>	<b>Diskussion .....</b>	<b>60</b>
4.1	Georeferenzierung.....	60
4.2	Beispieldatensatz .....	61
4.3	Methode .....	64
4.4	Training .....	64
4.5	Interferenz .....	64
4.6	Nachprozessierung.....	66
4.7	Landbedeckungsänderung .....	68
<b>5</b>	<b>Zusammenfassung.....</b>	<b>70</b>

## Abbildungsverzeichnis

Abbildung 1:	Anzahl der verfügbaren KH-Bilder für das Gebiet des heutigen Freistaates Sachsen. Die Qualität und Nutzbarkeit wurde visuell anhand der Vorschaubilder kategorisiert. ....	11
Abbildung 2:	Abdeckung mit "gut" (keine oder geringe Wolkenbedeckung) nutzbaren Bildern für Zeitschnitte von jeweils drei Jahren über dem Gebiet des Freistaates Sachsen. ....	12
Abbildung 3:	Beispiele für Kontrollpunkte .....	13
Abbildung 4:	Talsperre Eibenstock 1973 und 2018 .....	14
Abbildung 5:	Abbildung einer alten Feldgrenze (links) und korrespondierende CORONA-Bilddaten.....	14
Abbildung 6:	Abbildung aller Beispielgebiete .....	17
Abbildung 7:	Beispiel für eine Zerlegung eines Referenzbildes in Segmente für die Generierung von Referenzdaten. ....	19
Abbildung 8:	Abbildung des Beispielgebietes 10 und der korrespondierenden semantischen Maske.....	20
Abbildung 9:	Relative Verteilung der klassifizierten Beispielpixel in den Studiengebieten. ....	20
Abbildung 10:	Verteilung der panchromatischen Bildinformationen über alle Klassen. ....	21
Abbildung 11:	Darstellung der genutzten GLCM-Texturmaße für das Studiengebiet 10. 23	
Abbildung 12:	Korrelation der GLCM-Texturmaße für Studiengebiet 10 .....	23
Abbildung 13:	Beispiele für Eingabebildpatches (nur originale panchromatische Bildinformation) und der korrespondierenden semantischen Maske zur Lösung der Landbedeckungsklassifikation als semantische Segmentierung. ....	25
Abbildung 14:	Beispiele von Eingabebildpatches zur Patch-basierten Bildklassifikation. ....	26
Abbildung 15:	Ablauf der Interferenz zur semantischen Segmentierung einer Corona Szene.....	36
Abbildung 16:	Differenz der CORONA-Aufnahmen zu den Referenzdaten, Mittelwert 1965 = 6,47 m, Median 1965 = 4,12 m, Mittelwert 1975 = 5,17 m, Median 1975 = 3 m .....	41
Abbildung 17:	Trainingsverlauf eines ResNet18 mit und ohne Methoden der Datenerweiterung. ....	42
Abbildung 18:	Trainingsperformance der Architekturen zur semantischen Segmentierung der Landbedeckung. ....	43
Abbildung 19:	Beispiele für die semantische Segmentierung von Trainingsdaten mit einem Trans U-Net. ....	45
Abbildung 20:	Trainingsperformance der Architekturen zur semantischen Segmentierung der Landbedeckung. ....	46
Abbildung 21:	Beispielhafter Interferenztest für KH-Bilddaten aus dem Jahr 1975. ....	49
Abbildung 22:	Semantische Segmentierung einer CORONA-Szene von 1975. ....	50
Abbildung 23:	Vergleich von semantischen Masken und Interferenzergebnissen für 1975. ....	51
Abbildung 24:	Landbedeckungskarte Sachsen 1975. ....	52
Abbildung 25:	Veranschaulichung der Nachprozessierung der semantischen Karten für 1975 von 2 x 2 m auf 100 x 100 m. ....	53
Abbildung 26:	Veranschaulichung der Anpassung der semantischen Karten auf das LfULG-Referenzraster mit 100 x 100 m Auflösung. ....	53
Abbildung 27:	Auswirkung der Anwendung des Regelsatzes auf die Klassifikationsergebnisse der Jahre 1965 und 1975. ....	54

Abbildung 28: Beispiele für auffällige Änderungen in den Klassifikationsergebnissen ohne und mit Anwendung des Regelsatzes.....	55
Abbildung 29: Übersicht über Anzahl der Landbedeckungsänderungen über den gesamten Zeitraum (A) vor und (B) nach Anwendung des Regelsatzes und Flag-Karten von (C) unzulässigen Pixeln mit 3 oder mehr Landbedeckungsänderungen im Untersuchungszeitraum und (D) Pixeln, die im Zuge der Anwendung des Regelsatzes eine Änderung an mindestens einer Stelle in der Zeitreihe erfahren haben. ....	56
Abbildung 30: Zusammensetzung der Landbedeckung über alle Zeitschnitte. ....	57
Abbildung 31: Veränderung der Landbedeckung über die Zeit. ....	58
Abbildung 32: Modernes DOP vom GeoSN mit Daten des Straßennetzes von OpenStreetMap.....	60
Abbildung 33: Überlagerung der modernen Daten des Straßen und Schienennetzes von OpenStreetMap und CORONA-Bilddaten. ....	61
Abbildung 34: Beispielausschnitt aus den KH-Bilddaten über alle Zeitschnitte. ....	62
Abbildung 35: Beispielpatches mit den korrespondierenden semantischen Masken und den Vorhersagen des Trans U-Nets. ....	63
Abbildung 36: Beispiel für die Ergebnisse der semantischen Segmentierung in bewölkten Bereichen für den Zeitschnitt 1975. ....	65
Abbildung 37: Diversität von Gewässerflächen. ....	66

## Tabellenverzeichnis

Tabelle 1: Auszug aus der verwendeten CORINE Landbedeckungsnomenklatur. ....	16
Tabelle 2: Zuordnung der Studiengebiete zu den Zielzeitschnitten .....	18
Tabelle 3: Berechnung der GLCM-Texturmaße (Haralick et al., 1973). ....	22
Tabelle 4: Regelsatz für die Definition von plausiblen Landbedeckungsänderungen innerhalb eines Zeitschnitts ("von Klasse A nach Klasse B").....	39
Tabelle 5: Ergebnisse der semantischen Segmentierung der Testdaten. ....	44
Tabelle 6: Ergebnisse der Patch-basierten Landbedeckungsklassifikation. 47	
Tabelle 7: Flächenanteile pro Klasse und pro Zeitschnitt in Quadratkilometer .....	58

## Abkürzungsverzeichnis

AG	Attention Gates
CLC	CORINE Land Cover
CORINE	Coordination of Information on the Environment
CNN	Convolutional Neural Networks
DGM	Digitales Geländemodell
FCN	Fully Connected Networks
GAN	Generative Adversarial Networks
GeoSN	Staatsbetrieb Geobasisinformation und Vermessung Sachsen
GLCM	Gray Level Co-occurrence Matrix (Graustufen-Ko-Auftrittsmatrix)
IoU	Intersection over Union
KH	Keyhole
KI	Künstliche Intelligenz
LfULG	Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie
OS	Open Street Map
sGAN	Semi-Supervised Generative Adversarial Networks
USGS	United States Geological Service
UTM	Universal Transverse Mercator
VGG	Visual Geometry Group (Universität Oxford)
VRT	Virtual Raster Template
WMTS	World Map Tile Service

# 1 Einleitung

Die Kartierung historischer Landbedeckungs- und Landnutzungsänderungen ist essentiell für das Verständnis und die Modellierung von Veränderungen in der regionalen Landschaftsstruktur, Klima und Wasserhaushalt (Renner et al., 2014). Gezielte zivile Erdbeobachtungsprogramme für die Landoberflächenkartierung sind jedoch erst seit 1972 vorhanden (z. B. Landsat-1) (Hansen and Loveland, 2012). Zur Kartierung früherer Landbedeckungsänderungen können nur flugzeuggetragene Luftbilder aus der öffentlichen Vermessung oder von militärischen oder geheimdienstlichen Aufnahmen genutzt werden. Seit den 1960er Jahren gibt es beispielsweise Bilddaten von verschiedenen Satellitenspionagesystemen der Central Intelligence Agency (CIA) der Vereinigten Staaten. Dabei zeichneten verschiedene Satellitensysteme des sogenannten Keyhole-Programms (KH) erstmalig die Erdoberfläche großflächig auf analogen Filmstreifen auf (Song et al., 2015; Walz et al., 2004). Die KH-Kameras KH-1 (1959-1960), KH-2 (1960-1961), KH-3 (1961-1962), KH4 (1962-1963), KH-4A (1963-1969) und KH-4B (1967-1972) werden auch weitverbreitet als "CORONA" bezeichnet. "CORONA" dient auch oft als Synonym für das gesamte Beobachtungsprogramm und die dazugehörigen Bilddaten. Spätere Kameras haben jedoch andere Namen bekommen: KH-5 Argon (1961-1964), KH-6 Lanyard (1963), KH-7 Gambit (1963-1967), KH-8 Gambit-3 (1966-1984) und KH-9 Hexagon (1971-1986). Während diese Aufnahmen ursprünglich als streng geheim klassifiziert waren, wurden viele Aufnahmen in den 1990er und 2000er Jahren "deklassifiziert" und können seitdem beim United States Geological Survey (USGS) für wissenschaftliche Zwecke bezogen werden.

Allerdings ist die automatische Analyse dieser historischen Bilddaten der US-Satellitenspionagesysteme bisher nur eingeschränkt möglich und verlangt einen hohen Bedarf an grundlegender Forschung. Beispielsweise wurden diese Daten aus variablen Orbits und mit unterschiedlichen Kameramodellen aufgenommen und somit hat jeder Bilddatensatz eigene geometrische und radiometrische Eigenschaften. Die Bilder liegen nur panchromatisch (Graustufenbilder) vor. Damit können klassische Methoden der multispektralen Bildklassifikation (Richards, 2013) nicht einfach angewendet werden. Als Alternative haben Bildklassifikationsverfahren aus dem Bereich der künstlichen Intelligenz (KI; z. B. Convolutional Neural Networks oder andere Deep Learning-Methoden) das Potential dieser Daten in bisher unerreichter Qualität und Quantität auszuwerten (Deshpande et al., 2021; Zhang et al., 2020).

Das Ziel dieses Vorhabens besteht in der Kartierung der Landbedeckung und Landnutzung in Sachsen in den 1960er und 1970er Jahren anhand historischer Satellitendaten des Keyhole-Programms (z. B. Corona, Gambit, Hexagon). Dazu wird ein methodischer Ansatz anhand von Deep Learning-Methoden entwickelt und umgesetzt um die Landbedeckung und -nutzung zu klassifizieren. Im Ergebnis soll die Landnutzung entsprechend der CORINE-Landbedeckungsklassifikation als Datensatz für die Weiterverwendung für Umwelt- und Klimaanalysen bereitgestellt werden. Das Vorhaben hat einen hohen Anteil an ergebnisoffener Grundlagenforschung insbesondere in der Implementierung und Anwendung von Deep Learning-Methoden.

## 2 Daten und Methoden

Für die Landbedeckungsklassifikation mit Daten aus dem Keyhole-Programm und mit Deep Learning-Verfahren sind mehrere Arbeitsschritte nötig. Erstens müssen geeignete Daten für einzelne Jahre aus den 1960er und 1970er Jahren ausgewählt und beschafft werden, die sich gut für die Landbedeckungsklassifikation eignen. Zweitens müssen alle Daten georeferenziert werden, da sie nur als gescannte Bilddaten ohne Koordinatensystem geliefert werden. Drittens müssen ausreichend Referenzdaten für jede Landbedeckungsklasse gesammelt werden. Im Anschluss können, viertens, verschiedene Strukturen für Deep Learning-Modelle implementiert, trainiert und getestet werden. Im Rahmen dieser Test muss auch evaluiert werden inwieweit sich Modelle auf Bilddaten anderer Kameras, anderer geographischer Räume und anderer Jahreszeiten übertragen lassen. Dabei wird getestet, wie verschiedene Modellstrukturen, Parametrisierungen und die Verwendung verschiedener Konfigurationen von Trainingsdaten die Genauigkeit der Klassifikation einzelner Klassen und der Gesamtgenauigkeit beeinflussen. Wenn eine erfolgreiche Übertragung demonstriert werden kann, wird im Anschluss die Klassifikationsmethode auf alle Bilddaten angewendet um Landbedeckungskarten für verschiedene Jahre zu generieren. Die einzelnen methodischen Schritte werden im Folgenden beschrieben.

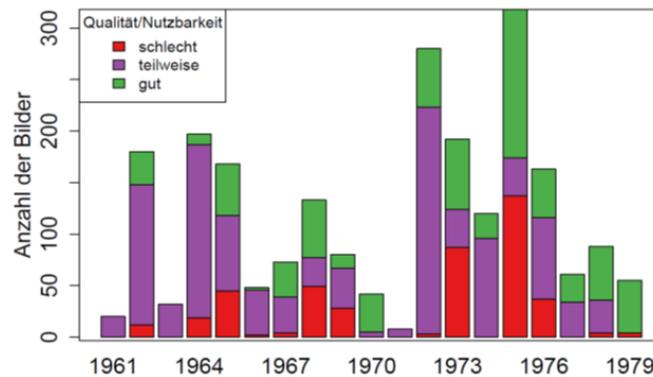
### 2.1 Verfügbarkeit und Auswahl der Bilddaten

Bilddaten der KH-Sensoren verschiedener Generationen wurden über den Earth Explorer des USGS<sup>1</sup> bezogen. Die verwendeten Bilddaten wurden 1996 (Declass 1), 2002 (Declass 2) und 2013 (Declass 3) freigegeben. Für die räumliche Ausdehnung des heutigen Freistaates Sachsen wurde eine Abfrage über alle verfügbaren Bilddatensätze durchgeführt. Bilddaten sind überwiegend aus Declass 1 (n = 973) und Declass 3 (n = 1308) verfügbar. Declass 2 umfasst nur 84 Aufnahmen, die nach visueller Begutachtung alle nicht geeignet sind und daher nicht weiter berücksichtigt wurden. Aus dieser Abfrage lassen sich grundlegende Metadaten zu jedem Bilddatensatz abrufen, wie z. B. Aufnahmedatum, Missionsnummer, Blickrichtung der Kamera relativ zur Flugrichtung (vorwärts oder rückwärts), Auflösung der Kamera und die Eckkoordinaten des Bildes. Informationen zum verwendeten Kameramodell (z. B. KH-4 Corona oder KH-7 Gambit) sind nicht verfügbar, können aber bei Bedarf aus der Missionsnummer abgeleitet werden. Die angegebenen Eckkoordinaten der Bilder sind nur grobe Annäherungen, in einzelnen Fällen lagen die heruntergeladenen Bilddaten auch bis zu 60 km weiter nördlich oder südlich.

---

<sup>1</sup> <https://earthexplorer.usgs.gov/>

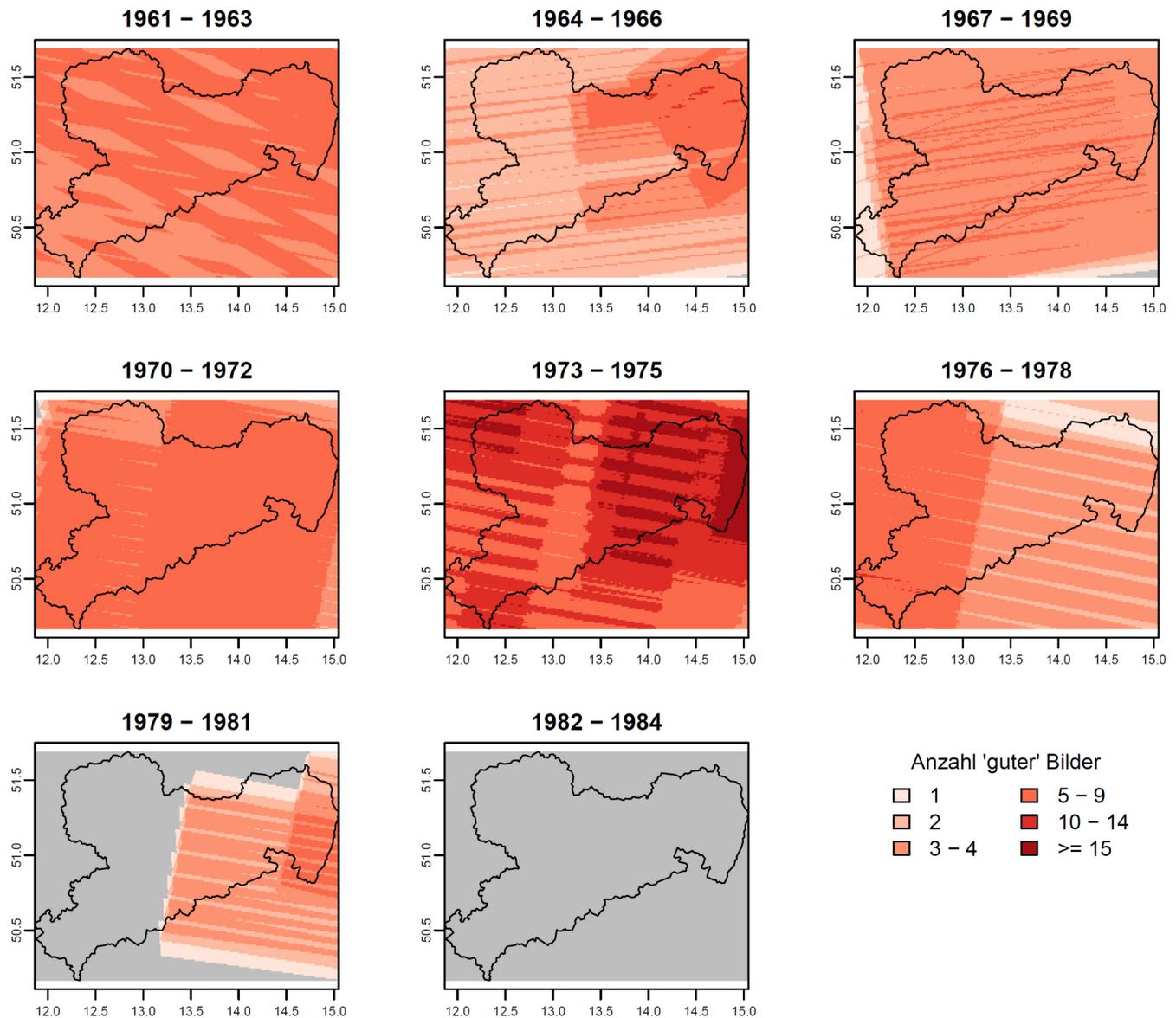
## Anzahl der verfügbaren KH-Bilddaten im Freistaat Sachsen



**Abbildung 1: Anzahl der verfügbaren KH-Bilder für das Gebiet des heutigen Freistaates Sachsen. Die Qualität und Nutzbarkeit wurde visuell anhand der Vorschaubilder kategorisiert.**

Die Qualität und damit die Nutzbarkeit der einzelnen Bilder wurde anhand von Vorschaubildern bewertet. Dabei wurde visuell die Wolken- und Schneebedeckung jedes Bildes abgeschätzt. Komplett wolkenfrei bzw. klar waren 366 Bilder, minimale Wolkenbedeckung gab es in 93 Bildern, 472 Bilder hatten große Felder von einzelnen Cumuluswolken und 656 Bilder waren nahezu komplett bewölkt. 336 Bilder waren vollständig oder teilweise mit Schnee bedeckt. Für 47 Bilder gab es keine Vorschau oder sie hatten visuell eine geringe Auflösung, waren unvollständig oder zeigten verschiedene Artefakte. Anhand dieser Einschätzung wurde die Nutzbarkeit der Bilder als "gut" (keine oder geringe Wolkenbedeckung), "teilweise" (wenige Wolken oder nur Wolken in Teilen des Bildes, Schneebedeckung ohne Wolken) und "schlecht" (vollständige Wolkenbedeckung oder große Wolkenfelder, keine Vorschaubilder verfügbar oder mit Artefakten) kategorisiert. Im Ergebnis haben 704 Bilder eine gute Nutzbarkeit, 431 Bilder sind teilweise nutzbar und 1123 sind schlecht nutzbar. Die Verfügbarkeit von gut nutzbaren Bildern unterscheidet sich dabei stark zwischen den einzelnen Jahren (Abbildung 1). Beispielsweise gibt es 144 gut nutzbare Bilder im Jahr 1975, aber keine nutzbaren Bilder in Jahren 1961, 1963, und 1971.

Die Verfügbarkeit von gut nutzbaren Bilddaten unterscheidet sich in den einzelnen Jahren auch räumlich teilweise stark (Abbildung 2). In den Zeiträumen 1961 - 1963, 1967 - 1969 und 1970 - 1972 gibt es eine nahezu gleichmäßige Abdeckung mit 3 bis 4 Bildern im gesamten Gebiet des Freistaates Sachsen. Im Zeitraum 1964 - 1966 gibt es eine häufigere Abdeckung in Ostsachsen als in Mittel-, Nord- oder Westsachsen. Die häufigste Abdeckung gibt es im Zeitraum 1973 - 1975 (v. a. im Jahr 1975), wo Teile Sachsens mit bis zu 20 Bildern abgedeckt werden. In den Zeiträumen 1979 - 1981 und 1982 - 1984 ist Sachsen nur unvollständig bzw. gar nicht mit Bilddaten abgedeckt. Ausgehend von dieser Analyse wurden 241 KH-Bilddaten verschiedener Missionen ausgewählt und beschafft. Diese decken die Jahre 1962 (Juli und September), 1965 (Mai und August), 1967 (Mai und November), 1968 (Dezember), 1969 (August), 1970 (November), 1972 (Juli), 1973 (Mai und August), 1975 (August und September), 1976 (August), 1978 (April und Mai) und 1979 (April, Mai und Juni) ab.



**Abbildung 2: Abdeckung mit "gut" (keine oder geringe Wolkenbedeckung) nutzbaren Bildern für Zeitschnitte von jeweils drei Jahren über dem Gebiet des Freistaates Sachsen. Gute Bilder beschreiben KH-Bilddaten welche nicht vorwiegend von Wolkenbedeckung geprägt sind, geometrisch stark verzerrt sind oder ein Rauschen aufweisen.**

## 2.2 Georeferenzierung

Die Georeferenzierung der Bilddaten wurde für Zone 33 des universalen transversalen Mercator-Koordinatensystems (UTM) auf WGS84 durchgeführt (EPSG:32633). Eine automatisierte Georeferenzierung der KH-Bilddaten ist nicht möglich, da die Flugbahnen der jeweiligen Raketen nicht präzise kontrolliert wurden und daher die Bilddaten verschiedene und unregelmäßige Verzerrungen aufweisen. Daher wurden manuell Passpunkte auf den KH-Bilddaten und auf den Referenzdaten ausgewählt und die Georeferenzierung mit einer hoch-flexiblen Transformationsmethode durchgeführt.

Die Bilder wurden in ArcGIS 10.8.1 unter Verwendung der Transformation Adjust georeferenziert. Ursprünglich wurde die Spline-Transformation gewählt, die sich jedoch aufgrund der großen Bildgröße und der Anzahl der benötigten Referenzpunkte als zu zeitaufwändig erwies.

Diese Streifen wurden vom USGS in Abschnitten geliefert, d.h. ein Bildstreifen wurde in vier oder mehr Einzelbildern, z. B. DS1019-1057DA027 a-d zur Verfügung gestellt. In einigen Fällen, insbesondere in den

späteren Jahren, wurden die Daten in bis zu dreizehn Einzelbildern geliefert, z. B. D3C1206-300319A48 a-m. Diese überschneiden sich dabei häufig mit Nachbarbildern. In je mehr Bilder der Bildstreifen aufgespalten wurde, desto mehr Zeit war für die Georeferenzierung erforderlich.

Mithilfe der vom USGS zur Verfügung gestellten Geometriedatei "footprints" (deutsch: Fußabdruck) konnte der ungefähre Standort jedes Bildes ermittelt werden. Dies führte zu einer Reihe von Problemen, da der Fußabdruck oft nicht mit dem Bild übereinstimmte. So konnte der Fußabdruck beispielsweise weiter östlich/westlich oder nördlich/südlich liegen als das Bild. In einigen Fällen lag dies in der Größenordnung von 10 Kilometern. Ein Datensatz war um über 90 km nach Süden verschoben. Weiterhin waren die Bilder häufig spiegelverkehrt, was dazu führte, dass es länger dauerte, den allgemeinen Standort des Bildes zu bestimmen. In diesem Fall wurden große Waldgebiete zur Standortbestimmung herangezogen. In ausgedehnten Agrarlandschaften war dieser erste Ansatz noch zeitaufwändiger.

Sobald die Position des Bildes gefunden war, wurde das Bild an die ungefähre Position verschoben. Anschließend wurde für jede der vier Ecken des Bildes ein geeigneter Bezugspunkt gefunden. Das Bild wurde dann über die Landschaft "verteilt" und an seinem ungefähren Standort angeheftet.

Zur Bereitstellung von Referenzpunkten, auf die die KH-Bilddaten georeferenziert werden konnte, wurde ein moderner Datensatz für das Straßen- und Schienennetz verwendet. Die verwendeten Daten sind OpenStreetMap-Daten, die von Geofabrik.de<sup>2</sup> heruntergeladen wurden. Darüber hinaus wurde ein modernes Orthofoto verwendet, ein WMTS, das über GeoSN, den sächsischen Geodaten-Viewer-Dienst, abgerufen wurde, der auf GeoSN<sup>3</sup> verfügbar ist. Die Kreuzungen von Straßen wurden als Kontrollpunkte verwendet, siehe Abbildung 3.

Beispielabbildungen für Kontrollpunkte



**Abbildung 3: Beispiele für Kontrollpunkte**

Die Verwendung eines modernen Datensatzes erforderte eine zweite Überprüfung, um sicherzustellen, dass die neuen Straßendaten mit denen des Bildes übereinstimmen, zum Beispiel durch die Verwendung

---

<sup>2</sup> <https://www.geofabrik.de/>

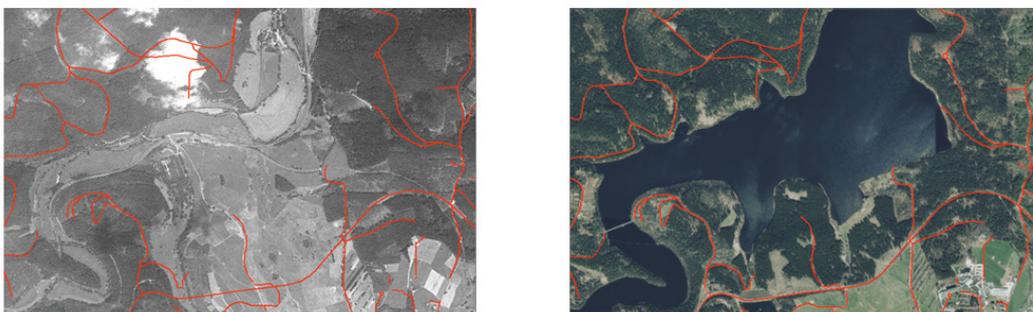
<sup>3</sup> [www. https://www.geosn.sachsen.de/](https://www.geosn.sachsen.de/)

moderner Orthofotos. In vielen Fällen unterschieden sich Straßenkreuzungen und sogar die Straßen selbst zwischen dem modernen Datensatz und den Bildern.

Insgesamt wurden für jedes Bild zwischen 800 und 3500 Passpunkte benötigt. Für Bilder in Flachlandgebieten waren weniger Passpunkte erforderlich, für Bilder in Gebirgsregionen oder in Gebieten mit tiefen Tälern wie dem Erzgebirge und dem Vogtland wurden teilweise deutlich mehr Kontrollpunkte benötigt.

Die sächsische Landschaft hat sich seit der Aufnahme der Corona-Bilder so stark verändert, dass einige Teile des Bildes nicht mehr referenziert werden konnten, z. B. die Tagebauabbaugebiete in der Lausitz oder bei Leipzig, wo auf den alten Bildern Dörfer und Wälder usw. zu sehen sind und heute keine Straßennetze mehr vorhanden sind (siehe Abbildung 4). In diesen Fällen wurde darauf geachtet, dass das umliegende Gebiet mit mehr Passpunkten als üblich referenziert wurde. Der gleiche Ansatz wurde in wolkenbedeckten Gebieten verfolgt.

Talsperre Eibenstock 1973 (links) and 2018 (rechts)



**Abbildung 4: Talsperre Eibenstock 1973 (links) und 2018 (rechts)**

In großen landwirtschaftlich genutzten Gebieten, in denen der Abstand zwischen straßenbasierten Kontrollpunkten erheblich war, wurden auch andere Landschaftselemente zur Georeferenzierung herangezogen. Zum Beispiel dort, wo Hecken aufeinandertreffen oder wo eine alte Feldgrenze in den modernen Orthofotos noch unter dem Boden zu sehen ist. In einigen Fällen dienten einzelne Feldbäume als Passpunkte in ausgedehnten Agrarlandschaften, dasselbe gilt für einige, wenn auch nicht alle, Strommasten und Gebäude. In Abbildung 5 ist die schwache Linie der alten Feldgrenze ist noch unter dem modernen Feld sichtbar.



**Abbildung 5: Abbildung einer alten Feldgrenze (links) und korrespondierende CORONA-Bilddaten**

Die Genauigkeit der Georeferenzierung wurde sowohl visuell als auch durch den Vergleich der Georeferenzbilder mit den heutigen Orthofotos überprüft. Für den Abgleich mit den Orthofotos wurde das folgende Verfahren angewandt.

- 1) Ein Satz von 100 zufällig generierten Punkten wurde in ArcMap mit dem Random Point Generator in der Arc Toolbox erstellt. Als Abstand zwischen den Punkten wurde ein Minimum von 1 km vorgegeben.
- 2) Da es sich um Zufallspunkte handelt, konnten diese Punkte überall liegen, z. B. auf einem Feld, in einem See oder unter einer Wolke. Um sicherzustellen, dass der Punkt auf den georeferenzierten Bildern gefunden werden konnte, wurde die nächstgelegene Straßenkreuzung oder ein anderes Merkmal als Kontrollpunkt gewählt. Diesen Punkten wurde dann eine Koordinate zugewiesen.
- 3) Anschließend wurde eine zweite Geometriedatei erstellt, in dem die Standorte dieser Straßenkreuzungen oder anderen Merkmale angezeigt wurden. Dazu wurde jeder Kontrollpunkt aufgesucht und ein Punkt an derselben Straßenkreuzung oder einem anderen Merkmal erstellt, dessen Koordinaten dann hinzugefügt wurden. Auf diese Weise wurde ein Datensatz mit bestimmten Standorten auf den modernen Orthofotos und ein Datensatz mit den entsprechenden Standorten auf den georeferenzierten Bildern erstellt. In Fällen, in denen die spezifische Straßenkreuzung oder ein anderes Merkmal auf den georeferenzierten Bildern nicht vorhanden war, wurde ein neuer Punkt erstellt, der dem Original so nahe wie möglich kam.
- 4) Der Abstand zwischen den Punkten wurde dann durch die Anwendung der folgenden Gleichung berechnet.

$$D = \sqrt{(dx^2 + dy^2)} \quad (1)$$

D ist der Abstand zwischen dem Kontrollpunkt und dem entsprechenden Ort auf dem georeferenzierten Bild, dx ist die Differenz zwischen der Ostrichtung des Kontrollpunkts und der entsprechenden Position auf dem georeferenzierten Bild, dy ist die Differenz zwischen dem Nordwert des Kontrollpunkts und dem entsprechenden Ort auf dem georeferenzierten Bild.

- 5) Das Verfahren wurde für die Jahre 1965 und 1975 durchgeführt und die Ergebnisse werden in Kapitel 3 vorgestellt.

### 2.3 Klassifikationsschema und Generierung von Referenzdaten

Landbedeckungsklassen können in einem hierarchischen Klassifizierungsschema organisiert werden (Congalton et al., 2014). Durch die Europäische Umweltagentur wurde 1998 ein solcher hierarchischer Ansatz unter dem Namen Coordination of Information on the Environment (CORINE) Land Cover (CLC) entwickelt. Das Schema besteht aus fünf großen Gruppen (künstliche Flächen, landwirtschaftliche Flächen, Wälder und naturnahe Flächen, Feuchtgebiete, Gewässer). Diese fünf Gruppen werden wiederum in zwei untergeordnete Ebenen mit 15 Klassen auf Ebene 2 und 42 Klassen auf Ebene 3 gegliedert. Ziel dieser Arbeit ist es mindestens die fünf Klassen der 1. Ebene sowie innerhalb der Klasse 300 Wälder und naturnahe Flächen auch die Klassen der 2. Ebene (310 Wälder, 320 Strauch- und Krautvegetation, 330 Offene Flächen ohne oder mit geringer Vegetation) zu klassifizieren (Tabelle 1). Wünschenswert wäre außerdem eine Unterscheidung in der 2. Ebene innerhalb der Klasse 200 Landwirtschaft und eine Unterscheidung in der 3. Ebene in der Klasse 310 Wälder (311 Laubwälder, 312 Nadelwälder, 313 Mischwälder). Die Referenzdaten werden zunächst mit der maximal tiefsten Klassengenauigkeit erstellt. Dies ist sinnvoll um mit dem Datensatz Folgeuntersuchungen zu ermöglichen und auch Klassifikationen über die Mindestanforderungen hinaus zu ermöglichen. In manchen Gebieten lässt die Bildqualität allerdings keine Unterscheidung auf dritter oder zweiter CORINE-Klassifikationsebene zu. Das CORINE Klassifikations-

schema lässt es zu die Klassen von tieferen Ebenen in Klassen übergeordneter Ebenen zusammen zu fassen. Der aktuell angewandte Datensatz wird zur Entwicklung der Architekturen auf die Mindestanforderungen des Projektes heruntergerechnet. Die nachfolgende Tabelle zeigt den Aufbau des CORINE-Klassifikationsschemas. Die fett hervorgehobenen Klassen bezeichnen die sieben Zielklassen.

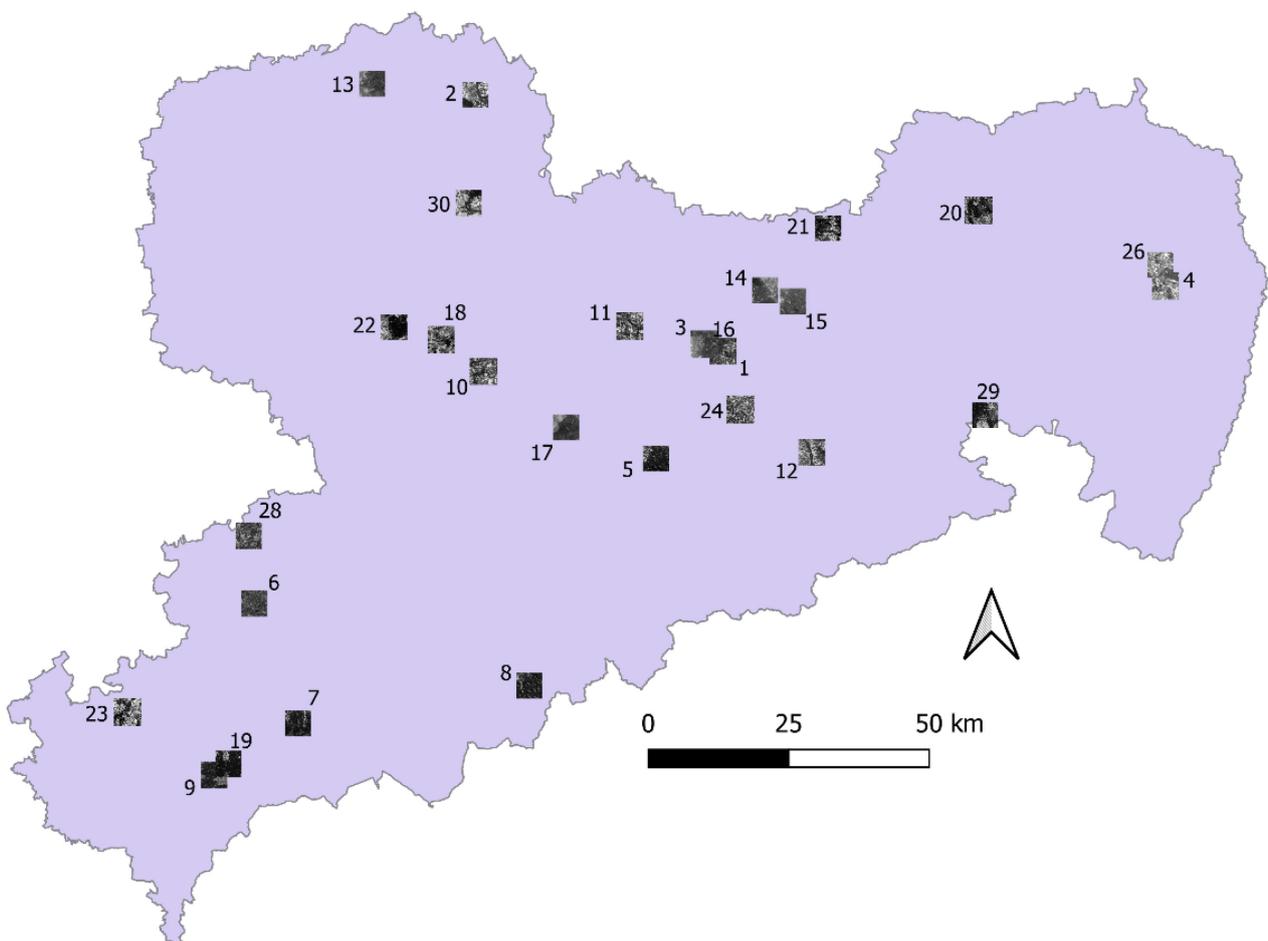
**Tabelle 1: Auszug aus der verwendeten CORINE Landbedeckungsnomenklatur. Fett geschriebene Klassen reflektieren die Minimalanforderungen und Klassen in normaler Schrift die mögliche Klassentiefe. Kursiv geschriebene Klassen lassen sich voraussichtlich anhand des Datenmaterials ebenfalls klassifizieren.**

<b>Ebene 1</b>	<b>Ebene 2</b>	<b>Ebene 3</b>
<b>100 Bebaute Flächen</b>	<i>110 Städtisch geprägte Flächen</i>	<i>111 Durchgängig städtische Prägung</i>
		<i>112 Nicht durchgängig städtische Prägung</i>
	<i>120 Industrie-, Gewerbe- und Verkehrsflächen</i>	<i>121 Industrie- und Gewerbeflächen, öffentliche Einrichtungen</i>
		<i>122 Straßen-, Eisenbahnnetze und funktionell zugeordnete Flächen</i>
	<i>130 Abbauflächen, Deponien und Baustellen</i>	
	<i>140 Künstlich angelegte, nicht landwirtschaftlich genutzte Grünflächen</i>	<i>141 Städtische Grünflächen</i>
<i>142 Sport- und Freizeitanlagen</i>		
<b>200 Landwirtschaftliche Flächen</b>	210 Ackerflächen	
	220 Dauerkulturen	
	230 Grünland	
	240 Heterogene landwirtschaftliche Flächen	
300 Wälder und naturnahe Flächen	<b>310 Wälder</b>	311 Laubwälder
		312 Nadelwälder
		313 Mischwälder
	<b>320 Strauch- und Krautvegetation</b>	
	<b>330 Offene Flächen ohne / mit geringer Vegetation</b>	<i>332 Felsflächen ohne Vegetation</i>
		<i>333 Flächen mit spärlicher Vegetation</i>
400 Feuchtflächen	<i>410 Feuchtflächen im Landesinneren</i>	
<b>500 Wasserflächen</b>	<i>510 Wasserflächen im Landesinneren</i>	511 Gewässerläufe
		512 Wasserflächen

Für das Training und Testen der Klassifikationsmethoden sind Referenzdaten für jede Klasse notwendig. Es wird davon ausgegangen, dass für die Berechnung der Landbedeckungsklassifikation als Patch-basierte Bildklassifikation mit mittlerer bis hoher Genauigkeit mindestens 5000 Trainingsflächen (100m x 100m) pro Klasse verfügbar sein müssen. Um die Landbedeckungsklassifikation als semantische Seg-

mentierung zu lösen, wird von einer doppelten Menge ausgegangen. Zusätzliche Testdaten werden benötigt, um die Genauigkeit der Landbedeckungsklassifikation anhand von unabhängigen Daten zu testen. Zur Generierung dieser Referenzdaten wurden Flächen auf den KH-Bilddaten ausgewiesen und der jeweiligen Klasse manuell zugeordnet. Die Zuordnung zu einzelnen Klassen erfolgte überwiegend durch visuelle Interpretation der KH-Bilddaten und unter Zuhilfenahme von historischen topographischen Karten und im Vergleich mit aktuellen digitalen Orthophotos. Dabei wurde berücksichtigt, dass die digitalen Orthophotos eine andere Landbedeckung zeigen können als in den historischen KH-Bilddaten. Aktuell besteht der Beispieldatensatz (Trainings- und Testdaten) aus 27 ausgewiesenen Untersuchungsgebieten welche eine Fläche von rund 570 km<sup>2</sup> abdecken. In der nachfolgenden Abbildung 6 wird die Lage alle 27 Studiengebiete zu denen eine semantische Maske erstellt wurde dargestellt.

Abbildung aller Beispielgebiete



**Abbildung 6: Abbildung aller Beispielgebiete. Aufgrund von zu schlechter Bildqualität oder Klassifikationsfehlern wurden die erstellen semantischen Masken für die Beispielgebiete 25, 27 und 28 nicht im Projekt berücksichtigt.**

In der Abbildung sind drei paarweise überlappende Untersuchungsgebiete zu finden, im Speziellen die Beispielgebiete 9 und 19, 3 und 16, 4 und 26. Diese überlappenden Trainingsgebiete wurden erstellt um die Netzwerke hinsichtlich ihrer Übertragbarkeit in der Entwicklungsphase der Architekturen zu testen und nicht ausschließlich für den Genauigkeitstest der Netzwerke. Außerdem wurden diese Überlappungen genutzt um zu überprüfen, ob und wie unterschiedlich die Landbedeckung im Erstellungsprozess von un-

terschiedlichen Personen klassifiziert wird, wenn verschiedene KH-Bildausschnitte vorliegen. Jedes in Abbildung 4 dargestellte Untersuchungsgebiet hat eine quadratische Form mit einer Kantenlänge von 4.608 m (2.304 Pixel). Insgesamt wurden somit 143.327.232 Pixel für den Beispieldatensatz klassifiziert.

Im Endprodukt dieses Projekts entstehen fünf Landbedeckungskarten von Sachsen. Jede Landbedeckungskarte repräsentiert dabei einen Zeitschnitt. Die Auswahl der Untersuchungsgebiete ist so konzipiert, dass für jeden Zeitschnitt eine ähnlich Anzahl von Beispielgebieten vorliegt. Dies ist ein notwendiger Schritt um eine vergleichbare Performance der Netzwerke in jedem Zeitschnitt erwarten zu können. Außerdem müssen die Beispielgebiete eines Zeitschnitts, räumlich gleichmäßig verteilt sein. Hinzukommt außerdem, dass die KH-Bilddaten innerhalb eines Zeitschnittes häufig eine sehr unterschiedliche Erscheinung haben. Auch diese Repräsentationsdiversität muss zum Erreichen einer homogenen Genauigkeit der Interferenz eingehalten werden. Als letztes müssen die Repräsentationen von verschiedenen Landbedeckungen über all diese Anforderungen ebenfalls gleichermaßen im Beispieldatensatz vorkommen.

Im Rahmen des Projektes war es aus wirtschaftlichen Gründen nicht möglich alle Anforderungen an den Beispieldatensatz zu erfüllen. Der Fokus wurde insbesondere auf die Diversität der Bilddaten und eine gleichmäßige räumliche Verteilung gelegt. Die Gleichverteilung der Beispieldaten über alle Zeitschnitte ist dabei in den Hintergrund gerückt. In der folgenden Tabelle wird die Verteilung der Beispielgebiete über die Zeitschnitte aufgeschlüsselt. Die Bezeichnungen der Trainingsgebiete decken sich mit denen der Karte aus Abbildung 6.

**Tabelle 2: Zuordnung der Studiengebiete zu den Zielzeitschnitten**

Zeitschnitt	Zugehörige Beispielgebiete
1965 - 1967	01, 02, 04, 06, 07, 09, 20, 30, 05
1967 - 1972	10, 11, 03, 26
1972 - 1975	13, 14, 15, 16, 17
1975 - 1978	18, 19, 20, 21, 22, 24, 24
> 1978	29, 30

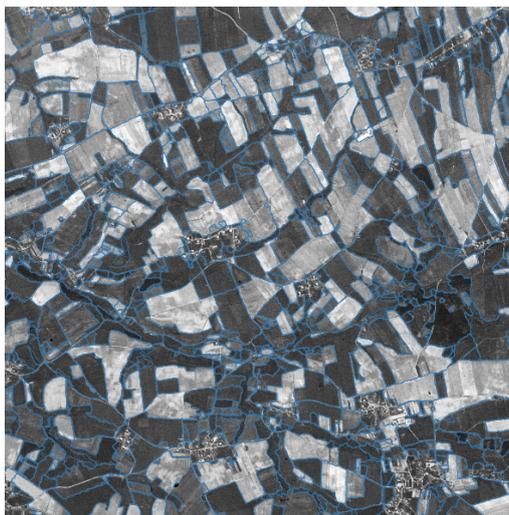
Im Weiteren wird auf die Klassenverteilung der Pixel im Beispieldatensatz eingegangen. Ideal sind Beispieldatensätze mit einer Gleichverteilung von Beispieldaten für jede Klasse. In der Praxis der Landbedeckungsklassifikation ist dies allerdings nicht möglich, da verschiedene Landbedeckungstypen (z. B. landwirtschaftliche Fläche oder Wald) deutlich häufiger vorkommen als beispielsweise Gewässer oder Strauchvegetation. Deswegen enthält der Beispieldatensatz drei überrepräsentierte Klassen (100 Künstliche Flächen, 200 Landwirtschaftliche Flächen und 310 Wald) und drei unterrepräsentierte Klassen (320 Strauchvegetation, 330 offene Flächen und 510 Gewässer). Ein unausgewogener Datensatz stellt für ein Deep Learning Projekt eine besondere Herausforderung dar. Auf die Methoden dieser Problematik entgegenzuwirken, wird in Kapitel 2.5.3 und 2.6 eingegangen.

Eine ursprüngliche Zielstellung des Projektes war die Klassifikation der Klasse 400 Feuchtflächen. Im Projektverlauf zeigte sich, dass für diese Klasse nicht genügend Beispielflächen gefunden oder zuverlässig als solche ausgewiesen werden konnten. Die Klassifikation der Feuchtgebiete war daher nicht möglich.

### 2.3.1 Erstellung von Referenzdaten

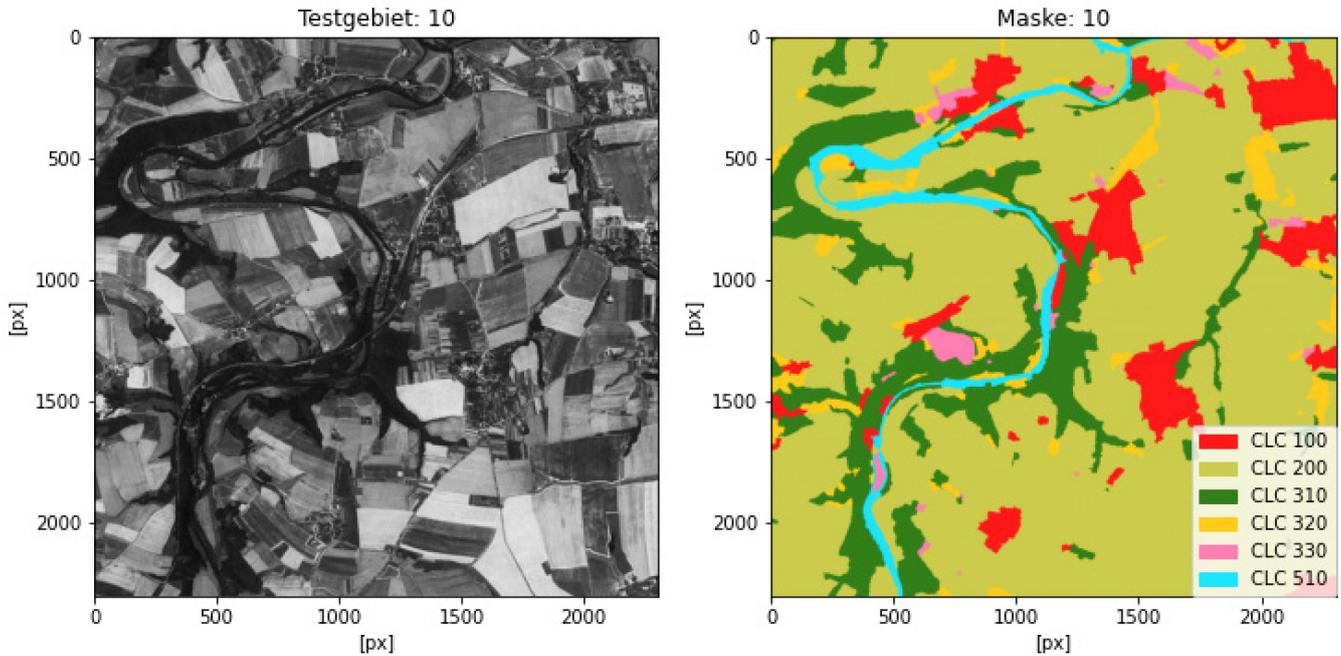
Dieses Kapitel beschreibt den Prozess und die Methoden zur Erstellung der korrespondierenden semantischen Masken zu den Studiengebieten.

Im ersten Schritt zur Erstellung der semantischen Masken werden die Studiengebiete mithilfe eines Quick Shift Algorithmus segmentiert. Für die Anwendung des Quick Shift Algorithmus (Vedaldi & Soatto, 2008) wurde für den ratio-Parameter, welcher die Gewichtung zwischen Farbraumnähe und Bildraumnähe steuert, der Wert 0,5 festgelegt, um beiden Extremen gleich viel Gewicht zu geben. Im Ergebnis der Vorsegmentierung entsteht ein Bild mit etwa 6.000-7.000 Segmenten. Der Anspruch der Vorsegmentierung bestand darin, sensitiv auf kleine Änderungen im Farb- und Bildraum zu reagieren und so kleine Segmente zu erzeugen. Diese kleinen Segmente werden im zweiten Schritt händisch zu homogenen Flächen zusammengeführt. Ein Untersuchungsgebiet besteht nach diesem Schritt aus rund 500 - 700 Segmenten (siehe Abbildung 7).



**Abbildung 7: Beispiel für eine Zerlegung eines Referenzbildes in Segmente (blaue Linien) für die Generierung von Referenzdaten.**

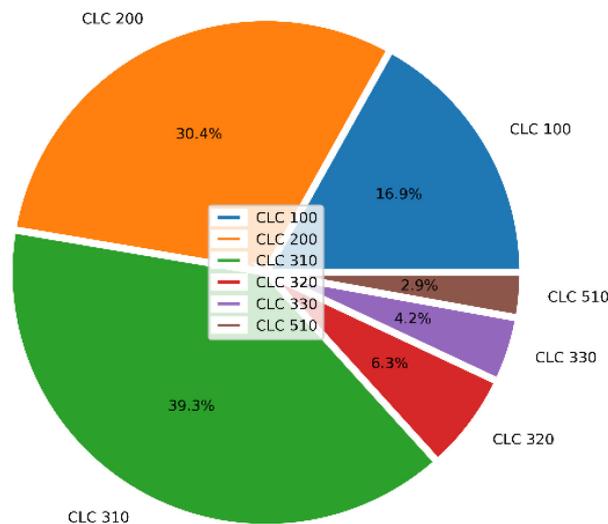
Insbesondere in urbanen Gebieten sind nach der Vorsegmentierung einige unerwünschte Ergebnisse entstanden. An diesen Stellen wurde die Form der Segmente händisch nachbearbeitet. Im letzten Schritt wurde jedem Segment eine Klassenbezeichnung gemäß Tabelle 1 zugeordnet. Am Ende gibt es eine semantische Karte pro Studiengebiet als Geometriedatei, welches in ein Rasterdateiformat konvertiert wird. Dadurch verschwinden die Segmente aus der Vorprozessierung und homogene Flächen gleichen Landbedeckungstyps verschmelzen zu einer Fläche. Beispielhaft wird eine semantische Karte und das korrespondierende Untersuchungsgebiet in Abbildung 8 dargestellt. Alle erstellten semantischen Masken haben wie die KH-Daten selbst eine Pixelgröße von 2 m.



**Abbildung 8: Abbildung des Beispielgebietes 10 und der korrespondierenden semantischen Maske. Die semantische Maske zeigt die sechs Zielklassen. CLC 100 = Künstliche Oberflächen, CLC 200 = Landwirtschaftliche Flächen, CLC 310 = Wald, CLC 320 = Strauchvegetation, CLC 330 = offene Flächen, CLC 510 = Gewässer.**

Die folgende Abbildung 9 zeigt die Verteilung der Landbedeckungsklassen der klassifizierten Pixel über alle Beispielgebiete.

Verteilung der Beispieldaten pro CORINE Landbedeckungsklasse



**Abbildung 9: Relative Verteilung der klassifizierten Beispieldaten in den Studiengengebieten. Deutlich sind drei überrepräsentierte Klassen erkennbar (100 Urbane Flächen, 200 Landwirtschaftliche Flächen, 310 Wald) und vier unterrepräsentierte Klassen (320 Strauchvegetation, 330 offene Flächen, 400 Feuchtflächen, 510 Gewässer)**

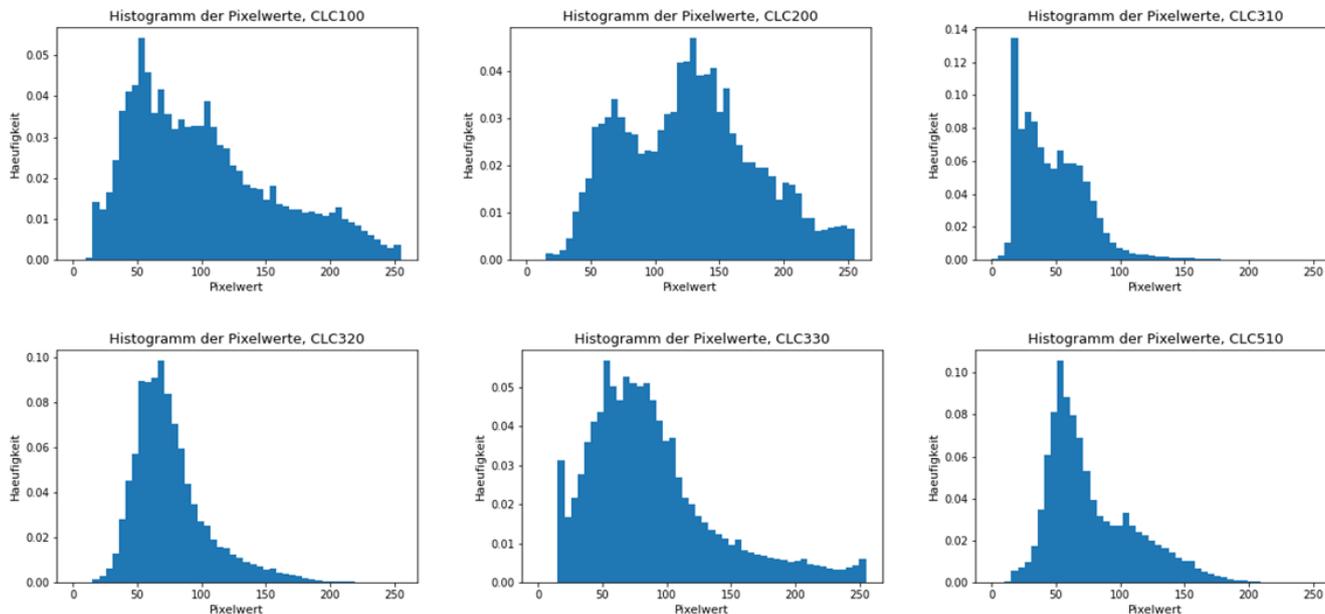
Es ist zu erkennen, dass die Klasse Wald mit über 39% den größten Datenanteil im Beispieldatensatz hat. Ebenfalls mit über 30 % ist die Klasse Landwirtschaftliche Flächen vertreten. Das letzte Drittel der Trainingsdaten bilden die Klassen Künstliche Oberflächen mit 16,9%, Strauchvegetation mit 6,7 % und Offene Flächen mit 4,2 %. Den geringsten Datenanteil hat die Klasse Gewässer mit 2,9 %.

### 2.3.2 Generierung von Bildmerkmalen

Aus den KH-Bilddaten wurden verschiedene Merkmale abgeleitet, die genutzt wurden um die Landbedeckung zu klassifizieren. Dazu wurden zwei verschiedene Merkmalstypen verwendet:

1. die originalen panchromatischen Bildinformationen (1 Band)
2. Texturmaße aus der Graustufen-Ko-Auftritts-Matrix (GLCM) mit verschiedenen Filtergrößen (40 Bänder)

Die originalen Pixelwerte repräsentieren die echten radiometrischen Verhältnisse der Aufnahme. Um dem Problem der explodierenden Gradienten beim Training neuronaler Netze entgegenzuwirken wurden die originalen Pixelwerte mit einem Wertebereich von  $W = [0, 255]$  auf einen Wertebereich mit  $W = [0, 1]$  normalisiert, indem die Pixelwerte durch 255 dividiert werden. In der nachfolgenden Abbildung 10 wird die Verteilung aller originalen panchromatischen Bildinformationen in den Studiengebieten über alle Klassen in sechs Histogrammen abgebildet. Es ist zu erkennen, dass alle sechs Zielklassen eine unterschiedliche Verteilung beschreiben. Aufgrund dessen lassen sich die Zielklassen anhand ihrer panchromatischen Bildinformationen voneinander trennen und damit klassifizieren.



**Abbildung 10: Verteilung der panchromatischen Bildinformationen über alle Klassen. Die Histogramme zeigen die Verteilung der panchromatischen Bildinformationen von allen Studiengebieten pro Klasse. Es gilt: CLC 100 = Künstliche Oberflächen, CLC 200 = Landwirtschaftliche Flächen, CLC 310 = Wald, CLC 320 = Strauchvegetation, CLC 330 = offene Flächen, CLC 510 = Gewässer.**

Die zweite Gruppe von Bildmerkmalen sind die Texturmaße. Die Berechnung von Texturmaßen ist ein statistisches Verfahren zur Untersuchung der räumlichen Beziehung von Pixeln. Der verwendete Ansatz wird als Graustufen-Ko-Auftrittsmatrix (GLCM) bezeichnet (Hall-Beyer, 2017). Die GLCM-Maße charakterisieren die Textur eines Bildes, indem sie berechnen, wie oft Pixelpaare mit bestimmten Werten in einer

bestimmten räumlichen Beziehung in einem Bild vorkommen. Die Texturmaße können keine Informationen über die Form eines Bildobjektes liefern (Haralick et al., 1973). In Tabelle 3 werden die berechneten Texturmaße beschrieben. Durch die Anwendung verschieden großer Filter bei der Berechnung der GLCM-Maße kann die Textur der KH-Bilddaten durch die Architekturen in räumlich unterschiedlich großen Kontexten verstanden werden. Deshalb werden bei der Erstellung der GLCM-Eingabedaten vier verschieden große Filter (5 x 5, 7 x 7, 9 x 9, 11 x 11) angewandt. Durch jeden Filter ergeben sich zehn neue Bänder für die Eingabedaten.

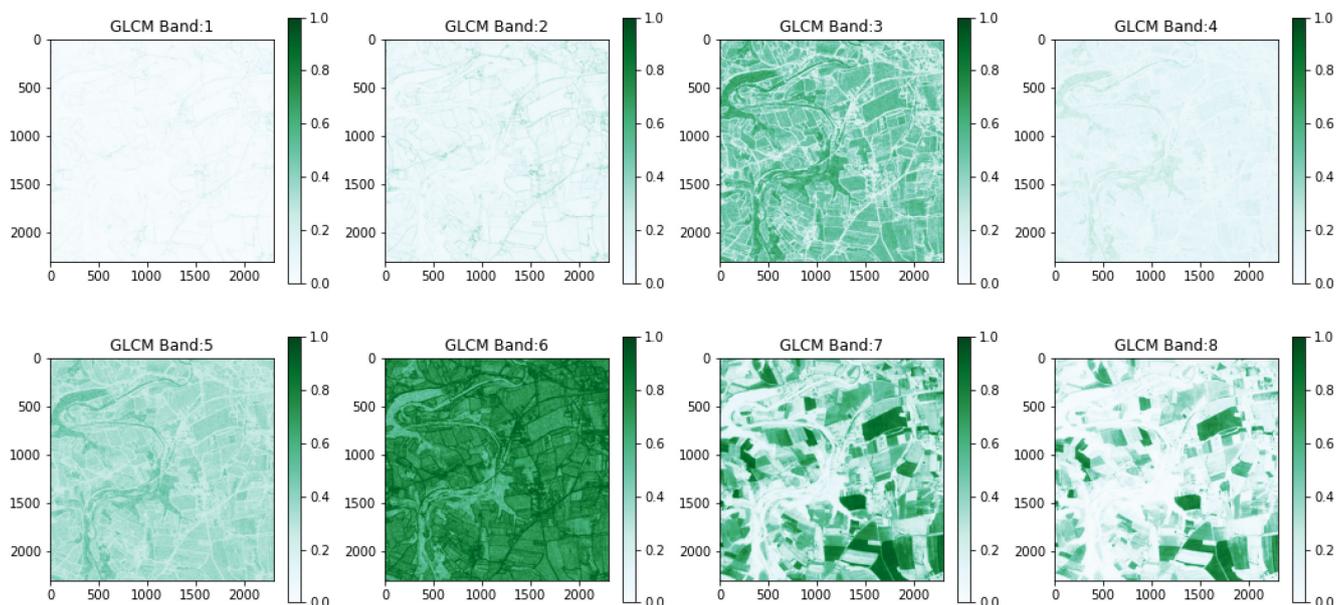
In Tabelle 3 werden die Berechnungsvorschriften zu den GLCM-Texturmaßen aufgezeigt. Automatisiert wurde die Berechnung der GLCM-Texturmaße mithilfe der API des ESA Programms SNAP.

**Tabelle 3: Berechnung der GLCM-Texturmaße (Haralick et al., 1973).** *i* beschreibt die Zeilen, *j* die Spalten,  $P_{i,j}$  beschreibt den Wahrscheinlichkeitswert einer Zelle, *N* ist die Anzahl der Spalten und Zeilen.

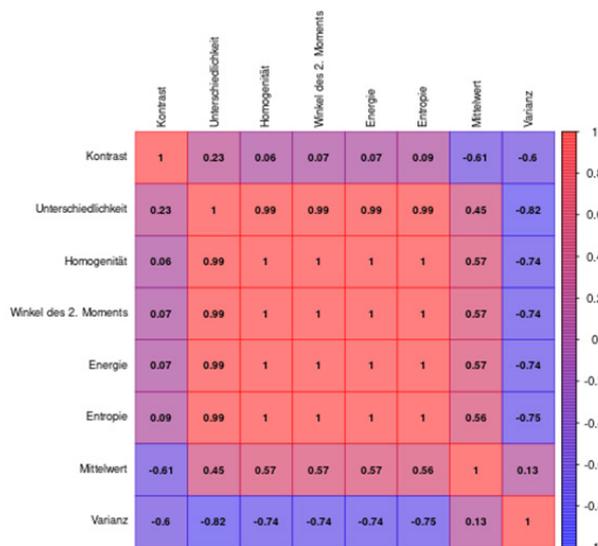
Band	Texturmaß	Berechnung
1	Kontrast	$Con = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
2	Unähnlichkeit	$Dis = \sum_{i,j=0}^{N-1}  P_{i,j} ^2$
3	Homogenität	$Hom = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
4	Winkel des zweiten Momentes	$ASM = \sum_{i,j=0}^{N-1} P_{i,j}^2$
5	Energie	$E = \sqrt{ASM}$
6	Maximalwert	$\max = \max(P_{i,j})$
7	Entropie	$Ent = \sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$
8	Mittlerer GLCM-Wert	$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})$
9	Varianz der Graustufenmatrix	$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2$
10	Korrelation der Graustufenmatrix	$Cor = \sum_{i,j=0}^{N-1} P_{i,j} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$

Voruntersuchungen zeigten das die Bänder 6 (Maximalwert) und 10 (Korrelation der Graustufenmatrix) einen deutlichen Performancenachteil bei der Klassifikation der KH-Daten aufweist weshalb diese in der weiteren Prozessierung nicht verwendet werden. Insgesamt entstehen für die Texturmaße somit noch 32

Bänder in den Eingabedaten. Analog zu dem Band der originalen Pixelwerte werden die Bänder der Texturmaße in einen Wertebereich zwischen 0 und 1 normalisiert. Durch die Verwendung der anderen Bänder konnte die beste Performance beobachtet werden. Trotz der Tatsache, dass die GLCM-Texturmaße untereinander teilweise sehr stark korreliert sind entstehen durch Vernachlässigung anderer Bänder immer kleine Performancenachteile, insbesondere in den unterrepräsentierten Klassen. In Abbildung 11 werden die GLCM-Texturmaße für das Studiengebiet 10 dargestellt. Die Korrelationen der Bänder untereinander werden in Abbildung 12 dargestellt.



**Abbildung 11: Darstellung der genutzten GLCM-Texturmaße für das Studiengebiet 10. Für die Abbildung gilt: GLCM Band 1 = Kontrast, GLCM Band 2 = Unähnlichkeit, GLCM Band 3 = Homogenität, GLCM Band 4 = Winkel des 2. Moments, GLCM Band 5 = Energie, GLCM Band 6 = Entropie, GLCM Band 7 = Mittelwert, GLCM Band 8 = Varianz.**



**Abbildung 12: Korrelation der GLCM-Texturmaße für Studiengebiet 10**

Aus Abbildung 12 ist eine starke Korrelation für die Bänder Unähnlichkeit, Homogenität, Winkel des 2. Moments, Energie und Entropie zu erkennen. Zum Mittelwert und zur Varianz zeigt der Kontrast eine mittlere negative Korrelation auf. Der Kontrast zeigt insgesamt die geringste Korrelation zu alle anderen GLCM-Texturmaßen. Der Mittelwert zeigt überwiegend eine mittlere bis niedrige Korrelation zu allen anderen Bändern auf. Für die Varianz wird überwiegend eine mittlere Korrelation zu den meisten GLCM-Texturmaßen verzeichnet. Eine geringe Korrelation hat der Mittelwert zur Varianz.

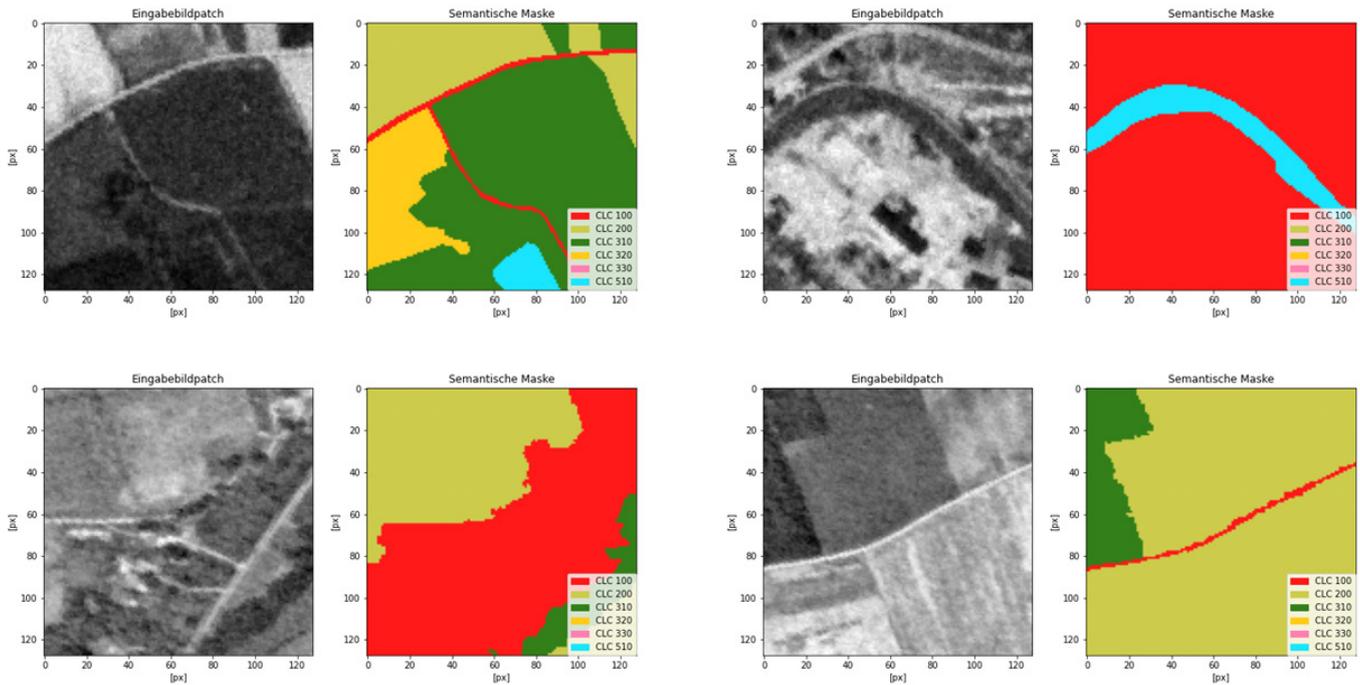
## **2.4 Prozessierung der Eingabedaten**

In diesem Unterkapitel wird auf die Vorprozessierung der Eingabedaten für die unterschiedlichen Netzwerkansätze eingegangen. Für die Beschreibung der Prozessierungskette wurden zwei Stränge aufgebaut. Der erste Strang bezieht sich auf die Klassifikation der Landbedeckung mit einer klassischen Patch-basierte Bildklassifikation. Dabei wird jedem quadratischen Bildpatch ein Klassenlabel zugewiesen. Der zweite Strang beschreibt die Prozessierung der Eingabedaten zur Klassifikation der Landbedeckung mit einer semantischen Segmentierung. Bei der semantischen Segmentierung wird jedem Pixel innerhalb eines quadratischen Bildpatches ein Klassenlabel zugewiesen. Mit der Entscheidung nach welchem Aufgabentyp die Landbedeckungsklassifikation durchgeführt wird entscheidet sich auch die Auflösung der Landbedeckungsklassifikation.

Bei der klassischen Bildklassifikation ergibt sich die Auflösung der Landbedeckung aus der Kantenlänge des gegebenen Bildpatches. Für die Klassifikation der Landbedeckung als semantische Segmentierung ist die Auflösung der Landbedeckung gleich der Kantenlänge eines Pixels (2 m).

### **2.4.1 Semantische Segmentierung**

Die Studiengebiete sind alle quadratisch und haben eine Kantenlänge von 4.608 m bzw. 2.304 Pixel. Für die semantische Segmentierung werden die Studiengebiete in kleinere quadratische Patches mit einer Kantenlänge von 128 Pixel zerlegt. Jeder Bildpatch hat damit eine Kantenlänge von 256 m. Die Patches haben bei der Zerlegung keine Überlappung. Entnimmt man die sechs Studiengebiete mit Überlappung aus dem Beispieldatensatz bleiben 21 Studiengebiete für den Beispieldatensatz. Aus diesen 21 Studiengebieten ergeben sich 6.804 Beispielpatches mit einer Kantenlänge von 128 Pixeln als Trainingsdaten, Validierungsdaten und Testdaten für die Netzwerke. Jeder Bildpatch hat 33 Eingabebänder (originaler Grauwert und 4 mal 8 GLCM-Texturmaße). Zu jedem Bildpatch existiert eine korrespondierende semantische Maske welche die Klassenlabel für jedes Pixel beinhaltet. Beispielhaft wird das Grauwertbild eines Eingabebildes und die dazugehörige korrespondierende semantische Maske in Abbildung 13 dargestellt.

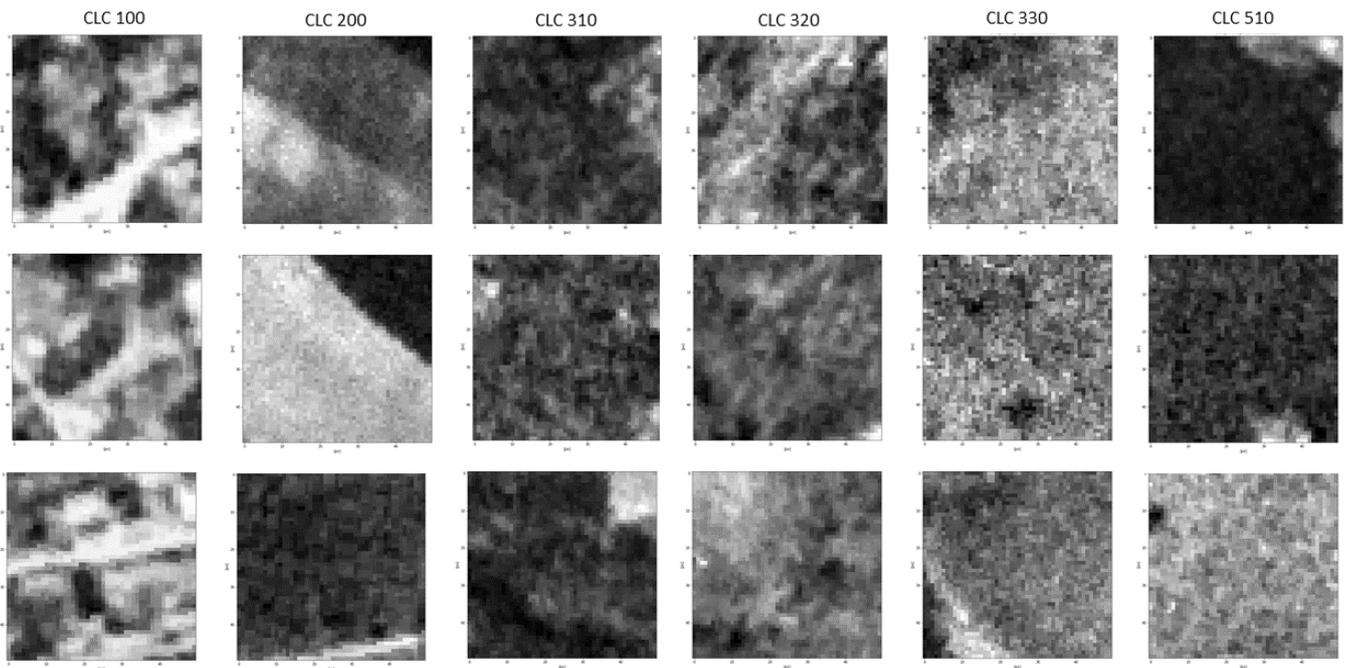


**Abbildung 13: Beispiele für Eingabebildpatches (nur originale panchromatische Bildinformation) und der korrespondierenden semantischen Maske zur Lösung der Landbedeckungsklassifikation als semantische Segmentierung. Es gilt: CLC 100 = Künstliche Oberflächen, CLC 200 = Landwirtschaftliche Flächen, CLC 310 = Wald, CLC 320 = Strauchvegetation, CLC 330 = offene Fläche, CLC 510 = Gewässer.**

## 2.4.2 Patch-basierte Bildklassifikation

Die Patch-basierte Klassifikation wurde anhand der gewünschten Ausgabeauflösung von 100 m entwickelt. Im ersten Schritt wurden die Studiengebiete im ersten Schritt auf eine Größe von 2.300 Pixeln Kantenlänge zugeschnitten. Dazu wurden an jeder Kante des Studienggebietes jeweils zwei Pixel entfernt. Anschließend wurden die Studiengebiete in nicht überlappende quadratische Patches mit einer Kantenlänge von 50 Pixel (100 m) zerlegt.

Aus jedem Studiengebiet ergeben sich 2.116 Patches mit korrespondierenden Masken. Insgesamt entstehen so aus den 21 Studiengebieten, für das Training und den Genauigkeitstest der Netzwerke, 44.436 Bildpatches. Die Eingabebilder haben analog zur semantischen Segmentierung ebenfalls 33 Bildkanäle. Das Klassenlabel eines Patches wird aus der dazugehörigen semantischen Maske abgeleitet. Das Label des Bildpatches ist gleich der Klasse mit dem häufigsten Klassenlabel in der semantischen Maske. Die semantische Maske bildet bei der Patch-basierten Klassifikation ein Vektor mit einer Länge von 44.436. Jeder Vektoreintrag korrespondiert mit einem Bildpatch und repräsentiert das Klassenlabel des Bildpatches. In Abbildung 14 werden für jede Klasse drei Beispielpatches abgebildet.



**Abbildung 14: Beispiele von Eingabebildpatches zur Patch-basierten Bildklassifikation. Jede Spalte zeigt je drei Repräsentationen für alle sechs Zielklassen. Von links nach rechts: CLC 100 = Künstliche Oberflächen, CLC 200 = Landwirtschaftliche Flächen, CLC 310 = Wald, CLC 320 = Strauchvegetation, CLC 330 = offene Fläche, CLC 510 = Gewässer.**

Für die Anwendung einer Patch-basierte Klassifikationsmethode ist die Verwendung eines gleichverteilten Trainingsdatensatzes notwendig. Dazu wurden für jede Zielklasse gleich viele Beispieleingaben pro Klasse zum Training genutzt. Umgesetzt ist dies indem zufällig 100 Beispielpatches mit den korrespondierenden Labels gezogen wurden.

### 2.4.3 Spaltung der Beispieldaten: Trainingsdaten, Testdaten, Validierungsdaten

Für das Training neuronaler Netze ist eine Aufspaltung der Beispieldaten in drei unabhängige nicht überlappende Subdatensätze notwendig. Diese drei Datensätze werden als Trainingsdaten, Testdaten und Validierungsdaten bezeichnet. Der Beispieldatensatz, aus den 21 Studiengebieten, wurde so aufgespalten, dass 70% der Beispieldaten pro Klasse in die Trainingsdaten fließen, 10% in die Validierungsdaten und 20% in den Testdatensatz. Um sowohl die Konsistenz der Datenspaltung als auch des Trainings sicherzustellen, wurde die Aufspaltung des Beispieldatensatzes mit einer 10-fachen zufälligen Kreuzvalidierung durchgeführt. Es entstehen somit zehn Trainingsdatensätze, Validierungsdatensätze und Testdatensätze. Umgesetzt wurde die Datenaufspaltung mithilfe des Pythonpaketes scikit-learn<sup>4</sup>.

### 2.4.4 Datenerweiterung

Unter Datenerweiterung versteht man in einem Deep Learning-Projekt die synthetische Erstellung von Trainingsdaten. Dabei werden grundlegend drei verschiedene methodische Ansätze unterschieden. Zum einen in geometrische, spektrale Methoden und generative Methoden mit generativen gegnerischen neuronalen Netzwerken. Aufgrund der begrenzten Projektlaufzeit und der verfügbaren wirtschaftlichen Ressourcen war es nicht möglich einen für Deep Learning-Methoden genügend großen Trainingsdatensatz durch manuelle Arbeit aufzubauen. Standarddatensätze wie zum Beispiel der BigEarth-Datensatz zur

<sup>4</sup> <https://scikit-learn.org/stable/>

Landbedeckungsklassifikation mit Sentinel-2 Szenen besteht aus rund 590.000 multispektralen Bildpatches. Der in diesem Projekt aufgebaute Datensatz bestehend aus etwa 44.000 Bildpatches. Allein aus dieser Differenz ergibt sich die Notwendigkeit Methoden der Datenerweiterung anzuwenden. Diese Notwendigkeit wird verstärkt, weil die KH-Bilddaten mit unterschiedlichen Sensoren aufgenommen wurden und weil die einzelnen Szenen enorme Qualitätsunterschiede aufweisen.

Es gilt im Folgenden wieder zwischen Patch-basierter Bildklassifikation und semantischer Segmentierung der Landbedeckung zu unterscheiden.

Für die Landbedeckungsklassifikation als semantische Segmentierung wurden drei verschiedene geometrische (1. - 3.) Datenerweiterungsmethoden und eine spektrale Datenerweiterungsmethode (4.) angewandt.

- 1) Rotation: dabei wird das Eingabebild und die semantische Maske zufällig mit einem Winkel zwischen  $0^\circ$  und  $20^\circ$  gedreht.
- 2) Spiegelung des Eingabebildes und der semantischen Maske
- 3) Zoom: dabei wird zufällig zwischen 0% und 10% in das Eingabebild und in die semantische Maske gezoomt
- 4) Zufällige Kontrasterhöhung oder Senkung des Studiengebietes um  $\pm 5\%$  (Berechnung der GLCM-Texturmaße nach der Kontrasteditierung).

Alle vier Methoden wurden sowohl auf den Trainingsdatensatz als auch den Validierungsdatensatz angewandt. Umgesetzt und implementiert wurde die Datenerweiterung mithilfe der Pythonpakete Scikit-image<sup>5</sup>, Keras<sup>6</sup> und Tensorflow<sup>7</sup>.

Für das Training der Netzwerke zur Landbedeckungsklassifikation als Patch-basierte Bildklassifikation wurden ebenfalls drei geometrische Datenerweiterungsmethoden und eine spektrale Datenerweiterungsmethode angewandt.

- 1) Rotation: dabei wird das Eingabebild und die semantische Maske zufällig mit einem Winkel zwischen  $0^\circ$  und  $20^\circ$  gedreht.
- 2) Spiegelung des Eingabebildes und der semantischen Maske
- 3) Zoom: dabei wird zufällig zwischen 0% und 10% in das Eingabebild und in die semantische Maske gezoomt
- 4) MixUp-Methode nach Zhang et al., 2018 mit  $\alpha = 0.2$ . Mit MixUp werden Bildmerkmale von Eingabepatches mit unterschiedlichen Klassenlabel systematisch vertauscht.

---

<sup>5</sup> <https://scikit-learn.org/stable/>

<sup>6</sup> <https://keras.io/>

<sup>7</sup> <https://www.tensorflow.org/>

## 2.5 Klassifikation mit neuronalen Netzen

Es wurden verschiedene Deep Learning-Methoden hinsichtlich ihrer Eignung für die Klassifikation historischer KH-Bilddaten evaluiert. Zum einen verstehen wir die Klassifikation der Landbedeckung als semantische Segmentierung bei der jedem Pixel eine Klasse zugewiesen wird. Beim zweiten Ansatz wird die Landbedeckungsklassifikation als Patch-basierte Bildklassifikation verstanden, bei der kleine Mosaikbilder (Bildpatches) der KH-Daten klassifiziert werden. Die Größe eines Bildpatches ist damit das Maß für die räumliche Auflösung der Landbedeckungskarte. Zur Lösung dieser zwei methodischen Ansätze der Landbedeckung kommen allgemein die Faltungsnetzwerke (Convolutional Neuronal Networks) und Transformer in Frage.

Für die semantische Segmentierung (siehe Kap. 2.5.1) wird eine spezielle Weiterentwicklung der CNNs verwendet, insbesondere die sogenannten vollständigen Faltungsnetzwerke (Long, Shelhamer, & Darrell, 2015). Dabei sind insbesondere unterschiedliche U-förmige Architekturen ein leistungsstarker Ansatz für die Herausforderungen der semantischen Segmentierung (Ulmas & Liiv, 2022; John & Zhang, 2022; Gu, et al., 2022). Ihre Stärke liegt insbesondere darin gute Segmentierungsergebnisse mit einer geringen Anzahl von Beispieldaten zu erreichen (Ronneberger, Fischer, & Brox, 2015).

Die im Rahmen des Projektes implementierten U-förmigen Netzwerkarchitekturen können in drei Gruppen unterteilt werden. Die erste Gruppe sind U-Nets (Ronneberger et al., 2015) mit einer unterschiedlichen Ansatz von Faltungsebenen und Konfigurationen der Faltungsschichten. Ziel dieses Ansatzes ist es herauszufinden, welche Modelkomplexität notwendig ist, um die Landschaftsbedeckung der CORONA-Daten zu lernen und mit guten Genauigkeiten zu klassifizieren. Bei der zweiten Gruppe handelt es sich um U-Nets mit verschiedenen populären Netzwerkarchitekturen zur Merkmalsidentifikation im Kontraktionspfad. Dazu wurde die ResNet50-Architektur (Kaiming, Zhang, Ren, & Sun, 2015), die ResNet101-Architektur (Kaiming, Zhang, Ren, & Sun, 2015) und die InceptionV3-Architektur (Szegedy, Vanhoucke, Ioffe, & Shlens, 2015) bzw. die InceptionResNetV2-Architektur (Szegedy, Ioffe, & Vanhoucke, 2016) zur Identifikation von Merkmalen verwendet. Die dritte Gruppe bilden spezielle Weiterentwicklungen der U-förmigen Netzwerkarchitekturen. Speziell wurde das U-Net2+ (Zhou, Rahman, Siddiquee, & T, 2018), das Attention U-Net (Oktay, Schlemper, & Folgoc, 2018), das U-Net3+ (Huang, Lin, Tong, & Hu, 2020), das Transformer U-Net (Chen, et al., 2021) und das SwinU-Net (Cao, et al., 2021) implementiert und zur semantischen Segmentierung getestet.

Für den zweiten Ansatz der Patch-basierten Bildklassifikation (vgl. Kapitel 2.5.2) werden verschiedene Architekturansätze implementiert. Zum einen wird experimentell ein eigenes CNN aufgebaut um insbesondere ein Verständnis dafür zu entwickeln welche Hyperparameter und Architektureigenschaften in der Lage sind die KH-Bilddaten besser oder schlechter zu modellieren. Außerdem wurden populäre Architekturen wie das ResNet50, ResNet101, ResNet152, InceptionV3 und InceptionResNetV2 zur patch-basierten Klassifikation der Landschaftsbedeckung zu Patch-basierter Landbedeckungsklassifikation getestet. Ausgehend von den Erkenntnissen des selbst aufgebauten CNNs werden die genannten populären Architekturen teilweise angepasst. Um dem Problem der geringen Verfügbarkeit von gelabelten Beispieldaten entgegenzuwirken, wurde außerdem ein halb überwacht generatives Gegenübermodell - ein sogenanntes sGAN (Chavdarova & Fleuret, 2018) verwendet.

In den folgenden Subkapiteln werden die Kerneigenschaften der angewandten Architekturen knapp aufgearbeitet. Für weiterführende Informationen wurde auf die Grundlagenliteratur verwiesen. Die Subkapitel 2.5.1 bis 2.5.3 beschreiben die relevanten Architekturen für die semantische Segmentierung. In den Subkapiteln 2.5.4 bis 2.5.6 werden die angewandten Architekturen zur Patch-basierenden Landbedeckungsklassifikation vorgestellt.

### 2.5.1 U-Net

Das U-Netz ist ein tiefes Neuronales Netz, welches ursprünglich zur semantischen Segmentierung von biomedizinischen Bildern entwickelt wurde (Ronneberger et al., 2015). Der Ansatz der Architektur beruht auf dem SegNet und kann den vollständigen Faltungsnetzwerken zugeordnet werden. Es gibt mehrere Gründe eine semantische Segmentierung mit einem U-Netz durchzuführen. Zum einen ist die Architektur nicht komplex aufgebaut, was das Netzwerk leicht implementierbar macht und auch schnell. Zum anderen können U-Netze mit einer geringen Anzahl Beispieldaten bei einer guten Performance trainiert werden (Torralba et al., 2005).

Der Name U-Net resultiert aus dem symmetrischen Aufbau der Architektur, die einer U-Form ähnlich ist. Die Architektur kann in drei Bestandteile zerlegt werden. Der linke Arm des U-Netzes wird als Kontraktionspfad bezeichnet. Der Kontraktionspfad (englisch: Encoder) ist eine alternierende Abfolge aus Faltungsschichten und Pooling-Schichten welche Merkmale aus der Eingabe extrahieren. Diese Funktionsweise ist analog zu klassischen Patch-basierten CNN-Ansätzen. Der Engpass bildet die Verbindung zwischen den zwei Armen. Der rechte Arm des U-Netz wird als Expansionspfad (englisch: Decoder) bezeichnet und erstellt basierend auf den extrahierten Merkmalen eine Segmentierungskarte durch eine abwechselnde Anwendung von transponierten Faltungsschichten, Faltungsschichten und Pooling-Schichten. Im Vergleich zum SegNet und anderen Encoder-Decoder-Architekturen werden in der U-Netz Architektur Sprungverbindungen (englisch: skip-connections) zwischen den beiden Pfaden verwendet, um Standortinformationen aus dem Kontraktionspfad mit den Kontextinformationen des Expansionspfads zu kombinieren. Aufgrund der Kopplung der beiden Pfade ist eine detailliertere und genauere Segmentierung möglich. Im Vergleich zu klassischen CNN-Ansätzen hat das U-Netz keine dichte Schicht, womit die Architektur nur die Gewichte des Kernels lernen muss. Daraus resultiert, dass das U-Netz unterschiedliche Bildgrößen verarbeiten kann.

Das in dieser Arbeit verwendete U-Net wurde von Iglovikov (Iglovikov et al., 2017) für die Klassifikation der Landschaftsbedeckung vorgestellt. Für die Anwendung in dieser Arbeit wird das Netzwerk noch in zwei Punkten verändert. Zum einen wurde die Größe der Filter auf allen Faltungsebenen zu 5x5 geändert, als auch die Anzahl der Faltungsschichten von fünf auf sechs vergrößert.

### 2.5.2 U-Nets mit verschiedenen Kontraktionspfaden

Der zweite Ansatz ist die Anwendung verschiedener bewährter Netzwerkarchitekturen zur Merkmalsidentifikation im Kontraktionspfad. In der vorliegenden Studie wird wie eingangs in 2.6 beschrieben das ResNet50, das ResNet101, das InceptionV3-Net und das InceptionResNetV2-Netzwerk als Kontraktionspfad in die U-förmige Netzwerkarchitektur eingesetzt. Alle genannten Architekturen wurden bereits in klassischen CNN-Ansätzen und als Rückgrate (englisch: Backbone) in Encoder-Decoder Architekturen erfolgreich zur Landbedeckungsklassifikation und vielen anderen Klassifikationsaufgaben und Segmentierungsaufgaben eingesetzt. Nach eingängiger Recherche konnte keine Studie gefunden werden, welche die Fähigkeit dieser prominenten Architekturen zur Landbedeckungsklassifikation mit CORONA-Daten untersucht.

In den folgenden Subkapiteln werden die grundlegenden Ideen und Methoden der Architekturen knapp beschrieben

#### 2.5.2.1 ResNet

Das ResNet wurde 2015 von Microsoft Research vorgeschlagen. Spätestens seit 2014 wurden immer tiefere Neuronale Netze entwickelt, gleichzeitig wurde die Überanpassung von Netzwerken ein immer relevanteres Thema. Die Überanpassung wird durch das Problem der verschwindenden Gradienten beschrieben. Infolge der Multiplikation der Gradienten durch mehrere Schichten, kann der Gradient durch eine zunehmende Anzahl von Schichten unendlich klein werden. Daraus resultiert eine Sättigung des

Netzwerks und eine Verschlechterung der Trainingsleistung. Ziel der ResNet-Entwicklung ist die Adressierung des Problems der verschwindenden Gradienten. Benannt werden die unterschiedlichen ResNet-Entwicklungen nach der Anzahl ihrer Schichten. Das ResNet50 besteht somit aus 50 Schichten. Am Anfang steht eine Faltung mit einer Schrittweite mit  $s = 2$  gefolgt von einer MaxPooling-Funktion, was zu einem starken Reduzieren der Eingabe führt. Anschließend folgt eine Einheit aus zwei Blöcken. Die zweite Einheit besteht ebenfalls aus zwei Blöcken, wobei die erste Faltungsoperation eine Schrittweite mit  $s = 2$  hat, was zu einer Verkleinerung der Merkmalskarten führt. Infolge der unterschiedlichen Anzahl von Filtern pro Faltung und der Reduzierung der Größe durch MaxPooling unterscheidet sich die Größe von Eingang und Ausgang des ersten Blocks der zweiten Einheit. Mithilfe einer  $1 \times 1$ -Faltung wird die Größe der Eingabe angepasst.

### **2.5.2.2 Inception**

Während andere berühmte Architekturen, wie VGG16 (Simonyan & Zisserman, 2015) auf eine Erhöhung der Faltungsschichten und Anzahl der Filter setzen um die Performance der Netzwerke zu verbessern, wird durch den Vorschlag der Inception-Netzwerke ein anderer Ansatz verfolgt. Die verschiedenen Merkmale der Landschaftsoberfläche können in ihrer Größe und Position stark variieren, aufgrund dieser Variation ist die Entscheidung der richtigen Filterfenstergröße nicht einfach und nicht eindeutig. Mit größeren Filtermatrizen können globale Merkmale und mit kleineren Filtermatrizen lokale Merkmale identifiziert werden. Im VGG-Netzwerk werden viele Faltungsoperationen angewandt, um Merkmale unterschiedlicher Größe zu identifizieren, dieser Ansatz führt zu einer enormen signifikanten Erhöhung der notwendigen Rechenleistung und begünstigt das Problem der verschwindenden Gradienten. Die Idee der Inception-Architektur besteht in der Einführung einer parallelen Verarbeitung von unterschiedlich großen Filtermatrizen auf einer Faltungsebene. Dies führt dazu, dass das Netzwerk etwas breiter statt durch viele Faltungsschichten tiefer wird.

Im ResNet50, ResNet101 und in der InceptionResNetV2-Architektur sind sogenannte Restblöcke implementiert welche für einen Informationsfluss über mehrere Faltungsschichten hinweg im Netzwerk sorgt. Aufgrund dieses effizienteren Informationsflusses ist das Netzwerk in der Lage dem Problem der verschwindenden Gradienten entgegenzuwirken welches beim Training von insbesondere tiefen neuronalen Netzen ( $>10$  Faltungsschichten) auftritt. In unseren Vorarbeiten wurden U-Nets mit einer unterschiedlichen Anzahl von Faltungsschichten getestet. Im Ergebnis wurde festgestellt, dass eine Erhöhung der Anzahl der Schichten zu besseren Ergebnissen der semantischen Segmentierung führt. Aufbauend auf diesen Erkenntnissen wurden im Rahmen des Projektes noch tiefere Architekturen mit Restblöcken mit 50 bzw. 101 Schichten verwendet. Aufgrund der starken Diversität von Landbedeckungsmerkmalen sowohl in Form und Größe wurde die InceptionV3 Architektur verwendet, welche in der Lage ist, Merkmale unterschiedlicher Größe auf einer Faltungsebene zu erfassen. Die InceptionV3-Architektur ist damit in der Lage die Merkmale der Landbedeckung effizienter zu identifizieren. Ausgehend von den Erfolgen, welche die ResNet-Architektur gezeigt hat, wurde das InceptionResNetV2 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) vorgeschlagen. Diese Architektur ist eine hybride Lösung welche das Inception Modul mit der Idee der Restblöcke vereint um die Vorteile aus beiden Methoden auszuschöpfen.

## **2.5.3 Spezielle Weiterentwicklungen U-förmiger Architekturen**

### **2.5.3.1 U-Net2+**

Das U-Net2+ kombiniert Ideen des DensNet (Huang, Liu, van der Maaten, & Weinberger, 2017) mit dem klassischen U-Net nach Ronneberger et al. (2015). Die Hauptidee des U-Net2+ ist es, die semantische Lücke zwischen den zwei Pfaden des U-Nets zu schließen. Dazu wurden sowohl verschachtelte (englisch:

nested) und dichte (englisch: dense) Sprungverbindungen mit Faltungsschichten implementiert. Mit anderen Worten, der dichte Faltungsblock bringt die Merkmalskarten des Kontraktionspfads näher an die der im Expansionspfad erwarteten Merkmalskarten. Die Einführung der dichten Faltungsblöcke an den Sprungverbindungen führt dazu, dass mehr Verlustbedingungen im Netzwerk eingeführt werden, was wiederum dazu führt, dass feinkörnigere Strukturen vom U-Net2+ besser erfasst und lokalisiert werden können. Diese Änderungen zur originalen U-Net-Architektur ziehen allerdings eine deutliche Erhöhung der Modelparameter nach sich zum Nachteil von Trainingszeiten und Interferenzzeiten.

### **2.5.3.2 U-Net3+**

Das U-Net3+ stellt eine Weiterentwicklung U-Net++ dar und adressiert unter anderen die Verbesserung der Kantenlokalisierung von Objekten. Im Unterschied zum klassischen U-Net und zum U-Net++ welche nicht Merkmalskarten im Expansionspfad von mehreren Maßstabsebenen und Verarbeitungsschichten erhält werden Informationen im Expansionspfad des U-Net3+ von allen Ebenen der Architektur verarbeitet. Jede Expansionspfad in U-Net3+ enthält Merkmalskarten im kleineren und gleichen Maßstab vom Kontraktionspfad und Merkmalskarten im größeren Maßstab vom Kontraktionspfad, die feinkörnige Details und grobkörnige Grenzsemantik im vollständigen Maßstab erfassen. Wie beim U-Net2+ wird auf jeder Ebene des Kontraktionspfads eine Vorhersage ausgegeben welche weitere Verlustbedingungen produziert und damit eine tiefere und robustere Überwachung des Trainings zulässt. Zusätzlich wird im U-Net3+ ein Modul (englisch: classification guided modul) implementiert welches eine binäre Klassifikationsaufgabe einführt und vorhersagt ob eine bestimmte Zielklasse überhaupt im Eingabebild zu finden ist. Mit diesem Modul wird der Übersegmentierung eines Eingabebildes entgegen gewirkt.

### **2.5.3.3 Attention U-Net**

Um die räumlichen Informationen aus dem Kontraktionspfad mit dem Expansionspfad zu kombinieren werden im klassischen U-Net Sprungverbindungen eingesetzt. Der Einsatz von Sprungverbinden führt jedoch zu vielen redundanten Merkmalskarten auf den niedrigen Ebenen des U-Nets, da die Merkmale in den Anfangsschichten nicht spezifisch und komplex genug sind. Umso den Trainingsprozess effektiver zu gestalten werden sogenannte Attention Gates (AG) verwendet (Oktay, Schlemper, & Folgoc, 2018). Diese Technik sorgt dafür, dass beim Training relevante Bildbereiche hervorgehoben, während andere Bereiche vernachlässigt werden. Dies wird umgesetzt indem unterschiedlichen Teilen eines Bildes ein unterschiedliches Gewicht gegeben wird – relevante Bildteile bekommen dabei ein größeres Gewicht und damit mehr Aufmerksamkeit und nicht relevante Bildteile ein kleines Gewicht. Die Gewichte sind auch Trainingsparameter und verändern sich im Laufe des Trainings, das bedeutet mit AGs lernt eine Netzwerkarchitektur immer besser bestimmte Merkmale besser zu identifizieren und zu lokalisieren. Die an den Sprungverbindungen implementierten AGs unterdrücken aktiv die Aktivierung von Neuronen in irrelevanten Bereichen des Eingabebildes, wodurch die Anzahl der übertragenen redundanten Merkmale reduziert wird.

#### 2.5.3.4 Transformer U-Net (Trans U-Net)

Das Trans U-Net (Chen, et al., 2021) ist die erste vorgeschlagene Transformerbasierte Architektur zur semantischen Segmentierung. Anders als das klassische U-Net und alle anderen bisher vorgestellten Architekturen verwendet Trans U-Net eine hybride CNN-Transformer-Architektur als Kontraktionspfad, um sowohl die hochauflösenden räumlichen Informationen durch CNNs als auch globale Kontextinformationen von Transformern zu lernen. Diese CNN-Transformerarchitektur wurde zuerst für die sogenannten Vision Transformer (Dosovitskiy, et al., 2021) zur Bildklassifikation vorgeschlagen. Während CNNs sehr gut in der Lage sind lokale Kontextinformationen aus Bildern zu lernen, liegen ihre Schwächen darin Bildinformationen über einen großen Abstand hinweg zueinander in Verbindung zu setzen. Transformer-Architekturen überwinden diese Problematik. Sie sind aufgrund ihrer Konzipierung geeigneter um Kontextinformationen über größere Distanzen aus einem Bild zu lernen. Der Expansionspfad ist wie im klassischen U-Net konzipiert. Im Rahmen des Projektes wurde das Trans U-Net weiter angepasst. Das angewandte Trans U-Net besteht aus vier Schichten und 6 Transformerblöcken. Außerdem wurden nur 6 Attention-Blöcke verwendet. Das Multi Layer Perceptron in den Transformerblöcken hat 1536 Neuronen, welche mit der Gaußschen Aktivierungsfunktion<sup>8</sup> (GELU) genutzt werden.

#### 2.5.3.5 Swin U-Net

Das Swin U-Net (Cao, et al., 2021) stellt eine Weiterentwicklung des Trans U-Nets dar und ist die erste Architektur und ist die erste U-förmige Architektur welche nur auf Methoden der Transformer beruht. Ein Transformer führt die Merkmalerkennung mithilfe des Attention-Mechanismus durch in dem das Eingabebild in kleinere Bildpatches zerlegt wird. Diese werden Token genannt. Das Swin U-Net verwendet eine Weiterentwicklung des VisionTransformer - den sogenannten SwinTransformer (Liu, et al., 2021). Diese Architektur lernt Bildinformationen anders als der VisionTransformer in einem hierarchischen Schema und wird im SwinU-Net zur Merkmalsidentifikation eingesetzt und fungiert wie ein Kontraktionspfad. Symmetrisch dazu hat die Architektur einen SwinTransformer-basierten Expansionspfad der die ursprüngliche räumliche Auflösung der Merkmalskarten wiederherstellt.

#### 2.5.4 CNN

Zur Patch-basierten Bildklassifikation der Landbedeckung wird wie eingangs in Kapitel 2.5 erwähnt experimentell ein eigenes CNN aufgebaut. Das eingesetzte CNN besteht aus sieben Faltungsschichten, einer vollständig verbundenen Schicht und aus insgesamt etwa 11 Millionen trainierbaren Parametern (Gewichten). Die ersten drei Faltungsschichten haben eine Faltungsoperation mit einer 7 x 7-Filtermatrix, alle anderen Faltungsschichten eine 5 x 5-Filtermatrix. Damit wird in tieferen Schichten der Berechnungsaufwand der Merkmalskarten verkleinert. Nach jeder Faltungsoperation wird die ReLu-Aktivierungsfunktion angewandt. Ebenfalls wird ab der zweiten Faltungsebene eine Komprimierung der Merkmalskarten nach MaxPooling mit einem Poolingfenster von 2 x 2 und einer Schrittweite von  $s = 2$  durchgeführt. Ab der dritten Faltungsschicht werden DropOuts mit einer Verteilung von 0,3 angewandt um eine möglichen Überanpassung entgegen zu wirken und um das Netzwerk robuster zu generalisieren.

---

<sup>8</sup> <https://arxiv.org/pdf/1606.08415v4.pdf>

### 2.5.5 ResNets & InceptionNets

Neben dem in 2.5.4 vorgestellten CNN wurden verschiedene prominente Bildklassifikationsarchitekturen zur Landbedeckungsklassifikation verwendet. Angewandt wurden unterschiedliche Versionen der in Kapitel 2.5.2 beschriebenen Architekturen. Spezifisch wurde ein ResNet50 (Kaiming, Zhang, Ren, & Sun, 2015), ein ResNet101 (Kaiming, Zhang, Ren, & Sun, 2015), ein ResNet152 (Kaiming, Zhang, Ren, & Sun, 2015), ein ResNet50V2 (He, Zhang, Ren, & Sun, 2016), ResNet101V2 (He, Zhang, Ren, & Sun, 2016), ein InceptionV3 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) und ein InceptionResNetV2 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) angewandt.

### 2.5.6 Semi Supervised Generative Adversarial Network (sGAN)

Das sGAN (Chavdarova & Fleuret, 2018) wird den halbüberwachten Klassifikationsmodellen zugeordnet. Es handelt sich bei der Architektur um eine Weiterentwicklung der Generativen Gegnerischen Netzwerke (GANs) (Goodfellow, et al., 2014). Aufgrund der zeitlichen und ökonomischen Herausforderung einen eigenen Trainingsdatensatz aufzubauen, ist der Einsatz einer Architektur, die gute Ergebnisse bei einer sehr geringen Anzahl von Trainingsdaten liefert, sinnvoll. Das sGAN adressiert das Klassifikationsproblem, bei dem es wenige Trainingsdaten und viele ungekennzeichnete Daten gibt. Der Diskriminator in einem traditionellen GAN ist darauf trainiert, vorherzusagen, ob ein Eingabepatch echt (aus dem Datensatz) oder gefälscht (generiert) ist, wodurch er Merkmale von unbeschrifteten Bildern lernen kann. Der Diskriminator wird durch das in 2.5.4 beschriebene CNN gebildet. Der Diskriminator kann dann über Transferlernen als Ausgangspunkt bei der Entwicklung eines Klassifikators für denselben Datensatz verwendet werden, so dass die überwachte Vorhersageaufgabe vom nicht überwachten Training des GAN profitieren kann. Beim sGAN wird das Diskriminatormodell trainiert, um  $K+1$  Klassen vorherzusagen, wobei  $K$  die Anzahl der Klassen der Landbedeckung ist und die zusätzliche Klassenbezeichnung für eine neue gefälschte Klasse hinzugefügt wird. Es beinhaltet das direkte Trainieren des Diskriminatormodells für die nicht überwachte Aufgabe des GANs und auch für die überwachte Klassifikation gleichzeitig. Als solcher wird der Diskriminator in zwei Varianten trainiert. Einer überwachten und einer nicht überwachten Variante. Bei der unüberwachten Variante wird der Diskriminator auf die gleiche Weise wie beim traditionellen GAN trainiert, um vorherzusagen, ob das Beispiel echt oder falsch ist. Im überwachten Modus wird der Diskriminator darauf trainiert, die Klassenbezeichnung echter Beispiele vorherzusagen. Das Training im nicht überwachten Modus ermöglicht es dem Modell, nützliche Merkmale aus einem großen Datensatz ohne Label zu erlernen, während das Training im überwachten Modus es dem Modell ermöglicht, die extrahierten Merkmale zu verwenden und Klassenlabels vorherzusagen. Das Ergebnis ist ein leistungsstarkes Klassifikationsmodell, welches an nur sehr wenigen gelabelten Daten trainiert wird. Außerdem kann der Trainingsprozess auch zu einer besseren Bildqualität führen, die vom Generatormodell ausgegeben wird. Der Generator wird aktuell von einem CNN mit sieben transponierten Faltungsoperationen gebildet. Auf allen Faltungsebenen wird ein  $3 \times 3$ -Filter angewandt. Die Anzahl der Filter wird auf jeder Faltungsschicht gegenüber der vorherigen verdoppelt.

## 2.6 Genauigkeitsanalyse

Die Genauigkeit der verschiedenen Netzwerkstrukturen zur semantischen Segmentierung und zur Patch-basierten Klassifikation wurde anhand einer 10-fachen Kreuzvalidierung evaluiert. Die Aufteilung der Daten erfolgt wie in Kapitel 2.3 beschrieben zufällig. Die Auswahl von passenden Metriken ist von herausragender Bedeutung für eine fundierte Leistungsbeurteilung des Klassifikators. Bei einem Klassifizierungsproblem, wo ein künstliches neuronales Netz zwischen  $C$  Klassen unterscheidet, ergibt sich in der Netz-

werkausgabe ein normalisierter Vektor  $y_p \in \mathbb{R}^C$ . Jedes Element  $y_{p,c}$  in  $y_p$  beschreibt die Wahrscheinlichkeit für eine Klasse  $C_i$ . Wird ein neuronales Netz für die semantische Segmentierung von Bildern verwendet, steht im Ergebnis eine Wahrscheinlichkeitskarte (Matrix)  $Y_p \in \mathbb{R}^{H \times W \times C}$ .  $H$  und  $W$  beschreiben die seitlichen Dimensionen des Bildes,  $C$  steht für die Anzahl der möglichen Klassen. Für die Berechnung der Metriken wird  $Y_p$  durch eine Verkettung der Zeilen in die Form  $Y_p \in \mathbb{R}^{H \cdot W \times C}$  umgeformt.  $Y_p$  enthält so  $C$  Vektoren mit  $H \cdot W$  Elementen ( $y_p$ ). Diese Vektoren werden in anschließend äquivalenten Referenzdaten verglichen. Im Folgenden werden die projektbezogenen Evaluationsmetriken vorgestellt und erläutert.

Üblicherweise wird für die Evaluierung einer pixelweisen semantischen Segmentierung die Metrik Intersection over Union (IoU) oder auch Jaccard Index ( $J$ ) genannt verwendet. Der IoU beschreibt das Maß der Überlappung zwischen Vorhersagemaske und Groundtruthmaske in einem Wertebereich zwischen 0 und 1. Für zwei Mengen  $A$  und  $B$  kann der IoU entsprechend Formel (2) berechnet werden.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

Dabei beschreibt  $A$  die Referenzmaske und  $B$  das Klassifikationsergebnis. Für den Jaccard-Index  $J_C$  pro Klasse  $C$  gilt:

$$J_C(y_{t,c}, y_{p,c}) = \frac{y_{t,c} \cdot y_{p,c}}{\sum_{i=1}^C y_{t,i} + \sum_{i=1}^C y_{p,i} - y_{t,c} \cdot y_{p,c}} \quad (3)$$

Wobei  $y_{t,c}$  bzw.  $y_{p,c}$  die Referenz bzw. die Vorhersage der Klasse  $C$  beinhalten. Für die Berechnung des Mittleren IoU über alle Klassen  $C$ , ohne Gewichtung der Klassen gilt:

$$J_M(y_t, y_p) = \frac{1}{C} \sum_{i=1}^C J_i(y_{t,i}, y_{p,i}) \quad (4)$$

Im Rahmen des Projektes stellt der IoU die primäre Metrik zur Beurteilung der Leistung der semantischen Segmentierung dar.

Zur Bewertung von Patch-basierten Bildklassifikationsmodellen wird die Metrik Genauigkeit (Accuracy) verwendet. Die Accuracy beschreibt den Anteil der Vorhersagen, den das Modell richtig klassifiziert hat. Formal ist Genauigkeit wie folgt definiert:

$$P = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

Zur Bewertung der Klassifikationsgenauigkeit werden ebenfalls Konfusionsmatrizen gebildet, um systematische Klassifikationsfehler in anderen Klassen zu untersuchen.

## 2.7 Training der Architekturen

Alle Architekturen zur semantischen Segmentierung (Kapitel 2.5.1 bis 2.5.3) wurden zwischen 100 und 150 Epochen trainiert. Im Training wurden die betreffenden Architekturen mit einem gewichteten Verlustensemble trainiert. Dieses Ensemble bestand aus einer vom IoU abgeleiteten Verlustfunktion (Jadon, 2020) und der gewichteten Kreuzentropie. Die Gewichtung des Ensembles wurde beim Training der Architekturen mehrfach individuell variiert um die beste Konfiguration zu finden.

Für die vom IoU abgeleitete Verlustfunktion gilt formal

$$L_{IOU} = \log(J_M). \quad (6)$$

Bei Formel (6) wird ein Logarithmus angewendet um einen geringen IoU mit einem hohen Verlustwert zu bestrafen.

Die am häufigsten angewandte Verlustfunktion für Klassifikationsaufgaben ist die Kreuzentropie (englisch: cross entropy). Dabei wird die Differenz zwischen der vorhergesagten und der wahren Wahrscheinlichkeitsverteilung gebildet. Für eine Klassifikationsaufgabe mit  $i$  verschiedenen Klassen wird die Kreuzentropie mathematisch nach Gleichung (7) beschrieben (Jadon, 2020).

$$L_{CE} = - \sum_{i=1}^C \beta_i y_{t,i} (y_{p,i}) \quad (7)$$

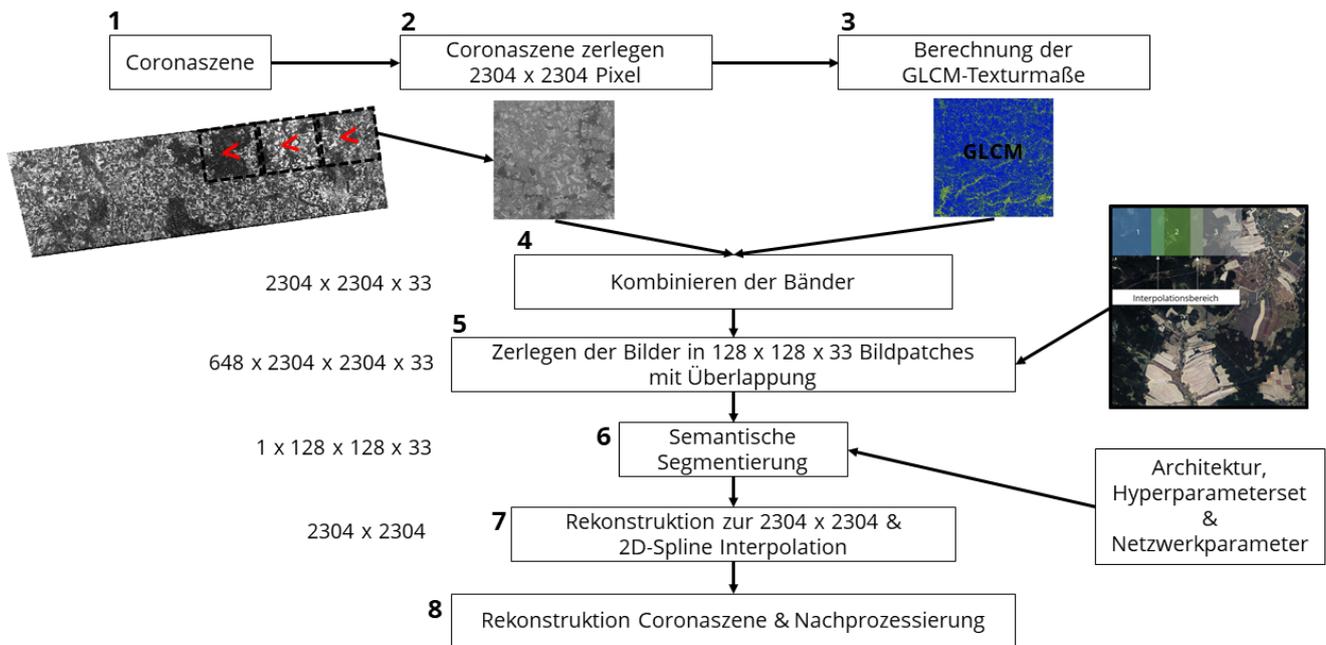
Dabei beschreibt  $C$  die Anzahl der Klassen,  $\beta$  die Gewichte der Klasse,  $y_{t,i}$  die wahre Klasse und  $y_{p,i}$  die vorhergesagte Klasse.

Für die Patch-basierte Landbedeckungsklassifikation wurde die gewichtete Kreuzentropie nach Gleichung (7) angewandt. Die Gewichte der Verlustfunktionen wurden aus der Verteilung der Trainingsdaten abgeleitet. Das geringste Gewicht bekommt die überrepräsentierteste Klasse (Wald) im Beispieldatensatz und das kleinste Gewicht die Klasse mit der geringsten Anzahl Beispieldaten (Gewässer). Für alle Trainings wurde eine adaptive Lernrate mit dem Adam Optimizer (Kingma & Ba, 2014) genutzt.

## 2.8 Interferenz

Im Rahmen des Abschlusses des Projektes wird aufgrund der in Kapitel 3 präsentierten Ergebnisse die Interferenz mit einer semantischen Segmentierung durchgeführt. Auf die Erklärung der geplanten Interferenzverfahrensweise für eine Patch-basierte Landbedeckungsklassifikation wird deshalb an dieser Stelle verzichtet. Für eine großflächige Landbedeckungsklassifikation der CORONA-Daten müssen die Architekturen mit deutlich größeren Bilddaten arbeiten können. Abbildung 15 zeigt den Ablauf der Interferenz für eine Coronaszene.

### Ablauf der Interferenz einer Coronaszene



**Abbildung 15: Ablauf der Interferenz zur semantischen Segmentierung einer Corona Szene. Die Schritte 1 bis 8 beschreiben schematisch den Ablauf der Interferenz. (2) Zunächst wird die Szene in 2304 x 2304 Pixel große Bildpatches zerlegt. (3) Für diese werden die entsprechenden GLCM-Texturmaße prozessiert. (4) Alle Bänder werden kombiniert in ein Bild mit 33 Kanälen. (5) Dieses Bild wird erneut in 128 x 128 x 33 große Bildpatches zerlegt welche die (6) Eingabe in die Architekturen bildet. (7) Es entstehen 128 x 128 Pixel große Landbedeckungskarten welche zur ursprünglichen Größe von 2304 x 2304 Pixel rekonstruiert werden. Im Überlappungsbereich wird während der Vorhersage interpoliert. In (8) wird durch Rekonstruktion der 2304 x 2304 Pixel großen Landbedeckungskarten die ursprüngliche Größe der Coronaszene wiederhergestellt.**

Zur semantischen Segmentierung der KH-Bilddaten können die U-förmigen Architekturen als bewegliches Fenster, wie in Abbildung 15 dargestellt, angewandt werden. Durch diesen Ansatz wird das große Bild iterativ verarbeitet. Diese iterative Verarbeitung ist notwendig aufgrund des enormen Speicherplatzes den die Daten einer Coronaszene einnehmen. Weiterhin wurden die Architekturen zur semantischen Segmentierung aufgrund des begrenzten GPU-Speichers auf eine Eingabegröße von 128 x 128 Pixel trainiert. Das Eingabefenster der U-förmigen Architekturen gleitet nun mit einer Schrittweite über das Coronabild und bestimmt die Landbedeckung in Form von Wahrscheinlichkeitskarten. Die resultierenden Subwahrscheinlichkeitskarten werden anschließend so miteinander verkettet, dass im Ergebnis eine Wahrscheinlichkeitskarte mit derselben lateralen Ausdehnung entsteht wie das ursprünglich große Bild (2.304 x 2.304 Pixel). Die Randbereiche einer Subwahrscheinlichkeitskarte zeigen meist verzerrte Ergebnisse mit schlechteren Genauigkeiten. Um diesen Effekt entgegenzuwirken wurde die Architektur als bewegliches Fenster mit einer kleineren Schrittweite als das Eingabefenster der U-förmigen Architektur bewegt. Daraus resultieren überlappende Vorhersagen an den Randbereichen der Wahrscheinlichkeitskarten. Für die Berechnung einer endgültigen Klassifikation innerhalb des Überlappungsbereichs wird eine Interpolation mit einer einfachen Spline-Fensterfunktion zweiter Ordnung durchgeführt. Anschließend werden die entstehenden Wahrscheinlichkeitskarten wieder zur ursprünglichen Größe der Coronaszene rekonstruiert. Für diesen Schritt werden alle 2.304 x 2.304 Karten pro Szene als virtuelle Raster (VRT) eingelesen, verarbeitet und schlussendlich als Mosaik zusammengesetzt. Mit dieser Methode wird sowohl Speicherplatz als auch Rechenzeit eingespart, da die endgültige Prozessierung erst dann ausgeführt wird, wenn sie auch wirklich benötigt wird und somit auf Zwischenprodukte verzichtet werden kann.

## 2.9 Nachprozessierung

Im Zuge der Nachprozessierung der Wahrscheinlichkeitskarten müssen diese zunächst zu einem zusammenhängenden Bild zusammengefügt werden. Da sich die Szenen allerdings in den Randbereichen jeweils überlappen, ist eine einfache Mosaikierung wie am Ende von Kapitel 2.8 beschrieben nicht möglich. Vielmehr muss eine Regel definiert werden die festlegt, welcher Pixelwert bei zwei sich überlappenden Pixeln in die finale Karte übernommen wird. Die breiten schwarzen Streifen an den Rändern der Corona-Szenen sind hierbei die Hauptfehlerquelle. Um diese beim Zusammensetzen der Karten zu beseitigen, wurden die Interferenzergebnisse zu Hilfe genommen. Da für die Streifenpixel zum allergrößten Teil die Klasse Wald vorhergesagt wurde (siehe Kapitel 3), wurde das Kriterium definiert, dass bei zwei unterschiedlichen Pixelwerten in den Überlappungsbereichen zunächst derjenige bevorzugt werden soll, der nicht der Klasse Wald entspricht. Analog wurde festgelegt, dass Klassen mit hoher Testgenauigkeit (Künstliche Oberflächen, Landwirtschaftliche Flächen) generell Vorrang gegenüber anderen Klassen bekommen sollen.

Nachdem die Mosaikierung auf Grundlage dieser Kriterien durchgeführt wurde, entsteht nun eine vollständige Landbedeckungskarte bestehend aus allen verfügbaren Corona-Szenen für ein Jahr mit 2 x 2 m Auflösung. In einem weiteren Nachprozessierungsschritt wurde diese Karte nun auf das vorgegebene Raster des LfULG mit 100 x 100 m Auflösung angepasst. Dazu wurde ein bereitgestelltes Template-Raster als Referenz für eine Resampling-Operation hinzugezogen. Als Resampling-Methode wurde Nearest Neighbor gewählt, um die kategorischen Pixelwerte zu erhalten. Im Anschluss wurde die gesamte Karte auf das Staatsgebiet Sachsens zugeschnitten und Ergebnisse für Gebiete außerhalb Sachsens wurden nicht weiter berücksichtigt. Im finalen Schritt wurden nun noch die Klassifikationsbezeichnungen von den zur vereinfachten Prozessierung genutzten Nummerwerten (0, 1, 2, ...) wieder ins CORINE-Klassifikationsschema (CLC 100, CLC 200, CLC 310, ...) überführt.

## 2.10 Konsistenz

Da die fertig prozessierten Landbedeckungskarten an einzelnen Stellen noch fehlerbehaftet sein können, in etwa durch Fehlklassifikationen, Wolken in den Originalaufnahmen oder Lücken in der Datenabdeckung, muss im Nachgang eine Strategie zur bestmöglichen Beseitigung dieser Fehler aufgestellt werden. Da davon ausgegangen werden kann, dass viele nicht konsistent in jedem Zeitschnitt auftreten, ist eine Betrachtung der Ergebnisse pro Pixel über die Zeit sinnvoll um deren Konsistenz zu gewährleisten. Um am Ende allerdings nur wahre Ausreißer zu beseitigen und nicht tatsächliche Landbedeckungsänderungen, wurde ein präzise definierter Regelsatz entworfen, welcher alle Eventualitäten in der Klassifikationszeitreihe berücksichtigt und sowohl einzelne Ausreißer detektiert und beseitigt, als auch potentielle Landbedeckungsänderungen auf ihre Wahrscheinlichkeit überprüft und unmögliche Änderungen nicht zulässt. Im Folgenden wird dieser Regelsatz beschrieben.

Wie bereits angedeutet, wird für den Regelsatz stets die Zeitreihe der Klassifikationswerte pro Pixel über alle Jahre hinzugezogen. Hier werden zunächst die Änderungen des Klassifikationswerts in der gesamten Zeitreihe gezählt. Für die Beseitigung von Ausreißern werden im Anschluss lediglich Pixel behandelt, die eine oder zwei Änderungen im Betrachtungszeitraum erfahren. Da es insgesamt lediglich fünf Zeitschnitte gibt, ist es bei drei oder vier Änderungen in der Zeitreihe nahezu unmöglich, die reale Dynamik zu erkennen ohne sie manuell in den Originalaufnahmen zu inspizieren. Gleichermaßen ist es unwahrscheinlich, dass diese Zeitreihen fehlerfrei sind, weswegen die betroffenen Pixel innerhalb des Regelsatzes nicht verändert werden und stattdessen geflaggt werden, um darauf hinzuweisen, dass die Ergebnisse der Klassifikation an dieser Stelle mit hoher Wahrscheinlichkeit inkonsistent und mit Vorsicht zu genießen sind.

Wird genau eine Landbedeckungsänderung in der Pixelzeitreihe festgestellt, wird diese nur als Ausreißer erkannt, wenn sie gleich zu Beginn auftritt, da die ersten beiden Zeitschnitte (1965 und 1967) zeitlich nah bei einander liegen und es sonst keine Möglichkeit gibt automatisch zu erkennen, ob der Wert aus dem Jahr 1965 wahr oder fehlerhaft ist. In diesem Fall wird der Wert dem Rest der Zeitreihe angeglichen. Tritt die Änderung am Ende auf, so wird zur Ausreißerprüfung die in Teilvorhaben 2 (TV2) erstellte Landbedeckungskarte aus dem Jahr 1985 als nächster Zeitschnitt hinzugezogen. Diese kann als Indikator für potentielle Änderungen am Ende des Beobachtungszeitraums fungieren. Im vorliegenden Fall wird der Klassenwert des Jahres 1978 nur dann beibehalten, wenn dieser dem Wert des Jahres 1985 an dieser Stelle entspricht. Tut er dies nicht, wird er als Ausreißer detektiert und den zuvorkommenden Werten angeglichen.

Bei Zeitreihen in denen zwei Landbedeckungsänderungen im gesamten Zeitraum auftreten wird wieder zunächst geprüft wo diese auftreten. Folgen sie direkt hintereinander wird der Wertebereich der Zeitreihe geprüft. Falls es nur zwei Werte gibt, kann davon ausgegangen werden, dass es sich um einen Ausreißer handelt. An dieser Stelle wird der Wert schließlich an den Rest der Reihe angeglichen. Eine Ausnahme besteht, wenn der Wert des Jahres 1975 der detektierte Ausreißer ist. In diesem Fall wird analog zum vorigen Absatz das Jahr 1985 hinzugezogen. Stimmen beide Werte überein, wird stattdessen der Wert des Jahres 1978 als der Ausreißer angesehen und zum Wert der beiden genannten Jahre gesetzt. Tritt die eine Änderung direkt am Anfang und die andere am Ende auf, so wird der erste Wert wiederum als Ausreißer erkannt und den Folgewerten angeglichen. Dies geschieht allgemein stets, wenn mindestens die zwei folgenden Werte der Jahre 1967 und 1972 gleich sind. Der letzte Wert (des Jahres 1978) wird analog zu oben unter Berücksichtigung des entsprechenden Wertes aus 1985 behandelt. Auch dieses Vorgehen wird in den meisten Fällen in denen der letzte Wert ein potentieller Ausreißer sein könnte angewandt. Das Jahr 1985 fungiert demnach lediglich als Zusatzinformation um auffällige Fehler im Jahr 1978 zu überprüfen und gegebenenfalls zu korrigieren, nicht aber als "wahrer Wert" für 1978. Mit all diesen Konfigurationen im Regelsatz soll sichergestellt werden, dass potentielle echte Änderungen im Beobachtungszeitraum erhalten bleiben, Fehler aber weitestgehend eliminiert werden. Der Regelsatz ist daher bewusst an manchen Stellen konservativ gehalten, da direkt in die Ergebnisse der Klassifikation eingegriffen wird und somit stets die Gefahr besteht, korrekt klassifizierte Pixel zu verändern.

Im nächsten Schritt des Regelsatzes wird auf die Klassifikationswerte direkt eingegangen. Grundsätzlich wird die Zeitreihe bei einer konsistenten permanenten Klassenänderung nicht verändert. Erscheint die Änderung allerdings unlogisch (beispielsweise von Künstliche Oberflächen zu Wald innerhalb weniger Jahre), so kann auch in diesem Fall von einem oder mehreren Fehlern in der Zeitreihe ausgegangen werden. Für die Definition von möglichen und unmöglichen Klassenänderungen innerhalb eines Zeitschnitts wurden alle Eventualitäten hinsichtlich ihrer Plausibilität überprüft. Diese werden in der nachfolgenden Tabelle 4 dargestellt.

**Tabelle 4: Regelsatz für die Definition von plausiblen Landbedeckungsänderungen innerhalb eines Zeitschnitts ("von Klasse A nach Klasse B")**

von \ nach	CLC 100	CLC 200	CLC 310	CLC 320	CLC 330	CLC 510
CLC 100	ja	nein	nein	ja	ja	nein
CLC 200	ja	ja	nein	ja	ja	nein
CLC 310	ja	ja	ja	ja	ja	nein
CLC 320	ja	ja	ja	ja	ja	nein
CLC 330	ja	ja	nein	ja	ja	ja
CLC 510	ja	ja	nein	ja	ja	ja

Im Umgang mit unzulässigen Klassenänderungen wird wieder zunächst die Anzahl der Änderungen in der jeweiligen Zeitreihe analysiert. Wenn eine einmalig vorkommende Landbedeckungsänderung in einer Zeitreihe gleichzeitig eine nicht-plausible ist, so wird diese beseitigt indem alle Werte die nicht dem Medianwert der Reihe entsprechen durch ebendiesen Wert ersetzt. Finden zwei Änderungen statt, wird das gleiche Vorgehen angewandt, wenn über die gesamte Zeitreihe lediglich zwei unterschiedliche Klassenwerte vorkommen. Ist dies nicht der Fall, wird analog zum bereits oben beschriebenen Vorgehen analysiert, wann die Änderungen auftreten. Passiert dies gleich zweimal zu Beginn oder am Ende der Zeitreihe, werden wieder alle Werte zum Medianwert der Reihe geändert. Treten die beiden Änderungen in der Mitte auf, kann diese Methode nicht angewandt werden, auch da hier eventuelle "wahre" Landbedeckungsänderungen wahrscheinlicher sind. Stattdessen wird wieder der korrespondierende Wert auf dem Jahr 1985 hinzugezogen um zu einer Entscheidung zu kommen. Je nachdem wann die zwei Änderungen in der Zeitreihe auftreten werden auf Grundlage dieses Wertes ein oder mehr Werte in der Zeitreihe angepasst oder nicht (meist, wenn die Klassen von 1978 und 1985 übereinstimmen). Meist wird dadurch nur eine Änderung in der Zeitreihe beseitigt, welche naturgemäß die unzulässige sein sollte. Stimmt der Wert aus 1985 allerdings mit mehreren Werten am Anfang der Zeitreihe überein, so wird wiederum die ganze Zeitreihe an diesen Wert angepasst. Auch nach Anwendung dieses zweiten Teils des Regelsatzes können demnach unzulässige Landbedeckungsänderungen in den Pixelzeitreihen verbleiben für die keine allgemeingültige Regel zur Beseitigung gefunden werden konnte. Diese werden letztendlich mit einem Flag versehen. Auch an dieser Stelle werden Zeitreihen mit drei oder vier Klassenänderungen nicht behandelt, da wie oben beschrieben nicht automatisch detektiert werden kann, welche Änderungen in diesen Fällen richtig sind und welche nicht und ebenso keine allgemeingültige Regel zur Verbesserung des Ergebnisses aufgestellt werden kann. Flags weisen auch hier auf die Unzulässigkeit solcher Pixel hin.

Schließlich müssen zusätzlich Regeln für Zeitreihen in denen Werte für einen oder mehrere Jahre fehlen aufgestellt werden. Im Regelsatz werden diese generell genauso behandelt wie andere Klassenwerte und werden dementsprechend auch meist durch die oben beschriebenen Festlegungen eliminiert. Lediglich in Fällen in denen eine Mehrzahl der Werte fehlen (Median = NA) werden diese gesondert behandelt. Bei nur einer Änderung in dieser Zeitreihe werden alle Werte zu dem jeweils nicht-fehlenden Wert gesetzt. Bei zwei Änderungen ist nicht ersichtlich mit welchem Wert die fehlenden Werte ersetzt werden sollten, daher werden in diesem Fall alle Werte als nicht vorhanden betrachtet, um keine falschen Landbedeckungsänderungen zu erzeugen. Es ist allerdings anzumerken, dass in den Ergebnissen äußerst selten mehr als zwei Werte in der Zeitreihe nicht vorhanden sind.

Nach der Anwendung des Regelsatzes werden Pixel in denen eine Klassenänderung an mindestens einem Zeitpunkt in der Zeitreihe vorgenommen wurde ebenfalls mit einem Flag versehen. Die Prozessierung der Landbedeckungskarten ist mit diesem Schritt abgeschlossen.

## 2.11 Implementierung und Software

Die Georeferenzierung wurde in der ESRI Software ArcGIS 10.2<sup>9</sup> durchgeführt. Zur Auswahl der Studiengebiete und zur Erstellung der semantischen Masken wurde die freie GIS Software QGIS<sup>10</sup> in der Version 3.2 verwendet. Alle Implementierungen der Architekturen und Prozesse zur Vorprozessierung der Beispieldaten wurden in Python 3.7 implementiert. Für alle Deep Learning bezogenen Prozesse wurden die Bibliotheken Tensorflow 2.6 und Keras 2.6 verwendet. Zur Nutzung der GPUs im Training der Architekturen und zur Interferenz wurde die NVIDIA Software CUDA<sup>11</sup> in der Version 10.1 genutzt. Für die Nachprozessierung wurden zur Verarbeitung großer Rasterdaten die Bibliotheken raster 3.6.3<sup>12</sup> und terra 1.6.17<sup>13</sup> in R 4.2.2 verwendet.

Zur Entwicklung und für das Training der Architekturen wurden zwei virtueller Servercomputer mit AMD EPYC 7513 Prozessoren und zwei NVIDIA A40-12Q GPUs verwendet. Die Interferenz wurde teilweise mit dem selbigen virtuellen Servercomputern durchgeführt, als auch mit HPC-Cluster der TU Dresden. Dazu wurden im Rahmen der Interferenz eine NVIDIA A100 GPUs verwendet.

---

<sup>9</sup> <https://www.arcgis.com/home/index.html>

<sup>10</sup> <https://www.qgis.org/de/site/>

<sup>11</sup> <https://developer.nvidia.com/cuda-toolkit>

<sup>12</sup> <https://cran.r-project.org/web/packages/raster/raster.pdf>

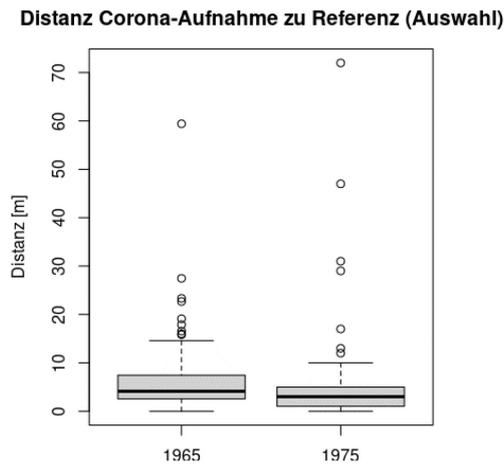
<sup>13</sup> <https://cran.r-project.org/web/packages/terra/index.html>

### 3 Ergebnisse

In den folgenden Subkapiteln werden die Projektergebnisse vorgestellt. Die Landbedeckungsklassifikation wurde für zwei unterschiedliche Aufgabentypen implementiert und getestet. Je nach Aufgabenstellung werden die Ergebnisse im Weiteren unabhängig voneinander vorgestellt.

#### 3.1 Georeferenzierung

Wie in Kapitel 2.2 beschrieben wird die Genauigkeit der Georeferenzierung anhand von 100 Kontrollpunkten überprüft. Umgesetzt wird dies beispielhaft für die Zeitschnitte 1965 und 1975. In der nachfolgenden Abbildung 16 werden die Ergebnisse dafür in einem Boxplot zusammengefasst.



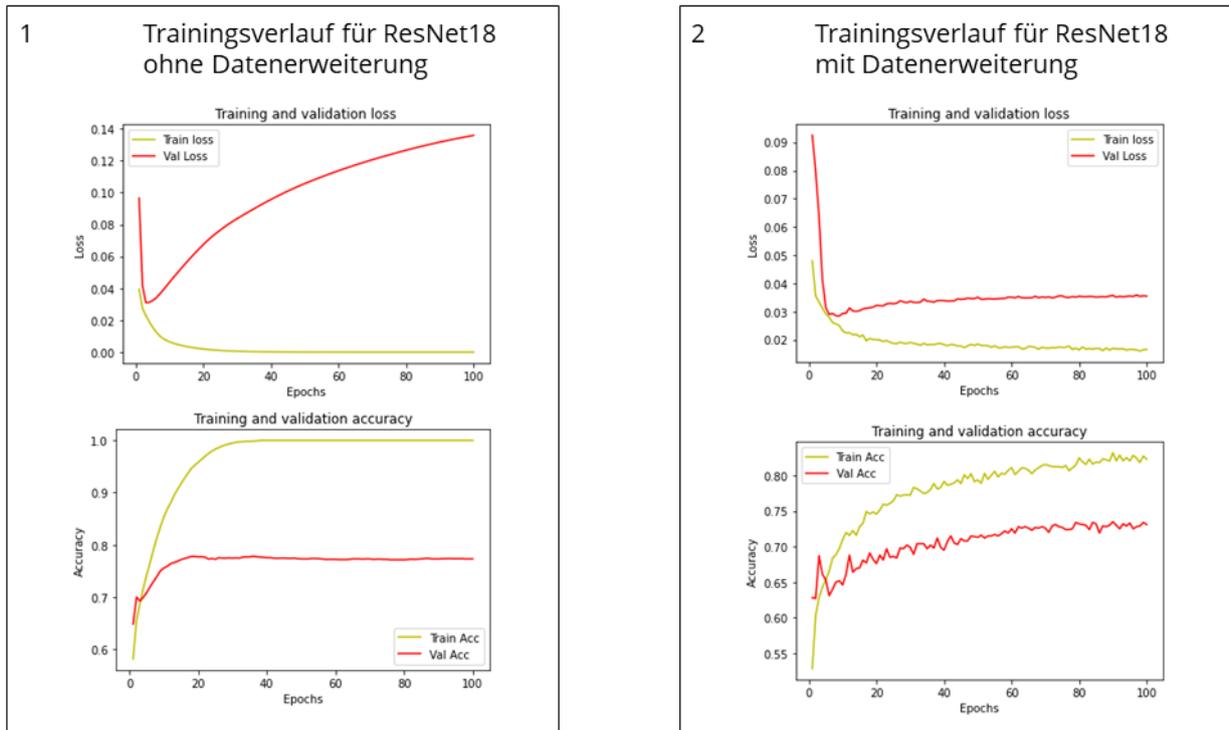
**Abbildung 16: Differenz der CORONA-Aufnahmen zu den Referenzdaten, Mittelwert 1965 = 6,47m, Median 1965 = 4,12m, Mittelwert 1975 = 5,17m, Median 1975 = 3m**

Aus Abbildung 16 geht hervor das 75% der Kontrollpunkte für 1965 eine Abweichung von weniger als 8 m aufweisen. Der Median der Differenz beträgt 4,12 m für 1965. Die größte Abweichung tritt für einen Ausreißer mit rund 60m auf. Die meisten Ausreißer haben eine Differenz zwischen 15m und rund 30m. Aufgrund der besseren Bildqualität für 1975 konnte für diesen Zeitschnitt ein besseres Ergebnis für die Georeferenzierung erreicht werden. Der Median betrug hier 3 m. Außerdem weisen 75% der Kontrollpunkte eine Differenz von weniger als 5 m auf. Die Ausreißer sind offener verteilt als für 1965. Die größte Abweichung beträgt über 70m.

Aus den Ergebnissen der Georeferenzierung kann damit abgeleitet werden, das mit steigender Bildqualität ebenfalls auch die Genauigkeit der Georeferenzierung gesteigert werden kann.

#### 3.2 Ergebnisse der Datenerweiterung

In diesem Subkapitel werden die Ergebnisse und die Auswirkung der Datenerweiterung beispielhaft anhand der Patch-basierten Klassifikation der Landbedeckung dargelegt. Die Notwendigkeit und die Vorteile der Datenerweiterung können insbesondere in den Verlustkurven und Genauigkeitskurven für die Trainingsdaten und Validierungsdaten während des Trainings erkannt werden. Dazu wird im Folgenden beispielhaft der Trainingsverlauf eines ResNet18 mit und ohne Datenerweiterung dargestellt.



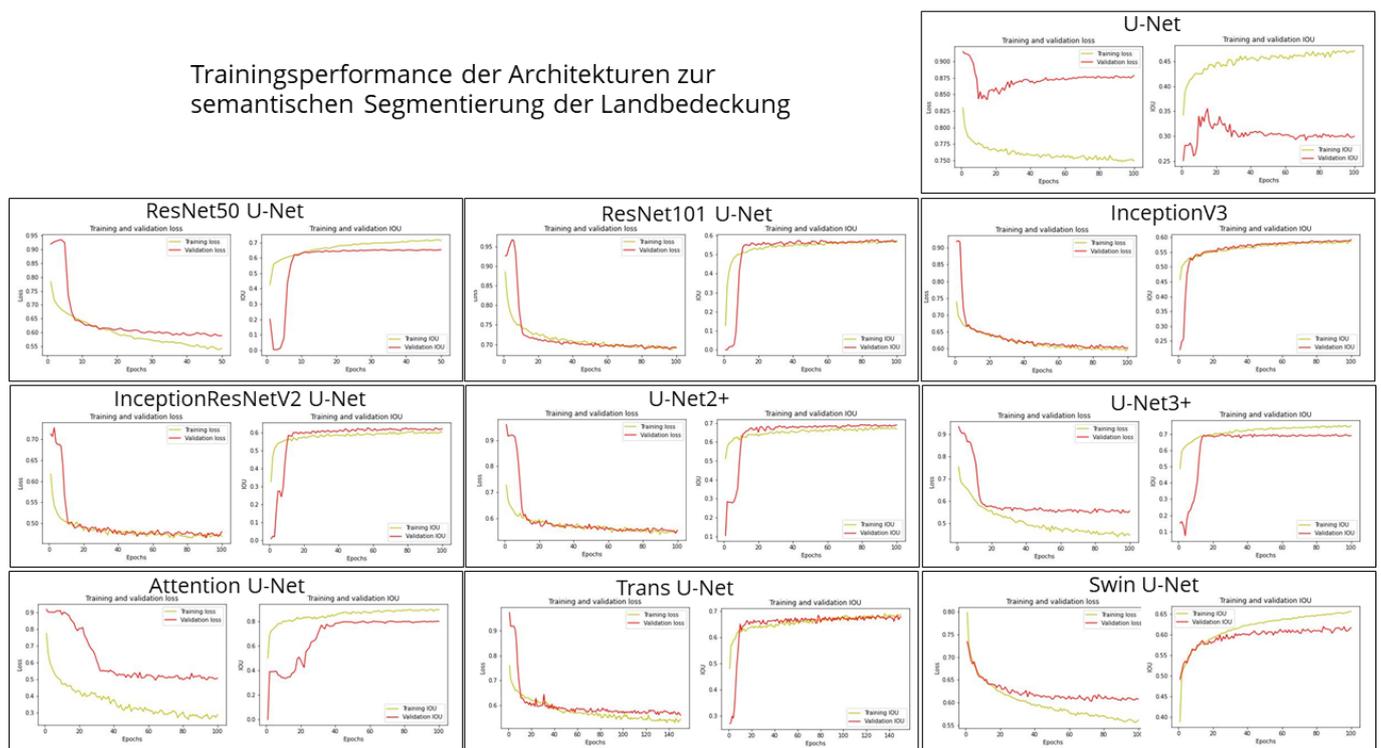
**Abbildung 17: Trainingsverlauf eines ResNet18 mit und ohne Methoden der Datenerweiterung. Die Abbildung besteht aus zwei Subabbildungen. Die linke zeigt den Trainingsverlauf eines ResNet18 ohne Datenerweiterung. Die rechte Subabbildung repräsentiert den Trainingsverlauf eines ResNet18 mit Anwendung der Methoden der Datenerweiterung (siehe Kapitel 2.5.4).**

Aus der linken Teilabbildungen von Abbildung 17 ist eine deutliche Überanpassung der Trainingsdaten gegenüber den Validierungsdaten zu erkennen. Dies äußert sich außerdem anhand der zwei Verlustkurven in der unteren Abbildung. Diese zeigen im Verlauf des Trainings einen immer größeren Abstand. Dieses Trainingsverhalten zeugt von einer schlechten Generalisierbarkeit der Architektur auf unbekannte Daten. Damit ist eine Architektur welche einen solchen Verlauf zeigt nicht für eine Interferenz geeignet. Im Training für die rechte Subabbildung wurden die in Kapitel 2.5.4 bis 2.5.6 für die Patch-basierte Landbedeckungsklassifikation genannten Methoden angewandt. Aus den Verlaufskurven wird deutlich, dass mit Datenerweiterung der Überanpassung entgegengewirkt werden kann. Die Kurven zum Verlust zeigen mit Anwendung der Datenerweiterung einen deutlich kleinere Überanpassung der Trainingsdaten. Dies äußert sich ebenfalls im gleichmäßigen Anstieg der Verlaufskurven in der oberen Abbildung, welche die Modellierungsgenauigkeit und Validierungsgenauigkeit während des Trainings zeigen. Aus Abbildung 17 geht ebenfalls hervor, dass die Methoden der Datenerweiterung für dafür eine spätere Sättigung der Architektur sorgt. Während das ResNet18 ohne Methoden der Datenerweiterung bereits ab 30 Epochen keine Zunahme der Validierungsgenauigkeit zeigt, steigt die Kurve der Modellierungsgenauigkeit und Validierungsgenauigkeit für das ResNet18 mit Datenerweiterung kontinuierlich an. Durch die Anwendung der Datenerweiterung ergibt sich eine schlechtere Validierungsgenauigkeit beim Training. Ohne Datenerweiterung steigt die Validierungsgenauigkeit auf über 75% an, mit Datenerweiterung nur bis maximal 71%.

Die Notwendigkeit der Datenerweiterung ergibt sich trotz der negativen Auswirkungen auf die Validierungsgenauigkeit, aus Notwendigkeit der Vermeidung einer Überanpassung um die Anwendbarkeit der Architekturen auf unbekannte Testdaten sicher zu stellen.

### 3.3 Training der semantischen Segmentierungs-Methoden

Nach der finalen Hyperparametereinstellung für jede Architektur wurden im Rahmen des Projektes für die abschließende Genauigkeitsanalyse insgesamt 100 Modelle trainiert. Dabei entfallen je zehn Modelle auf jede Architektur. Für die Architektur mit der besten Trainingsperformance über alle Kreuzvalidierungen werden im Weiteren die Verlaufskurven für den Verlust und für den IoU während des Trainings abgebildet. Aus diesen Kurven kann abgeleitet werden ob eine Architektur in der Lage ist die Beispieldaten zu modellieren und damit sinnvolle Merkmale aus den Daten extrahieren kann oder nicht. Außerdem können Überanpassungen und Unteranpassungen der Architekturen an die Daten erkannt werden. Dies wichtig um die Anwendbarkeit einer Architektur sicher zu stellen. Es können ebenfalls Indizien, dafür gefunden werden, ob ein Trainingsdatensatz beispielweise nicht repräsentativ für den Testdatensatz ist. Dies äußert sich beispielsweise durch einen großen Abstand zwischen den Kurven.



**Abbildung 18: Trainingsperformance der Architekturen zur semantischen Segmentierung der Landbedeckung. Abgebildet wird je das beste Trainingsergebnis einer Architektur, also das Ergebnis einer Kreuzvalidierung. Für jede Architektur werden zwei Subabbildungen gezeigt. Die rechte zeigt den Verlauf der Modellierungsgenauigkeit bzw. Validierungsgenauigkeit, die linke den Verlust der Trainingsdaten und Validierungsdaten. Die gelben Kurven korrespondieren zum Trainingsverlauf, die roten Kurven zum Validierungsverlauf. Es ist anzumerken, dass die Achsen nicht einheitlich skaliert sind.**

Alle verwendeten Architekturen zeigen aufgrund ihrer steigenden Genauigkeitskurven und fallenden Verlustkurven vom Training die Fähigkeit die Beispieldaten zu modellieren und damit KH-Bilddaten semantisch zu segmentieren. Das InceptionResNetV2 U-Net und das U-Net2+ zeigen eine leichte Unteranpassung der Beispieldaten was bedeutet, dass der Trainingsdatensatz für die Architekturen schwerer zu modellieren als die Validierungsdaten vorherzusagen. Kennzeichnend wird das an der Validierungskurve der Genauigkeit, welche über der Trainingskurve der Genauigkeit liegt. Alle Architekturen zeigen ebenfalls einen schnellen Modellierungserfolg. Über 80% der Performance wird von allen Architekturen in den ersten 20 Epochen erreicht. Die besten Ergebnisse gemessen an der Modellierungsgenauigkeit und Validierungsgenauigkeit zeigt das Trans

U-Net, das Attention U-Net und das U-Net3+. Die höchste Modellierungsgenauigkeit erreichte das Attention U-Net von über 80%, dicht gefolgt vom U-Net3+ und vom Trans U-Net. Die gleichmäßigste Modellierung der Beispieldaten zeigt das Trans U-Net und das InceptionV3 U-Net mit einem fast identischen Kurvenverlauf über die gesamte Länge des Trainings. Das Attention U-Net und das U-Net3+ zeigen hingegen in den ersten Epochen Sprünge in den Kurven was auf eine zu kleine Batch-Size hindeuten kann. Aufgrund des begrenzten GPU-Speichers war eine Erhöhung der Batch-Size nicht möglich. Insgesamt zeigt das Trans U-Net die beste Mischung aus gleichmäßiger Modellierung, Trainingsgenauigkeit und Validierungsgenauigkeit.

### 3.4 Testergebnisse der semantischen Segmentierung

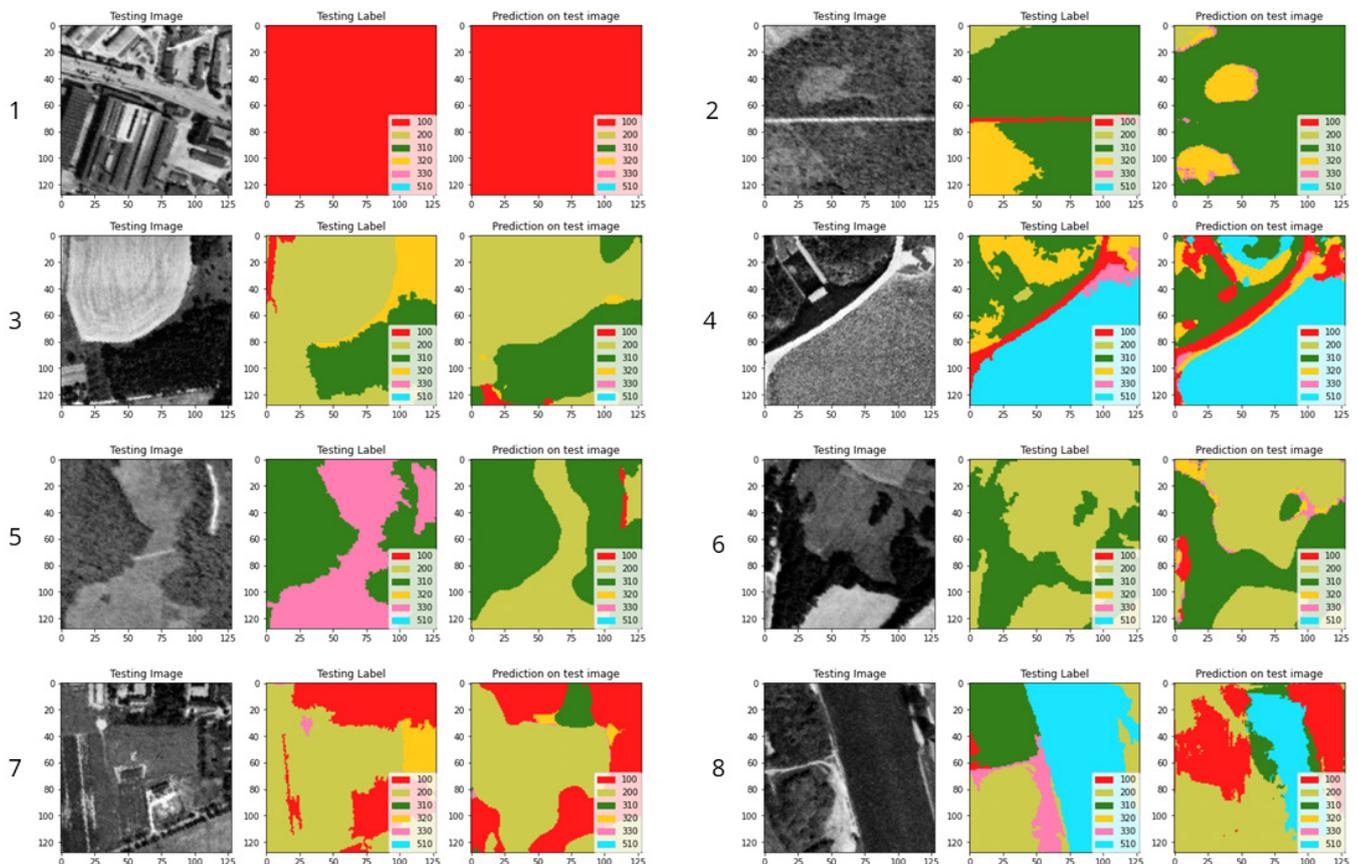
In diesem Subkapitel werden die Ergebnisse der semantischen Segmentierung der Testdaten aufgezeigt. Die nachfolgende Tabelle 5 gibt die mittleren IoU Ergebnisse je Klasse für alle Architekturen an.

**Tabelle 5: Ergebnisse der semantischen Segmentierung der Testdaten. Die Ergebnisse pro Klasse sind je die Mittelwerte aus den Kreuzvalidierungen für jede Architektur. Der mittlere IoU für eine Architektur ergibt sich aus dem Mittelwert der Klassen. CLC 100 = Künstliche Oberflächen, CLC 200 = Landwirtschaftliche Flächen, CLC 310 = Wald, CLC 320 = Strauchvegetation, CLC 330 = offene Fläche, CLC 510 = Gewässer. Alle Angaben in der Tabelle geben den IoU wieder.**

	CLC 100	CLC 200	CLC 310	CLC 320	CLC 330	CLC 510	Mittlerer IoU
U-Net	57,07 %	63,47 %	73,10 %	0,00 %	0,00 %	0,00 %	32,22 %
ResNet50 U-Net	53,54 %	62,07 %	75,31 %	3,86 %	10,45 %	39,37 %	40,77 %
ResNet101 U-Net	57,05 %	63,02 %	74,42 %	12,74 %	6,63 %	34,98 %	41,13 %
InceptionV3 U-Net	59,14 %	67,82 %	78,47 %	3,11 %	0,68 %	53,81 %	43,84 %
InceptionResNetV2 U-Net	60,97 %	70,26 %	84,27 %	17,27 %	12,13 %	61,83 %	51,12 %
U-Net2+	61,25 %	70,21 %	82,76 %	14,27 %	9,26 %	50,02 %	47,95 %
U-Net3+	61,58 %	71,99 %	82,22 %	14,82 %	7,67 %	57,89 %	49,36 %
Attention U-Net	53,24 %	62,85 %	75,38 %	3,32 %	19,94 %	39,25 %	42,33 %
Trans U-Net	68,83 %	74,21 %	83,29 %	21,82 %	22,97 %	66,49 %	56,26 %
Swin U-Net	62,91 %	71,27 %	84,67 %	14,54 %	16,41 %	52,78 %	52,10 %

Der beste mittlere IoU über alle Klassen wurde von den beiden Transformerarchitekturen Trans U-Net und Swin U-Net erreicht. Das Trans U-Net erreicht einen mittlere IoU von 56,26% und das Swin U-Net einen mittleren IoU von 52,10%. Ebenfalls eine sehr gute Testperformance zeigt das InceptionResNetV2 U-Net mit 51,12%. An dieser Stelle gilt es zu berücksichtigen, dass das InceptionResNetV2 in sieben von zehn Kreuzvalidierungen eine Unteranpassung der Trainingsdaten gezeigt hat und damit die Testergebnisse nicht verallgemeinert auf ähnliche Interferenzdaten werden kann. Alle anderen Architekturen zeigen einen mittleren IoU über alle Klassen von unter 50%. Das Trans U-Net zeigt außerdem bei einer Einzelbetrachtung der Klassen einer Ausnahme die beste Performance in allen Klassen. Nur in der Klasse Wald erreichte das InceptionResNetV2 U-Net eine bessere Testgenauigkeit von 84,27% gegenüber den 83,29% des Trans U-Nets. Das Trans U-Net ist außerdem die einzige Architektur welche in allen drei unterrepräsentieren Klassen einen mittleren IoU von über 20% aufzeigt. Den schlechtesten mittleren IoU über alle Klassen zeigt das klassische U-Net mit einer Testgenauigkeit von 32,22%. Abschließend ergibt sich im

Rahmen des genutzten Beispieldatensatzes, das das Trans U-Net am besten zur semantischen Segmentierung von KH-Bilddaten geeignet ist. Im Folgenden werden die Ergebnisse der semantischen Segmentierung auf die Testdaten in einigen Abbildungen dargestellt.



**Abbildung 19: Beispiele für die semantische Segmentierung von Trainingsdaten mit einem Trans U-Net. Die Subabbildungen 1 bis 8 zeigen von links nach rechts je das panchromatische Band des Testpatches, die semantische Maske und recht die Vorhersage für den Testpatch. Die Klassenlabel haben folgende Bedeutung: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

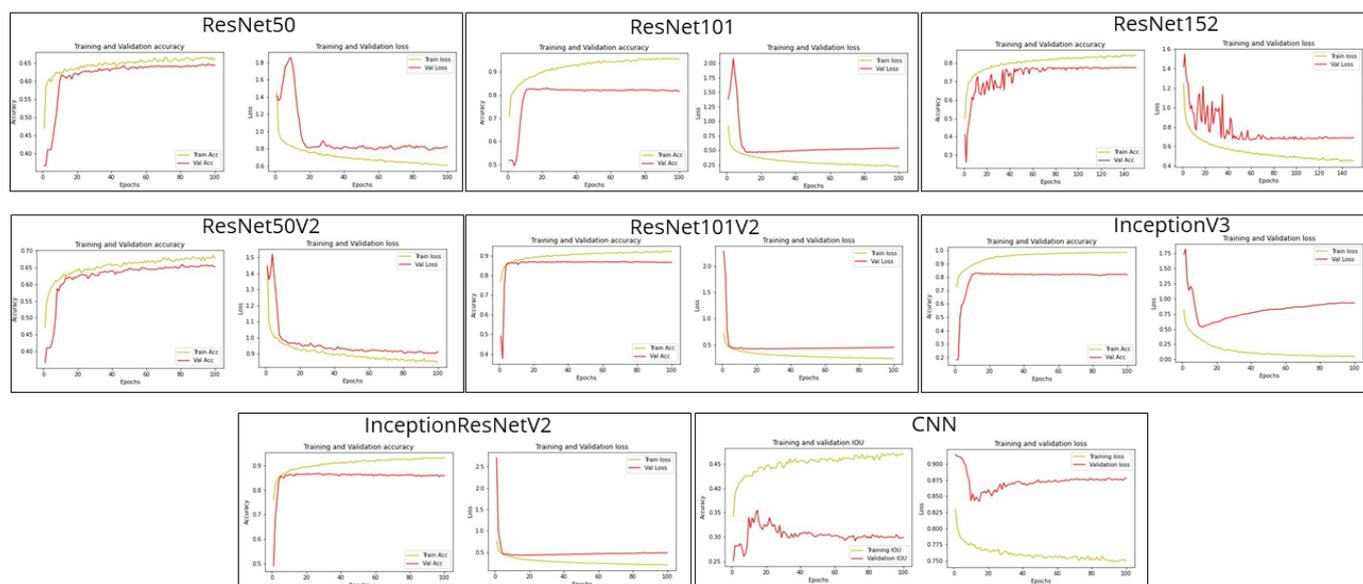
Die Vorhersagen von Abbildung 19 zeigen eine gute Genauigkeit der semantischen Segmentierung für alle Klassen. Die erste Subabbildung zeigt vollständig eine Künstliche Oberfläche welche auch so vollständig richtig klassifiziert wurde. In der zweiten Subabbildung ist überwiegend Wald gemischt mit Strauchvegetation zu sehen. Beide Klassen wurden überwiegend gut segmentiert. Es fällt allerdings auf, dass in der Mitte der Waldfläche eine Fläche richtig als Strauchvegetation klassifiziert wurde, welche in der semantischen Maske nicht gekennzeichnet ist. Hier führt das Trans U-Net eine semantisch richtige Klassifikation durch welche allerdings die Klassifikationsgenauigkeit aufgrund des Fehlers in der semantischen Maske verfälscht. Ein ähnliches Beispiel ist in Subabbildung 5 zu erkennen. Bei dieser wurde die offene Fläche in der semantischen Maske in der Vorhersage richtig als Klasse Landwirtschaftliche Fläche segmentiert. In der Subabbildung 3 wurde die Strauchvegetation links im originalen panchromatischen Bild falsch als Landwirtschaftliche Fläche gekennzeichnet. Weitere Fehler des Trans U-Nets sind in der Subabbildung 8 zu erkennen. Dort hat das Trans U-Net sowohl den Wald links im originalen panchromatischen Bild und den Flussverlauf nicht richtig segmentiert. Der Flussverlauf wurde dabei nicht durchgängig segmentiert und bei den Waldflächen kam es zur Fehlklassifikation in die Klasse Künstliche Oberflächen und Landwirtschaftliche Fläche.

Betont werden muss an dieser Stelle, dass Fehler im Trainingsdatensatz wie sie Abbildung 19 in aufzeigt wurden sowohl zu einer fehlerhaften Modellierung von Klassen während des Trainings führen und damit zu Fehler in der Interferenz als auch zu unzuverlässigeren Genauigkeitsangaben für die Tests der Architekturen.

### 3.5 Training der Patch-basierten Bildklassifikationsmethoden

Die Ergebnisse der Patch-basierten Bildklassifikation der Landschaftsbedeckung in einer Rasterweite von 100 m basieren auf einer 10-fachen Kreuzvalidierung des Beispieldatensatzes in Trainingsdaten, Testdaten und Validierungsdaten. Die Prozessierung der Ergebnisse für das sGAN beruhen auf dem gleichverteilten Datensatz mit je 100 Beispielen pro Klasse. Nach der finalen Hyperparametereinstellung für jede Architektur wurden im Rahmen des Projektes für die abschließende Genauigkeitsanalyse der Patch-basierten Landbedeckungsklassifikation insgesamt 80 Modelle trainiert - zehn Modelle pro Architektur. Für die Architektur mit der besten Trainingsperformance über alle Kreuzvalidierungen werden im Weiteren die Verlaufskurven für den Verlust und für die Genauigkeit während des Trainings abgebildet.

Trainingsperformance der Architekturen zur Patch-basierten Bildklassifikation der Landbedeckung



**Abbildung 20:** Trainingsperformance der Architekturen zur semantischen Segmentierung der Landbedeckung. Abgebildet wird je das beste Trainingsergebnis einer Architektur, also das Ergebnis einer Kreuzvalidierung. Für jede Architektur werden zwei Subabbildungen gezeigt. Die linke zeigt den Verlauf der Modellierungsgenauigkeit bzw. Validierungsgenauigkeit, die rechte den Verlust der Trainingsdaten und Validierungsdaten. Die gelben Kurven korrespondieren zum Trainingsverlauf, die roten Kurven zum Validierungsverlauf.

Aus Abbildung 20 ist anhand der steigenden Kurven der Modellierungsgenauigkeit und der fallenden Verlustkurven zum Training zu erkennen, dass alle Architekturen in der Lage sind die Beispieldaten erfolgreich zu modellieren. Die besten Modellierungsgenauigkeiten weist das InceptionV3, das ResNet101, das ResNet101V2 und das InceptionResNetV2 auf. Alle vier Architekturen zeigen eine Modellierungsgenauigkeit der Trainingsdaten von über 90%. Alle anderen Architekturen zeigen eine geringere Fähigkeit die Trainingsdaten zu modellieren. Die besten Ergebnisse zur Validierungsgenauigkeit beim Training zeigen das InceptionResNetV2 und das ResNet101V2. Beide Architekturen zeigen im Maximum eine Validierungsgenauigkeit von 84%. Das InceptionV3 und das ResNet101, sowie alle anderen Architekturen zeigen eine

geringere Performance für die Validierungsdaten. Außerdem zeigen das InceptionResNetV2 und das ResNet101V2 die gleichmäßigste Trainingsperformance. Dies ist insbesondere an den glatten Verlaufskurven zum Verlust der Architekturen zu erkennen. Aus den Verlaufskurven des Verlustes ist allerdings auch zu erkennen, dass beide Architekturen eine leichte Übermodellierung der Trainingsdaten über den Verlauf des Trainings zeigen. Dies ist aus dem leicht zunehmenden Abstand der Verlaufskurven des Trainings und der Validierung zum Verlust zu erkennen. Diese Übermodellierung ist beim InceptionResNetV2 etwas größer als beim ResNet101V2. Insgesamt zeigt damit das ResNet101V2 die beste Trainingsperformance sowohl für die Fähigkeit, die Beispieldaten gleichmäßig auf hohem Niveau zu modellieren, als auch für die Validierungsgenauigkeit. Alle anderen Architekturen zeigen im Verlauf des Trainings teilweise größere Sprünge. Besonders ist dies beim ResNet152 zum Beginn des Trainings bis Epoche 40 zu erkennen. Diese Sprünge deuten in diesem Fall aufgrund der glatten Verlaufskurven des Trainings eine anfängliche Übermodellierung der Trainingsdaten an.

### 3.6 Klassifikationsergebnisse der Patch-basierten Bildklassifikation

In diesem Subkapitel werden die Ergebnisse der Anwendung der Architekturen zur Patch-basierten Landbedeckungsklassifikation auf die Testdaten aufgezeigt. Dabei wurde sich zur Genauigkeitsbeurteilung auf die Mittelwerte aus allen zehn Kreuzvalidierungen bezogen. Die mittleren Genauigkeiten pro Klasse und über alle Klassen für jede angewandte Architektur sind in Tabelle 6 dargestellt.

**Tabelle 6: Ergebnisse der Patch-basierten Landbedeckungsklassifikation. Alle Ergebnisse beziehen sich auf die Berechnung der Genauigkeit nach Gleichung 5. Die in der Tabelle enthaltenen Werte geben die Mittelwerte der Klassengenauigkeiten über alle zehn Kreuzvalidierungen wieder. Außerdem werden in der letzten Spalte der Tabelle die Mittelwerte der Klassen gemittelt zur Berechnung der Gesamtgenauigkeit einer Architektur.**

	CLC 100	CLC 200	CLC 310	CLC 320	CLC 330	CLC 510	OA Acc
CNN	43,27 %	61,89 %	58,33 %	0,00 %	0,00 %	0,00 %	27,24 %
ResNet50	76,00 %	85,20 %	62,33 %	1,40 %	9,34 %	81,14 %	52,56 %
ResNet101	77,91 %	52,00 %	85,40%	0,00 %	7,51 %	7,12 %	38,32 %
ResNet152	73,26 %	76,02 %	81,45 %	2,37 %	4,12 %	52,23 %	48,24 %
ResNet50V2	81,14 %	77,37 %	83,29 %	5,54 %	3,92 %	69,62 %	53,31 %
ResNet101V2	82,12 %	81,33 %	92,86 %	11,77 %	55,32 %	69,87 %	65,54 %
InceptionRes-NetV2	83,89 %	80,69 %	76,50 %	5,80 %	2,10 %	82,50 %	55,25 %
sGAN	66,34 %	72,43 %	55,92 %	0,00 %	12,00 %	76,33 %	47,17 %

Die beste Testgenauigkeit wurde für das ResNet101V2 mit 65,54% erlangt (Tabelle 6). Die zweitbeste Testgenauigkeit wurde vom InceptionResNetV2 mit 55,25%, gefolgt vom ResNet50V2 mit 53,31% erzielt. Ebenfalls eine Testgenauigkeit von über 50% hat das ResNet50 erreicht. Alle vier Architekturen zeigten ebenfalls im Training eine gute Trainingsperformance. Alle anderen Architekturen erreichen eine Testgenauigkeit von unter 50%. Betrachtet man die Klassen einzeln, zeigt sich keine Architektur als dominant über alle Klassen. Über alle Klassen zeigen die drei im Beispieldatensatz überrepräsentierten Klassen einen deutlichen Testgenauigkeitsvorsprung gegenüber den unterrepräsentierten Klasse. Die überrepräsentierten Klassen werden mit einer Ausnahme für jede Architektur mit einer Genauigkeit von über 50% klassifiziert. Für die Klasse Künstliche Oberflächen zeigt das InceptionResNetV2 mit über 83% die beste Testgenauigkeit - dicht gefolgt vom ResNet101V2 und ResNet50V2. Alle anderen Architekturen zeigen

eine Testgenauigkeit von unter 80% für Künstliche Oberflächen. Die beste Testgenauigkeit für Landwirtschaftliche Flächen zeigt das ResNet101V2 mit 81,33%. Nur das InceptionResNetV2 zeigt für Landwirtschaftliche Flächen auch eine Testgenauigkeit von über 80%. Für die Klasse Wald hat das ResNet101V2 als einzige Architektur eine Testgenauigkeit von über 90% erreicht. Alle anderen Architekturen zeigen eine Testgenauigkeit von maximal 83%. Ebenfalls für die Klasse Strauchvegetation hat das ResNet101V2 mit einer Testgenauigkeit von 11,77% das beste Ergebnis erzielt. Das selbe gilt für die Klasse offene Flächen. Hier zeigt das ResNet101V2 mit einer Testgenauigkeit von 55,32% einen deutlichen Vorsprung gegenüber allen anderen Architekturen. Die beste Testgenauigkeit für Gewässer zeigt erneut das InceptionResNetV2 mit über 80%. Ebenfalls eine gute Klassifikationsgenauigkeit für Gewässer zeigen das ResNet50, mit ebenfalls über 80% und das sGAN mit rund 76%. Das ResNet101V2 zeigt für Gewässer eine Testgenauigkeit von rund 69%.

Insgesamt zeigt sich im Rahmen der genutzten Beispieldaten, dass das ResNet101V2 die beste Architektur zur Patch-basierten Bildklassifikation von KH-Bilddaten ist.

### **3.7 Anwendung der Architekturen**

In diesem Subkapitel werden die Ergebnisse der Datenprozessierung für die Interferenz, die Entscheidung der Interferenzarchitektur und die Ergebnisse aus der Durchführung der Interferenz aufgezeigt.

#### **3.7.1 Datenprozessierung**

Zur Durchführung der Interferenz mussten die GLCM-Texturmaße für alle vier angewandten Filtergrößen für alle vorliegenden KH-Bilddaten berechnet werden. Pro Szene entsteht in Abhängigkeit von ihrer Größe so ein Prozessierungsvolumen zwischen 80GB und 150GB. Der gesamte Prozessierungsumfang für die GLCM-Texturmaße über alle Zeitschnitte beträgt rund 45TB.

Zur Interferenz wurden alle Szenen zunächst in 2.304 x 2.304 Pixel große Bildpatches zerlegt. Anschließend wurde die Interferenz wie in Kapitel 2.8 erläutert für jede CORONA-Szene durchgeführt.

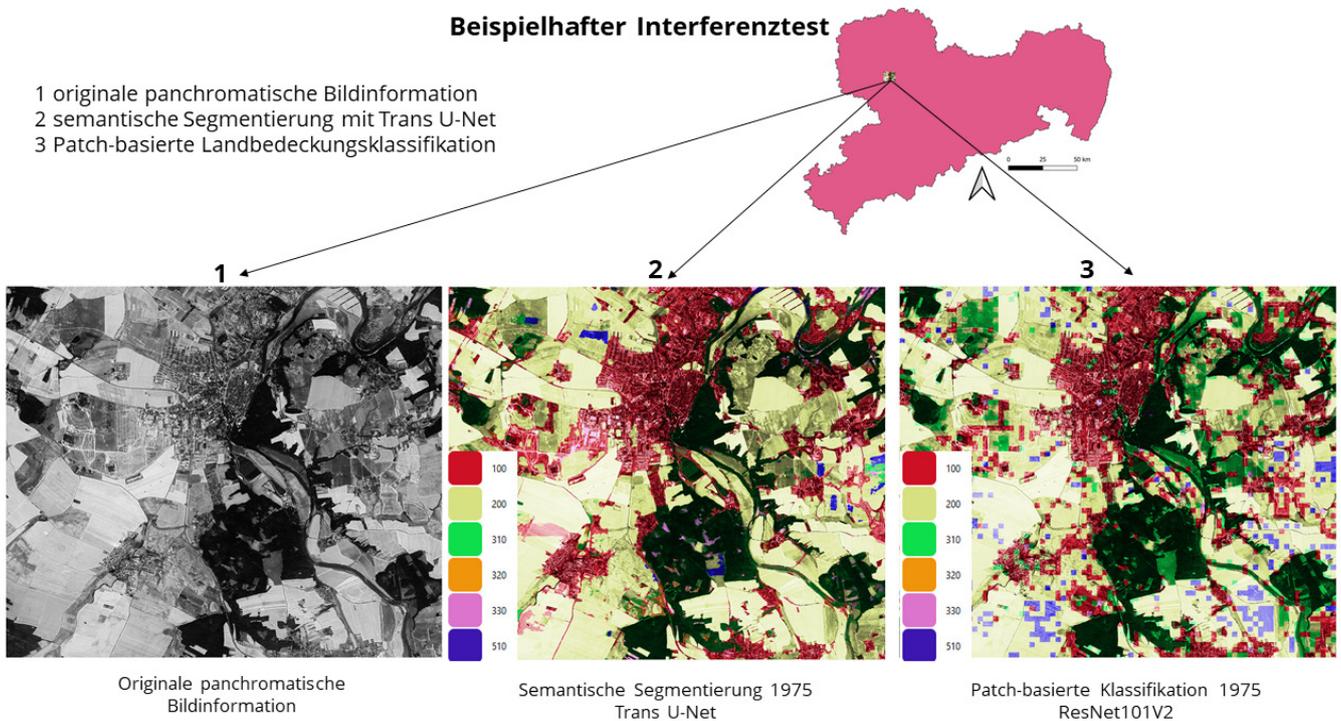
#### **3.7.2 Methodische Entscheidung zur Interferenz**

Bevor die Landbedeckung für ganz Sachsen für alle Zeitschnitte durchgeführt wurde, wurden Interferenztests für verschiedene Gebiete in Sachsen vorgenommen um die Leistung der Interferenz zu beurteilen. Die Interferenztests wurden für die Patch-basierte Landbedeckungsklassifikation mit dem ResNet101V2 getestet und für die semantische Segmentierung mit dem Trans U-Net.

Da für die methodische Interferenzentscheidung keine Testdaten vorliegen kann die Beurteilung nur visuell erfolgen. Im Weiteren wird für eine Beispielausschnitt von Sachsen mit KH-Bilddaten aus dem 1975 ein Interferenztest durchgeführt. Die Ergebnisse des Interferenztests für die semantische Segmentierung wird mit einer Auflösung von

2 x 2 m dargestellt, für die Patch-basierte Landbedeckungsklassifikation ergeben sich 100 x 100 m.

## Beispielhafter Interferenztest



**Abbildung 21: Beispielhafter Interferenztest für KH-Bilddaten aus dem Jahr 1975. (1) zeigt die originalen panchromatischen Bilddaten, (2) die Ergebnisse der semantischen Segmentierung und (3) die Ergebnisse der Patch-basierten Landbedeckungsklassifikation. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

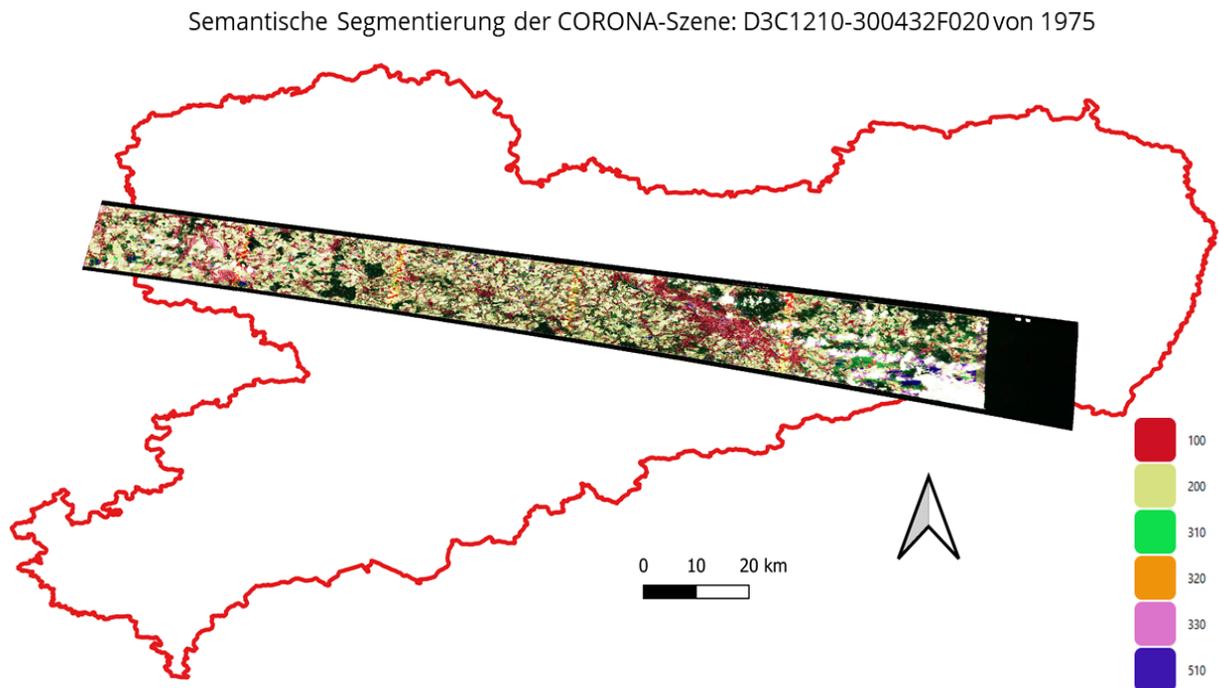
Aus Abbildung 21 ist zu erkennen, dass die Ergebnisse der Landbedeckung für die semantische Segmentierung (2) eine viel konsistentere Erscheinung haben. In den Ergebnissen zur Patch-basierten Landbedeckungsklassifikation (3) sind beispielweise auf den Landwirtschaftlichen Flächen immer wieder kleine Fehlklassifikationen zu erkennen. Insbesondere kommt es zu Fehlklassifikationen von einzelnen Bildpatches in die Klasse Künstliche Oberflächen und Gewässer. Ebenfalls ist zu erkennen, dass die Grenzen des Waldes in (3) sehr grob klassifiziert werden. Diese Beobachtungen konnten in mehreren Interferenztests über mehrere Zeitschnitte gleich beobachtet werden.

Neben der Möglichkeit eine viel höhere Auflösung für die Landbedeckungsklassifikation zu erzielen, zeigt die semantische Segmentierung auch konsistentere Segmentierungsergebnisse. Ein Nachteil die Landbedeckungsklassifikation als semantische Segmentierung durchzuführen liegt in der Prozessierungszeit. Die Vorhersage mit dem Trans U-Net dauert etwa dreimal so lang wie die Vorhersage mit dem ResNet101V2 für eine CORONA-Szene. Die Semantische Segmentierung benötigt zur Vorhersage zwischen 1 Stunde und 45 Minuten bis 3 Stunden, während die Vorhersage einer Szene mit Patch-basierter Bildklassifikation maximal eine Stunde benötigt. Der Grund ist insbesondere die Vorhersage für die Überlappungsbereiche und die damit verbundene Interpolation. Weiterhin kommt hinzu, dass das Trans U-Net komplexer aufgebaut ist und mehr Modellparameter hat als das ResNet101V2.

Im Rahmen der Bearbeitung des Projektes wurden die wirtschaftlichen Nachteile der semantischen Segmentierung zu gunsten der besseren Interferenzperformance vernachlässigt. Die Interferenz auf alle KH-Bilddaten über alle Zeitschnitte wurde mit einer semantischen Segmentierung und dem Trans U-Net durchgeführt.

### 3.8 Interferenz

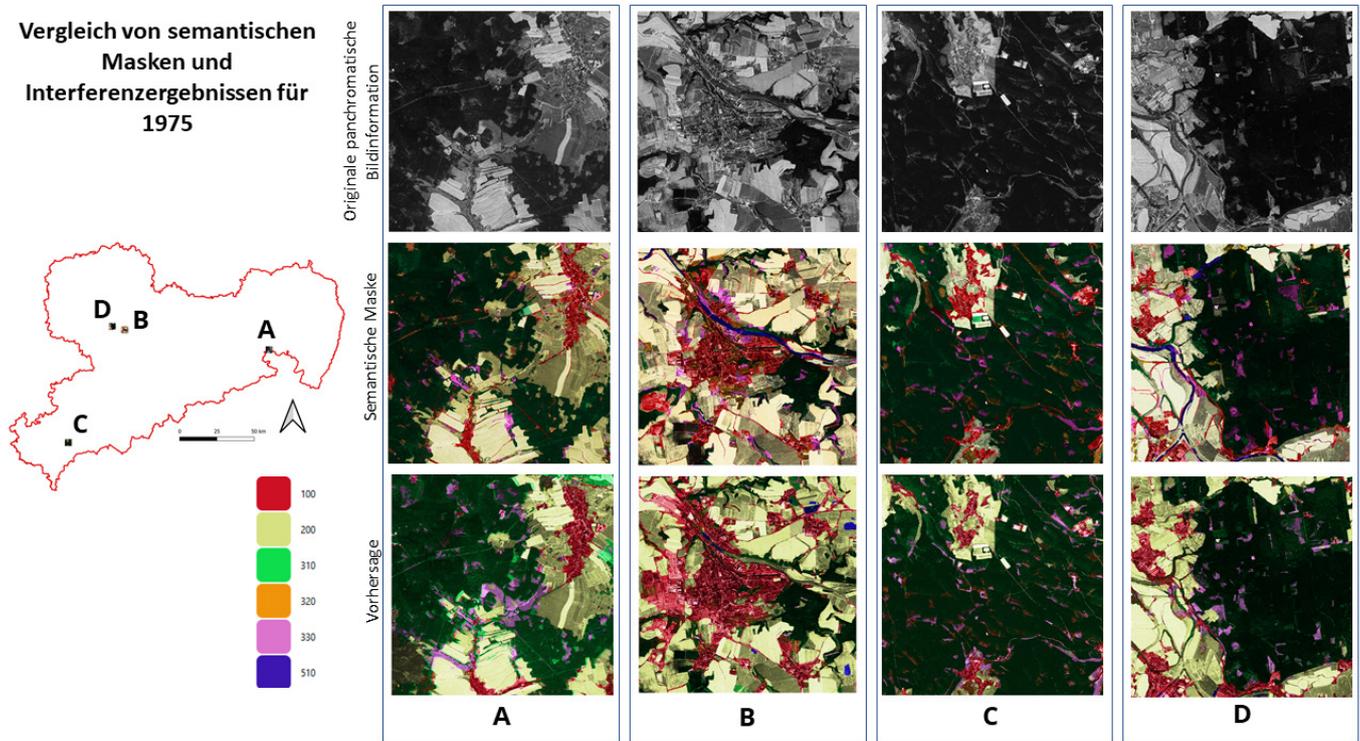
Die Interferenz wurde nach dem in Kapitel 2.8 erläuterten Interferenzschema durchgeführt. Beispielhaft wird in Abbildung 22 für die Interferenz über eine CORONA-Szene von 1975 dargestellt. Die Auflösung der Landbedeckungsklassifikation nach der Interferenz beträgt 2 x 2 m. Diese Interferenz wurde mit dem Trans U-Net durchgeführt.



**Abbildung 22: Semantische Segmentierung einer CORONA-Szene von 1975. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

Um die Güte der semantischen Segmentierung zu verstehen wurden die Vorhersagen des Trans U-Nets mit den semantischen Masken der Studiengebiete verglichen. Beispielhaft sind im Folgenden vier Studiengebiete von 1975 mit ihren korrespondierenden semantischen Masken und Vorhersagen abgebildet. Es ist an dieser Stelle darauf hinzuweisen, dass ein Teil dieser Daten für das Training und die Validierung verwendet wurden.

Vergleich von semantischen Masken und Interferenzergebnissen für 1975



**Abbildung 23: Vergleich von semantischen Masken und Interferenzergebnissen für 1975. Alle Studienggebiete korrespondieren zu Beispieldaten für den Zeitschnitt von 1975. Alle panchromatischen Bildinformationen und semantischen Karten haben eine Kantenlänge von 4.608 m. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

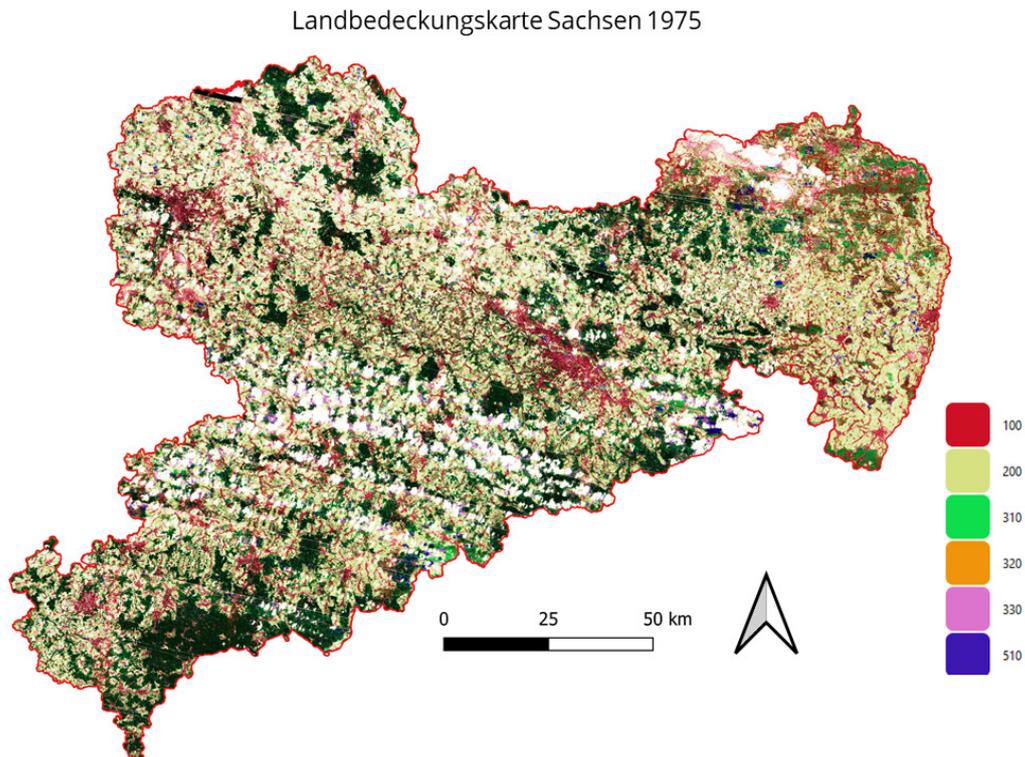
Aus Abbildung 23 ist für alle vier Beispielgebiete eine hohe Segmentierungsgenauigkeit zu erkennen. Insbesondere die drei dominanten Klassen Künstliche Oberflächen, Landwirtschaftliche Flächen und Wald zeigen eine sehr gute Übereinstimmung. Für die unterrepräsentierten Klassen sind vereinzelt Segmentierungsfehler zu erkennen. Beispielsweise zeigt Studienggebiet A im zentralen Bildbereich eine Übersegmentierung der Klasse offene Flächen. Ebenfalls im Norden von A wurde teilweise Landwirtschaftliche Fläche falsch als Wald semantisch segmentiert. Ein ähnlich kleines Beispiel lässt in C für eine kleine Fläche erkennen. Auffällig ist über die Interferenz von 1975 insgesamt, dass Flussverläufe und Seen häufig falsch klassifiziert werden. Diese Problematik ist in B veranschaulicht. Dort wurde nur ein Teil des Flussverlaufes richtig semantisch segmentiert. Es zeigt sich in dem Beispiel und auch in der gesamten Interferenzkarte von 1975 eine systematische Verwechslung der Gewässer mit der Klasse Wald. Die Zuverlässigkeit wird außerdem durch Fehler in den semantischen Masken beeinflusst. In A wurde westlich von der künstlichen Oberfläche eine Landwirtschaftliche Fläche falsch als Wald gekennzeichnet. Außerdem wurden Strauchflächen und offene Flächen im Wald von A mehrfach falsch als Landwirtschaftliche Fläche in der semantischen Maske gekennzeichnet. In B ist zu erkennen, dass der Flussverlauf auch im südlichen Teil des Studienggebietes nicht durchgängig richtig gekennzeichnet wurde. In C ist in der semantischen Maske eine Landwirtschaftliche Fläche, welche falsch als Wald gekennzeichnet wurde, zu erkennen. Diese wurde in der Vorhersage richtig als Landwirtschaftliche Fläche gekennzeichnet. Außerdem ist in der semantischen Maske von C eine Künstliche Oberfläche innerhalb des Waldes zu groß gekennzeichnet wurden. Die Vorhersage zeigt eine konsistentere Auflösung und grenzt die Ortschaft deutlicher von den umliegenden Landwirtschaftlichen Flächen ab. Ein weiteres solches Beispiel ist in D zu erkennen. Dort wurde im Norden der semantischen Maske eine Gewässerfläche gekennzeichnet. In der Vorhersage wurde diese Fläche richtig dem Wald und der Strauchvegetation zugeordnet. Solche Fehler in den semantischen Masken führen zu einer Verringerung der Zuverlässigkeit der Testergebnisse.

### 3.9 Nachprozessierung

In diesem Subkapitel werden die Ergebnisse aus den Schritten der Nachprozessierung beschrieben.

#### 3.9.1 Kombination der Interferenzergebnisse

Wenn die Interferenz für alle Szenen für einen Zeitschnitt abgeschlossen ist erfolgt die Nachprozessierung. Dazu wurden im ersten Schritt die semantischen Masken aller CORONA-Daten zu einer Landbedeckungskarte zusammengeführt. Beispielhaft wird diese Landbedeckungskarte für das Jahr 1975 in Abbildung 24 dargestellt. Die bewölkten Bereiche der CORONA-Szenen werden in diesem Prozessierungsschritt noch als Teil der Landbedeckung klassifiziert und sind noch nicht gekennzeichnet. Überwiegend werden bewölkte Bereiche als Landwirtschaftliche Fläche gekennzeichnet.



**Abbildung 24: Landbedeckungskarte Sachsen 1975. Die dargestellte Landbedeckungskarte hat eine Auflösung von 2 x 2 m und teilt die Landbedeckung in sechs Klassen. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

#### 3.9.2 Prozessierung der Zielauflösung der Landbedeckung

Im letzten Schritt der Nachprozessierung werden die Landbedeckungskarten pro Zeitschnitt in die Zielauflösung von 100 x 100 m überführt und an das Referenz-Raster des LfULG angeglichen. Umgesetzt wurden diese Schritte nach den in 2.9 beschriebenen Methoden. Abbildung 25 und Abbildung 26 zeigen die Unterschiede in den Ergebnissen zwischen den Auflösungen und die Konsistenz der Ergebnisse mit dem projektbezogenen Referenzraster. In Abbildung 25 wird durch einen Beispielausschnitt die Änderung der Landbedeckungsauflösung von 2 x 2 m zu 100 x 100 m dargestellt.

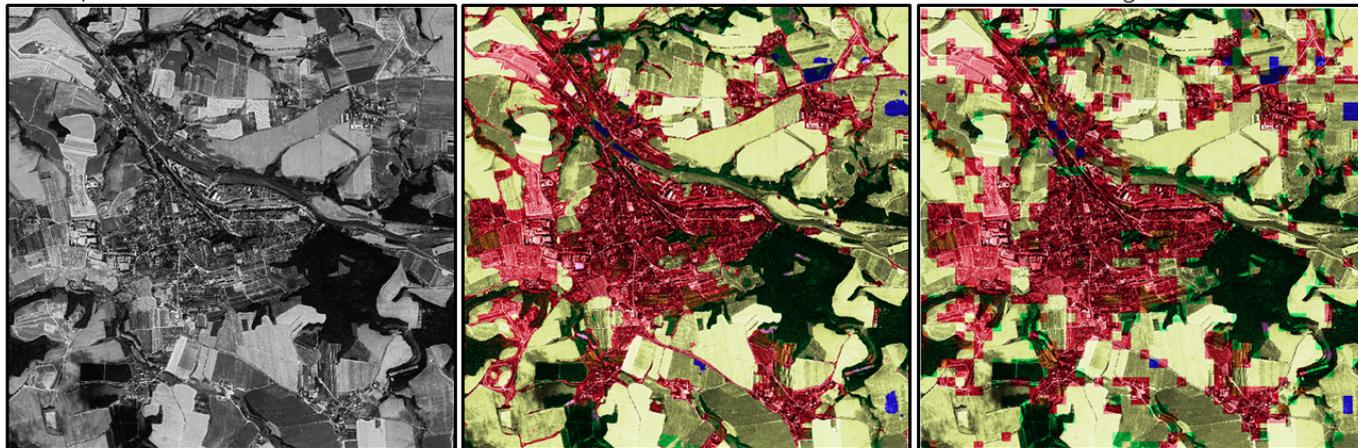
**Nachprozessierung der semantischen Karten: Herstellung der Zielauflösung von 100 x 100m**



A: panchromatische Bildinformation

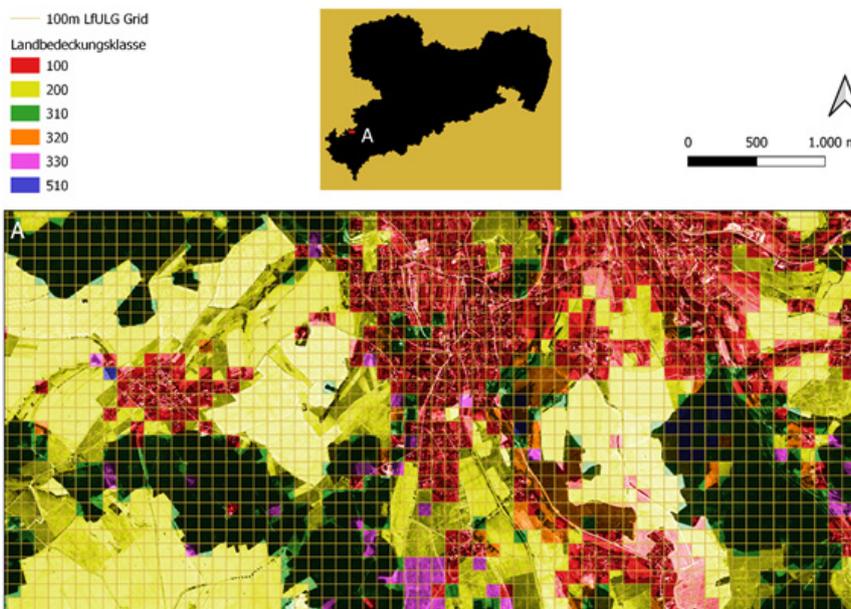
A: Vorhersage semantische Segmentierung 2x2m

A: Vorhersage 100 x 100m



**Abbildung 25: Veranschaulichung der Nachprozessierung der semantischen Karten für 1975 von 2 x 2 m auf 100 x 100 m. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

**Rasteranpassung auf Zielauflösung**

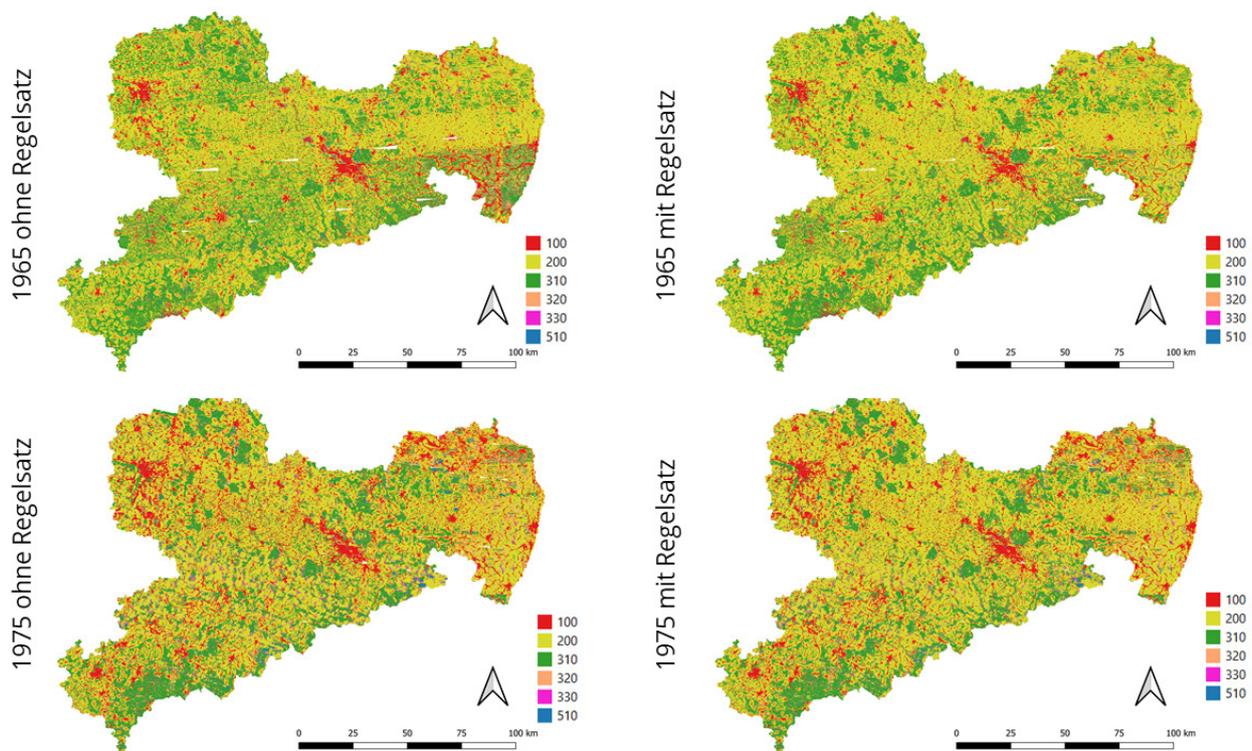


**Abbildung 26: Veranschaulichung der Anpassung der semantischen Karten auf das LfULG-Referenzraster mit 100 x 100 m Auflösung. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

Ein Resampling von 2 x 2 m auf 100 x 100 m stellt zwar einen Informationsverlust dar, beseitigt allerdings gleichermaßen kleinräumliche Ungenauigkeiten in der Klassifikation. Beide Fälle sind in den oben gezeigten Abbildungen gut zu erkennen. Die Anpassung an ein vorgegebenes Referenz-Raster stellt zudem eine direkte Kompatibilität mit anderen Produkten sicher.

### 3.9.3 Anwendung des Regelsatzes

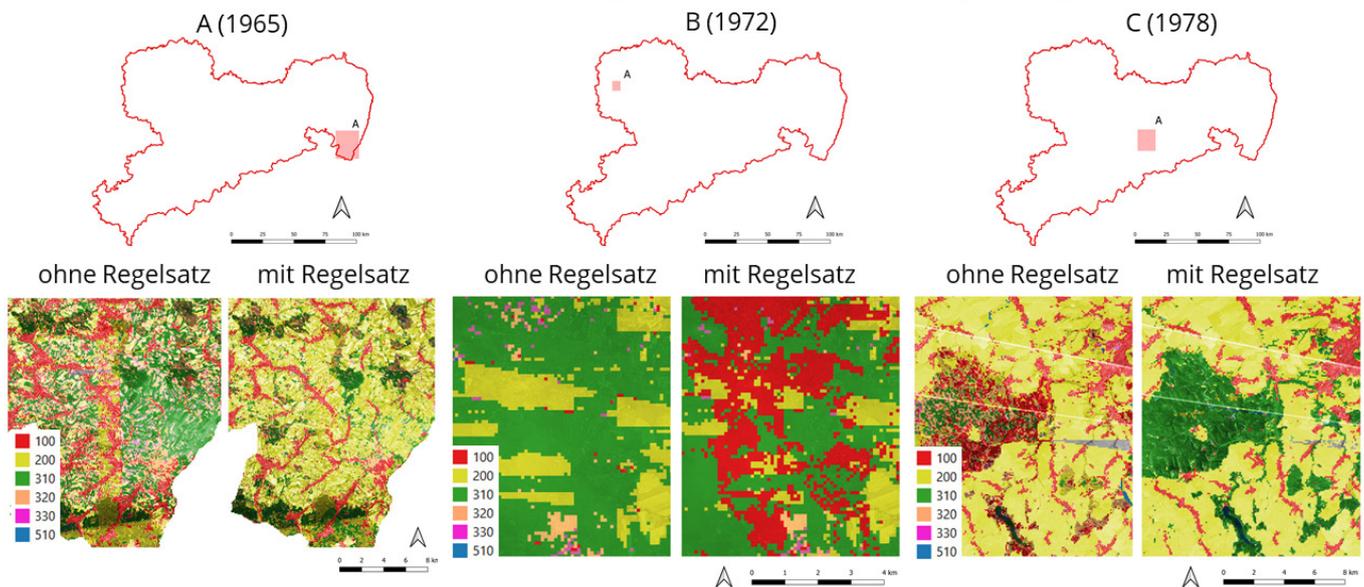
Wie in Kapitel 2.10 beschrieben, wurden die finalen Landbedeckungskarten schlussendlich einer Konsistenzprüfung nach einem genau definierten Regelsatz unterzogen. Ausreißer sollen somit beseitigt und unlogische Klassenänderungen vermieden werden. Außerdem sollen Datenlücken bestmöglich gefüllt werden. Nach Anwendung dieses Regelsatzes sind in allen Jahren Verbesserungen in der Landbedeckungsklassifikation zu erkennen, in manchen sogar deutlich. In Abbildung 27 werden beispielhaft die Auswirkungen der Konsistenzprüfung für die Jahre 1965 und 1975 gezeigt.



**Abbildung 27: Auswirkung der Anwendung des Regelsatzes auf die Klassifikationsergebnisse der Jahre 1965 und 1975. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

Während an manchen Stellen lediglich einzelne Pixelwerte durch den Regelsatz geändert werden, sind an anderen Stellen in einigen Jahren auch großflächige Unterschiede zu erkennen. In Abbildung 28 werden drei solcher Fälle aufgezeigt.

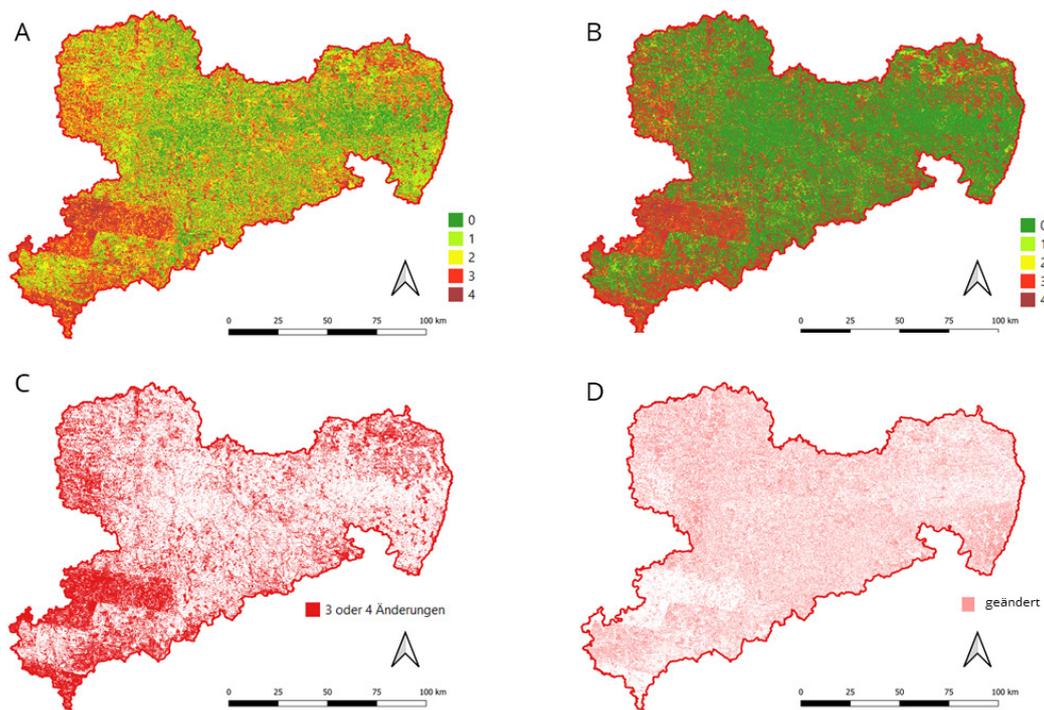
Beispielausschnitte für Vorhersagen ohne und mit Anwendung des Regelsatzes



**Abbildung 28: Beispiele für auffällige Änderungen in den Klassifikationsergebnissen ohne und mit Anwendung des Regelsatzes. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

Es wird deutlich, dass der Regelsatz diverse Schwachstellen der bisherigen Prozessierungsschritte, resultierend aus unterschiedlicher Bildqualität, Güte der Trainingsdaten oder Wolkenbedeckung beseitigen kann. Selbst in Fällen bei denen die Klassifikation aufgrund von schlechter Bildqualität komplett fehlschlug, führt die Anwendung des Regelsatzes zu einer zumindest leichten Verbesserung (siehe Beispiel B in Abbildung 28).

Um einen Überblick über die durch den Regelsatz entstandenen Anpassungen zu bekommen, wurden diverse Übersichts- und Flag-Karten erzeugt, die in Abbildung 29 aufgezeigt sind.



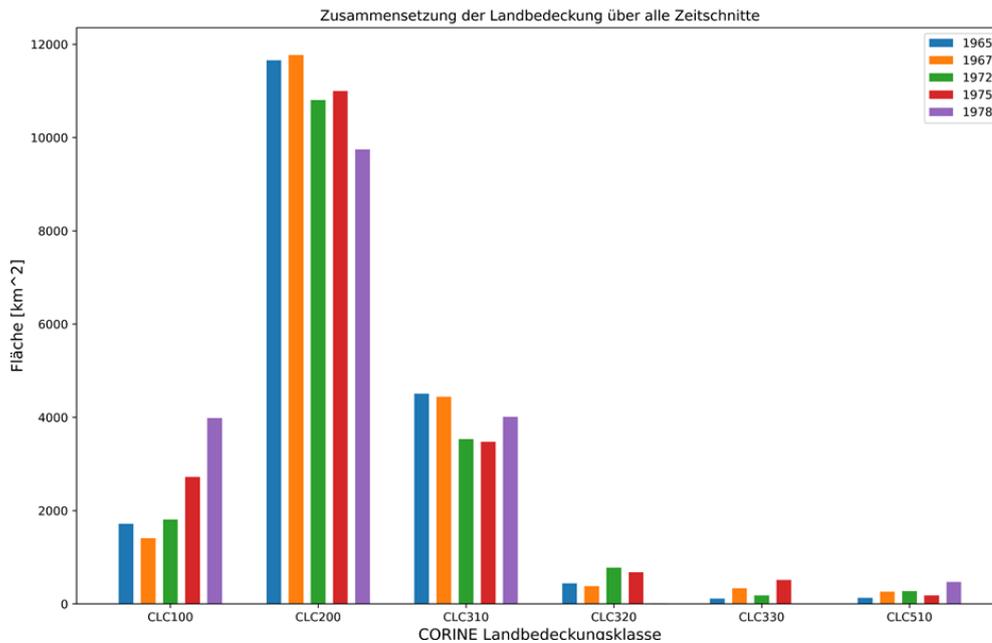
**Abbildung 29: Übersicht über Anzahl der Landbedeckungsänderungen über den gesamten Zeitraum (A) vor und (B) nach Anwendung des Regelsatzes und Flag-Karten von (C) unzulässigen Pixeln mit 3 oder mehr Landbedeckungsänderungen im Untersuchungszeitraum und (D) Pixeln die im Zuge der Anwendung des Regelsatzes eine Änderung an mindestens einer Stelle in der Zeitreihe erfahren haben.**

Mit Blick auf die Anzahl der Änderungen über den gesamten Beobachtungszeitraum jeweils vor und nach Anwendung des Regelsatzes (siehe Abbildung 29A und B), wird deutlich, dass nach der Anpassung bedeutend weniger Landbedeckungsänderungen über die Zeit verbleiben. Ebenso wird ersichtlich, dass es räumliche Unterschiede gibt. Viele Änderungen treten vor allem in der Lausitz, in der Region Leipzig und in (Süd-)Westsachsen auf, wobei diese bei letzteren in vielen Fällen auf Lücken im Datensatz oder unzureichender Bildqualität zurückzuführen sind. Wie in Kapitel 2.10 beschrieben, wurden solche Pixel vom Regelsatz nicht beachtet, wenn sie drei oder mehr Änderungen über die gesamten fünf Zeitschnitte erfahren. Diese wurden stattdessen als unzuverlässige Pixel geflaggt (siehe Abbildung 29C). Schlussendlich zeigt die Übersicht der durch den Regelsatz angepassten Pixel, dass eine Mehrheit der zuverlässigen Pixel in mindestens einem Jahr eine Änderung der Klassifikation durch den Regelsatz erfahren hat (siehe Abbildung 29D).

### 3.10 Analyse der Landbedeckung

In diesem Subkapitel werden die Ergebnisse der Landbedeckungskarten hinsichtlich ihrer Klassenverteilung pro Zeitschnitt und der Veränderung über die Zeit analysiert. Die Analyse der Landbedeckung bezieht sich auf die nachprozessierten Daten infolge der Anwendung des Regelsatzes aus Kapitel 3.9.3

In Abbildung 30 ist die Zusammensetzung der Landbedeckung über alle Zeitschnitte dargestellt.



**Abbildung 30: Zusammensetzung der Landbedeckung über alle Zeitschnitte.** Das Balkendiagramm zeigt entlang der x-Achse die CORINE Zielklassen. Über die y-Achse wird der Flächenanteil der Klasse abgebildet. Für jede Klasse gibt es fünf Balken. Jeder Balken repräsentiert einen Zeitschnitt.

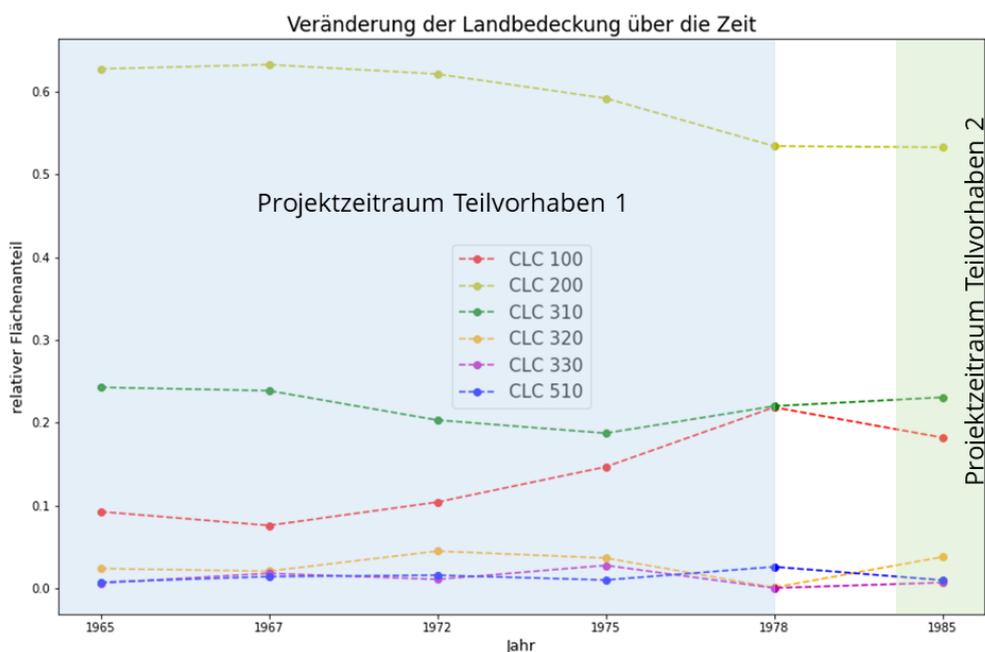
Aus dieser Abbildung ist zu erkennen welche Klassen am häufigsten und wenigsten in Sachsen zwischen 1965 und 1978 auftreten. Es wird deutlich, dass die drei überrepräsentierten Klassen im Beispieldatensatz auch die häufigsten Klassen der vorhergesagten Landbedeckung sind. Den mit Abstand größten Flächenanteil über alle Zeitschnitte bilden die Landwirtschaftlichen Flächen. Landwirtschaftliche Flächen wurden in allen Zeitschnitten mit einer Fläche von mindestens  $9.000 \text{ km}^2$  bestimmt. Es ist dabei zu erkennen, dass der Anteil der Landwirtschaftlichen Fläche in den Jahren 1965 und 1967 am größten ist. Den zweitgrößten Anteil der Landbedeckung bildet die Klasse Wald. Aus Abbildung 30 ist zu erkennen, dass die meisten Waldflächen in den Jahren 1965 und 1967 vorhergesagt wurden. Dort waren noch über  $4.000 \text{ km}^2$  der Fläche von Sachsen von Waldflächen bedeckt. In den restlichen Zeitschnitten zeigt der Wald einen Flächenanteil von rund  $3.400\text{-}4.400 \text{ km}^2$ . Die urbanen Flächen zeigen im Untersuchungszeitraum einen Flächenanteil zwischen rund  $1.400 \text{ km}^2$  und rund  $4.000 \text{ km}^2$ . Die unterrepräsentierten Klassen Strauchvegetation, offene Flächen und Gewässer zeigen durchschnittlich über alle Zeitschnitte einen Flächenanteil von weniger als  $1.000 \text{ km}^2$ . Die Flächen haben damit nur einen Anteil von knapp über 5% der Gesamtfläche von Sachsen. Die Strauchvegetation hat im Zeitschnitt von 1978 eine Fläche von  $1.467,00 \text{ km}^2$ . Das ist der größte gemessene Anteil Strauchvegetation im Untersuchungszeitraum. Der durchschnittliche Anteil der offenen Flächen liegt knapp unter der Strauchvegetation. Der größte Anteil offener Flächen wurde im Zeitschnitt von 1975 gemessen mit  $510 \text{ km}^2$ . Der geringste Anteil Strauchvegetation wurde für 1965 bestimmt. Gewässerflächen haben den kleinsten Flächenanteil in Sachsen. Im Maximum beträgt ihre Fläche  $470 \text{ km}^2$ . Dieser Wert wurde im Zeitschnitt von 1978 gemessen. Der geringste Anteil Gewässerflächen wurde im Jahr 1965 mit  $112,88 \text{ km}^2$  verzeichnet.

In der nachfolgenden Tabelle werden die absoluten Flächenanteile pro Klasse und Zeitschnitt aufgezeigt. Es gilt hierbei zu beachten, dass nicht klassifizierte Flächen aufgrund fehlender Daten nicht berücksichtigt werden konnten.

**Tabelle 7: Flächenanteile pro Klasse und pro Zeitschnitt in Quadratkilometer**

CLC Klasse	1965 [ $km^2$ ]	1967 [ $km^2$ ]	1972 [ $km^2$ ]	1975 [ $km^2$ ]	1978 [ $km^2$ ]
CLC 100	1.716,78	1.409,37	1.808,11	2.722,43	3.986,50
CLC 200	11.655,14	11.768,32	10.805,03	10.996,77	9.744,30
CLC 310	4.507,28	4.438,45	3.532,15	3.478,38	4.013,06
CLC 320	439,33	380,81	779,16	676,26	1.467,00
CLC 330	112,88	334,01	185,03	510,61	315,00
CLC 510	132,60	260,31	272,45	182,37	470,26

In der nächsten Abbildung wird die Veränderung der Landbedeckung pro Klasse über die Zeit abgebildet.



**Abbildung 31: Veränderung der Landbedeckung über die Zeit. Die x-Achse beschreibt den zeitlichen Verlauf. Die y-Achse gibt den Flächenanteil wieder. Für jede Zielklasse ist ein Graph abgebildet welcher die Veränderung der Landbedeckung angibt. Die gestrichelten Linien sollen die Erkennbarkeit von Veränderungen unterstützen, nicht aber einen linearen Zusammenhang implizieren. In der Abbildung werden die Ergebnisse von Teilvorhaben 1 und 2 zusammengeführt.**

Abbildung 31 zeigt den relativen Flächenanteil pro Klasse über die pro Jahr analysierte Gesamtfläche. Wie bereits in Abbildung 30 angedeutet bilden Landwirtschaftliche Flächen den größten Anteil an der Gesamtfläche über alle Zeitschnitte. Über die Zeit ist eindeutig eine Abnahme der Landwirtschaftlichen Flächen zu verzeichnen. Während Landwirtschaftliche Flächen im Zeitschnitt von 1965 noch mehr als 60% der Gesamtfläche ausmachen, beträgt ihr Anteil 1978 nur noch rund 53%. Für die Klasse Wald ist ebenfalls eine Flächenabnahme über die Zeit zu verzeichnen. Im Zeitschnitt 1965 betrug der Anteil des Waldes noch rund 24% in Sachsen, im Zeitschnitt von 1978 aber nur noch rund 21%. Der geringste Flächenanteil für den Wald wurde im Zeitschnitt von 1975 mit rund 19% gemessen. Für Künstliche Oberflächen ist eine starke Flächenzunahme über die Zeit verzeichnet. Während der Flächenanteil 1965 nur rund 10% beträgt, liegt er am Ende des Untersuchungszeitraums 1978 bei rund 20%.

Damit hat sich der Anteil der Künstlichen Oberflächen in 12 Jahren verdoppelt. Der geringste Anteil Künstlicher Oberflächen wurde im Zeitschnitt von 1967 mit 8,6% verzeichnet. Die Strauchvegetation zeigt keinen eindeutigen Trend wie die drei überrepräsentierten Klassen. Der größte Flächenanteil für Strauchvegetation wurde im Zeitschnitt von 1972 mit einem Flächenanteil von 0,4% gemessen. Der geringste Anteil wurde im Zeitschnitt von 1978 mit 0,08% verzeichnet. Die Klasse offene Flächen zeigt ebenfalls keinen eindeutigen Veränderungstrend. Der größte Flächenanteil für offene Flächen wurde 1975 mit einem relativen Gesamtflächenanteil von rund 0,03% bestimmt. Der geringste Wert wurde im Zeitschnitt 1978 gemessen mit einem Anteil von 0,02%. Die Gewässerflächen zeigen wie die beiden anderen unterrepräsentierten Klassen Schwankungen im Untersuchungszeitraum. Während im Zeitschnitt von 1975 eine Abnahme der Gewässerflächen verzeichnet wird, nimmt der Anteil der Gewässerflächen über alle anderen Zeitschnitte chronologisch geordnet zu. Es ist somit eine geringe Zunahme der Gewässerflächen zu verzeichnen. Der geringste Anteil Gewässerflächen wurde im Jahr 1965 mit 0,6% gemessen. Im Jahr 1978 lag der Anteil der Gewässerflächen bei rund 2,6%.

## 4 Diskussion

In dem Kapitel der Diskussion wird auf verschiedene projektbezogene Teilaufgaben eingegangen. Insbesondere werden dabei die Ergebnisse der einzelnen Teilschritte diskutiert und deren Auswirkung auf das Endergebnis des Projektes geschildert.

### 4.1 Georeferenzierung

Der Prozess des Verstehens des Unterschieds zwischen den modernen Orthofotos und den georeferenzierten Bildern unterliegt menschlichem Versagen, d.h. der Benutzer muss entscheiden, wo die Mitte der Straße liegt, wenn er einen Datensatz mit den auf dem georeferenzierten Bild gezeigten Orten erstellt. Die Auflösung von 2 m und die Grauskala des Bildes erfordern ein gewisses Maß an Urteilsvermögen auf Seiten des Benutzers.

Diese Methode beruht auch darauf, dass die modernen Orthofotos und das ursprüngliche Straßen-/Schienen-Geometriedatei übereinstimmen. Ein Beispiel für ein solches Problem ist in Abbildung 32 dargestellt. Die Bilder wurden in erster Linie anhand des Straßen- und Schienendatensatzes an die heutige Landschaft angepasst, sodass sich etwaige Unterschiede zwischen den beiden in der Genauigkeit der Georeferenzierung widerspiegeln.



**Abbildung 32: Modernes DOP vom GeoSN mit Daten des Straßennetzes von OpenStreetMap**

Ferner wird die Problematik der Georeferenzierung in Bezug auf die Übereinstimmung der CORONA-Bilddaten mit den modernen Daten des Straßen- und Schienennetzes in einer weiteren Abbildung verdeutlicht.



**Abbildung 33: Überlagerung der modernen Daten des Straßen und Schienennetzes von OpenStreetMap und CORONA-Bilddaten.**

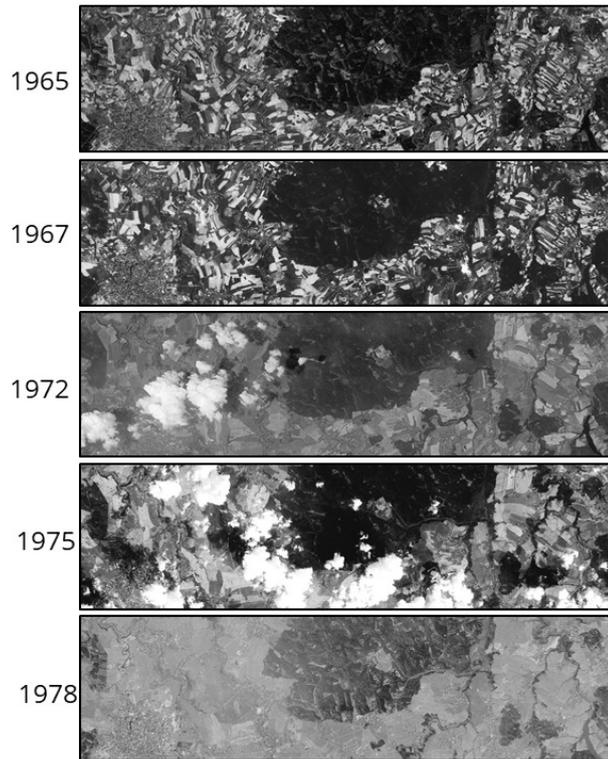
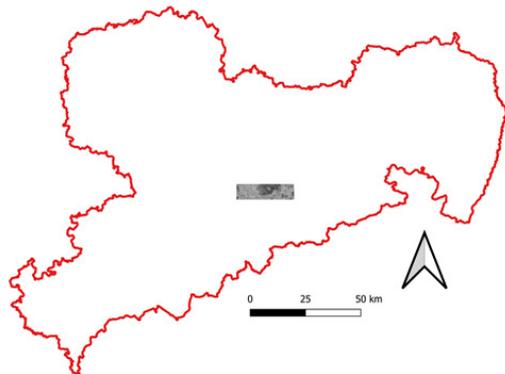
Aus Abbildung 33 geht hervor, dass im Untersuchungszeitraum viele Wege und Straßen noch nicht vorhanden waren, aber auch das Straßen umgelegt oder verschoben wurden. Insbesondere letzteres erschwert die Georeferenzierung da ein Neubau von einer veralteten Straße unterschieden werden muss. Ebenfalls wird die Georeferenzierung erschwert in Gebieten in denen keine Straßen oder Schienen vorhanden waren. Dies tritt insbesondere im landwirtschaftlich geprägten ländlichen Raum auf.

Insgesamt stellte sich die Durchführung der Georeferenzierung als äußerst langwieriger Prozess heraus, der eine große Menge an Arbeitskraft erfordert. Durch die Größe der Szenen und den aufgrund der Aufnahmetechnik enthaltenen Deformationen sind pro CORONA-Szene bis zu über 3.000 Referenzpunkte zu setzen. Hügeliges Terrain und schlechte Bildqualität können diesen Prozess erschweren und führen dazu, dass kleinere Deformationen an einzelnen Stellen im Bild verbleiben, trotz einer ansonsten großen Zahl an Referenzpunkten. Der immense Arbeitsaufwand ist schlussendlich auch der Grund für Datenlücken an Stellen an denen bis zum Ende des Projekts keine Georeferenzierung für CORONA-Szenen mehr durchgeführt werden konnte (siehe 1972).

## 4.2 Beispieldatensatz

Die wichtigste Grundlage zum Gelingen eines Deep Learning Projektes ist ein repräsentativer Beispieldatensatz welcher die Trainingsdaten, Testdaten und Validierungsdaten liefert. Die CORONA-Daten zeigen unter anderem aufgrund der unterschiedlichen Sensortechnologien über die Zeit eine unterschiedliche Erscheinung. Diese Unterschiede werden in Abbildung 34 beispielhaft für ein Gebiet in Sachsen gezeigt. Es ist zu erkennen, dass jeder Zeitschnitt die Landbedeckung anders zeigt. Weitere Einflüsse auf die Erscheinung der Bilder haben die Tageszeit der Aufnahme, die Jahreszeit und der Sonnenstand.

### Beispielausschnitt aus den KH-Bilddaten über alle Zeitschnitte



**Abbildung 34: Beispielausschnitt aus den KH-Bilddaten über alle Zeitschnitte.**

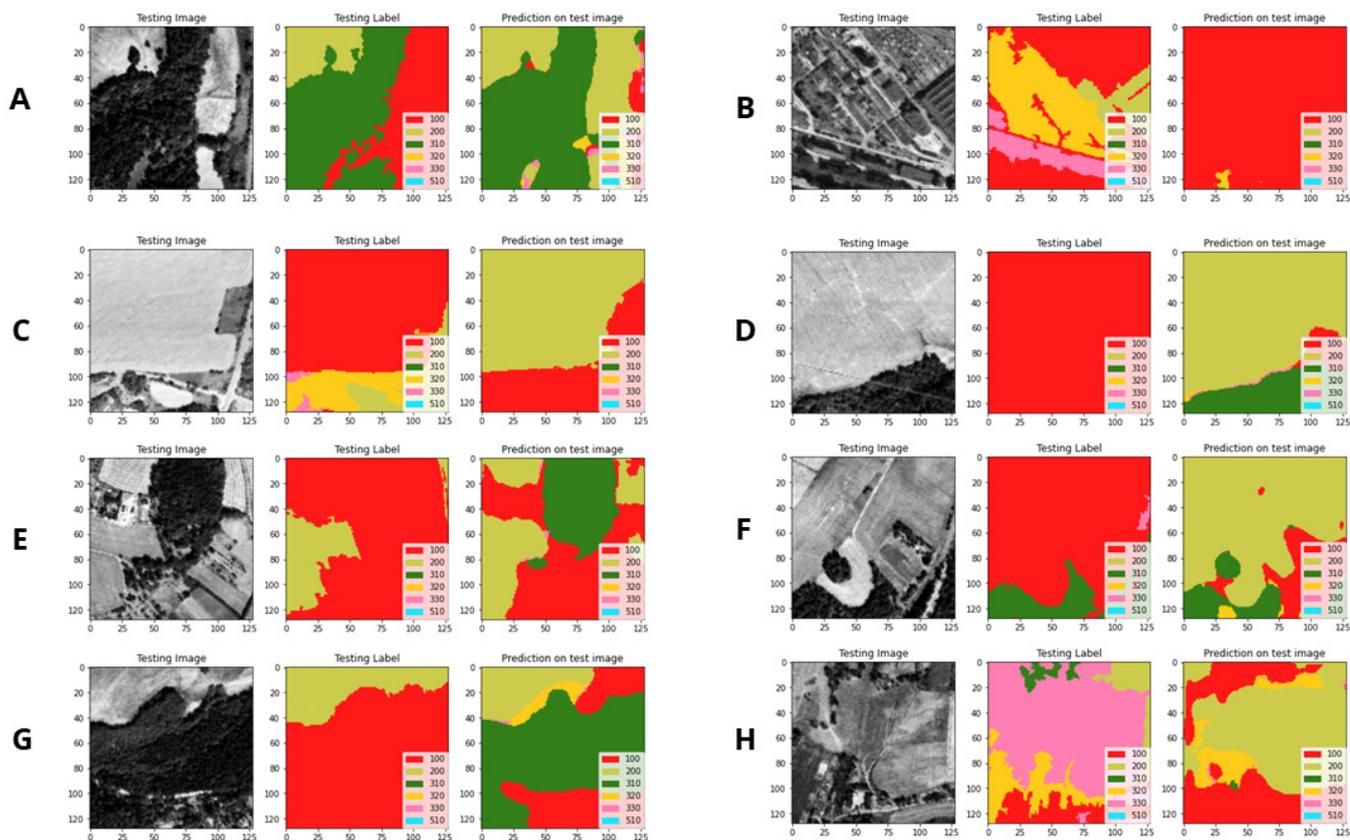
Aufgrund dieser Diversität der Daten bestand eine Herausforderung des Projektes darin einen genügend großen, diversen und repräsentativen Trainingsdatensatz aufzubauen. Dazu wurden aus jedem Zeitschnitt Studienggebiete aus unterschiedlichen Regionen Sachsens ausgewählt. Bei der Auswahl wurde der Fokus auf eine diverse Erscheinung der KH-Bilddaten gelegt, zum Nachteil einer idealen räumlichen Verteilung der Daten. Die Erstellung der semantischen Masken der Daten wurde teilweise von Studierenden der TU Dresden vorgenommen. Im Rahmen der Projektlaufzeit war es nicht möglich einen ausreichend großen Datensatz aufzubauen, der für alle Zeitschnitte genügend viele und diverse Beispieldaten beinhaltet. Um diesem Problem entgegen zu wirken wurden Methoden der Datenerweiterung angewandt. Diese Methoden können das Problem der Unterdiversifizierung allerdings nicht vollständig beheben.

Eine weitere Schwierigkeit im Aufbau des Datensatzes besteht darin, dass ein idealer Datensatz eine gleichverteilte Menge Trainingsdaten über alle Klassen besitzt. Aufgrund der Tatsache, dass die Landbedeckung nicht gleichmäßig verteilt ist kann für eine semantische Segmentierung der Landbedeckung kein gleichverteilter Datensatz aufgebaut werden. Um diesem Problem entgegen zu wirken wurden gewichtete Verlustfunktionen und Verlustensembles eingesetzt. Im Rahmen der Projektentwicklung wurde erkannt, dass diese Methoden dem Problem entgegenwirken indem sich die Testgenauigkeit für unterrepräsentierte Klassen vergrößert hat. Die resultierenden Probleme des enormen Ungleichgewichts können allerdings durch diese Methoden nicht vollständig behoben werden.

Ebenfalls muss die Größe des Beispieldatensatzes diskutiert werden. Der Standarddatensatz zur Landbedeckungsklassifikation DeepGlobe hat beispielsweise über 500.000 gekennzeichnete Bildpatches. Der im Projekt aufgebaute Datensatz ist etwas kleiner als ein Zehntel davon. Insbesondere in Kombination mit dem Problem der Diversität der Daten war es eine besondere Herausforderung gute Ergebnisse für die semantische Segmentierung zu erzielen. Aufgrund des Mangels von Beispieldaten für verschiedene Zeitschnitte und Regionen waren bereits nach dem Aufbau des Beispieldatensatzes größere Ungenauigkeiten

für verschiedene Zeitschnitte und Regionen zu erwarten. Auf Methoden der Überabtastung wurde im Rahmen des Projektes bewusst verzichtet um die Testergebnisse so wenig wie möglich zu beeinflussen.

In Kapitel 3.6 zur Interferenz wurde auf einzelne Fehler in den semantischen Masken hingewiesen. Diese Fehler führen zu zwei Problemen. Zum einen werden für Klassen falsche Merkmale durch die Architekturen gelernt, welche dann wiederum zu falschen Vorhersagen auf unbekannte Interferenzdaten führt. Weiterhin werden die Testergebnisse unzuverlässiger, weil häufig angeblich falsche Vorhersagen der Netzwerke auf eine falsche Kennzeichnung in den semantischen Masken zurückgeführt werden kann. Häufig ist in den Testdaten zu erkennen, dass die Netzwerke Flächen trotz der Verschiebung im Netzwerk richtig segmentieren. Veranschaulicht wird diese Problematik durch Abbildung 35.



**Abbildung 35: Beispielpatches mit den korrespondierenden semantischen Masken und den Vorhersagen des Trans U-Nets. Die Abbildungen A bis H sollen auf Fehler in den semantischen Masken aufmerksam machen. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

In allen Subabbildungen von Abbildung 35 sind Fehler in den semantischen Masken zu erkennen. In einigen Fällen gibt jedoch das Trans U-Net eine wahre Vorhersage der Klasse, welche in den Trainingsdaten falsch klassifiziert wurde und damit die Testergebnisse negativ beeinflusst. In A ist beispielsweise in der semantischen Maske ein Streifen als Künstliche Oberflächen markiert. In der Vorhersage wurde dieser richtig als Landwirtschaftliche Fläche gekennzeichnet. Außerdem wurden kleine Fläche im Wald(-rand) richtig als Strauchvegetation und offene Fläche gekennzeichnet. Akuter werden falsche Kennzeichnungen von Künstlichen Oberflächen in C, D, F und G gezeigt. Dort wurden große Flächen als Künstliche Oberfläche markiert, welche andere Landbedeckungsklassen zeigen. In all diesen Subabbildungen zeigt das Trans U-Net wahre Vorhersagen. In B und H wurden Flächen der Strauchvegetation und offene Flächen in den

semantischen Masken falsch gekennzeichnet. Das Trans U-Net zeigt in B die Flächen richtig als Künstliche Oberflächen segmentiert. In H wurden die entsprechenden Flächen richtig als Landwirtschaftliche Flächen gekennzeichnet. Ähnliche Fehler in unterschiedlichen Ausmaßen sind in fast allen Studiengebieten zu finden.

### **4.3 Methode**

Es soll hier nicht in Frage gestellt werden, ob die Anwendung von Deep Learning für die Herausforderung der Studie sinnvoll ist. Vielmehr gilt es anzumerken, dass Deep Learning Projekte konzeptionell enorm durch die menschliche Fehlbarkeit beeinflusst werden. Die menschliche Fehlbarkeit steckt in allen Entscheidungen eines Deep Learning Projektes. Begonnen bei der Datenrecherche und Datenauswahl, dem Erstellen von Referenzdaten, der Auswahl der Zielmetriken und der Modellentwicklung, welche letztlich das bisherige Verständnis der Daten durch den Menschen spiegelt (Dillon & Collett, 2019). Es kann damit nie ausgeschlossen werden, dass die getroffenen Entscheidungen ein glaubhaftes Ergebnis zeigen, welches aber aufgrund falscher Entscheidungen nicht die Realität abbildet. Besonders kritisch ist es, wenn Architekturen zur Übertragung oder Interferenz genutzt werden, da so das Risiko zur Reproduktion von Fehlern erhöht wird. Die Auswahl und Anwendung der Architekturen wurde durch einen Autor entwickelt und durchgeführt was zu einer undemokratischen Projektkonzeptionierung führt und einen nicht untersuchbaren menschlichen Bias verursacht.

Die Diskussion eines Deep Learning Projektes wird weiterhin durch die intransparenten Entscheidungen eines Netzwerkes erschwert. Aus einer Architektur kann nicht eindeutig entnommen werden, wie und unter welchen Gesichtspunkten eine Entscheidung getroffen wird. Damit kann die Möglichkeit, dass besonders in den tiefen Schichten einer Architektur falsche semantische Zusammenhänge gelernt und reproduziert werden, nicht ausgeschlossen oder untersucht werden.

### **4.4 Training**

Fast alle vorgestellten Architekturen zeigen nach etwa 50 Epochen keine Zunahme der Validierungsgenauigkeit. Es ist daher möglich die Trainingszeiten zu verkürzen um Ergebnisse mit ähnlicher Genauigkeit bei kürzeren Trainingszeiten zu erlangen. Mithilfe eines kürzeren Trainings würde beispielsweise der Überanpassung der InceptionV3- und ResNet101-Architekturen zur Patch-basierten Landbedeckungsklassifikation entgegengewirkt werden. Das Swin U-Net zeigt nach 100 Trainingsepochen kein konvergentes Verhalten. Im Rahmen des Projektes wurde die Architektur auch für 300 Epochen trainiert. Allerdings konnte auch nach dieser Trainingszeit kein Genauigkeitsvorteil gegenüber dem Trans U-Net erreicht werden.

Alle Architekturen wurden auf eine Eingabefenstergröße von 128 x 128 Pixel trainiert. Es ist möglich, dass eine weitere Vergrößerung des Eingabefensters die Fähigkeit der Architekturen verbessert, den räumlichen Zusammenhang unter Klassen zu lernen. Im Rahmen des Projektes konnte aufgrund des begrenzten GPU Speicherplatzes keine Vergrößerung der Fenstergröße und deren Auswirkung untersucht werden.

### **4.5 Interferenz**

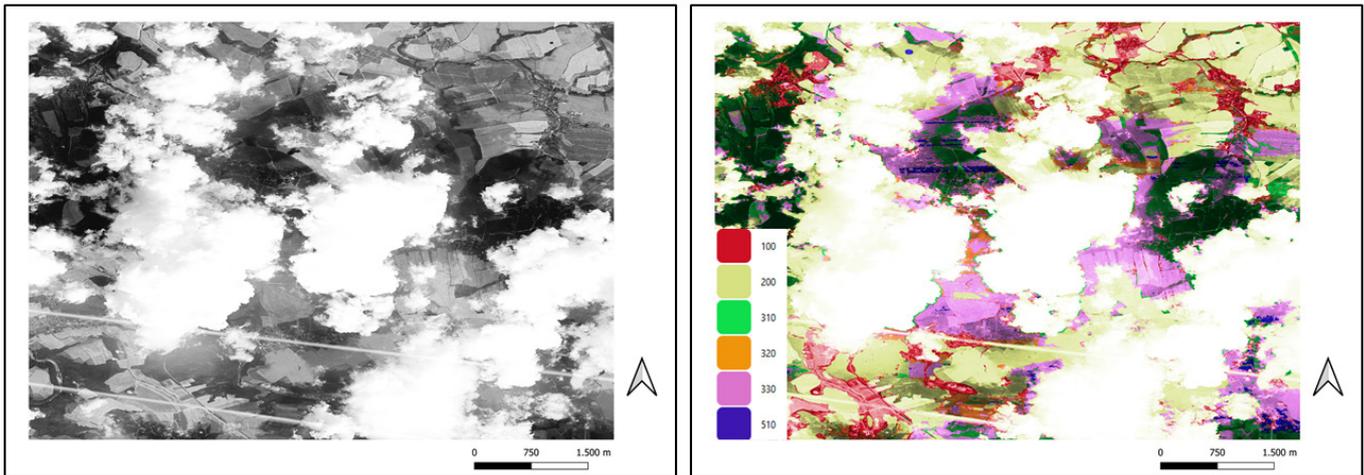
Für die Interferenz wurde das Eingabefenster in das Trans U-Net als bewegliches Fenster betrachtet. Das Fenster wurde dazu mit einer Schrittweite kleiner der Größe des Eingabefensters über das Bild bewegt. Diese Notwendigkeit entstand aus ersten Interferenztests in der Projektkonzipierung. Es war auffällig, dass die Vorhersagen am Rand eines Bildpatches deutlich schlechter sind als die im Zentrum eines Bildpatches. Dies liegt vor allem am fehlenden Kontextinformationen für umliegende Gebiete im Randbereich. Um dieses Problem zu beheben wurden die Überlappungsbereiche der Bildpatches mit einer 2D-Spline Interpolation vorhergesagt.

Bei der Betrachtung der Interferenzergebnisse wurde deutlich, dass in Bereichen um Wolken eine ungenauere semantische Segmentierung vorliegt. Diese Fehler sind ebenfalls wie im Randbereich auf fehlende Bildinformationen im umliegenden Bereich der Wolken zurückzuführen. Auffällig ist ebenfalls, dass die Bereiche um eine Wolke oft verrauscht sind. Diese verrauschten Bildinformationen führten häufig zu einer verrauschten und inkonsistenten Vorhersage der Klassen. Abbildung 36 soll die Problematik der Landbedeckungsklassifikation in bewölkten Bereichen aufzeigen.

### Ergebnisse der semantische Segmentierung in bewölkten Bildbereichen von 1975

Originale panchromatische Bildinformation

Vorhersage der Landbedeckung



**Abbildung 36: Beispiel für die Ergebnisse der semantischen Segmentierung in bewölkten Bereichen für den Zeitschnitt 1975. Bedeutung der Label: 100 = Künstliche Oberflächen, 200 = Landwirtschaftliche Flächen, 310 = Wald, 320 = Strauchvegetation, 330 = offene Fläche, 510 = Gewässer.**

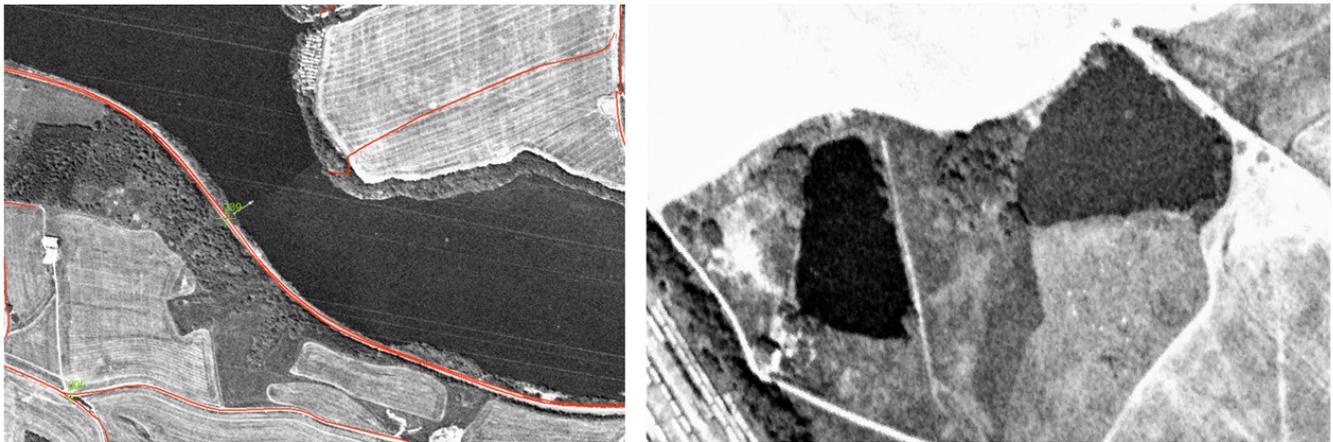
Aus Abbildung 36 ist zu erkennen, dass es in bewölkten Bereichen der KH-Bilddaten zu Fehlern in den Vorhersagen kommt. Ebenfalls zeigt sich, dass die Schatten der Wolken ebenfalls teilweise zu fehlerhaften Vorhersagen führen. In den Randbereichen der Wolke kommt es oft zu einer gemischten semantischen Segmentierung der Daten. Überwiegend werden Randbereiche der Wolken als offene Flächen klassifiziert. Ebenso kommt es zur Klassifikation von Gewässerflächen in den Randbereichen. Die Kanten der Wolken werden häufig der Klasse Wald zugeordnet. Die Abbildung legt dar, dass in Schattenbereichen die Landbedeckung manchmal der Klasse Wald zugeordnet wird. Die Genauigkeit der richtigen Zuordnung von Künstlichen Oberflächen wird in den Randbereichen der Wolken kaum beeinflusst. Diese werden stets auf hohem Niveau richtig semantisch segmentiert.

Für die weitere Prozessierung der Daten gilt es die Wolkenflächen und die Schattenbereiche zu kennzeichnen und deren Vorhersagen mit Achtsamkeit zu nutzen. Außerdem gilt anzumerken, dass die die Vollständigkeit der Daten durch die Wolken beeinflusst wird.

In den Interferenzergebnissen über alle Zeitschnitte ist eine geringe Genauigkeit für die semantische Segmentierung von Gewässerflächen zu verzeichnen. Die schlechte Genauigkeit kann auf drei Hauptgründe zurückgeführt werden. Zum einen haben die Gewässer die geringste Anzahl Beispielinformationen. Dies führt dazu, dass nicht genug repräsentative Merkmale für die Klasse von den Architekturen gelernt werden können. In Konsequenz sind die Architekturen nicht genügend generalisiert für die Vorhersage der Gewässer was letztlich zu einer geringen Segmentierungsgenauigkeit für unbekannte Daten führt. Der zweite

Grund ist die teilweise ungenaue Kennzeichnung der Gewässer in den semantischen Masken, welche das Lernen von repräsentativen Merkmalen zusätzlich erschwert. Der letzte Grund sind die diversen Erscheinungen von Gewässern. Die Erscheinung eines Flussverlaufes ändert sich in den KH-Bilddaten beispielsweise durch Richtungsänderungen oder durch Wolkenschatten. Ebenfalls werden die Gewässerflächen stark von Wind beeinflusst, was für eine starke Diversität in der Textur führt. Es ist außerdem zu beobachten das gewässernahe Bäume den Flussverlauf überragen und es so zur Klassifikation von Waldflächen kommt. In der folgenden Abbildung werden zur Verdeutlichung des Problems zwei Extrembeispiele aufgezeigt.

Diversität der Gewässerflächen



Teil eines Sees aus den KH-Bilddaten von 1975

Zwei Seen aus den KH-Bilddaten von 1972

**Abbildung 37: Diversität von Gewässerflächen. Die Abbildung links zeigt einen See von einem Bild. In der Mitte des Sees ist eine glatte Kante in der Gewässerfläche zu erkennen. Der obere Teil des Sees wirkt kontrastreicher und texturärmer als der untere Teil des Sees. In der rechten Abbildung sind zwei Seen zu erkennen. Beide liegen sehr nah beieinander dennoch ist ihre Erscheinung in den KH-Bilddaten sehr unterschiedliche.**

Mithilfe der Abbildung 37 soll die Komplexität zur Klassifikation der Gewässerflächen noch einmal herausgestellt werden.

#### 4.6 Nachprozessierung

Innerhalb der Nachprozessierung der Landbedeckungskarten soll sichergestellt werden, dass die Klassifikationsergebnisse auf Szenen-Level sowohl räumlich als auch zeitlich konsistent miteinander betrachtet werden können. Die in den Kapiteln 2.9 und 2.10 beschriebenen Schritte erreichen dieses Ziel und sind dabei weitestgehend robust gegenüber potentiellen Fehlprozessierungen, allerdings verbleiben einige Restungenauigkeiten. Bei der Mosaikierung der Szenen eines Zeitschnitts sind die Hauptfehlerquelle die Überlappungsflächen und dabei vor allem die schwarzen Randstreifen, die zum festen Bestandteil jeder analogen KH-Aufnahme gehören. Auch wenn durch gezielte Kriterien versucht wird, jeweils den sinnvollsten Pixelwert (derjenige der nicht zu einem Randstreifen gehört oder der mit der höheren Sicherheit) in das finale Szenen-Mosaik zu übernehmen, so kann dies von Szene zu Szene unterschiedlich sein, vor allem da die Randstreifen stets unterschiedlich lang und die Überlappungsbereiche unterschiedlich groß sind. In Einzelfällen wird für die Randbereiche auch ein fehlender Wert vorhergesagt (siehe 1978). Diese Vorhersagen bleiben leider in der implementierten Mosaikierung bestehen und führend somit auch in der Gesamtkarte zu Patches ohne Datenwert. Da es sich um Bereiche handelt, in denen die Güte der Daten ohnehin niedrig ist, ist dies für die finale Analyse der Landbedeckung nicht schwerwiegend, allerdings sind

Lücken in den finalen Karten stets unansehnlich und unterbrechen eine ansonsten konsistente Pixelzeitreihe. Auch wenn durch die Aggregation der Ergebnisse von 2 x 2 m auf 100 x 100 m einige Fehler der Mosakierungsstrategie beseitigt werden (beispielsweise werden die Szenenkennungen in den schwarzen Randbereichen bei der Mosaikierung nicht beseitigt, deren Einfluss verschwindet jedoch komplett bei der Aggregation), ist diese Problematik bei der Interpretation der Landbedeckungskarten zu berücksichtigen. Zusätzlich ist zu erwähnen, dass in einigen Jahren kleine Datenlücken zwischen Szenen entstehen, wenn es an bestimmten Stellen durch inkonsistente Flugbahnen keinen Überlappungsbereich gibt (siehe beispielsweise 1965 und 1975). Diese wurden versucht durch den Regelsatz bestmöglich zu füllen.

Die Konsistenzprüfung mittels eines Regelsatzes soll die Klassifikationsergebnisse dahingehend verbessern, dass einzelne Fehler (entstanden beispielsweise aufgrund von schlechter Bildqualität durch unvorteilhafte Lichtverhältnisse oder Wolkenbedeckung) ausgebügelt und unrealistische Klassensprünge auf temporaler Ebene verhindert werden. Da durch diese Strategie direkt in die Ergebnisse eingegriffen und diese potentiell verändert werden, ist eine passgenaue Konfiguration dieses Regelsatzes unerlässlich um vor allem eine Verschlechterung der Ergebnisse auszuschließen. Die in 2.10 beschriebenen Schritte wurden sorgfältig diskutiert und vor ihrer Implementierung mehrfach unter Zuhilfenahme von Testdaten überprüft. Da es in einer Zeitreihe von Klassenwerten zu sehr vielen verschiedenen Abfolgen kommen kann, können allerdings nicht alle Fälle zuverlässig verarbeitet werden. Dies zeigt sich vor allem bei Zeitreihen mit drei oder vier Klassenänderungen im Beobachtungszeitraum. In diesen Fällen gibt es effektiv keine Möglichkeit automatisch zu erkennen, welche Werte in der Reihe korrekt und welche falsch klassifiziert wurden, weswegen von einem Eingriff in diese Pixelzeitreihen abgesehen wurde. Leider treten diese Fälle in einigen Gebieten durchaus häufiger auf, hauptsächlich in West- und Südwest-Sachsen. Dies ist sowohl auf die vorhandenen Datenlücken in diesen Regionen zurückzuführen, als auch auf die inkonsistente Bildqualität. So gibt es Bereiche in denen im Jahr 1972 eine Datenlücke existiert und in den Jahren 1967 und 1978 die Bildqualität durch Wolken und unterschiedliche Kamerakonfiguration (im Vergleich zu Ostsachsen) stark beeinträchtigt ist. Durch diese sich durchziehenden Fehlereinflüsse ist eine nachträgliche Verbesserung der Klassifikationen nahezu unmöglich. Im Gegensatz dazu gelingt es mithilfe des Regelsatzes, einfache Ausreißer und mutmaßliche Fehlklassifikationen in Zeitreihen mit wenig Fluktuation (maximal zwei Änderungen) in angemessenem Maße zu eliminieren. Die Zuhilfenahme der Landbedeckungskarte der ersten Zeitschnitts in Teilvorhaben 2 des Projekts aus dem Jahr 1985 als Zusatzinformation hat sich hierbei als nützlich erwiesen. Gerade in den Jahren 1965 und 1978 konnten sichtlich bessere Ergebnisse in weiten Teilen der Karten erzielt werden. Auch in den aufgrund von schlechter Bildqualität mangelhaft klassifizierten Gegenden im Westen Sachsens in den Jahren 1967 und 1972 konnten einige Pixel zumindest im Großen und Ganzen merklich verbessert werden, auch wenn offensichtlich viele Fehler verbleiben. Gleiches gilt für die nicht vorhandenen Pixel in den Jahren 1972 und 1978, von denen immerhin ein Teil wiederhergestellt werden konnte. Auch die Evaluation von vorher festgelegten unlogischen Klassenänderungen innerhalb zwei aufeinanderfolgender Zeitschnitte zeigte einige Fehler in der Klassifikation auf. Auch diese wurden durch die Anwendung des Regelsatzes weitestgehend beseitigt. Es ist anzumerken, dass in wenigen Fällen keine Anpassung der Werte durch den Regelsatz ausgelöst wird, da selbst bei nur zwei Änderungen in der gesamten Zeitreihe Fälle auftreten können, die nicht eindeutig gelöst werden können. Dies wurde vor allem definiert, um eine weitere Verschlechterung des Ergebnisses zu vermeiden. Der Einsatz von Flags soll auf Pixel hinweisen, die von diesen komplizierten Fällen betroffen sind. Des Weiteren können auch vermeintlich einfache Fälle mit wenigen oder isolierten Einzelabweichungen innerhalb der Pixelzeitreihe falsch verarbeitet werden, wenn die Mehrheit der Szenen an dieser Stelle beispiels-

weise Wolkenbedeckung aufweist oder aus anderen Gründen falsch klassifiziert wurde. Solche Fälle können von Natur aus vor allem in Gebirgsregionen wie den Tälern des Erzgebirges auftreten, auch wenn diese in den vorliegenden fünf Gesamtkarten äußerst selten auftreten. Im Allgemeinen kann gesagt werden, dass die Betrachtung der Pixelzeitreihe zur Fehlerbeseitigung auch in dem Sinne nicht immer robust ist, da ein Pixel in jedem Zeitschnitt unterschiedlichen Einflüssen unterliegt und demnach nicht alle Klassenwerte auf der gleichen Grundlage betrachtet werden sollten. Eine bestmögliche Prüfung der Konsistenz in den Ergebnissen wäre demnach nur manuell möglich, was offensichtlich in der Realität nicht zu realisieren ist. Dass auch die in Tabelle 4 festgelegten plausiblen Änderungen nach eigenem Ermessen aufgestellt wurden, zeigt außerdem den Interpretationsspielraum, der beim Erstellen und Anwenden dieser Strategie existiert. So gesehen ist der hier implementierte Regelsatz ein Konzept mit dem ein bedeutender Teil der Fehlklassifikationen eliminiert werden kann, kompliziertere Fälle allerdings grundsätzlich nicht verbessert werden können. Die Beachtung der Flag-Karten sind für die Analyse der Landbedeckungsklassifikationen demnach unerlässlich.

#### **4.7 Landbedeckungsänderung**

Die Landbedeckungsanalyse und Veränderungsanalyse wurde im Rahmen des Projektes basierend auf den nachprozessierten Ergebnissen aus der semantischen Segmentierung durchgeführt. In Kapitel 3.10 wurden die Ergebnisse zur Veränderung der Landbedeckung im Untersuchungszeitraum vorgestellt. Aus den Ergebnissen konnten für die überrepräsentierten Klassen der Landbedeckung klare Entwicklungstrends abgelesen werden. Es hat sich gezeigt, dass der Anteil Landwirtschaftlicher Flächen und Wald abnimmt, während sich der Anteil der Künstlichen Oberflächen im Projektzeitraum verdoppelt hat. Für die unterrepräsentierten Klassen der Landbedeckung (Strauchvegetation, offene Flächen und Gewässer) konnten keine signifikanten Entwicklungstrends erkannt werden.

Die Zuverlässigkeit der Landbedeckung oder der beobachteten Entwicklungstrends zu beurteilen gestaltet sich aus verschiedenen Gründen allerdings schwierig. Gute Genauigkeiten zur Vorhersage der Landbedeckung (siehe Kapitel 3.4) wurden nur für die drei überrepräsentierten Klassen (Künstliche Oberflächen, Landwirtschaftliche Flächen und Wald) erreicht. Die unterrepräsentierten Klassen (Strauchvegetation, offene Flächen und Gewässer) zeigen mittlere Genauigkeiten. Ausgehend davon können nur für überrepräsentierte Klassen gute Ergebnisse für die Interferenz und damit für die Analyse der Landbedeckung erwartet werden. Aufgrund des geringeren Auftretens und der mittleren Genauigkeit für unterrepräsentierte Klassen wird für deren Vorhersage eine geringere Zuverlässigkeit angenommen. Nach visueller Betrachtung der Ergebnisse wird angenommen, dass die Vorhersagen der Strauchvegetation und offenen Flächen eine bessere Zuverlässigkeit aufweisen als die Vorhersagen der Gewässer. Insbesondere in den Ergebnissen für 1975 wurden Strauchvegetationen und offene Flächen auf gutem Niveau vorhergesagt. Die Beurteilung der Zuverlässigkeit von Landbedeckungsänderungen wird zusätzlich erschwert aufgrund der in Kapitel 4.2 beschriebenen Fehler im Beispieldatensatz. Diese Fehler traten insbesondere für unterrepräsentierte Klassen auf und erschweren eine Veränderungsanalyse der Landbedeckung hier erneut. Die Fehler im Beispieldatensatz verzerren aber auch die Ergebnisse der überrepräsentierten Klassen und verringern die Zuverlässigkeit der erkannten Trends der Landbedeckungsänderung. Es wurde allerdings an vielen Stellen beobachtet, dass das Trans U-Net richtige Vorhersagen trifft, auch wenn die semantischen Masken falsch sind, woraus eine bessere Segmentierungsgenauigkeit angenommen werden kann. Insbesondere in den Ergebnissen für 1975 konnte durch visuelle Beobachtung festgestellt werden, dass die Ergebnisse überdurchschnittlich gut sind und als besser als die Testergebnisse angenommen werden. Somit werden insbesondere die Vorhersagen von 1975 welche nicht von Wolken beeinflusst sind als be-

sonders zuverlässig angenommen. Aufgrund des häufigen Auftretens von Künstlichen Oberflächen, Landwirtschaftlichen Flächen und Wald wird für die Klassen insgesamt eine höhere Zuverlässigkeit angenommen. Insbesondere für Künstliche Oberflächen wurde durch visuelle Beobachtung eine sehr gute Segmentierungsgenauigkeit erkannt. Ausgehend davon wird dem Entwicklungstrend für Künstliche Oberflächen eine gute Zuverlässigkeit zugeschrieben. Ebenfalls eine gute Segmentierungsgenauigkeit zeigen die beiden anderen überrepräsentierten Klassen womit auch ihre Entwicklungstrends als zuverlässig angenommen werden.

Um den Wahrheitsgehalt der Entwicklungstrends weiter zu diskutieren wurden die Ergebnisse der Landbedeckung von Teilvorhaben 2 für 1985 herangezogen (siehe Abbildung 31). Aus den Ergebnissen konnte abgeleitet werden, dass die Entwicklungstrends für die überrepräsentierten Klassen zwischen den beiden Projektteilvorhaben übereinstimmen. Zwischen 1978 und 1985 wurde eine weitere Zunahme der Künstlichen Oberflächen erkannt im Vergleich zu den Zeitschnitten 1965 bis 1975. Auffällig war bei dem Vergleich das für den Zeitschnitt von 1978 zuviel Künstliche Oberfläche vorhergesagt wurde. Ebenfalls wurde bei dem Vergleich festgestellt, dass die Abnahme von Waldflächen sich bis 1985 fortsetzt. Die Trendentwicklung der Landwirtschaftlichen Flächen konnte auch mithilfe der Landbedeckungsdaten von 1985 bestätigt werden. Es hat sich hier gezeigt, dass zwischen 1978 und 1985 allerdings kaum eine Veränderung gemessen werden konnte.

Zusammenfassend ergibt sich daraus, dass die Ergebnisse der Landbedeckungsanalyse und der Veränderungsanalyse für die überrepräsentierten Klassen als zuverlässig angenommen werden können. Die Veränderungen der unterrepräsentierten Klassen können nur teilweise als zuverlässig angenommen werden. Aufgrund der guten Ergebnisse im Zeitschnitt von 1975 können dort insbesondere die Strauchvegetation und die offenen Flächen als zuverlässige Vorhersage angenommen werden.

## 5 Zusammenfassung

In dieser Schrift wurde eine Methode auf Basis von Künstlicher Intelligenz vorgestellt, trainiert, validiert und angewendet um die historische Landnutzung in Sachsen für die 1960er und 1970er Jahre aus historischen CORONA-Spionageaufnahmen der CIA zu kartieren. Die Unterschiede in der Bildqualität und in den geometrischen und radiometrischen Eigenschaften einzelner CORONA-Aufnahmen machen eine automatische Verarbeitung und Landbedeckungsklassifikation besonders schwierig. Um die Klassifikation überhaupt durchführen zu können, wurde erstmalig ein Trainingsdatensatz der Landbedeckung auf Basis von CORONA-Aufnahmen zusammengestellt.

Verschiedene Deep Learning-Methoden der Patch-basierten Bildklassifikation und der semantischen Segmentierung wurden dann anhand der Trainingsdaten trainiert und mit unabhängigen Validierungsdaten getestet. Unter den Patch-basierten Methoden erreichte das ResNet101V2 mit einer Gesamtgenauigkeit von 66% das beste Ergebnis. Diese Methode ermöglicht die Kartierung der Landbedeckung für Bild-Ausschnitte von 100 x 100 m. Unter den Methoden der semantischen Segmentierung erreichte das Trans U-Net mit 56% IoU die höchste Gesamtgenauigkeit. Diese Methode ermöglicht eine Kartierung der Landbedeckung auf einer Auflösung von 2 x 2 m anhand der genutzten CORONA-Daten. Beide Methoden erreichen insbesondere gute Genauigkeiten bei der Kartierung von Landwirtschaftlichen Flächen, Wäldern und Künstlichen Oberflächen. Klassen mit geringeren Flächenanteilen wie Strauchvegetation und offene Flächen werden in allen Methoden nur mit geringer Genauigkeit klassifiziert. Die Genauigkeit der Klassifikation von Gewässern ist von der Methode abhängig.

Im Anschluss wurde das Trans U-Net auf alle CORONA-Bilddaten des Freistaates Sachsen für die Jahre 1965, 1967, 1972, 1975 und 1978 angewendet. Aus der Anwendung der Architekturen ist eine gute Segmentierungsgenauigkeit für die überrepräsentierten Klassen (Künstliche Oberflächen, Landwirtschaftliche Flächen und Wald) zu erkennen. Für die Segmentierung der Klassen Strauchvegetation und offene Flächen wird eine mittlere Genauigkeit verzeichnet. Die schlechteste Segmentierungsgenauigkeit wurde für Gewässer verzeichnet. Für Ungenauigkeiten sind insbesondere der unvollständige Beispieldatensatz und die darin enthaltenen Fehler verantwortlich.

Mit Fertigstellung des Datensatzes sind vielfältige Anwendungen denkbar. Der bedeutendste Mehrwert des Datensatzes besteht in der Verlängerung der Zeitreihe der historischen Landbedeckungsänderung.

Die vorliegende Arbeit hat im weltweiten Vergleich einen absoluten Pioniercharakter. Noch nie wurden für ein so großes Untersuchungsgebiet und für mehrere Jahre CORONA-Daten genutzt um die Landbedeckung zu klassifizieren. Trotz der vorhandenen Mängel der Landbedeckungskarten, stellt diese Arbeit ein Meilenstein in der fernerkundlichen Landbeckungsklassifikation dar

## Literaturverzeichnis

- BADRINARAYANAN, V., KENDALL, A., & CIPOLLA, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495.
- CAO, H., WANG, Y., CHEN, J., JIANG, D., ZHANG, X., TIAN, Q., & WANG, M. (12. 05 2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv-Online*.
- CHAVDAROVA, T., & FLEURET, F. (2018). SGAN: An Alternative Training of Generative Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (S. 9407-9415).
- CHEN, J., LU, Y., YU, Q., LUO, X., WANG, X., WANG, Y., .ZHOU, Y. (08. 02 2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *ArXiv-Online*.
- DILLON, S., & COLLETT, C. (2019). *AI and Gender: Four Proposals for Future Research*. Cambridge: University of Cambridge.
- DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., WEISSENBORN, D., ZHAI, X., UNTERTHINER, T., HOULSBY, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *Proc. Int. Conf. Learn. Reinforcement (ICLR)*, (S. 1-22).
- DROZDZAL, M., VORONTSOV, E., CHARTRAND, G., KADOURY, S., & PAL, C. (2016). The Importance of Skip Connections in Biomedical Image Segmentation. In *Deep Learning and Data Labeling for Medical Applications*. Springer.
- GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., BENGIO, Y. (10. 06 2014). Generative Adversarial Nets. *Universite de Montreal*, S. 9.
- GROUP, S.-I. *Scikit-Image*. Abgerufen am 2021. 05 2021 von <https://scikit-image.org>
- Gu, X., Li, S., Renac, S., Zheng, H., Fan, C., & Xua, H. (09 2022). Adaptive enhanced swin transformer with U-net for remote sensing image segmentation. *Computers and Electrical Engineering*.
- HALL-BEYER, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38(5), 1312-1338.
- HARALICK, R., & SHANMUGAM, K. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 6, 610-621.
- HE, K., ZHANG, X., REN, S., & SUN, J. (2016). Identity Mappings in Deep Residual Networks. *Computer Vision - ECCV 2016*, (S. 630-645).
- HUANG, G., LIU, Z., VAN DER MAATEN, L., & WEINBERGER, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (S. 2261-2269).
- HUANG, H., LIN, L., TONG, R., & HU, H. (19. 04 2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *ArXiv-Online*.
- IGLOVIKOV, V., MUSHINSKIY, S., & OSIN, V. (2017). Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition.
- JADON, S. (2020). A survey of loss functions for semantic segmentation. *Proceedings of 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, (S. 1-7).
- JOHN, D., & ZHANG, C. (03 2022). An attention-based U-Net for detecting deforestation within satellite sensor imagery. *International Journal of Applied Earth Observation and Geoinformation*.

- KAIMING , H., ZHANG, X., REN, S., & SUN, J. (10. 12 2015). Deep Residual Learning from Image Recognition. *arXiv Online*.
- KINGMA, D. P., & BA, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
- KRIZHEVSKY, A., SUTSKEVER, I., & HINTON, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* , 25(2).
- LIU, Z., LIN, Y., CAO, Y., HU, H., WEI, Y., ZHANG, Z., GUO, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (S. 10012-10022).
- OKTAY, O., SCHLEMPER, J., & FOLGOC, L. (2018). Attention U-Net: Learning Where to Look for the Pancreas . *ArXiv Online* .
- OTTO, S. (2020). *CAPABILITY OF CNNs FOR ANALYZING LAND COVER USING HISTORICAL CORONA-DATA*. Dresden: TU Dresden.
- RONNEBERGER, O., FISCHER, P., & BROX, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (S. 234-241). Freiburg: Springer.
- SIMONYAN, K., & ZISSERMAN, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proc. 3rd Int. Conf. Learning Representations ICLR 2015*.
- SZEGEDY, C., IOFFE, S., VANHOUCHE, V., & ALEMI, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, (S. 4278-4284).
- SZEGEDY, C., VANHOUCHE, V., IOFFE, S., & SHLENS, J. (11. 12 2015). Rethinking the Inception Architecture for Computer Vision. *arXiv-Online*.
- ULMAS, P., & LIIV, I. (05. 03 2022). Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification. *arXiv-Online*.
- VASWANI, A., SHAZEER, N., PARMAR, N., & USZKOREIT, J. (2017). Attention Is All You Need. *ArXiv Online*.
- VEDALDI, A., & SOATTO, S. (2008). Quick Shift and Kernel Methods for Mode Seeking. In D. Forsyth, P. Torr, & A. Zisserman, *Computer Vision - ECCV 2008* (S. 705-718). Berlin, Heidelberg: Springer.
- ZHOU, Z., RAHMAN , M. M., SIDDIQUEE, R., & T, N. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. : *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* , (S. 3-11). Granada, Spanien.

**Herausgeber:**

Sächsisches Landesamt für Umwelt, Landwirtschaft und  
Geologie (LfULG)  
Pillnitzer Platz 3, 01326 Dresden  
Telefon: +49 351 2612-0  
Telefax: +49 351 2612-1099  
E-Mail: lfulg@smekul.sachsen.de  
www.lfulg.sachsen.de

**Autoren:**

Lucas Kugler, Christopher Marrs, Eric Kosczor  
und JProf. Dr. Matthias Forkel  
Technische Universität Dresden  
Institut für Photogrammetrie und Fernerkundung  
Telefon: +49 351 463-32270  
Telefax: +49 351 463-37284  
E-Mail: matthias.forkel@tu-dresden.de

**Redaktion:**

Sebastian Goihl  
Abteilung 2: Grundsatzangelegenheiten Umwelt,  
Landwirtschaft, Ländliche Entwicklung  
Referat 22: GAP, Informationsmanagement  
August-Böckstiegel-Straße 3, 01326 Dresden  
Telefon: +49 351 2612 2202  
Telefax: +49 351 2612 1099  
E-Mail: sebastian.goihl@smekul.sachsen.de

**Redaktionsschluss:**

30.11.2022

**Auflage:**

1. Auflage

**ISSN:**

1867-2868

**Hinweis:**

Die Broschüre steht nicht als Printmedium zur Verfügung,  
kann aber als PDF-Datei unter  
<https://publikationen.sachsen.de> heruntergeladen werden.

**Verteilerhinweis**

Diese Informationsschrift wird von der Sächsischen  
Staatsregierung im Rahmen ihrer verfassungsmäßigen  
Verpflichtung zur Information der Öffentlichkeit  
herausgegeben.

Sie darf weder von Parteien noch von deren Kandidaten  
oder Helfern zum Zwecke der Wahlwerbung verwendet  
werden. Dies gilt für alle Wahlen. Missbräuchlich ist  
insbesondere die Verteilung auf Wahlveranstaltungen,  
an Informationsständen der Parteien sowie das Einlegen,  
Aufdrucken oder Aufkleben parteipolitischer Informationen  
oder Werbemittel. Untersagt ist auch die Weitergabe an  
Dritte zur Verwendung bei der Wahlwerbung.

*Täglich für  
ein gutes Leben.*

www.lfulg.sachsen.de