

Untersuchung von «Deep Learning»- Modellen in ArcGIS Pro zur Extraktion von Merkmalen aus Historischen Karten: Objekt- und Pixelklassifizierung

Bachelor-Arbeit

Autorin:

Aline Pironato, BSc Raumbezogene Ingenieurwissenschaften
apironato@student.ethz.ch

Leitung:

Prof. Dr. Lorenz Hurni

Betreuung:

Dr. Magnus Heitzler
Chenjing Jiao
Xue Xia

Frühjahrssemester 2022

10. Juni 2022



Vorwort

Vor nun gut einem halben Jahr als ich mich mit der Themenwahl meiner Bachelor-Arbeit auseinandersetzen musste, war alles etwas überwältigend. Die Anzahl zur Auswahl stehender Themen war immens und meine Angst, eine unpassende Wahl zu treffen und mit einem Arbeitsthema zu enden, dass sich weder mit meinen Interessen noch meinen Fähigkeiten deckte, war enorm.

Rückblickend waren diese Sorgen und Bedenken zum Glück völlig unbegründet, denn mit meiner Entscheidung für dieses Thema, nämlich die Untersuchung von «Deep Learning»-Modellen, fand ich eine Aufgabe, die mir wirklich entsprach. Ich durfte mich ein Semester lang mit zukunftsgerichteten Themen befassen und ein vertieftes Verständnis für die Anwendung von Deep Learning und Neuronalen Netzen im Geoinformationsbereich gewinnen.

Somit möchte ich mich an dieser Stelle bei dem Institut für Kartografie und Geoinformation und Herrn Professor Dr. Lorenz Hurni für das Angebot sowie die Leitung dieser Bachelorarbeit bedanken. Weiteren Dank möchte ich auch meinen Betreuern Dr. Magnus Heitzler, Chenjing Jiao und Xue Xia aussprechen, die mich im Laufe des Arbeitsprozesses bei Bedarf immer tatkräftig unterstützen und mit Rat zur Seite standen.

Zusammenfassung

Die Extraktion von Merkmalen aus historischen Karten ist eine Aufgabe, mit der sich Experten aus Fachkreisen der Bild- und Geodatenverarbeitung schon lange beschäftigen. Ein vielversprechender Ansatz, um diesen Prozess zu optimieren, können die «Deep Learning»-Modelle sein, die auf Convolutional Neural Networks aufgebaut sind. In ArcGIS Pro werden bereits heute eine Auswahl solcher «Deep Learning»- Modellen zur Verfügung gestellt,

Mit historischen Daten der Schweizer Landeskarte und einem Feature-Layer der Schweizer Feuchtgebiete werden in dieser Arbeit verschiedene «Deep Learning»-Modelle der Pixel- und Objektklassifizierung dahingehend untersucht, ob sie sich dazu eignen, Merkmale in Form von Wetland-Polygonen aus den historischen Karten zu extrahieren. Der Fokus der Arbeit liegt dabei auf der Entwicklung eines optimalen Workflows, der Analyse und Vergleiche der trainierten Modelle und deren Extraktionen der Wetland-Features.

Im Detail werden in dieser Arbeit das U-Net-Modell, das PSPNet-Modell und das DeepLabV3-Modell untersucht. Im Zuge dieser Arbeit kann festgehalten werden, dass der in ArcGIS Pro etablierte Workflow in Kombination mit dem U-Net oder dem DeepLabV3-Modell und dem Parameter-Setup 10 zu einer optimierten Extraktion von Wetland-Features führt.

Inhaltsverzeichnis

1	Einleitung	9
1.1	Aufgabenstellung	9
1.2	Zielsetzung	9
2	Theorie	10
2.1	Grundlagen	10
2.1.1	Extraktion von Merkmalen aus Historische Karten	10
2.1.2	Deep Learning	10
2.2	Objektklassifizierung	12
2.2.1	Feature Classifier	12
2.3	Pixelklassifizierung	12
2.3.1	U-Net	12
2.3.2	Pyramid Scene Parsing Network	13
2.3.3	DeepLabV3	13
2.3.4	Bi-Directional Cascade Network for Edge Detector	14
2.3.5	Holistically-Nested Edge Detector	15
2.3.6	ConnectNet & Multi Task Road Extractor	15
2.3.7	Change Detector	15
3	Methoden und Vorgehen	16
3.1	Vorauswahl der Modelle.....	16
3.1.1	Aussortierte Modelle	16
3.1.2	Ausgewählte Modelle	18
3.2	Grundlagedaten	18
3.2.1	Historische Geodaten	18
3.2.2	Feature-Layer	19
3.3	Grundstruktur	20
3.4	Hardware & Software	21
3.4.1	Hardware	21
3.4.2	Software	21
3.5	Exportieren der Trainingsdaten.....	Fehler! Textmarke nicht definiert.
3.6	Training der Modelle	Fehler! Textmarke nicht definiert.
3.6.1	Parameter	Fehler! Textmarke nicht definiert.
3.6.1.1	Model Arguments	Fehler! Textmarke nicht definiert.
3.7	Klassifizierung	Fehler! Textmarke nicht definiert.
3.8	Evaluation	22
3.8.1	Statistische Gütemasse	22
3.8.2	Gütemasse in ArcGIS Pro	24
3.8.3	Gütemasse der Pixelklassifizierungen	25
3.8.4	Verwendete Gütemasse	26
4	Ergebnisse	26
4.1	Workflows	26
4.2	Auswertung	28
4.2.1	Setups	28
4.2.1.1	Variation der Modelltypen	28
4.2.1.2	Variation der Hyperparameter	29
4.2.1.3	Variation der Kartenblätter	30
4.2.1.4	Übersicht der Einteilung	32
4.3	Vergleich	33

4.3.1	Vergleich innerhalb der Modelle	34
4.3.1.1	U-Net	34
4.3.1.2	PSPNet	40
4.3.1.3	DeepLabV3	46
4.3.2	Vergleich nach Setup	52
4.3.3	Vergleich nach Kartenblatt	55
4.4	Empfehlung	60
5	Diskussion	60
5.1	Interpretation	60
5.2	Einschränkungen	61
6	Literaturverzeichnis	63
Anhang 66		
Diagramme	Fehler! Textmarke nicht definiert.	Fehler!
	Textmarke nicht definiert.	Fehler!
	Textmarke nicht definiert.	Fehler!
	Textmarke nicht definiert.	Fehler!
	Textmarke nicht definiert.	Fehler!
	Textmarke nicht definiert.	Fehler!
Eigenständigkeitserklärung		66

Titelbild

Quelle: Bundesamt für Landestopografie swisstopo (1952). Swiss Map Raster 25. Ausschnitt aus dem Kartenblatt 1169.

Abbildungsverzeichnis

Abbildung 1 Vereinfachte Illustration eines tiefen neuronalen Netzwerks mit mehreren Layern (Schindler, 2021)	11
Abbildung 2: Beispiel eines Convolutional Neural Network	11
Abbildung 3: U-Net Architektur (Ronneberger et al. 2015)	12
Abbildung 4: Atrous-Faltung. (Zhao, Shi, Qi, Wang, & Jia, 2016).....	13
Abbildung 5: PSPNet Modell Struktur. (Zhao, Shi, Qi, Wang, & Jia, 2016)	13
Abbildung 6: DCNN und der Einfluss von Atrous-Faltung. (Chen, Papandreou, Schroff, & Hartwig, 2017).....	14
Abbildung 7: Atrous Spatial Pyramid Pooling network. (Chen, Papandreou, Schroff, & Hartwig, 2017)	13
Abbildung 8: Architektur des Bi-Directional Cascade Network for Edge Detector. (He, Zhang, Yang, Shan, & Huang, 2019).....	14
Abbildung 9: Architektur des Holistically Nested Edge Detection. (Xie & Tu, 2015).....	15
Abbildung 10: STANet Network Architektur (Chen & Shi, 2020).....	16
Abbildung 11: Output Feature Classifier: Grüne Kacheln: Wetlands, Rosa Kacheln: Keine Wetlands	16
Abbildung 12 Pixelklassifizierung mit BDCN Edge Detection	17
Abbildung 13 Pixelklassifizierung mit HED.....	17
Abbildung 14: Pixelklassifizierung mit ConnectNet	17
Abbildung 15: Pixelklassifizierung mit Multi Task Roas Extractor.....	17
Abbildung 16: PSP Pixelklassifizierung.....	18
Abbildung 17: U-Net Pixelklassifizierung.....	18
Abbildung 18: DL Pixelklassifizierung	18
Abbildung 19: Historische Karte	19
Abbildung 20: Datensatz Wetlands	19
Abbildung 21: Manuell erfasste Wetlands	19
Abbildung 22: Illustration der Grundstruktur für das Trainieren von DL-Modellen zur Pixelklassifizierung in ArcGis Pro.....	20
Abbildung 23: Verlustfunktion des PSP9 Modells während dem durchlaufen der Trainingsepochen ..	25
Abbildung 24: Optischer Vergleich der Groundtruth-Daten (l.) mit den vom PSP9 klassifizierten Image-Chips (r.).....	24
Abbildung 25: Klassische Metrik mit Präzision, Recall und F1-Score des PSP9 Modells	25
Abbildung 26: Wetland-Groundtruth in pink	25
Abbildung 27: Pixelklassifizierung des DL1-Modells in orange mit Groundtruth-Outline in pink	25
Abbildung 28: Wetland-Groundtruth in schwarz, True Negative Pixel in blau, False Positive in orange	26
Abbildung 29: Nicht-Wetland in schwarz, True Positive in orange, False Negative in blau.....	26
Abbildung 30: Legende.....	27
Abbildung 31: Workflow ohne Parameter.....	27
Abbildung 32: Workflow mit Parametern.....	28
Abbildung 33: Ausschnitt des Kartenblatts 1111.....	30
Abbildung 34: Ausschnitt des Kartenblatts 1112.....	30
Abbildung 35: Ausschnitt des Kartenblatts 1113.....	30
Abbildung 36: Ausschnitt des Kartenblatt 1184	31
Abbildung 37: Ausschnitt des Kartenblatts 1169.....	31
Abbildung 38: Ausschnitt des Kartenblatts 1264.....	31
Abbildung 39: Ausschnitt des Kartenblatts 1203.....	32
Abbildung 40: Ausschnitte zweier Klassifizierungen durch UNET 3 mit der Groundtruth.Outline	36
Abbildung 41: Ausschnitte zweier Klassifizierungen durch UNET 2 mit der Groundtruth-Outline	36
Abbildung 42: Ausschnitte zweier Klassifizierungen durch UNET 9 mit der Groundtruth-Outline	37
Abbildung 43: Ausschnitte zweier Klassifizierungen durch UNET 11 mit der Groundtruth-Outline	39

<i>Abbildung 44: Ausschnitte zweier Klassifizierungen durch UNET 10 mit der Groundtruth-Outline</i>	<i>39</i>
<i>Abbildung 45: Ausschnitte zweier Klassifizierungen durch PSP 5 mit der Groundtruth-Outline.....</i>	<i>41</i>
<i>Abbildung 46: Fehlerhafte Klassifizierungen von PSP 5.....</i>	<i>42</i>
<i>Abbildung 47: Ausschnitte zweier Klassifizierungen durch PSP 1 mit der Groundtruth-Outline.....</i>	<i>42</i>
<i>Abbildung 48: Ausschnitte zweier Klassifizierungen durch PSP 10 mit der Groundtruth-Outline.....</i>	<i>42</i>
<i>Abbildung 49: Ausschnitte zweier Klassifizierungen durch PSP 7 mit der Groundtruth-Outline.....</i>	<i>43</i>
<i>Abbildung 50: Ausschnitte zweier Klassifizierungen durch PSP 6 mit der Groundtruth-Outline.....</i>	<i>45</i>
<i>Abbildung 51: Ausschnitte zweier Klassifizierungen durch PSP 3 mit der Groundtruth-Outline.....</i>	<i>45</i>
<i>Abbildung 52: Fehlerhafte Klassifizierungen von PSP 3.....</i>	<i>45</i>
<i>Abbildung 53: Loss-Funktion von PSP 10.....</i>	<i>46</i>
<i>Abbildung 54: Loss-Funktion von PSP 3.....</i>	<i>46</i>
<i>Abbildung 55: Ausschnitte zweier Klassifizierungen durch DL 3 mit der Groundtruth-Outline</i>	<i>48</i>
<i>Abbildung 56: Ausschnitte zweier Klassifizierungen durch DL 5 mit der Groundtruth-Outline</i>	<i>48</i>
<i>Abbildung 57: Fehlerhafte Klassifizierungen von DL 5</i>	<i>48</i>
<i>Abbildung 58: Ausschnitte zweier Klassifizierungen durch DL 2 mit der Groundtruth-Outline</i>	<i>49</i>
<i>Abbildung 59: Ausschnitte zweier Klassifizierungen durch DL 1 mit der Groundtruth-Outline</i>	<i>49</i>
<i>Abbildung 60: Ausschnitte zweier Klassifizierungen durch DL 10 mit der Groundtruth-Outline</i>	<i>51</i>
<i>Abbildung 61: Ausschnitte zweier Klassifizierungen durch DL 6 mit der Groundtruth-Outline</i>	<i>51</i>
<i>Abbildung 62: Ausschnitte fehlerbehaftete Groundtruth-Daten auf dem Tile 5.....</i>	<i>56</i>

Tabellenverzeichnis

Tabelle 1: Wahrheitsmatrix.....	22
Tabelle 2: Einteilung der U-Net-Modelle	32
Tabelle 3: Einteilung der PSPNet-Modelle	32
Tabelle 4: Einteilung der DL-Modelle	32
Tabelle 5: Einteilung nach Kartenblatt.....	33
Tabelle 6: Einteilung nach Setup.....	33
Tabelle 7: U-Net-Modell Metriken der Trainingsresultate mit den regulären Daten.....	34
Tabelle 8 U-Net-Modell Metriken der Trainingsresultate mit den korrigierten Daten	35
Tabelle 9 U-Net Klassifizierung originaler Kartenblätter	37
Tabelle 10 U-Net Klassifizierung korrigierter Kartenblätter	38
Tabelle 11: PSPNet-Modell Metriken der Trainingsresultate mit den regulären Daten	40
Tabelle 12 PSPNet-Modell Metriken der Trainingsresultate mit den korrigierten Daten	40
Tabelle 13 PSPNet Klassifizierung originaler Kartenblätter	43
Tabelle 14 PSPNet Klassifizierung korrigierter Kartenblätter.....	44
Tabelle 15 DeepLab-Modell Metriken der Trainingsresultate mit den regulären Daten	46
Tabelle 16 DeepLab-Modell Metriken der Trainingsresultate mit den korrigierten Daten.....	47
Tabelle 17 DeepLab Klassifizierung originaler Kartenblätter	50
Tabelle 18 DeepLab Klassifizierung korrigierter Kartenblätter.....	50

Plotverzeichnis

Plot 1: Balkendiagramm zum Vergleich der IOUs der drei Modelltypen nach Setup.....	52
Plot 3: Balkendiagramm zum Vergleich der Präzisionen der drei Modelltypen nach Setup	53
Plot 3: Balkendiagramm zum Vergleich der Recalls der drei Modelltypen nach Setup	54
Plot 4: Balkendiagramm zum Vergleich der F1-Scores der drei Modelltypen nach Setup	54
Plot 5: Balkendiagramm zum Vergleich der Durchschnitte der drei Modelltypen nach Setup	55
Plot 7: Durchschnittlicher IOU nach Kartenblatt	56
Plot 8: Median IOU nach Kartenblatt.....	56
Plot 9: Median der Präzision nach Kartenblatt	57
Plot 10: Durchschnittlicher Präzision nach Kartenblatt	57
Plot 11: Median des Recalls nach Kartenblatt.....	58
Plot 12: Durchschnittlicher Recall nach Kartenblatt	58
Plot 13: Durchschnittlicher F1-Score nach Kartenblatt	58
Plot 14: Median des F1-Scores nach Kartenblatt.....	58

1 Einleitung

1.1 Aufgabenstellung

Historische Kartenwerke sind einzigartige Quellen für vergangene raumbezogene Informationen. Während sich das Landschaftsbild im Laufe der Jahre unter einem stetig anhaltenden Wandel verändert hat, ermöglichen solch überlieferte Karten einen Blick zurück in die Vergangenheit.

Das Bundesamt für Landestopografie swisstopo besitzt mit über 50'000 Karten die umfangreichste Sammlung von Schweizer Geodaten, die heute als wertvolles Kulturgut gelten und einen wesentlichen Teil zum Schweizer Landschaftsgedächtnis beitragen. Speziell auch dadurch, dass diese Geodaten digitalisiert wurden und nun somit der Bevölkerung in Form von hochaufgelösten Scans zur Verfügung stehen (Bundesamt für Landestopografie swisstopo, 2022).

Retrospektive raumbezogenen Daten, wie sie sich aus historischen Karten extrahieren lassen, finden bereits seit geraumer Zeit in den verschiedensten Forschungsthemengebieten Verwendung. Die gewonnenen Daten dienen dann oft der Analyse des generellen Landschaftswandel wie bei Liu et al. (2018), aber auch für spezifische Geländerekonstruktionen wie die des Baumbestands (Kaim, et al., 2016; Vieilledent, et al., 2018) oder der Küstenlinie (Alberico, et al., 2017).

Um solche Analysen und Untersuchungen durchführen zu können, müssen die historischen Daten zuerst von der Karte extrahiert und zum Beispiel via Geoinformationssystem zugänglich gemacht werden. Erst dadurch lassen sich die gesuchten Geodaten abfragen, extrahieren und entsprechend in weitere Projekte integrieren (Chiang, Duan, Leyk, Uhl, & Knoblock, 2020, S. 11; Freeman & Pieroni, 1982, wie zitiert in Chiang, Leyk, & Knoblock, 2014, S.2).

Im Rahmen dieser Arbeit wird untersucht, wie raumbezogene Merkmale mithilfe von «Deep Learning»-Modellen aus historischen Karten entnommen werden können. In diesem konkreten Fall soll die Wetland-Featureklasse von der historischen Landeskarte der Schweiz im Massstab 1:25'000 extrahiert werden. Dabei wird auf einen Ansatz mit Deep Learning gesetzt, da die jüngsten Fortschritte im Forschungsgebiet der automatischen Bildverarbeitungstechnologien jeweils mit diesem Teilbereich des Machine Learnings in Zusammenhang gebracht wurden (Hurni, Heitzler, Jiao, & Xia, 2021).

Beim Framework stützt man sich im Rahmen dieser Arbeit auf die bereits etablierten und implementierten «Deep Learning»-Modelle von ArcGIS Pro. Die angebotenen «Deep-Learning»-Modelltypen lassen sich nach vier verschiedenen Anwendungstypen einteilen: Pixelklassifizierung, Objektklassifizierung, Objektdetektion und Bildübersetzung. Im Rahmen dieser Arbeit werden nur die Modelltypen betrachtet, die zu den Kategorien Pixel- oder Objektklassifizierung gehören. Die Analyse dieser Modelle soll Aufschluss darüber geben, wie es um die Anwendungsbereiche steht und was mit allfälligen Variationen der Hyperparameter erreicht werden kann, um die Extraktion von Merkmalen aus historischen Karten zu optimieren.

1.2 Zielsetzung

Die Zielsetzung dieser Arbeit umfasst in einem ersten Schritt das Erarbeiten von passenden Workflows für die zu untersuchenden «Deep Learning»-Modelle. Anhand von Vergleichen der Metriken, Testwerten und der Anwendung der trainierten Modelle auf die eigentlichen Kartenblätter soll eine Empfehlung dazu gemacht werden können, welches Modell mit welchem Workflow sich am besten für die Extraktion von Wetland-Features eignet.

Die Arbeit soll also eine Antwort auf die folgende Frage liefern:

***Welches «Deep Learning»-Modell, basierend auf Pixel- oder Objektklassifizierung, aus dem ArcGIS Pro Katalog liefert die besten Ergebnisse bei der Extraktion von Feuchtgebieten aus der historischen Landeskarte der Schweiz?
Wie sieht der dazugehörige Workflow aus?***

2 Theorie

2.1 Grundlagen

2.1.1 Extraktion von Merkmalen aus Historische Karten

Historische Karten sind wichtige und einzigartige Informationsquellen für raumbezogene Daten. Durch die fortschreitende Digitalisierung der historischen Kartensammlungen und die daraus resultierende exponentielle Steigerung der Zugänglichkeit zu diesen historischen Datenreihen stieg das Interesse an Geodaten stark an (Chiang, Duan, Leyk, Uhl, & Knoblock, 2020, S. 11).

So liegt es nun nicht mehr nur im Interesse der Geoinformationswissenschaftler, die in historischen Karten eingeschlossene Information durch Extraktion und Überführung in ein GIS brauchbar zu machen. Dies ist aber kein neues Anliegen. Technologien für die Extraktion und Erkennung von Inhalten von Karten in Raster Form werden seit den 1980er Jahren entwickelt. Die daraus resultierenden Kartenverarbeitungstechnologien sind aber vor allem in Belangen der Datenmenge noch limitiert (Chiang, Duan, Leyk, Uhl, & Knoblock, 2020, S. 11).

In bisherigen Extraktionen von Wetland-Features wurde zum Beispiel mit dem SymbolRecognizer-Framework gearbeitet, welches grafische Symbolisierungen in Karten aufspüren soll. Dabei wird dem Modell ein Bild mit der Wetland-Symbolisierung zugeführt, sodass das Muster in der Karte wiedererkannt werden soll (Chiang, Duan, Leyk, Uhl, & Knoblock, 2020, S. 21). Beim Extrahieren der Wetland-Features mit diesem Framework wird aber noch nicht von «Deep Learning» Gebrauch gemacht.

Extraktionen aus historischen Karten, die sich auf Deep Learning stützen, gibt es aber bereits. Garcia-Molsosa, et al. (2021) untersuchten das Potenzial von Deep Learning Segmentierung für die Extraktion von Archäologischen Features aus historischen Kartenreihen. Dabei wurden zwar nicht explizit die Identifizierung von Wetland-Features untersucht, aber es lassen sich gewisse Parallelen feststellen. Die Autoren zeigen auch auf, wieso das Potenzial der auf Deep Learning basierenden Segmentierung historische Karten bis jetzt noch nicht voll ausgeschöpft wurde. Einerseits wird darauf eingegangen, dass je nach Kartewahl, wesentlicher Aufwand durch Preprocessing generiert wird. Weiter wird beschrieben, dass Karten subjektive Werke sind und die Signaturen vom Kartografen abhängig ist. Und letztlich wird erwähnt, dass man es mit uneinheitlichen und unregelmässigen Formen zu tun hat, was das korrekte Klassifizieren weiter erschwert. Bei der Segmentierung von Wetlands werden diese Punkte ebenfalls relevant sein. Aber durch die Fähigkeit künstlicher neuronaler Netzwerke mit solchen Irregularitäten umzugehen und aus ihnen zu lernen, sollte dies eine robuste Grundstruktur abgeben.

2.1.2 Deep Learning

«Deep Learning»-Methoden zeichnen sich dadurch aus, dass sie sich aus mehreren Schichten, aus mehreren Layern zusammensetzen, um mithilfe mehrerer Abstraktionsebenen die Merkmale von Daten zu erlernen (LeCun et al., 2015, wie zitiert in Minar & Naher, 2018, S. 1). Das Deep Learning macht es dem Computer möglich, die komplizierten Konzepte zu erlernen, in dem sie sich aus einfacheren Konzepten zusammensetzen lassen (Goodfellow et al., 2016, wie zitiert in Minar & Naher, 2018, S. 1). Je schwieriger es für den Computer ist, dass Konzepte zu lernen, umso sollte die Architekturen der künstlichen neuronalen Netzwerke ausfallen (Minar & Naher, 2018, S. 1).

Tiefes überwachtes Lernen

Tiefe neurale Netzwerke können vor allem im Bereich des überwachten Lernens grosse Erfolge verzeichnen (Minar & Naher, 2018, S. 4). Überwachtes Lernen liegt vor, wenn die funktionale Form der Prädiktionfunktion bereits bekannt ist oder angenommen werden kann. Man kennt also die

Zielklassen, denen die Trainingsdaten zugewiesen werden. Die Trainingsdaten mit den bekannten Klassifikationen werden dann eingesetzt, um die Funktionsparameter zu schätzen. Die Parameter werden so lange angepasst, bis der aus den fehlerhaften Klassifikationen resultierende Fehler klein genug ist (Schindler, 2021a). In dieser Arbeit wird nur überwachtes maschinelles Lernen zur Anwendung kommen, da das zu extrahierende Feature bereits bekannt ist.

Tiefe neuronale Netzwerke

Die Funktion neuronale Netzwerke ist dieser des menschlichen Gehirns nachempfunden. Sie setzen sich aus Neuronen und Verbindungen zusammen. Wenn von tiefen neuronalen Netzwerken die Rede ist, so wird impliziert, dass es eine beträchtliche Anzahl versteckter Schichten, sogenannter «hidden Layers» gibt, die eine Maschine braucht, um die komplexen Funktionen zu lernen (Minar & Naher, 2018, S. 5).

Ein tiefes künstliches neuronales Netzwerk kann auch bei einer Architektur mit der Tiefe von einem einzigen «hidden Layer» bleiben, aber je nach Komplexität der Aufgabe muss der Layer in der Folge sehr breit sein. Generell gilt, dass tiefe, aber dafür weniger breite Netze das Lernen der gesuchten Parameter erleichtern (Schindler, Multivariate Statistik und Machine Learning, 2021b).

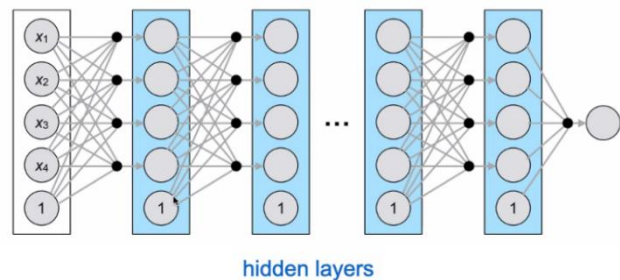


Abbildung 1 Vereinfachte Illustration eines tiefen neuronalen Netzwerks mit mehreren Layern (Schindler, 2021)

Convolutional Neural Network

18-Layer ResNet

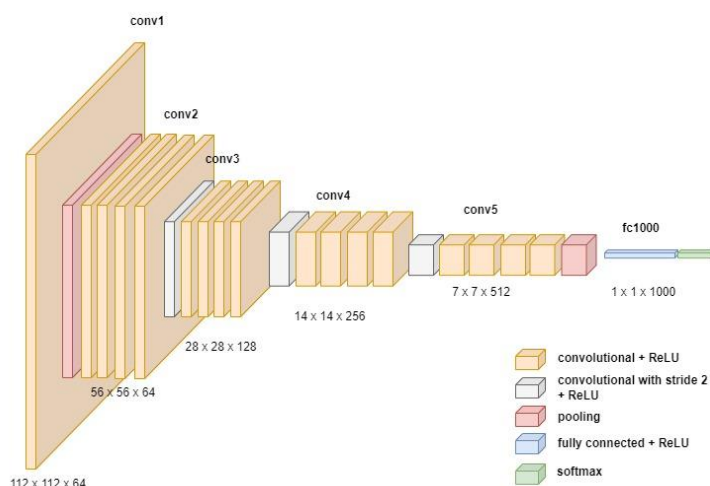


Abbildung 2: Beispiel eines Convolutional Neural Network

Ebenfalls ausschlaggebend sind die Pooling Layer, die den Output dezimieren, wie sie auch in der Abbildung 2 illustriert sind. Sie verursachen einen Informationsengpass, was das Modell dazu zwingt, die relevanten Informationen mit weniger Zahlen bestmöglich zu erhalten. Und als letzte Kernidee gilt das Verwenden vieler Layer (LeCun, Bengio, & Hinton, 2015, S. 4; Schindler 2021b).

Convolutional Neural Networks (CNN), also faltende neuronale Netzwerke zeichnen sich dadurch aus, dass die Layer anstatt mit Matrixmultiplikationen durch Faltungen miteinander kombiniert werden. CNNs stützen sich auf vier Kernideen. Erstens hat man eine Lokalität. Man kann die Räumlichkeit der Daten ausnützen und mit dem ersten Gesetz der Geografie, Tobler's Law, bestimmen, welche Daten grösseren Einfluss nehmen. Es werden nur die Werte miteinander kombiniert, die auch nahe beieinander liegen. Eine weitere Idee ist das «weight sharing», also das Teilen von derselben Gewichtung an den

2.2 Objektklassifizierung

2.2.1 Feature Classifier

Der Feature Classifier ist das einzige Objektklassifizierungsmodell, welches zum jetzigen Zeitpunkt von ArcGIS Pro zur Verfügung gestellt wird. Bei der Objektklassifikation sollen die Objekte, also die einzelnen Wetland-Features, einer Klasse zugewiesen werden. Input braucht der Feature Classifier in Form von einem Eingabe-Raster mit den Spektralbändern und ein Feature-Class-Layer für die Position der Features (Esri, 2022a).

Im Gegensatz zu den «Deep Learning»-Modellen zur Pixelklassifizierung wird hier keine spezielle Modellarchitektur eingesetzt. Es handelt sich damit um eines der Standardprobleme der Bildverarbeitung, bei dem ein Bild einer oder mehreren Klassen zugewiesen werden soll. Die Architektur des Feature Classifiers hängt damit vollkommen von der Wahl des Backbone-Modells ab. Hierbei wird aber empfohlen, auf eines der vielen zur Auswahl stehenden Convolutional Neural Networks zurückzugreifen. Dabei hat man die Auswahl aus verschiedenen DenseNet Versionen, MobileNet -Version 2, verschiedenen ResNet Versionen, verschiedenen VGG Versionen, verschiedenen Reid Versionen und DarkNet-53 (Esri, 2022a; Esri, 2022c).

Die Trainingsdaten, die sich per ArcGIS Pro Werkzeug «Export Training Data for Deep Learning» generieren lassen, müssen vom Typ Labeled_Tiles oder 'MultiLabeled_Tiles' sein, damit das Modell damit trainiert werden kann (Esri, 2022b).

2.3 Pixelklassifizierung

Wie bei der Objektklassifizierung braucht es zwei Input Layer. Analog zum Feature Classifier braucht man ein Raster Layer der Historischen Karte und ein Layer der Klassifizierten Feuchtgebiete. Damit Trainingsdaten, die sich mithilfe des ArcGIS Deep Learning Toolsets exportieren und aufbereiten lassen, von allen nachfolgenden Pixelklassifizierungsmodellen verwenden lassen, müssen sie mit den Metadaten vom Typ Classified_Tiles extrahiert werden

Je nach Modell sind auch hier spezifische Backbone CNN Modelle vorgesehen.

2.3.1 U-Net

Die U-Net-Architektur besteht aus zwei symmetrischen Hälften in Form von Encoder und Decoder, die damit ihre typische U-Form bilden, wie sie auch in Abbildung 3 zu sehen ist. Der Encoder ist in der Regel ein normales Convolutional Neural Network mit zusätzlichem Maxpool Downsampling. Sobald der Switch zum Decoder gemacht wird, versucht man eine möglichst dichte Klassifikation der Pixel zu erreichen. Es wird dabei wieder soweit hoch gesampelt, sodass die Auflösung wieder gleich hoch ist wie zu Beginn und damit erhalten bleibt (Esri, 2022d).

Das U-Net Model zeichnet sich dadurch aus, dass es nicht wie viele herkömmliche

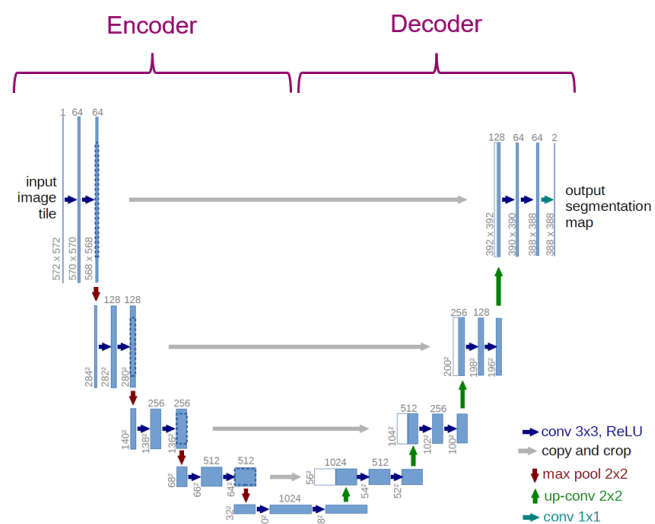


Abbildung 3: U-Net Architektur (Ronneberger et al. 2015)

Convolutional Networks einem Bild eine einzelne Klasse zuweist, sondern diese Klassifizierung pixelweise vornimmt. Es handelt sich bei der U-Net-Architektur um ein «fully convolutional network», da alle Schichten per Faltungsoperation miteinander verbunden sind. Beim Hochsamen im Decoder ist wichtig, dass viele Feature Kanäle vorhanden sind, denn diese propagieren die Kontext Information in die Schichten mit den höheren Auflösungen hinauf. Gerade für Aufgabenstellungen mit wenig vorhan-

denen Trainingsdaten ist U-Net eine geeignete Architektur, da sie mit wenig Trainingsdaten gute Resultate erzielen kann, falls diese angemessen augmentiert werden. Jedoch gilt es bei dieser Architektur zu bedenken, dass jeder Schritt in ein tieferes Layer im Encoder dazu führt, dass im Decoder durch das Upsampling etwas Information verloren geht. Für Modelle mit tiefen neuronalen Netzwerken muss beachtet werden, dass die Gewichte, die Parameter der Prädiktionsfunktion gut gewählt sind. Sonst kann es zu unausgebalancierten Modellen kommen (Ronneberger, Fischer, & Brox, 2015). Das U-Net-Modell ist nach wie vor eine sehr verbreitete Architektur und wird auch als Basis für viele der anderen Pixelklassifizierungsmodelle als Grundgerüst verwendet.

2.3.2 Pyramid Scene Parsing Network

PSPNet zeichnet sich dadurch aus, dass der globale Kontext des Bildes bei der Pixelklassifizierung in Betracht gezogen wird. PSPNet kann zum Beispiel schlussfolgern, dass autoähnliche Pixel bei/in einem Fluss nicht als Auto, sondern als Boot klassifiziert werden sollen, da ein Auto im Wasser nicht logisch wäre. Der Begriff Pyramid rührt daher, dass beim Pooling während dem Encoden vier verschiedene Grössen verwendet werden. Je grösser der Filter, umso grösser ist die räumliche Auflösung. Kleine Filter mit tieferer Auflösung erkennen somit grössere Features, während die grösseren Filter auch detailliertere Features erkennen können. Der Decoder dieser Architektur entspricht bei der ArcGIS Implementation ungefähr dem der U-Net Architektur (Esri, 2022e).

Wie die U-Net-Architektur ist auch PSPNet ein Fully Convolutional Neural Network und somit nur über Faltungsoperationen zwischen den Layern verbunden. Und ebenfalls wie beim U-Net-Modell wird die Auflösung des Bildes beibehalten. Um das Sichtfeld der bei dem Encoden möglichst lange möglichst gross beizubehalten, um auch so viele Informationen zum räumlichen Kontext wie möglich mitzunehmen, wird bei PSPNet mit Atrous-Faltung gearbeitet. Dabei liegen die Filtergewichte nicht eng nebeneinander, sondern mit Abstand, um die Auflösung nicht so schnell zu verkleinern. Darum handelt es sich auch um ein Modell, welches sich für tiefe Modell-Architekturen eignet, da auch bei grossen Tiefen noch genug räumliche Auflösung vorhanden ist (Zhao, Shi, Qi, Wang, & Jia, 2016).

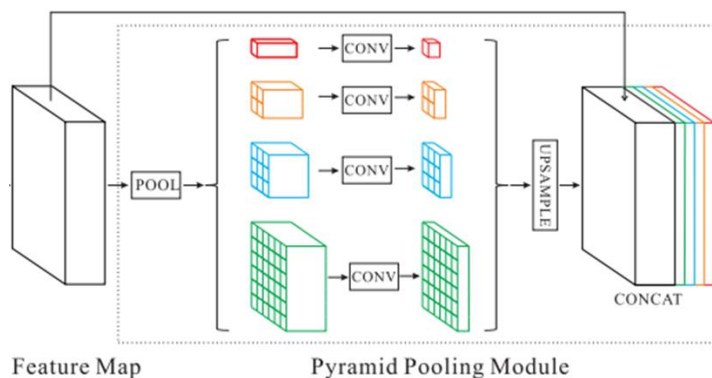


Abbildung 5: PSPNet Modell Struktur. (Zhao, Shi, Qi, Wang, & Jia, 2016)

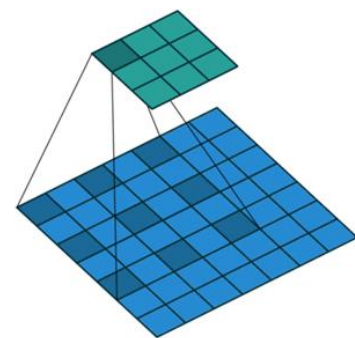


Abbildung 4: Atrous-Faltung. (Zhao, Shi, Qi, Wang, & Jia, 2016)

2.3.3 DeepLabV3

Im Gegensatz zu PSP- oder U-Net gibt es auch Modelle, deren Output nicht dieselbe Grösse aufweist, wie der Input. Durch all die Convolutions nimmt die Auflösung mit jedem Block ab. Auch weil es sich bei DeepLabV3 ebenfalls um ein Fully Convolutional Neural Network handelt. Wird dies nicht via Decoder wieder hochgesampelt, bleibt die räumliche Auflösung reduziert. Bei der Anwendung des Deep LabV3 Modells wird die Ausgangsauflösung nicht wiederhergestellt, aber durch Atrous-Faltung, wie sie auch bei PSPNet eingesetzt wird, bleibt ein gewisses Mass an räumlicher Auflösung erhalten. Der Filter wird

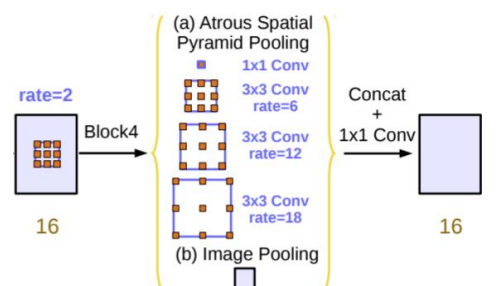


Abbildung 6: Atrous Spatial Pyramid Pooling network. (Chen, Papandreou, Schroff, & Hartwig, 2017)

dabei bei jeder Iteration so weit ausgedehnt, dass die Auflösung gleichbleibt. Der Bereich, der damit abgedeckt wird, wird somit grösser, was bessere Resultate liefern soll. Der genaue Einfluss wird in Abbildung 7 visualisiert. Speziell bei sehr tiefen Netzen und vielen Faltungen kann so eine Massnahme nützlich sein (Esri, 2022f).

Nebst der Atrous-Faltung teilen sich die beiden «Deep Learning»-Modelle auch das Pooling via Pyramiden Modul, wie es in Abbildung 6 gezeigt ist (Chen, Papandreou, Schroff, & Hartwig, 2017).

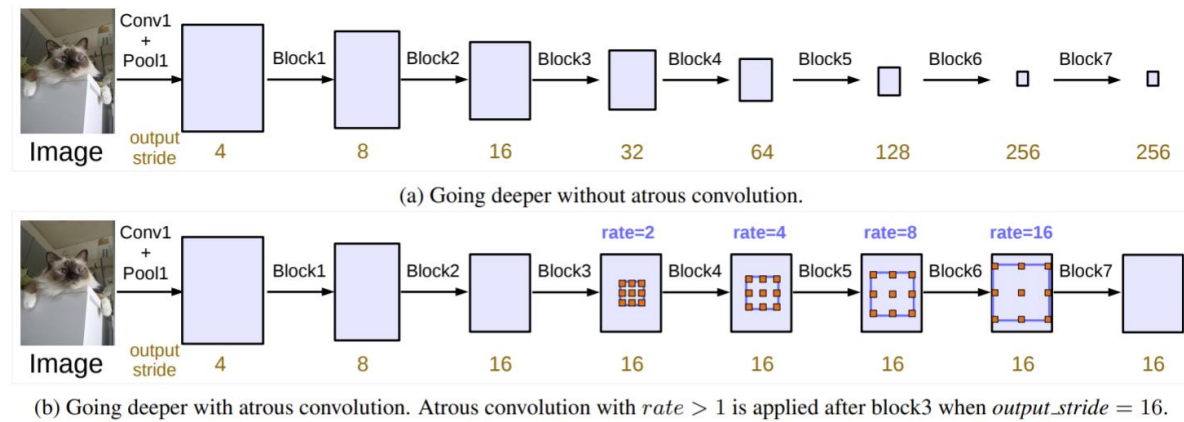


Abbildung 7: DCNN und der Einfluss von Atrous-Faltung. (Chen, Papandreou, Schroff, & Hartwig, 2017)

2.3.4 Bi-Directional Cascade Network for Edge Detector

Bei diesem Modell liegt der Fokus auf der Edge Detection von Objekten mit variierender Grösse. Die einzelnen Layer werden hierbei nicht in einer einzigen Standardgrösse betrachtet, sondern bei jedem Layer entsprechend angepasst (Esri, 2022g).

Je nach Tiefe im Modell werden andere Eigenschaften hervor-gehoben, wie in der Abbildung 8 angedeutet wird. Für die oberflächennahen Schichten werden eher die Details erkannt, zum Beispiel die Streifen eines Zebras oder das Muster auf einem Schmetterlingsflügel, Während tiefere Schichten eher die Umrisse dieser Tiere erkennen würden (He, Zhang, Yang, Shan, & Huang, 2019).

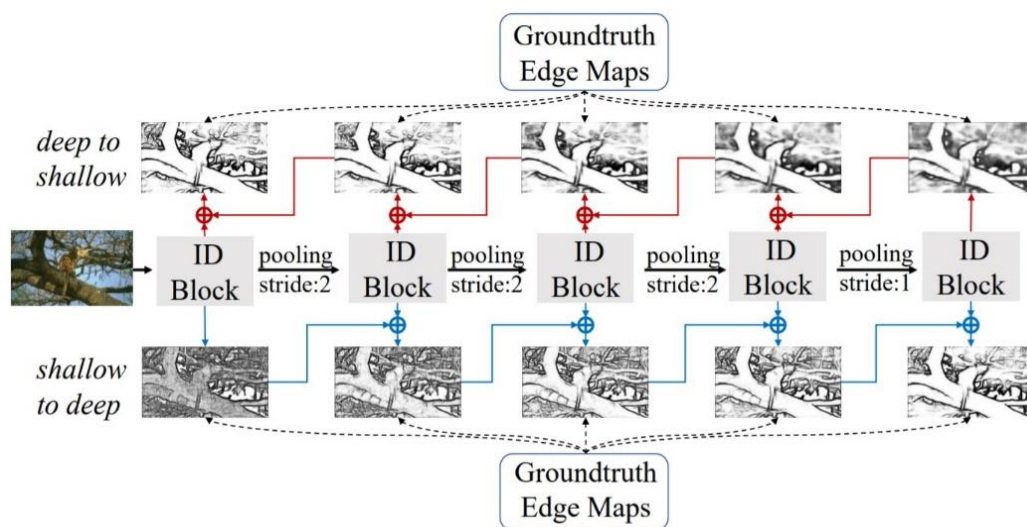


Abbildung 8: Architektur des Bi-Directional Cascade Network for Edge Detector. (He, Zhang, Yang, Shan, & Huang, 2019)

2.3.5 Hollistically-Nested Edge Detector

Der Hollistically-Nested Edge Detector ist der Vorläufer von BDCN und war einer der ersten CNN-basierten Edge-Detektor. Das hereingegebene Bild unterscheidet sich ganzheitlich vom Output. Das Modell baut auf fünf Convolution Blocks auf, wie in der Abbildung 9, die alle einen Side-Output hervorbringen. Diese werden verschmolzen, was dem endgültigen Output entspricht (Esri, 2022g).

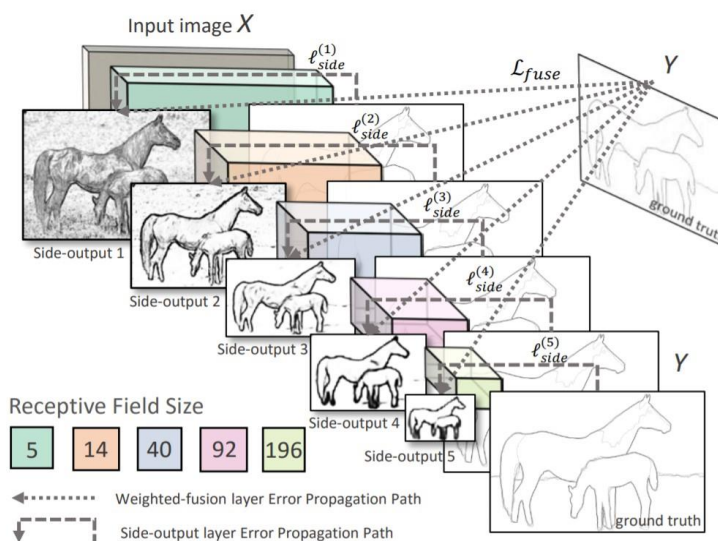


Abbildung 9: Architektur des Hollistically Nested Edge Detection. (Xie & Tu, 2015)

2.3.6 ConnectNet & Multi Task Road Extractor

Da diese Modelle beide auf derselben Architektur des Papers «Improved Road Connectivity by Joint Learning of Orientation and Segmentation» (Batra, et al., 2019) und die jeweiligen Einträge in der «Arc GIS API for Python»-Dokumentation identisch sind, kann davon ausgegangen werden, dass die Modelle ähnliche Resultate hervorbringen. Der Schwerpunkt der Modelle liegt darin, einzelne Segmente einer kontinuierlichen Linie zu einem Ganzen zusammenzufügen. So sollen zum Beispiel Strassen-segmente, die durch störende Vegetation nicht als zusammenhängendes Stück erfasst werden, entsprechend ausgebessert werden. Neben dem Identifizieren von Strassenpixeln wird zusätzlich auch die wahrscheinlichste Richtung/Orientierung der Strasse festgehalten, sodass die Connectivity der Strassen besser abgeschätzt werden kann.

Die Modell Architektur basiert auf stacked multi-branch Encoder-Decoder Blöcken. Die Stacks teilen sich zu Beginn einen Encoder, aber der Decoder, der den Output bestimmt, ist für die verschiedenen Stacks individuell. Dadurch auch die Modell-bezeichnung mit Multi-Task (Esri, 2022h).

2.3.7 Change Detector

Mit dem Change Detector kann ein Modell ausgeben, ob sich der Pixel verändert hat. Ein klassischer Input für dieses Modell wären also zwei Bilder, die an verschiedenen Zeitpunkten aufgenommen wurden. So können zum Beispiel Neubauten auf Anrieb identifiziert werden. Der Output ist aber limitiert auf die beiden Werte *change* oder *no change*. Es können nur binäre Veränderungen dokumentiert werden (Esri, 2022i).

Die Architektur des Modells greift dabei ebenfalls Elemente des U-Net-Aufbaus, in Form von En- und Decoder, basiert aber grösstenteils auf dem STANet Network, die in Abbildung 10 veranschaulicht wird. (Chen & Shi, 2020)

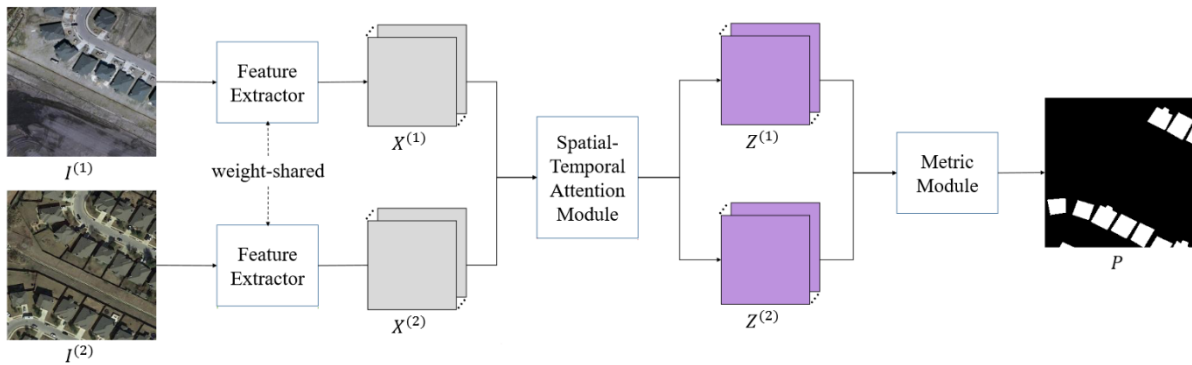


Abbildung 10: STANet Network Architektur (Chen & Shi, 2020)

3 Methoden und Vorgehen

3.1 Vorauswahl der Modelle

Um die Übersichtlichkeit zu erhalten und einen gewissen Fokus beim Trainieren der Modelle setzen zu können, war es im Rahmen dieser Bachelorarbeit nicht möglich, aber auch nicht angedacht, alle neubetrachteten «Deep Learning»-Modelle vertieft zu untersuchen und zu optimieren.

Um eine sinnvolle Vorauswahl treffen zu können, wurden die relevanten «Deep Learning»-Modell alle unter vereinfachten Bedingungen auf ihre Eignung bezüglich der Extraktion von Wetland-Merkmalen aus historischen Karten überprüft. Dabei wurde zwar die Grundstruktur wie sie in Abbildung 22 dargestellt ist, eingehalten, dabei sollte die Komplexität des Modells sollte möglichst tief bleiben. So wurden die neun «Deep Learning»-Modelle nach ihrem Standardsetup über 20 Epochen mit einem Trainingsatz à 969 Image-Chips trainiert und anschliessend klassifiziert.

3.1.1 Aussortierte Modelle

Feature Classifier

Der Feature Classifier von ArcGIS Pro bietet im Grunde keine Möglichkeit, die Wetland-Features aus den historischen Kartenwerken zu extrahieren. Der Anwendungszweck dieses Modells liegt in der Klassifizierung von Bildern, nicht dem Lokalisieren von Merkmalen. Durch die Information, dass sich irgendwo auf einer Bildkachel ein Stück Wetland-Feature befindet, lassen sich die gesuchten Features nicht extrahieren.

Es wäre aber durchaus vorstellbar, ein Modell vom Typ Objektklassifizierung einzusetzen, um die von anderen Modellen als Wetland klassifizierten Features wiederum zu klassifizieren, die Übereinstimmung zu prüfen und allfällige Fehlklassifizierungen festzustellen und zu beheben. Mit dem rudimentären Training ist die Qualität des Objektklassifizierungen aber noch nicht zufriedenstellend, wie auch der Abbildung 11 entnommen werden kann.



Abbildung 11: Output Feature Classifier: Grüne Kacheln: Wetlands, Rosa Kacheln: Keine Wetlands

Bi-Directional Cascade Network for Edge Detector & Holistically-Nested Edge Detector

Auch die beiden Edge Detektoren sind nicht die optimale Modellwahl, um die Wetland-Feature-Pixel korrekt zu klassifizieren. Betrachtet man Abbildung 12 und Abbildung 13, so lassen sich vor allem bei dem BDCN Edge Detektor gute Ansätze feststellen. Der Anteil an Falschklassifikationen ist jedoch bei beiden Modellen hoch. Der angedachte Anwendungsbereich der Edge Detektoren ist auch das Extrahieren von Kanten. Falls also Flüsse, Strassen oder andere linienähnlichen Objekte von Karten extrahiert werden sollen, würde sich dieses Modell sicherlich zur erneuten Eruiierung der Performance bei einer anderen Aufgabenstellung anbieten.



Abbildung 12 Pixelklassifizierung mit BDCN Edge Detection

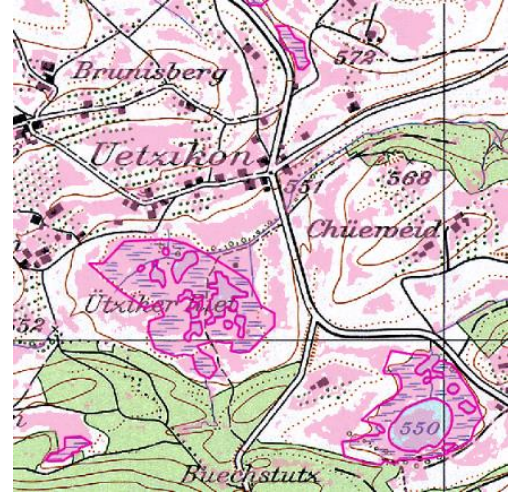


Abbildung 13 Pixelklassifizierung mit HED

ConnectNet & Multi Task Road Extractor

Auch die Auseinandersetzung mit den Modellen ConnectNet und Multi Task Road Extractor enden hier durch die gescheiterte Vorselektion. Auch hier zeigen die Modelle klares Potenzial, aber auch hier zeigt sich in den Abbildung 14 und Abbildung 15, dass diese Modelle für andere Aufgabentypen entwickelt wurden. ConnectNet und Multi Task Road Extractor sind ebenfalls dafür ausgelegt, linienhafte Merkmale zu klassifizieren.

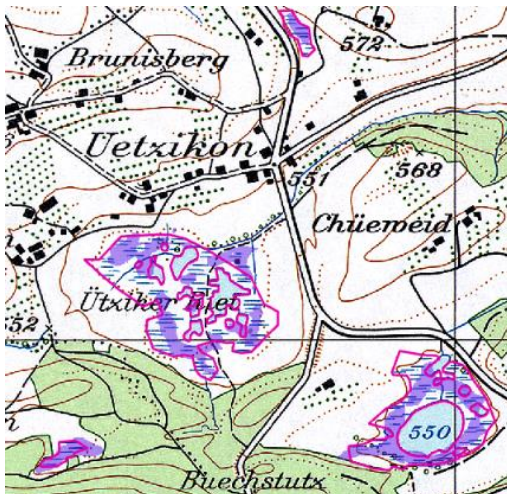


Abbildung 14: Pixelklassifizierung mit ConnectNet

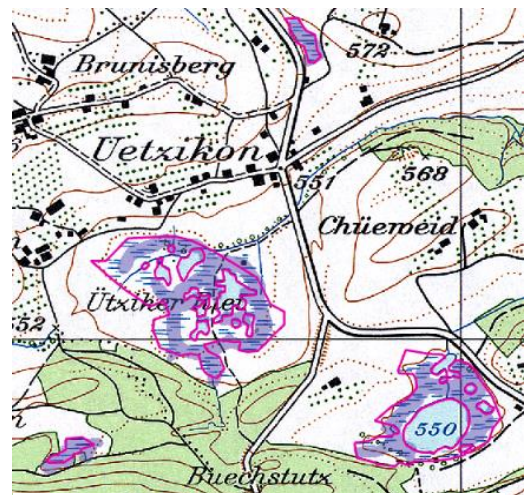


Abbildung 15: Pixelklassifizierung mit Multi Task Road Extractor

ChangeDetection

Mit ChangeDetection können keine Merkmale aus Karten extrahiert werden. Dieses Modell ist dafür ausgelegt, die binären Unterschiede zweier Raster-Bilder aufzuzeigen, die am selben Ort, aber zu verschiedenen Zeiten aufgenommen wurden. Eine alternative Anwendung dieses «Deep Learning»-Modells wäre zum Beispiel dadurch möglich, dass sich so aufzeigen liesse, wie sich der Bestand der Wetlands in einer bestimmten Zeitspanne verändert hat. Entweder in Form Vergleichen verschiedener historischer Zeitpunkte oder der Anwendung des ChangeDetectors auf verschiedene gerasterten Feature Layern.

3.1.2 Ausgewählte Modelle

Die vertiefte Untersuchung und Auseinandersetzung mit den verschiedenen «Deep Learning»-Modellen wird sich im Zuge dieser Arbeit also auf die drei Modelle U-Net, Pyramid Scene Parsing Network und DeepLabV3 fokussieren. Bereits mit wenig Trainingsdaten und dem unveränderten Standardsetting zeigt sich in den Abbildung 16, Abbildung 17 Abbildung 18 bereits ein grosses Mass an Präzision. Die Falschklassifizierungen sind bei allen drei Modellen kaum auszumachen. Ähnliches lässt sich zwar auch für die Klassifikationen in den Abbildungen Abbildung 15 Abbildung 14 sagen, aber so wie sich die klassifizierte Wetland-Feature der selektierten Modelle präsentieren, wird klar, dass sie die zu erwartende Form solcher Merkmale auf natürlichere Art und Weise emulieren, die Kanten und Ausdehnungen der Flächen realistischer ausgestaltet scheinen.

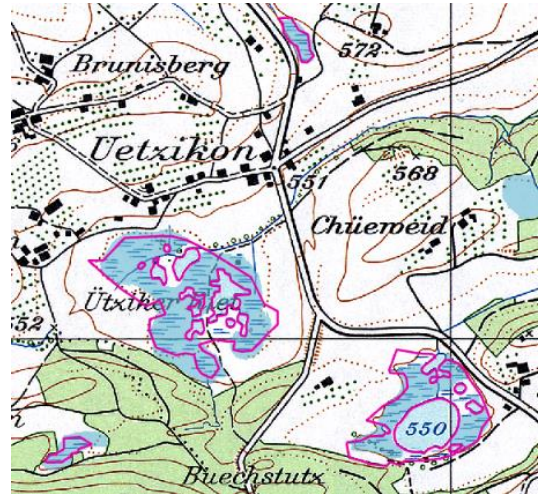


Abbildung 16: PSP Pixelklassifizierung

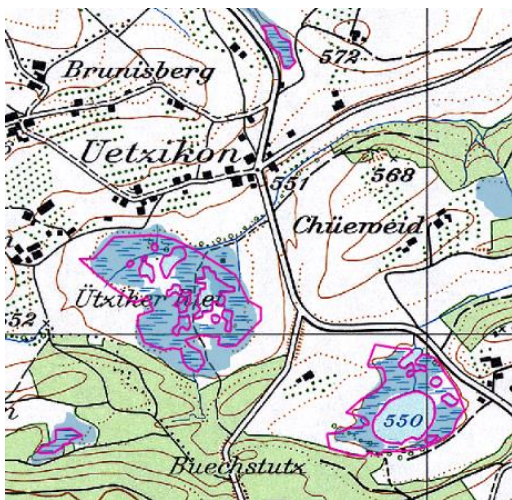


Abbildung 18: DL Pixelklassifizierung

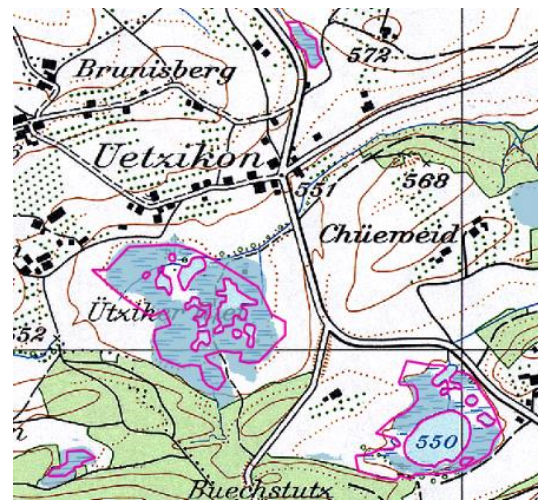


Abbildung 17: U-Net Pixelklassifizierung

3.2 Grundlagedaten

3.2.1 Historische Geodaten

Für die Extrahierung von Merkmalen aus historischen Karten stehen im Zuge dieser Arbeit 288 Kartenblätter der Schweizer Landeskarte im Massstab 1:25'000 im Tagged Image File Format (.TIF) zur freien Verfügung. Diese enthalten jedoch auch die Blätter ganz am Rande der Schweiz, auf denen nur noch ganz kleine Landesteile zu sehen sind, die der Grenze entlanglaufen. Die Erstaussgabe dieses Kartensatzes war 1952. Die TIF-Kacheln, die in dieser Arbeit verwendet werden, sind aber bereits Nachführungen aus den 1980er und 1990er Jahren (Bundesamt für Landestopografie swisstopo,

2022). Die Kartenkacheln haben alle eine Dimension von 14'000 x 9'600 Pixel und eine Auflösung von 508 dpi. 1 Pixel auf der Karte entspricht somit 1.25 Meter. Die Farbtiefe liegt bei 24, was heisst, dass über 16 Millionen verschiedenen Farbwerte dargestellt werden können.

3.2.2 Feature-Layer

Bei den Merkmalen, die aus den Ausschnitten der Schweizer Landeskarten extrahiert werden sollen, handelt es sich um die Wetland-Feature, also Sumpf- und Feuchtgebiete.

Für das Trainieren der Daten steht eine Shapefile-Datei (.shp) dieser Schweizer Wetlands zur Verfügung, die den Bestand im Jahr 2010 abbildet. Die Rekonstruktion dieser Feuchtgebiete von Stuber & Bürgi (2018) basiert dabei auf den digitalisierten Kartenblättern der Schweizer Landeskarte. Die Erfassung und Extraktion von Merkmalen ist also auf die Featureklasse der Wetlands limitiert.

In den historischen Landeskarten werden die Feuchtgebiete mit blauen kurzen variierenden parallelverlaufenden, aber versetzten Strichen symbolisiert (Bundesamt für Landestopografie swisstopo, 2022). Dabei gilt es zu erwähnen, dass durch die grosse zeitliche Spannweite zwischen den nachgetragenen Landeskartenausschnitten der 1950er und dem Wetland Datensatz aus der jüngsten Vergangenheit zu gewissen fehleranfälligkeiten kommt, da das Training mit dem fehlerbehafteten Feature Layer ausgeführt wird.¹

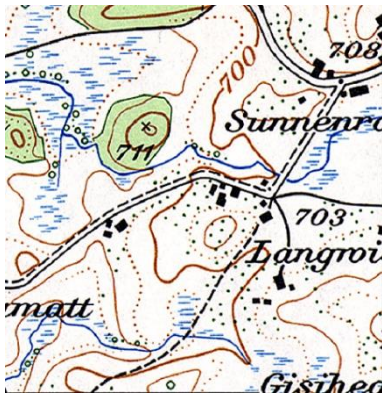


Abbildung 19: Historische Karte

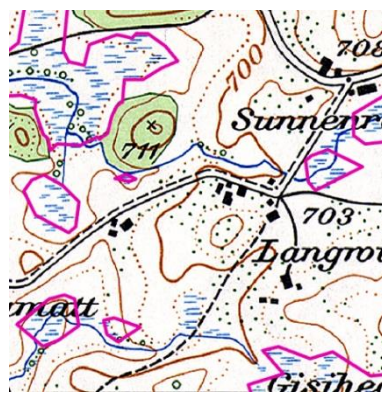


Abbildung 20: Datensatz Wetlands



Abbildung 21: Manuell erfasste Wetlands

¹ Dies ist darauf zurückzuführen, dass es mir nicht bewusst war, dass wir Zugriff auf einen Fehlerlosen Datensatz haben. Das korrigierte Feature Layer spielt somit erst ab der Klassifikation der Pixel eine Rolle.

3.3 Grundstruktur

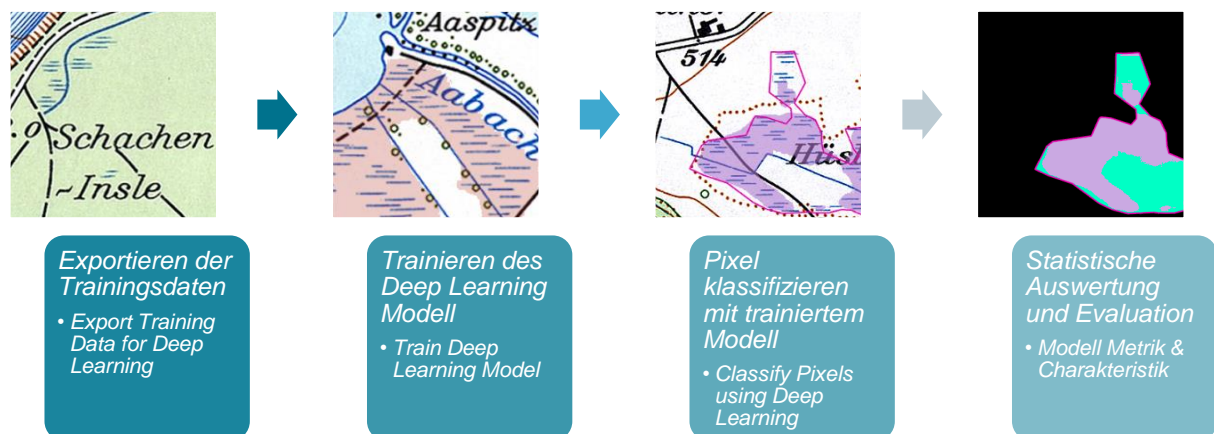


Abbildung 22: Illustration der Grundstruktur für das Trainieren von DL-Modellen zur Pixelklassifizierung in ArcGIS Pro

Der Ablauf rund um das Training eines «Deep Learning»- Modells in ArcGIS orientiert sich, zumindest im Hinblick dieser Arbeit, immer rund um den Ablauf, wie er in Abbildung 22 abgebildet ist. Je nach Workflow, Modelltyp und Anwendungszweck, variieren natürlich Einflussgrößen wie die Wahl der Parameter, die verlangten Inputformate oder der Output, aber der Grundaufbau bleibt bestehen.

Mit den Grundlagedaten in Form des Wetland-Feature-Layers als Shape-File und den historischen Kartenblättern der Schweizer Landeskarte von 1952 im TIF-Format werden die Trainingsdaten mithilfe des «Export Training Data for Deep Learning»-Werkzeugs erstellt. Je nach Modelltyp muss hier das Metadatenformat der resultierenden Trainings-Image-Chips angepasst werden, aber die erzeugten Trainingsdaten behalten das TIF-Format jeweils bei. Die Grösse der dabei erzeugten Quadrate hängt dabei von der gewählten Kachelgrösse ab. Die generierten Image-Chips bilden dann die Grundlage für den nächsten und gleichzeitig zentralsten Schritt der Grundstruktur, dem eigentlichen Training des gewählten Modells. Dabei durchlaufen die Modelle die Trainingsdaten entweder bis eine vorher festgelegte maximale Epochenanzahl erreicht ist, oder bis die Veränderungen der Parameterwerte so klein sind, dass bei weitem absolvierten Epochen keine weitere Verbesserung des Modells mehr zu erwarten ist. Diese Abschätzung wird jeweils mithilfe der Validierungsdaten, also dem festgelegten Prozentsatz der Trainings-Chips, die nur für die Modellüberprüfung verwendet werden und somit nur implizit in das eigentliche Training des Modells hineinwirken. Als Output erhält man nach Beendigung des Trainingsprozesses ein Ordner mit einem Deep Learning Model Package (.dlpk) welches die Dateien und Daten enthält, die erforderlich sind, um allfällige Deep-Learning-Inferencing-Tools auszuführen. Ebenfalls braucht es noch eine Esri Model Definition Datei (.emd) und die Datei mit dem trainierten Modell. Hier hängt das Ausgabeformat wiederum vom gewählten Setup ab. Um die Güte des trainierten Modells abschätzen zu können, stellt ArcGIS auch die generierten Metriken zur Verfügung. Was alles ausgegeben wird, ist wiederum von den gewählten Parametern abhängig.

Das trainierte Modell kann nun entweder direkt für die Klassifizierung von Pixeln eingesetzt werden, oder mit zusätzlichen Trainingsdaten weiter verbessert werden. Der Input für die Klassifizierung ergibt sich nun aus dem Deep Learning Model Package und dem Raster, dessen Pixel klassifiziert werden sollen. Als Output erhält man ein Raster-Layer mit den entsprechend klassifizierten Pixeln.

Der Letzte Baustein der Grundstruktur, die Statistische Auswertung und Evaluation der Pixelklassifizierung, erlaubt eine differenziertere Analyse der eigentlichen Eignung des Modells und dem dazugehörigen Setup, als auch der Vergleich zwischen der realen erreichten Genauigkeit der Klassifikation und den von ArcGIS generierten Metriken der Trainingsphase.

3.4 Hardware & Software

3.4.1 Hardware

Die Zeit, die für den Trainingsprozess eines Modells in Anspruch genommen wird, soll für die grösstmögliche Effizienz möglichst kurz sein. So ist es von Vorteil, über eine leistungsstarke Maschine mit einem schnellen Prozessor und einer daraus resultierenden hohen Rechenleistung zu verfügen. Falls der Prozessor des PCs allein nicht reicht oder die Geoverarbeitungen zu lange dauern, wird zeitgleich auch vom Grafikprozessors der Maschine Gebrauch gemacht, um die Verarbeitungsdauern zu verkürzen (Esri, 2022m).

Das Ausführen der «Deep Learning»-Operationen mit überdurchschnittlich hohem Bedarf an Verarbeitungsleistung verlangt nach einem leistungsintensiveren PC, wie dem IKG TX 1070 der ETH Zürich. Dieser verfügt über einen Intel Xeon Prozessor mit 32GB Random Access Memory. Auch die Parallelverarbeitung mit dem zusätzlichen Grafikprozessor, einem NVIDIA GeForce GTX 1070 mit 8 GB GPU Memory. Prozesse, die sich auf herkömmlichen Laptop über Tage erstrecken können, sind dank adäquater Rechenleistung in wenigen Stunden fertig ausgeführt.

3.4.2 Software

Mit ArcGIS Pro 2.9.0 steht ein Geoinformationssystem zur Verfügung, das sehr viele Fähigkeiten in einer Software vereint. Für die gesamte hier umschriebene Bachelorarbeit gibt es kaum Arbeitsschritte, die sich nicht mit einem Tool des ArcGIS Frameworks lösen lassen.

ArcGIS Image Analyst Toolbox & ArcGIS Deep Learning Toolset

Als Grundlage dieser Arbeit dient die ArcGIS Pro Erweiterung *Image Analyst*. Die Toolbox ist speziell für die Bild- und Geodatenanalysten ausgerichtet.

Für diese Arbeit ist das zur Erweiterung dazugehörige Toolset zum Themengebiet Deep Learning ausschlaggebend. Denn mit den darin enthaltenen Deep-Learning-Workflows und den zur Verfügung gestellten Geoverarbeitungswerkzeugen lassen sich viele komplexe Funktionen und Abläufe externer Deep-Learning-Frameworks wie pyTorch oder TensorFlow über das User Interface von ArcGIS zu steuern sobald, man das die benötigten Deep Learning Framework in Form von den verlangten Deep Learning Libraries² installiert und als neue Python-Umgebung in ArcGIS Pro integriert. Dieses Integrieren externer Open Source Tools führt auch dazu, dass die Bindung ans das ArcGIS API nicht ganz so stark ist und man sich auch bei Bedarf vom User Interface lösen kann, um gewisse Anpassungen vorzunehmen, die einem sonst verwehrt blieben.

Die Deep-Learning-Werkzeuge, so wie sie für die Erkennung oder Klassifizierungen von Features genutzt werden können, sind die folgenden:

- Classify Objects Using Deep Learning
- *Classify Pixels Using Deep Learning*
- Compute Accuracy For Object Detection
- Detect Change Using Deep Learning
- Detect Objects Using Deep Learning
- *Export Training Data For Deep Learning*
- Non Maximum Suppression
- *Train Deep Learning Model*

Relevant für den Arbeitsprozess des Trainierens von «Deep Learning»-Modellen für die Pixelklassifizierung sind dabei der Export der Trainingsdaten, das Trainieren des Modelles und das Klassifizieren der Pixel.

² [Deep Learning Libraries](#)

Für ein besseres Verständnis des ArcGIS Deep Learning Toolsets, und damit auch einem potenziell besseren Endprodukt, ist es notwendig, sich näher mit der arcgis API vertraut zu machen. Einerseits lassen sich allfällige Probleme einfacher lösen, da die Anforderungen an Input und Hyperparameter festgehalten sind. Ebenfalls erschliesst es dem User Optionen, die nicht aus dem User Interface hervorgehen. Je nach Versiertheit und Knowhow lässt sich so die Qualität des Workflows noch einmal anheben.

3.5 Evaluation

3.5.1 Statistische Gütemasse

Um eine optimale Empfehlung abgeben zu können, bedarf es einer ausführlichen und differenzierten Analyse der vorliegenden Metriken. Diese werden anhand von Vergleichen der Vorhersagen mit den Groundtruth-Daten vorgenommen.

Da es sich in dieser Arbeit um eine binäre Klassifikation handelt, eignen sich die zahlreichen Evaluationsmasse der binären Klassifikationen. Die Terminologie ist dabei an die Wahrheitsmatrix (siehe Tabelle 1) angelehnt. Stimmt die Klassifizierung eines Pixels mit der Groundtruth überein, so wird dieser als True Positive oder True Negative gewertet. Handelt es sich um eine fehlerhafte Klassifizierung, so spricht man bei diesem Pixel von False Negative oder False Positive. Positive bezieht sich jeweils auf die Klasse der Wetlands, Negative auf Nicht-Wetlands.

Reale positive Pixel: **P**

Reale negative Pixel: **N**

Korrekt positiv klassifizierte Pixel: true positive **TP**

Korrekt negativ klassifizierte Pixel true negative: **TN**

Fälschlicherweise positiv klassifizierte Pixel: false positive **FP**

Fälschlicherweise negativ klassifizierte Pixel: false negative **FN**

Tabelle 1: Wahrheitsmatrix

	P	N
P _{Classified}	TP	FP
N _{Classified}	FN	TN

Recall

Der Recall beschreibt das Verhältnis der Hits oder Treffer, also der wahren positiven Werten im Vergleich zu den Falschen negativen, also den verpassten Wetland-Pixeln. Ein Recall von 0 bedeutet, dass kein Pixel richtig als Wetland klassifiziert wurde, und 1, dass kein Wetland-Pixel als Nicht-Wetland klassifiziert wurde. Dabei ist ein hoher Recall Wert nicht automatisch als gut einzustufen. Falls ein Modell sehr viele grossräumige Flächen fälschlicherweise als Wetland klassifiziert, so erscheint der Recall Wert durchaus hoch, da nur wenige Pixel verpasst wurden, aber das Bild, dass die klassifizierte Karte abgibt, kann sehr fehlerbehaftet sein.

Im Sinne der Aufgabenstellung und der Extraktion von Wetlands ist es aber sicher im Interesse, des Users, dass möglichst viele Wetlands in ihrer Vollständigkeit aus den Geodaten extrahiert werden können.

$$Recall = \frac{TP}{TP + FN}$$

(Powers, 2020, S. 2)

Selektivität

Die Selektivität ist das Gegenstück zum Recall. Hier wird betrachtet, wie selektiv das Modell in seiner Klassifikation von positiven Werten, also von Wetlands vorgeht. Ein Wert von 1 bedeutet hier wiederum, dass das trainierte Modell kein Pixel als Wetland klassifiziert hat, welches auch keines ist. Eine 0

bedeutet hier, dass alle Nicht-Wetland-Pixel fälschlicherweise als Wetland identifiziert wurden. Durch das stark gestörte Gleichgewicht von Wetland und Nicht-Wetland-Pixeln ist dies je nach Kartenblatt eher eine Metrik mit einer tieferen Aussagekraft. Die wahren negativen Werte werden immer klar in der Überzahl sein, so dass auch eine Häufung von False Positives nicht ins Gewicht fallen

$$Selectivity = \frac{TN}{TN + FP}$$

(Powers, 2020, S. 3)

Präzision

Die Präzision spiegelt wider, wie das Verhältnis zwischen richtig und falsch klassifizierten Wetland-Pixeln aussieht. Eine 1 bedeutet, dass alle Pixel, die als Wetland klassifiziert werden, auch diesen wahren Wert haben. Eine 0 bedeutet, dass keine Wetland-Klassifikation korrekt ist.

$$Precision = \frac{TP}{TP + FP}$$

(Powers, 2020, S. 2)

Accuracy und Balanced Accuracy

Bei den Accuracies, den Genauigkeiten, gibt es andererseits die reguläre Accuracy, die zeigt, wie viele der gesamten Anzahl Pixel auch der richtigen Klasse zugeordnet wurden

Da auch hier die vielen leeren Pixel viel stärker ins Gewicht fallen, und auch eine schlechte Klassifikationsbilanz gut aussehen lassen, gibt es auch die Ausgewogene Accuracy, die diesem Ungleichgewicht entgegenwirken soll.

$$Accuracy = \frac{TP + TN}{P + N}$$

$$Balanced Accuracy = \frac{Recall + Selectivity}{2}$$

(scikit-learn, 2022)

Phi-Koeffizient

Der Phi-Koeffizient berücksichtigt die alle erhaltenen Werte und stellt diese auf eine ausgewogenen dar. Im Gegensatz zu den anderen binären Metriken können die erreichten Werte auch negativ ausfallen. Dabei beschreibt die 1 eine perfekte Vorhersage, während 0 dem entspricht, was bei einer zufälligen Klassifikation erreicht wird. Eine -1 ist die Inverse der Klassen und bedeutet, dass jeder Pixel falsch klassifiziert ist (scikit-learn, 2022).

$$\varphi = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Jaccard Index/ Intersection over Union

Bei der Intersection over Union wird angegeben, wie gross die Übereinstimmung zwischen zwei Gebieten ist. Hier in dieser Form wird die Anzahl Hits abgewogen gegenüber den fälschlicherweise ausgelassenen, und den fälschlicherweise klassifizierten Pixeln. Die FN sollten ja zur Fläche dazu gehören, wären der Flächenanteil der FP weg müsste. Bei 1 ist der Overlap komplett, bei 0 nicht vorhanden.

$$\text{Intersection Over Union IOU} = \frac{TP}{TP + FN + FP}$$

(Powers, 2020, S. 3)

F1-Score

Der Wert des F1-Scores kann als harmonischer Schnitt von Recall und Präzision interpretiert werden (Powers, 2020, S. 4).

$$F1 \text{ Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

False positive und False negative rate

Es lässt sich auch noch separat betrachten, wie die falsch klassifizierten Pixel ins Gewicht fallen.

$$\text{False positive rate} = \frac{FP}{N} \quad \text{False negative rate} = \frac{FN}{P}$$

(Powers, 2020, S. 3)

False discovery und false omission rate

Die False discovery & omission rate hält fest, in welchem Masse Pixel fälschlicherweise als Wetlands klassifiziert wurden, respektive wie die weggelassenen Wetland Pixel ins Gewicht fallen.

$$\text{False discovery rate} = \frac{FP}{FP + TP} \quad \text{False omission rate} = \frac{FN}{FN + TP}$$

(Powers, 2020, S. 3)

3.5.2 Gütemasse in ArcGIS Pro

Für jedes «Deep Learning»-Modell, das trainiert wird, werden in ArcGIS Pro kontinuierlich Daten gesammelt, Metriken überwacht und ausgewertet.

Diese Gütemasse zur Einstufung der Modelleistung werden dem User, wenn auch etwas limitiert, zur Verfügung gestellt.

Dabei variiert die Modellbewertung in Zugänglichkeit und Komplexität. Während die klassifizierten Validationsdaten im direkten Vergleich, wie in Abbildung 23 beispielhaft demonstriert, vermitteln, ob ein Modell die Aufgabenstellung erfüllt oder nicht. Rein optisch lässt sich ein Modell aber kaum von Auge direkt einschätzen.

Für die Einschätzung des Trainingsverlaufs wird auch die Verlustfunktion über alle Epochen aufgezeigt. Idealerweise ist diese weder zu steil noch zu flach. So wie sie in der Abbildung 24 geplottet wurde, entspricht etwa der angepeilten Verlaufskurve.

Die klassischen Gütemasse Präzision, Recall und F1-Score werden wie in Abbildung 25 gezeigt, für jedes Modell zur Verfügung gestellt. Für die Identifikation des optimalen Modells eignen sich diese bewährten Modellcharakteristiken am besten, da so ein Modellübergreifender Vergleich möglich wird. Auch eine Einschätzung nach Modelleigenschaften wird so möglich. Während die durchschnittliche IOU nicht zu den automatisch berechneten Standardmetriken gehört, kann diese zusätzlich mithilfe der Funktionen des `arcgis.learn`-Moduls berechnet werden.

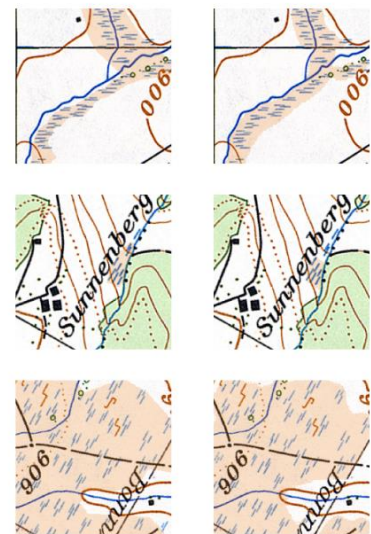


Abbildung 23: Optischer Vergleich der Groundtruth-Daten (l.) mit den vom PSP9 klassifizierten Image-Chips (r.)

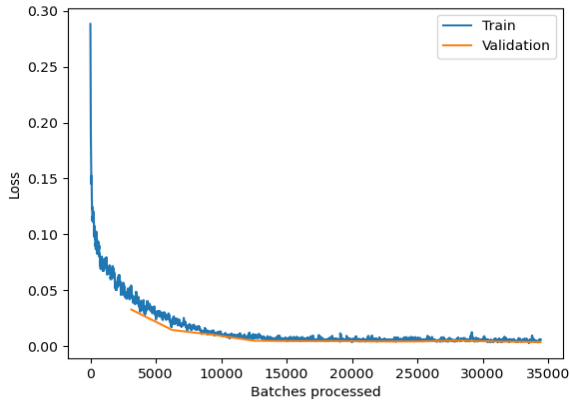


Abbildung 24: Verlustfunktion des PSP9 Modells während dem durchlaufen der Trainingsepochen

Per class metrics:

	NoData	Undefined
precision	0.964863	0.849533
recall	0.980594	0.754192
f1	0.972665	0.799028

Abbildung 25: Klassische Metrik mit Präzision, Recall und F1-Score des PSP9 Modells

3.5.3 Gütemasse der Pixelklassifizierungen

Um die eigentliche Pixelklassifizierung auszuwerten, die mit den trainierten Modellen vorgenommen werden können, gibt es in ArcGIS Pro kein Werkzeug. Um nebst den ArcGIS Metriken einen weiteren Überblick darüber zu gewinnen, wie sich die Klassifizierung der historischen Kartenblätter in statistischen Werten widerspiegelt, müssen die Rasterdaten mit der Image Analyst Toolbox weiterverarbeitet werden.

Um die Werte für die Wahrheitsmatrix zu erlangen, wie sie in Tabelle 1 aufgeführt sind, wird die Groundtruth, wie zum Beispiel in Abbildung 26, mit der Klassifizierung eines Modelles abgeglichen, wie sie in Abbildung 27 gezeigt wird. Dies wird erreicht durch das Bilden von Differenzen zwischen den klassifizierten Pixeln und der Groundtruth-Rasterdaten, als auch den Nicht-Wetland-Pixeln. So resultiert, wie beispielhaft in den Abbildungen Abbildung 29 und Abbildung 28 gezeigt, die Einteilung der Pixel in True Negative und False Positive, respektive True Positive und False Negative. Wie in 3.5.1 festgehalten, lassen sich anhand der Anzahl Pixel nach jeweiliger Einteilung die aufgelisteten statistischen Masse berechnen.

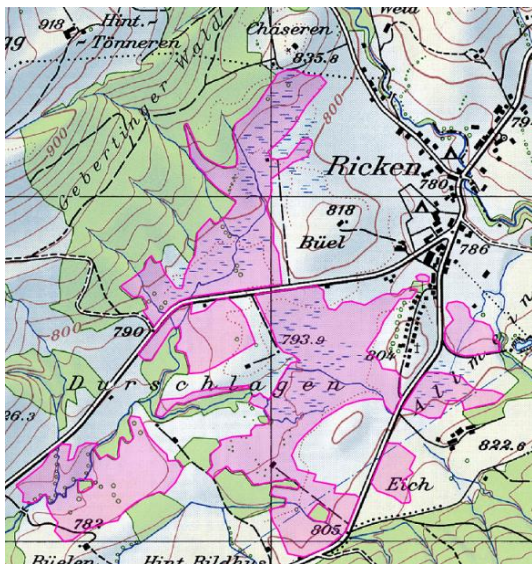


Abbildung 26: Wetland-Groundtruth in pink

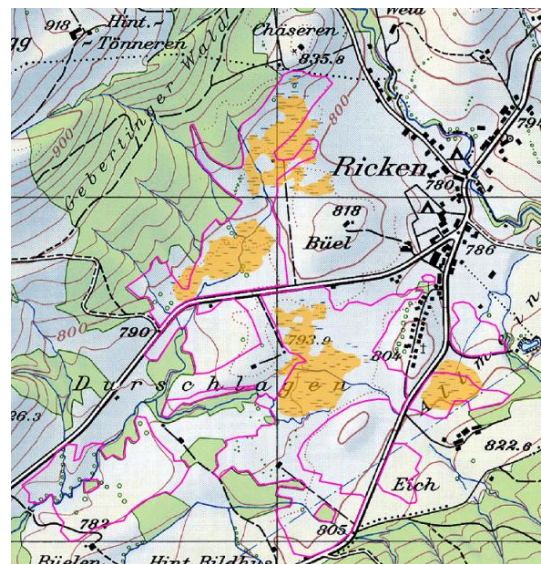


Abbildung 27: Pixelklassifizierung des DL1-Modells in orange mit Groundtruth-Outline in pink



Abbildung 29: Wetland-Groundtruth in schwarz, True Negative Pixel in blau, False Positive in orange



Abbildung 28: Nicht-Wetland in schwarz, True Positive in orange, False Negative in blau

3.5.4 Verwendete Gütemasse

Die Gütemasse, die in die Entscheidungsfindung einfließen können, werden dadurch limitiert, das ArcGIS nur eine beschränkte Selektion von Metriken zur Verfügung stellt. Um einen Vergleich mit den Auswertungen der Klassifizierungen der historischen Rasterkarten zu gewährleisten, wird der Fokus in dieser Arbeit auf den Recall, die Präzision, den F1-Score und den IOU gelegt. Diese Metriken werden vor allem in Bezug auf die Wetland-Klasse betrachtet. Durch das starke Klassenungleichgewicht sind die Gütemasse der Nicht-Wetland-Klassifizierungen immer deutlich höher und nehmen auch durch schlecht kalibrierte Modelle nur wenig ab. Der Kontrast in den Gütemassen lässt sich auch in Abbildung 25 gut veranschaulichen. Zusätzlich wird in der Auswertung auch der Durchschnitt dieser vier Metriken betrachtet. Während dies kein legitimes Standardmass ist, so kann es helfen, Modelle schneller zu identifizieren, die zwar bei keiner Metrik den Höchstwert vorzeigen können, aber in jeder Kategorie gut abschneiden. Je nach Präferenz können so andere Modelle besser oder schlechter eingestuft werden.

4 Ergebnisse

4.1 Workflows

Die grundlegenden Ergebnisse dieser Arbeit sollten die erarbeiteten Workflows sein. Diese bestehen durchaus, und können so auch als Ergebnis gewertet werden. Sie sind aber insoweit nicht leicht darzustellen, da die Variabilität zwischen den Workflows erst sichtbar wird, wenn die ModelBuilder-Dateien in ArcGIS Pro angewandt werden können, und die bestimmten Hyperparameter, durch die sich die verschiedenen Workflows voneinander abgrenzen, auch in Vorschein treten. Durch die kleine Anzahl an eingesetzten Tools, auch dadurch, dass alle betrachteten Modelle der Anwendung für Pixelklassifizierung entsprechen, ist der Grundaufbau aller Workflows gleich. Abgebildet werden hier der Workflow ohne türkise Parameter in Abbildung 31 und den kompletten Workflow mit den hier relevanten Parametern in Abbildung 32.

Anstelle der Workflows werden die jeweiligen Metriken präsentiert.

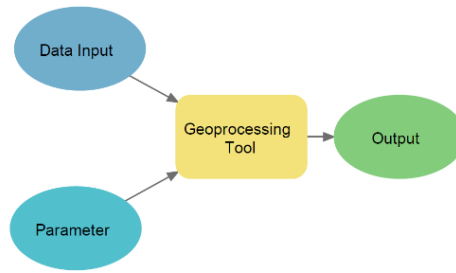


Abbildung 30: Legende

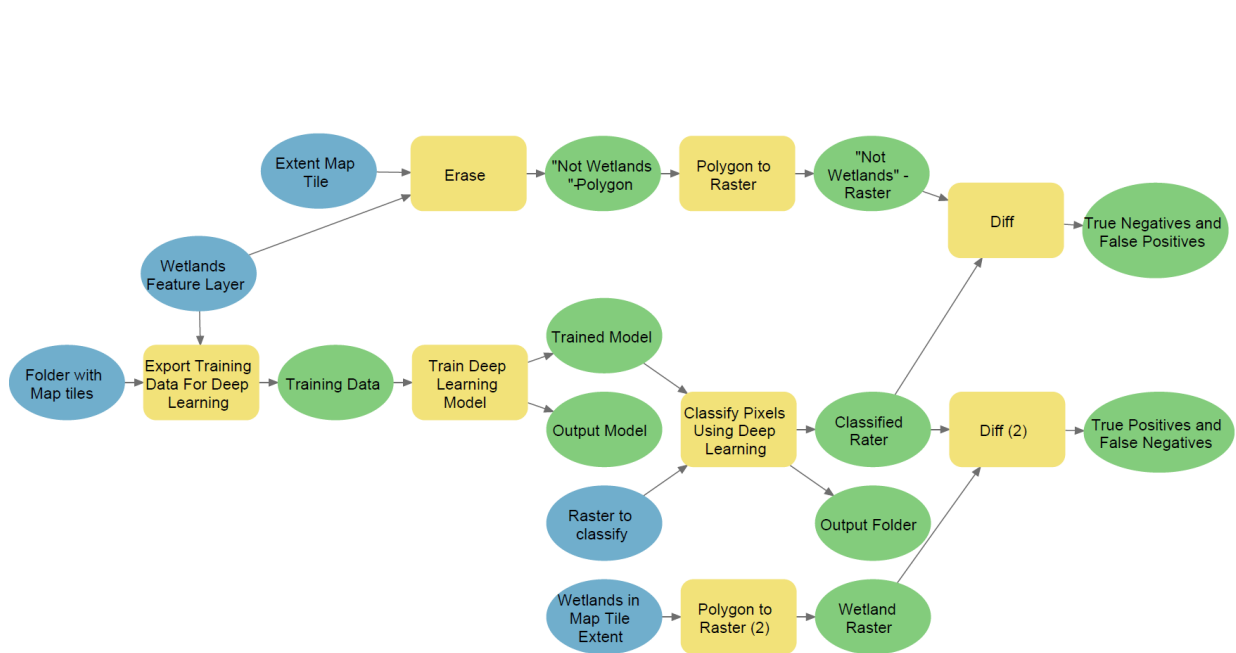


Abbildung 31: Workflow ohne Parameter

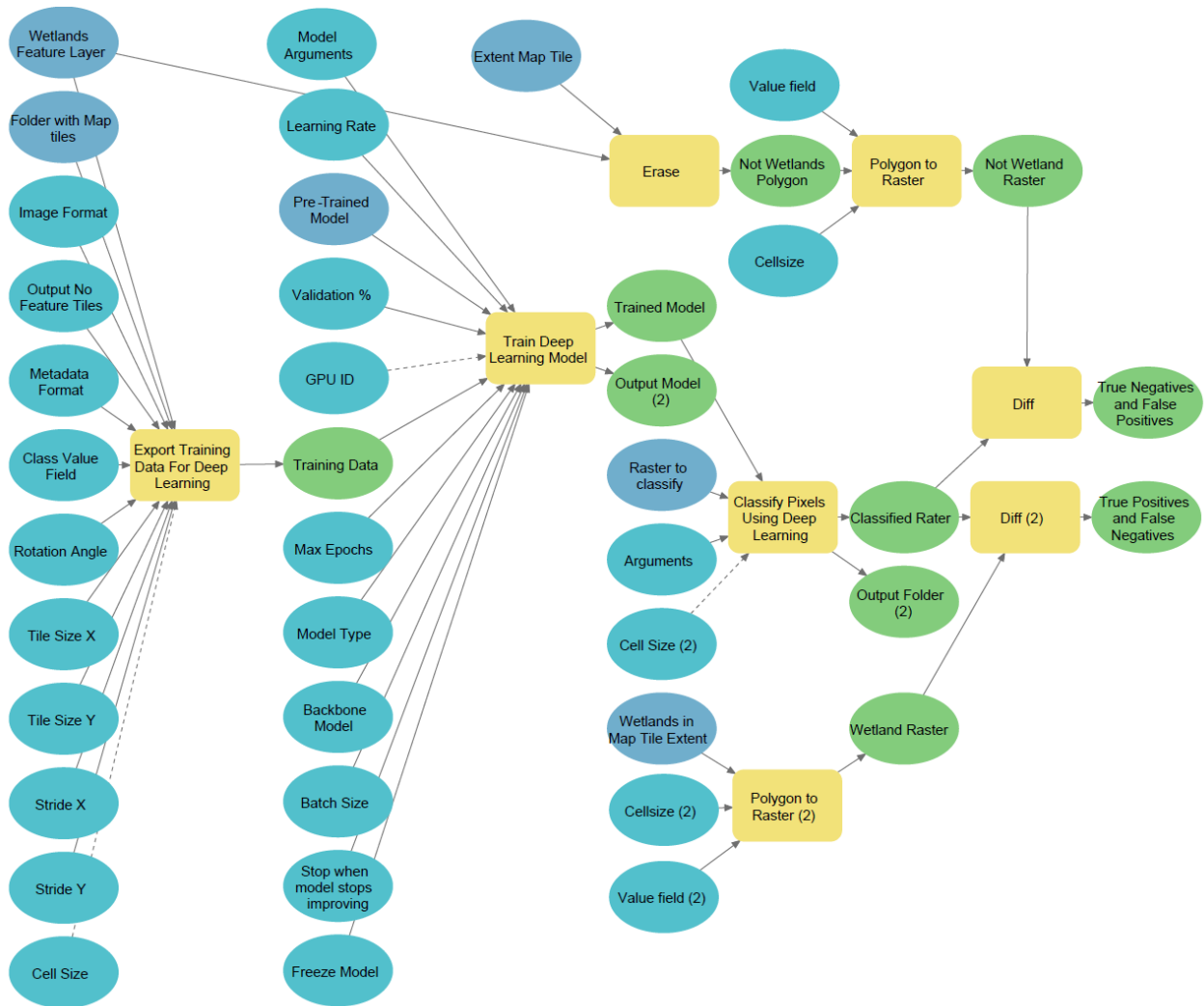


Abbildung 32: Workflow mit Parametern

4.2 Auswertung

4.2.1 Setups

Die Variation der Workflows kommt von den gewählten Modelltypen und den variierten Hyperparametern. Gewisse Unterschiede zeigen sich aber auch erst in der angewandten Klassifizierung von Kartenblättern mit unterschiedlichen Landschaftstypen. In diesem Sinne lässt sich kurz erklären, welche Aspekte in welchen Setups betrachtet wurden. Bei den Trainingsdaten handelt es sich immer um denselben Datensatz à 10 Kartenblättern mit insgesamt 27'817 Trainings-Chips mit Wetland-Anteil.

4.2.1.1 Variation der Modelltypen

Die grösste Unterscheidung zwischen den trainierten Modellen liegt dem Modelltypen zugrunde. Die Vergleichbarkeit unter diesen unterschiedlichen Modellarchitekturen ist aber dadurch gegeben, dass die Modelle alle zur Pixelklassifizierung verwendet werden. Zusätzliche Verbindungen können dadurch hergestellt werden, dass der UNetClassifier und der PSPNetClassifier die UNet-Architektur teilen, also beide auf der Encoder-Decoder-Struktur aufgebaut sind. Somit liefern diese Modelltypen Outputs mit derselben räumlichen Auflösung, wie der Input, die Image Chips die für das Training zur Verfügung gestellt werden. Als Kontrast dazu bleibt DeepLab, welches die räumliche Auflösung nicht

im gleichen Masse beibehält, aber diesen durch die Atrous Faltung minimiert und sich den räumlichen Kontext durch ein Pyramid Pooling beschafft. Diese Eigenschaften teilt es sich wiederum mit dem PSPNetClassifier, der ebenfalls auf Pyramid Pooling und Atrous Faltung aufbaut. So zeichnen sich auch beide Modellstrukturen durch eine Tauglichkeit für Modellskelette mit hohen Tiefen aus. Allfällige Stärken oder Schwächen der Modelltypen die in den Gruppierungen UNet & PSPNet oder PSPNet & DL auftreten, lassen sich damit, sofern es der Kontext erlaubt, also diesen Eigenschaften zuordnen.

4.2.1.2 Variation der Hyperparameter

Da die Modellarchitektur der drei Modelltypen von ArcGIS Pro festgelegt wird, sich nicht einsehen lässt, und auch nur beschränkt beeinflussbar ist bleiben die Hyperparameter die einzigen Größen, die das Training der verschiedenen Modelltypen beeinflussen können.

Im Rahmen dieser Arbeit werden nicht die gänzlichen variablen Parameter untersucht, die für jeden Modelltyp zur Verfügung stehen. Die folgenden Hyperparameter werden in verschiedenen Kombinationen und Variation im Trainingsprozess hervorgehoben, untersucht und analysiert. Die nicht aufgeführten Parameter werden als invariant angenommen und behalten ihren Standardwert bei.

Default

Im Default Setup wurden alle Einstellungen so belassen wie sie vom Trainings-Tool selbst ausgefüllt werden. Je nach Modell und Parameter ist dieser fix, wie zum Beispiel die maximale Epochenanzahl von 20. Andere Werte wie die Batch Size hängen davon ab, wie kompliziert das gewählte Backbone- oder «Deep Learning»- Modell ist.

Mixup

Mixup ist ein Lernprinzip, welches verhindern soll, dass sich neuronale Netze unerwünschte Verhaltensweisen wie Auswendiglernen oder das Scheitern an ungünstigen Beispielen aneignen. Um dem entgegenzuwirken, trainiert man bei diesem Data-Augmentation-Modell mit neu generierten Trainingsdaten. Diese werden generiert, indem zwei zufällige, bestehende Image-Chips mit Wetland-Anteil miteinander gemixt und kombiniert werden. Diese neuen synthetischen Trainingsdaten mit gewichteten Groundtruth-Labels sollen das Modell dazu bringen, bei den Trainingsdaten, die nicht von durch Mixup generiert wurden, die einfacheren Klassifizierungen zu wählen, um das Modell möglichst generell zu halten und eine Spezialisierung zu vermeiden (Zhang, Cisse, Dauphin, & Lopez-Paz, 2018, S. 1).

Unfreeze

Unfreeze führt als Parameter dazu, dass die Gewichte und Biases der Backbone-Modelle sich nun ebenfalls verändern lassen. So können sich die Parameter der Backbone-Modelle, die sonst während dem Trainingszyklus konstant bleiben, so verändert werden, dass sie besser zu den gezeigten Trainingsdaten passen. Während dies zwar den Rechenaufwand erhöht, sollten die resultierenden Klassifizierungen besser sein (Esri, 2022h).

Focal loss

Bei Focal loss wird versucht, gegen Klassenungleichgewicht vorzugehen. Dabei wird der cross entropy loss einer wenig vertretenen Klasse einem Beispiel zugewiesen, dass sich bereits als gut klassifiziert erwiesen hat. Da sowieso ein Ungleichgewicht herrscht, fällt dieser Loss bei der überrepräsentierten Klasse nicht stark ins Gewicht, belastet aber auch nicht die andere Klasse (Lin, Goyal, Girshick, He, & Dollár, 2018, S. 1).

Learning Rate 0.0002

ArcGIS rechnet jeweils vor der ersten Trainingsepoche das optimale Intervall für die Learning rate aus, sodass mit jeder Epoche die idealen Anpassungen an den Modellparametern vorgenommen werden. Je nach Rate wird nach jeder durchlaufenen Epoche mehr oder weniger der bestehenden

gesammelten Informationen mit neuen Erkenntnissen überschrieben. Mit der optimalen Learning rate sollte die Verlustfunktion langsam abflachen.

ResNet 18, ResNet 50 und ResNet 101

Hier wird mit den Backbone Modellen variiert. Als Default wird jeweils ResNet-34 eingesetzt. Ein CNN mit 18, 34 und entsprechend 50 Convolutional Layern. ResNet 18 ist somit weniger tief und komplex als der Default, ResNet 50 und 101 sind komplexer.

4.2.1.3 Variation der Kartenblätter

Um die Einsatzfähigkeit der resultierenden Modelle einordnen zu können, unabhängig von den reinen Modellmetriken, die ArcGIS aus den Klassifizierungen der Validationsdaten berechnet, werden alle trainierten Modelle die Pixel von sieben ausgewählten Kartenblättern à 210 km², respektive 134'400'000 Pixeln, klassifizieren.

Kartenblatt 1111

Das Kartenblatt 1111 zeigt einen Ausschnitt der durchschnittlichen Schweizer Landschaft mit dichten Siedlungsgebieten, Seen, Grünflächen und Flüssen (siehe Abbildung 33). Die gemäss Groundtruth vorliegenden Wetlands decken ein breites Spektrum von Grössen ab und liegen in Siedlungsnähe, als auch in naturdominierten Teilen der Kachel. Der Anteil an von Wetland-Fläche macht mit 2.6km² gut 1% aus.

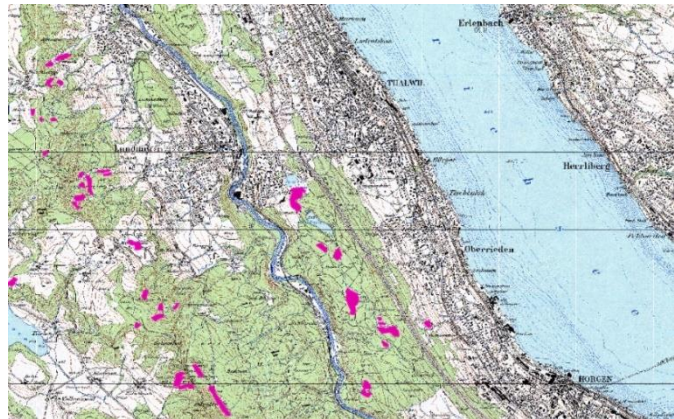


Abbildung 33: Ausschnitt des Kartenblatts 1111

Kartenblatt 1112

Im Kartenblatt 1112 kommen vorwiegend Wetland-Features vor, die sich an Waldrändern oder rund um Seen verteilen. Die Flächen fallen dabei grösstenteils in mittelgrossen Flächen an (siehe Abbildung 34). Generell ist das Gebiet mit dem Kartenblatt 1111 vergleichbar, aber beinhaltet mehr Siedlungsgebiete und Felder. Dabei machen die Wetlands mit 1.8 km² nicht ganz 1% der Teilfläche aus.

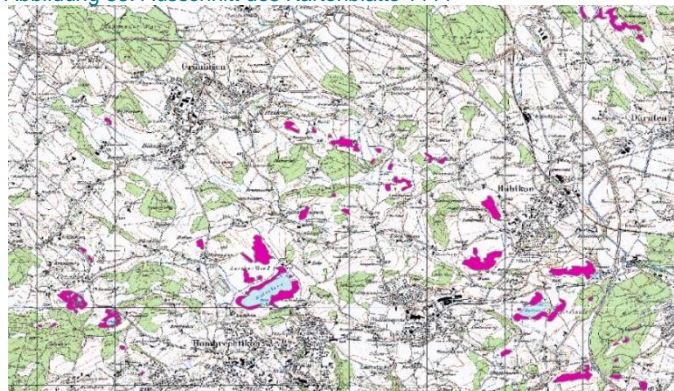


Abbildung 34: Ausschnitt des Kartenblatts 1112

Kartenblatt 1113

Die Kachel 1113 ist geografisch ähnlich zu verorten wie 1111 und 1112 und ist zum Teil auch noch stark von Siedlungen geprägt (siehe Abbildung 35). Der wesentliche Charakter des Blattes kommt aber vom grossen bergigen Anteil, der das Kartenblatt umfasst. Die darin enthaltenen Wetlands nehmen teilweise grosse, zusammenhängende Flächen ein, wie Hochmoore oder in Naturschutzgebieten liegende Feuchtgebiete. Die zu identifizierende Fläche beläuft sich so auf 2.76 km².

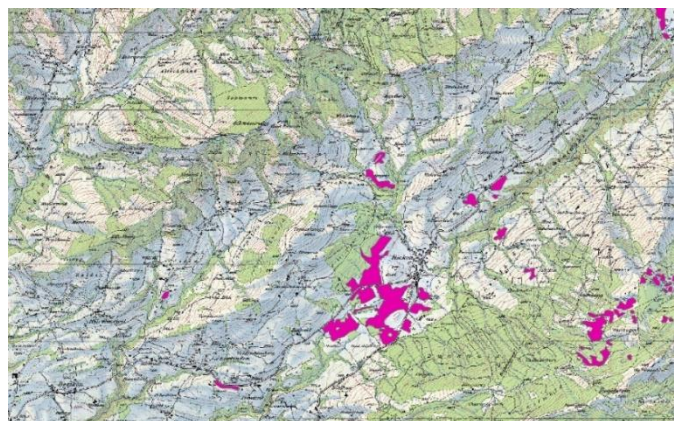


Abbildung 35: Ausschnitt des Kartenblatts 1113

Kartenblatt 1184

Für den Ausschnitt 1184 wird eine Landschaft gewählt, die vor allem von leeren Feldern geprägt ist. Wetland-Areale beschränken sich auf die Seeufergebiete, sind dort aber grossangelegt (siehe Abbildung 36). Da die Ausbreitung der Wetland-Features stark begrenzt ist, fällt auch die Fläche mit 1.6 km² kleiner aus.

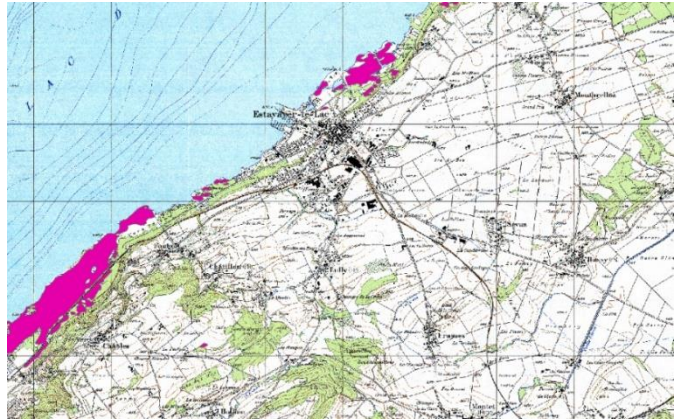


Abbildung 36: Ausschnitt des Kartenblatt 1184

Kartenblatt 1169

Das Kartenblatt 1169 umfasst eine grossflächige Moorlandschaft, die aber im Groundtruth-Layer nicht als zusammenhängendes Gebiet, sondern als viele einzelne Features dokumentiert wird (siehe Abbildung 37). Durch viele kleinflächige Wetland-Gebiete werden mit 2.5 km² mehr als 1% der Kachel durch diese abgedeckt.

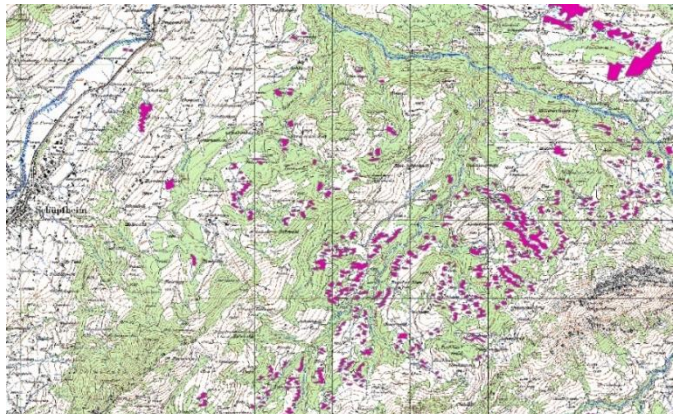


Abbildung 37: Ausschnitt des Kartenblatts 1169

Kartenblatt 1264

Das Kartenblatt 1264 zeigt, ähnlich wie 1184, ein grosses, zusammenhängendes Feuchtgebiet am Ufer eines grossen Sees (siehe Abbildung 38). Die Fläche läuft hier aber nicht nur entlang dem Ufer, sondern auch den Flüssen entlang in das Siedlungsgebiet hinein. Und nebst der flachen Siedlungsfläche bildet der Grossteil der Kachel Grünflächen in Berggebieten ab, die von kleineren Wetland-Features durchzogen sind. Hier beläuft sich die das

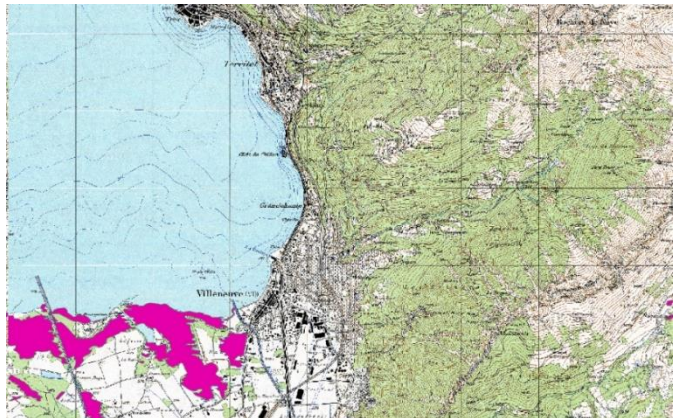


Abbildung 38: Ausschnitt des Kartenblatts 1264

Wetland-Gebiet auf 2.35 km².

Kartenblatt 1203

Zuletzt ist mit der Kachel 1203 auch noch ein Auszug der Schweiz dabei, der nur ganz wenig Wetland-Flächen aufweist. Wie bei dem Blatt 1184 ist die Charakteristik gegeben durch Siedlungen und dahinter folgende weite, offene Felder (siehe Abbildung 39). Die vorliegenden Wetlands befinden sich wiederum am Seeufer. Der Anteil an Wetlands fällt mit 0.4 km² entsprechend gering aus.

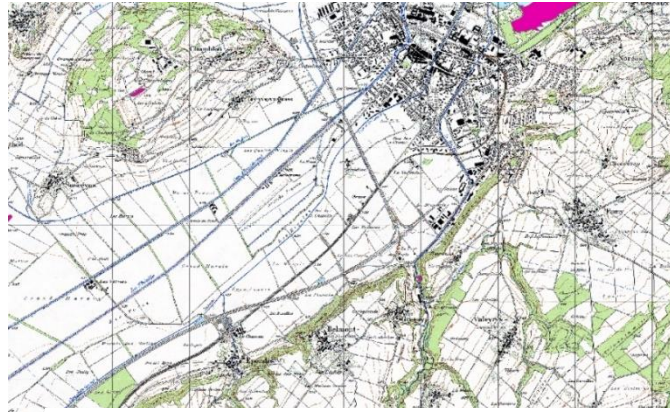


Abbildung 39: Ausschnitt des Kartenblatts 1203

4.2.1.4 Übersicht der Einteilung

Einteilung nach Modell

Tabelle 2: Einteilung der U-Net-Modelle

Modell	Setup	Anz. Epochen
UNET 1	Default	20/20
UNET 2	Mixup	20/20
UNET 3	Unfreeze	20/20
UNET 4	Focal loss	20/20
UNET 5	Learning Rate 0.00002	20/20
UNET 6	ResNet-18	20/20
UNET 7	ResNet-50	20/20
UNET 8	Focal loss & unfreeze	29/80
UNET 9	Mixup & unfreeze	27/80
UNET 10	Mixup, focal loss & unfreeze	35/80
UNET 11	ResNet-18, mixup, focal loss & unfreeze	24/80

Tabelle 3: Einteilung der PSPNet-Modelle

Modell	Setup	Anz. Epochen
PSP 1	Default	20/20
PSP 2	Mixup	20/20
PSP 3	Unfreeze	11/20
PSP 4	Focal loss	14/20
PSP 5	Learning Rate 0.00002	20/20
PSP 6	ResNet 18	20/20
PSP 7	ResNet 50	9/20
PSP 8	ResNet 101	20/20
PSP 9	Focal loss & unfreeze	11/80
PSP 10	Mixup & unfreeze	14/80
PSP 11	Mixup, focal loss & unfreeze	9/80
PSP 12	ResNet 18, mixup, focal loss & unfreeze	9/80
PSP 13	ResNet 50, mixup, focal loss & unfreeze	9/80

Tabelle 4: Einteilung der DL-Modelle

Modell	Setup	Anz. Epochen
DL 1	Default	20/20
DL 2	Mixup	20/20
DL 3	Unfreeze	20/20

DL 4	Focal loss	20/20
DL 5	Learning rate 0.00002	20/20
DL 6	ResNet 18	20/20
DL 7	ResNet 50	9/20
DL 8	ResNet 101	10/20
DL 9	Focal loss & unfreeze	9/80
DL 10	Mixup & unfreeze	13/80
DL 11	Mixup, focal loss & unfreeze	18/80
DL 12	ResNet 18, mixup, focal loss & unfreeze	20/80
DL 13	ResNet 50, mixup, focal loss & unfreeze	9/80

Einteilung nach Setup

Tabelle 6: Einteilung nach Setup

Setup 1	Default
Setup 2	Mixup
Setup 3	Unfreeze
Setup 4	Focal loss
Setup 5	Learning rate
Setup 6	ResNet 18
Setup 7	ResNet 50
Setup 8	ResNet 101
Setup 9	Focal loss & unfreeze
Setup 10	Mixup & unfreeze
Setup 11	Mixup, focal loss & unfreeze
Setup 12	ResNet 18, mixup, focal loss & unfreeze
Setup 13	ResNet 50, mixup, focal loss & unfreeze

Einteilung nach Kartenblatt

Tabelle 5: Einteilung nach Kartenblatt

Tile 1	Kartenblatt 1111
Tile 2	Kartenblatt 1112
Tile 3	Kartenblatt 1113
Tile 4	Kartenblatt 1184
Tile 5	Kartenblatt 1169
Tile 6	Kartenblatt 1264
Tile 7	Kartenblatt 1203

4.3 Vergleich

Vergleiche der Auswertungen dienen im Rahmen dieser Arbeit als wichtiges Werkzeug, um eine Empfehlung abgeben zu können. Auf der Grössenebenen der Pixel in Kombination mit der Extraktion der Wetlands ist dies aber von einer gewissen Ambivalenz geprägt. Da die Wetlands nicht durch klare Grenzen beschränkt sind, wie dies zum Beispiel bei Häusern oder Seen der Fall ist, hängt es vom individuellen Anwendungsbereich der Pixelklassifizierung ab, ob und wieviel Puffer rund um die einzelnen Symbolisierungen gewünscht ist, respektive als «richtig» taxiert wird. Auch wie klein der Ab-

stand zwischen zwei Wetland-Symbolen sein muss, um eine kontinuierliche Klassifizierung zwischen den zweien zu rechtfertigen, kann je nach Präferenz variieren. Es gibt in der Pixelklassifizierung von Wetlands viel subjektiven Input der Individuen, begonnen bei den Groundtruth-Daten, bei denen ebenfalls objektiv bestimmt werden musste, wie der Extent der Wetlands definiert wird. Modelle mit schlechteren Metriken klassifizieren also nicht zwangsläufig falsch, sondern etwas grosszügiger oder restriktiver als in den Validierungsdaten.

4.3.1 Vergleich innerhalb der Modelle

In den folgenden Vergleichen der Auswertung der Metriken ist der Aufbau so, dass jeweils erst die Trainingsresultate des ursprünglichen Datensatzes dargestellt werden. Diese massgebenden Werte werden dann wiederum mit den erlangten Gütemassen derselben Modelle verglichen, die aber noch 10 zusätzliche Epochen mit den Trainingsdaten des korrigierten Datensatzes durchlaufen haben.

Dabei soll gezeigt werden, wie sich die ausgegebenen Metriken verändern, wenn die Validierungsdaten, anhand derer die Gütemasse berechnet werden, auch fehlerfrei sind.

Für die Vergleiche der Resultate der klassifizierten Kartenblätter gilt ein ähnliches Schema. Erst wird gezeigt, was für Werte auf den Kartenausschnitten mit Bezug auf die regulären Wetland-Daten erreicht wurden. Danach wird aufgezeigt, wie sich diese Masswerte verändern, wenn die Feature-Daten fast fehlerfrei sind.

Zur besseren Übersicht werden die besten zwei Werte pro Metrik mit Farbe hervorgehoben. Die dunkelste Einfärbung kennzeichnet dabei den besten Wert. Die etwas hellere Einfärbung den zweitbesten Wert. Die aufgeführten Metriken sind an dem orientiert, was ArcGIS für die Evaluation der Modelle zur Verfügung stellt. Der Durchschnitt, der hier ebenfalls aufgeführt ist, bedient sich keiner Statistischen Relevanz. Es ist lediglich eine gewisse Stütze, um sich ein Bild der Modellqualität zu machen, ohne von Ausreissern einzelner Gütemasse blenden zu lassen. Oft weisen zum Beispiel die sehr tiefen Setups eine sehr hohe Präzision aus, haben dann aber einen sehr tiefen Recall-Wert. Je nach Bedürfnis des Users kann dies die Eignung des Modells deutlich auf- oder abwerten.

4.3.1.1 U-Net

Wie der Tabelle 2 entnommen werden kann, wurde das U-Net-Modell mit 11 aufgelisteten Setups trainiert. Für die Setups 8 und 13, wie sie in der Tabelle 6 vermerkt sind, war die Rechenkapazität des IKG 1070 zu niedrig. Somit fließen diese nicht in die Auswertung ein.

U-Net: Metriken der Trainingsphase

Tabelle 7: U-Net-Modell Metriken der Trainingsresultate mit den regulären Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
UNET 1	0.66167091	0.852696	0.789692	0.819986	0.88021576
UNET 2	0.68241409	0.859120	0.805293	0.831336	0.88863525
UNET 3	0.79471027	0.869979	0.835983	0.852642	0.91023532
UNET 4	0.67631222	0.857739	0.806043	0.831088	0.88674551
UNET 5	0.65657180	0.846180	0.788384	0.816260	0.87937705
UNET 6	0.62181304	0.826755	0.762424	0.793287	0.86694142
UNET 7	0.61408140	0.841727	0.746840	0.791450	0.85674897
UNET 8	0.67115014	0.857548	0.803994	0.829908	0.88713968
UNET 9	0.65700957	0.877033	0.764192	0.816733	0.88022169
UNET 10	0.65505651	0.852138	0.775258	0.811882	0.87779273
UNET 11	0.64770753	0.869927	0.752918	0.807205	0.87583616

Tabelle 8 U-Net-Modell Metriken der Trainingsresultate mit den korrigierten Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
UNET 1	0.84410148	0.918776	0.926451	0.922598	0.94251661
UNET 2	0.86216850	0.927870	0.934408	0.931128	0.948954565
UNET 3	0.86443618	0.934529	0.930139	0.932329	0.949813944
UNET 4	0.84835342	0.933242	0.917437	0.925272	0.944443012
UNET 5	0.84019503	0.915724	0.925369	0.920521	0.94100598
UNET 6	0.80162235	0.889815	0.911713	0.900631	0.926356434
UNET 7	0.78429995	0.914722	0.877962	0.895965	0.922078144
UNET 8	0.863751374	0.933604	0.927837	0.930712	0.94902648
UNET 9	0.849338765	0.931093	0.940649	0.935846	0.9492003
UNET 10	0.872176168	0.929647	0.919714	0.924654	0.94759036
UNET 11	0.85829122	0.934182	0.924915	0.929525	0.94771845

In den Tabelle 7 und Tabelle 8 zeigt sich von Beginn weg, dass durch die Verwendung der korrigierten Daten die Werte aller Metriken steigen. Das zu bevorzugende Modell ist über beide verschiedenen Trainingsresultate hinaus UNET 3, das Modell mit dem Unfreeze-Setup. Für das nächstbeste Modell fehlt dann aber die Übereinstimmung zwischen den verschiedenen Validierungsdaten. Bei den regulären Daten ist es UNET 2 mit dem Mixup-Setup, bei den Modellen mit der Validierung auf den korrigierten Kacheln liegt die Präferenz bei UNET 9, der Kombination von Mixup und Unfreeze.

Dass UNET 3 mit dem angewandten Unfreeze-Prinzip gut abschneidet, lässt sich sicher damit erklären, dass die Gewichte und Biases, die Parameter die vom Neuronalen Netzwerk gelernt und angepasst werden, des gewählten Backbone-Layers, hier ResNet 34, so verändert werden, dass sie den Trainingssamples möglichst gut angepasst sind. Dabei dauert der Trainingsprozess durch den Mehraufwand in der Regel länger, sollte aber auch bessere Resultate ermöglichen (Esri, 2022h). So ist dieser Parameter dem Standardsetting, als auch Modifikatoren wie Mixup oder Focal loss überlegen. Auch das Modell mit dem tieferen Backbone-Modell und folglich auch mehr Faltungslayern, UNET 7, bringt keine so gute Gütemasse wie die Möglichkeit zur nachträglichen Anpassung der Backbone-Layern. Was weiter für das UNET 3 Modell spricht, ist, dass es in allen hier betrachteten Metriken zu überzeugen vermag und in keinem Teilbereich Schwächen zu erahnen sind.

Dadurch, dass der Anteil an Nicht-Wetland-Pixel auch bei den Trainings-Chips deutlich höher ist, weisen Modelle, die zu einer verfrühten, und folglich schnell falschen Klassifikation der Pixel als Wetland neigen, trotzdem hohe Güterwerte auf, da die Fehlklassifikation von Nicht-Wetland-Pixel durch das grössere Vorkommen nicht so ins Gewicht fallen. Gerade im Falle der fehlerhaften Groundtruth-Daten, die auch Gebiete umfassen, die nicht mit Wetland-Illustrationen versehen sind, führen eigentlich fehlerhafte Klassifikationen solcher Gebiete zu einem verzerrten Abbild der Modelgüte. Es liegt also nahe zu vermuten, dass UNET 2, welches bei den regulären Trainingsdaten besser abschneidet als UNET 9, die Kombination der beiden andere Setups, einen Pixel tendenziell schneller als Wetland klassifiziert, als UNET 9. Dies widerspiegelt sich auch in der Präzision, die bei UNET 2 jeweils tiefer ausfällt als bei UNET 9, was sich auf ein schlechteres Verhältnis zwischen True Positives und False Positives zurückführen lässt.

Unter Umständen ist auch einfach die Übereinstimmung des Verständnisses, wie die Ausdehnung der Wetlands anhand der Symbolisierung aussehen soll, zwischen demjenigen, der die Datenkorrektur vornahm und dem UNET 9 Modell höher, was zu der besseren Güte führt.

Grundsätzlich ist die Performance der drei Modelle aber ähnlich einzustufen. Dies zeigt sich auch in den eigentlichen Klassifizierungen, wie sie in Abbildungen 40,41 und 42 zu sehen sind.

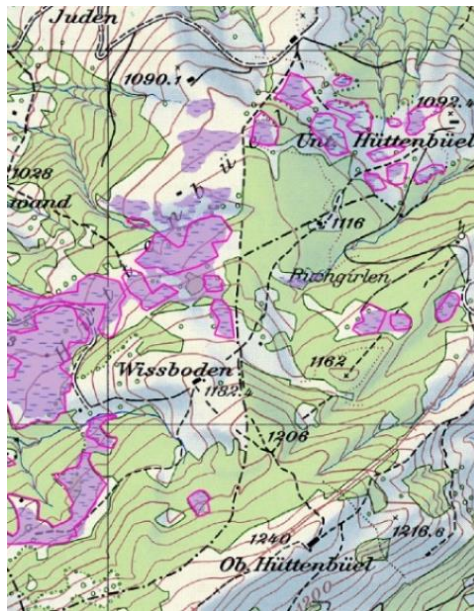
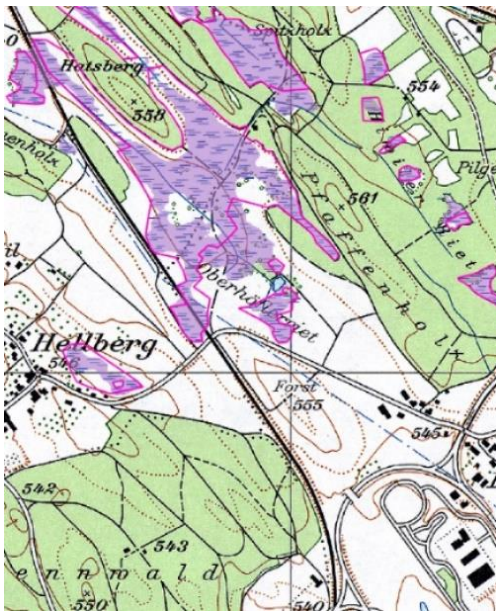


Abbildung 40: Ausschnitte zweier Klassifizierungen durch UNET 3 mit der Groundtruth-Outline

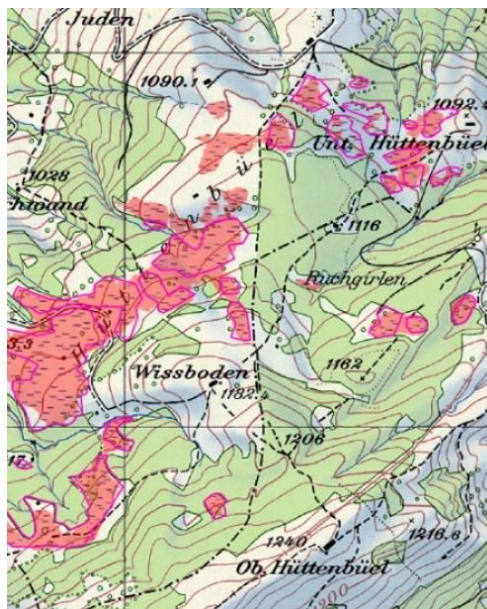
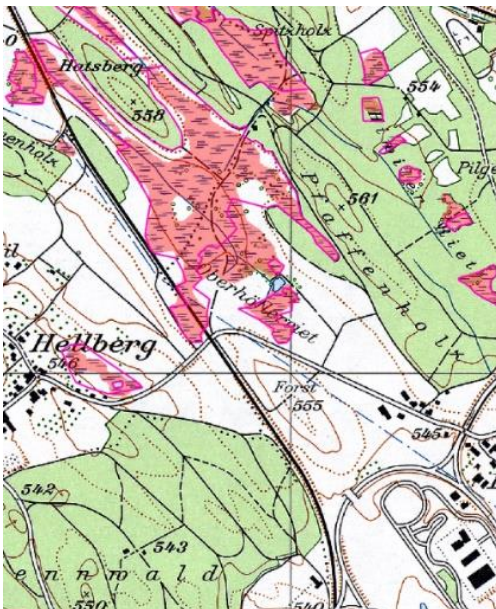


Abbildung 41: Ausschnitte zweier Klassifizierungen durch UNET 2 mit der Groundtruth-Outline

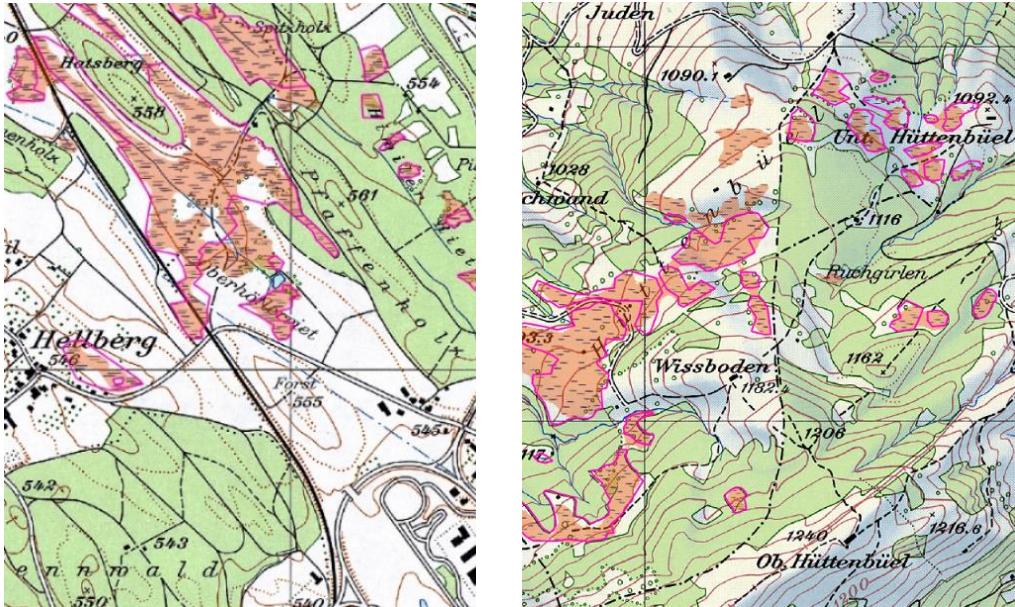


Abbildung 42: Ausschnitte zweier Klassifizierungen durch UNET 9 mit der Groundtruth-Outline

U-Net: Metriken der Pixelklassifizierung

Tabelle 9 U-Net Klassifizierung originaler Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
UNET 1	0.591114861	0.787254555	0.44596865	0.733077707	0.639353943
UNET 2	0.596397089	0.77673857	0.458715645	0.738987115	0.642709605
UNET 3	0.510778793	0.686346923	0.432908052	0.670087572	0.575030335
UNET 4	0.506370793	0.668619601	0.444407094	0.668746295	0.572035946
UNET 5	0.59097447	0.774623923	0.4525569	0.732689745	0.637711259
UNET 6	0.558311386	0.750700112	0.430967555	0.705051913	0.611257742
UNET 7	0.548081489	0.783120873	0.411211807	0.698221569	0.610158934
UNET 8	0.423474898	0.54783932	0.436387408	0.590944162	0.499661447
UNET 9	0.597079477	0.811964584	0.440719754	0.737543026	0.64682671
UNET 10	0.596249758	0.77309105	0.459918073	0.736565965	0.641456211
UNET 11	0.594634658	0.797989895	0.446907964	0.734306256	0.643459693

Tabelle 10 U-Net Klassifizierung korrigierter Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
UNET 1	0.821865567	0.916354559	0.567562651	0.901087996	0.80171769
UNET 2	0.835511335	0.912956013	0.580003105	0.909542512	0.80950324
UNET 3	0.733015784	0.820799291	0.553066846	0.841187351	0.73701732
UNET 4	0.78025083	0.863136297	0.565641904	0.87218356	0.77030315
UNET 5	0.802619168	0.892888389	0.566693315	0.889142816	0.78783592
UNET 6	0.792928051	0.899774224	0.554450274	0.881971587	0.78228103
UNET 7	0.77606819	0.912308464	0.534933117	0.872155182	0.77386624
UNET 8	0.678973923	0.753365185	0.551162134	0.797413845	0.69522877
UNET 9	0.82779434	0.956973543	0.550210362	0.904909815	0.80997201
UNET 10	0.828717051	0.911511977	0.57546699	0.905194316	0.80522258
UNET 11	0.809960342	0.926634543	0.554527322	0.894691543	0.79645344

Die manuell berechneten Metriken, die aus der Klassifizierung der Kartenblätter resultieren, lassen die Modelle teilweise anders abschneiden als die Gütemasse der Trainingsphase. Primär wird wieder angemerkt, dass eine klare Steigerung von den Werten der Tabelle 9 zur Tabelle 10 auszumachen ist. Bei der Einordnung dieser Metriken hat das Klassenungleichgewicht nochmals einen grösseren Einfluss. Bei der Validierung mit den Trainingsdaten werden die Auswertungen nur mit Bildausschnitten vorgenommen, auf denen auch Wetlands vorkommen. Bei den grossen Kartenblättern ist der Gesamtanteil der Wetlands sehr gering und macht meist nur 1% aus.

UNET 9 schneidet hier auch bei den regulären, nicht nur bei den korrigierten Groundtruth-Daten sehr gut ab. Bei den korrigierten Kartenblättern aber hauptsächlich deshalb, weil die Präzision sehr hoch ausfällt. Beim Recall, dem Mass, welches anzeigen soll, wie viele der Wetland-Pixel auch identifiziert werden, schneidet es unterdurchschnittlich ab.

Auch UNET 2 kann für beide betrachteten Datensätze gute Ergebnisse vorweisen.

Neue Modelle, die für dieser Art der Auswertung gute Resultate erzielen, sind UNET 11 und 10. Im Vergleich zu den anderen Modellen fallen sie nicht durch Höchstwerte in den verschiedenen Metriken auf, sondern durch ihre Beständigkeit. Die beiden Modelle teilen sich die Trainingsbasis von Mixup, Focal loss und Unfreeze und unterscheiden sich nur in der Tiefe des Backbone-Modells. Wobei UNET 10 mit ResNet 34 und UNET 11 mit ResNet 18 trainiert werden. Das Unfreeze und Mixup Wirkung zeigen, offenbarte sich bereits bei der Auswertung der UNET-Trainingswerte. Focal Loss, das Ungleichgewichte in den vorhandenen Klassen überwinden soll, wird hier ebenfalls zum relevanten Faktor. Der positive Effekt dieses Parameters könnte durch das im Vergleich zu den Image-Chips grössere Ungleichgewicht verstärkt und hervorgehoben werden. Während UNET 11 für präzisere Klassifikationen mit weniger False Positives bevorzugt werden kann, bietet UNET 10 mit seinem höheren Recall und IOU eine vollständigere Identifizierung der Wetlands. Dies äussert sich auch so in den Abbildungen Abbildung 43 und Abbildung 44.

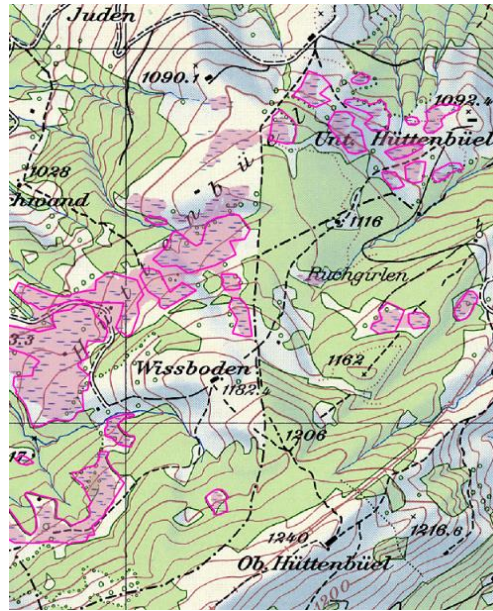
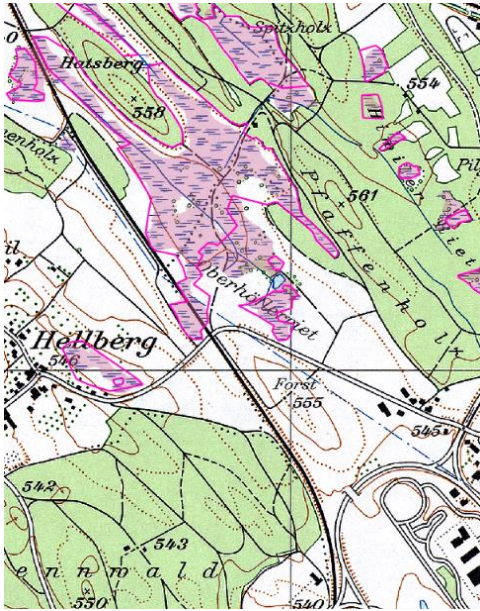


Abbildung 43: Ausschnitte zweier Klassifizierungen durch UNET 11 mit der Groundtruth-Outline

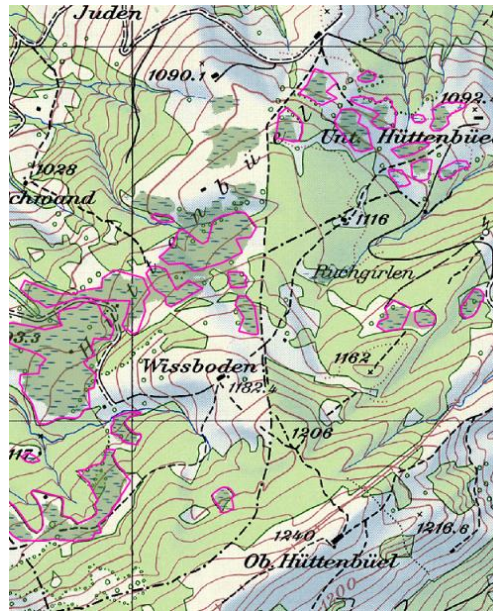
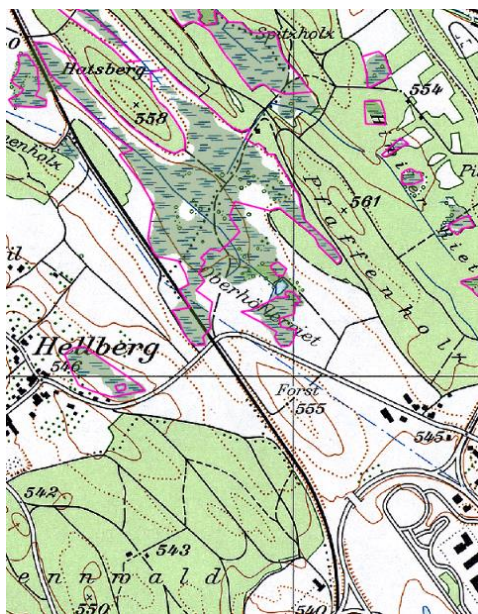


Abbildung 44: Ausschnitte zweier Klassifizierungen durch UNET 10 mit der Groundtruth-Outline

4.3.1.2 PSPNet

PSPNet: Metriken der Trainingsphase

Tabelle 11: PSPNet-Modell Metriken der Trainingsresultate mit den regulären Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
PSP 1	0.686791172	0.849774	0.822534	0.835932	0.884936041
PSP 2	0.679769467	0.858822	0.803136	0.830046	0.88175486
PSP 3	0.598046199	0.876183	0.704456	0.780991	0.852287228
PSP 4	0.635524917	0.848400	0.764031	0.804008	0.864842986
PSP 5	0.694788133	0.852385	0.832904	0.842532	0.888896214
PSP 6	0.661885719	0.875463	0.777202	0.823412	0.87736498
PSP 7	0.60939062	0.844298	0.742719	0.790258	0.855677951
PSP 8	0.505299881	0.859859	0.674177	0.755780	0.829374422
PSP 9	0.628911933	0.849533	0.754192	0.799028	0.862082223
PSP 10	0.637366209	0.866988	0.745014	0.801387	0.864962362
PSP 11	0.587685326	0.875871	0.701452	0.779018	0.850307772
PSP 12	0.586395345	0.874108	0.693369	0.773318	0.847932916
PSP 13	0.61977287	0.878545	0.727138	0.795703	0.861026236

Tabelle 12 PSPNet-Modell Metriken der Trainingsresultate mit den korrigierten Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
PSP 1	0.825996155	0.893636	0.933002	0.912895	0.93540129
PSP 2	0.847712532	0.937579	0.911406	0.924307	0.94399043
PSP 3	0.841519589	0.938796	0.904902	0.921537	0.94191536
PSP 4	0.824312598	0.938718	0.892372	0.914958	0.93664281
PSP 5	0.849140122	0.913525	0.928597	0.920999	0.94248521
PSP 6	0.843189177	0.944493	0.907879	0.925824	0.94407969
PSP 7	0.850494416	0.932474	0.924068	0.928252	0.94604202
PSP 8	0.768842707	0.953778	0.841344	0.894040	0.92026417
PSP 9	0.824865106	0.937874	0.900341	0.918724	0.93829572
PSP 10	0.853682781	0.939582	0.919056	0.929206	0.94700551
PSP 11	0.836314739	0.942004	0.898187	0.919574	0.94036071
PSP 12	0.847643974	0.921785	0.921465	0.921625	0.94256579
PSP 13	0.844076602	0.937088	0.909483	0.923079	0.94289342

Gemäss der Auswertung der Trainingsphase mit den regulären Daten (siehe Tabelle 11) liefern die PSP Modelle 1 und 5 die besten Klassifizierungen. Sie weisen in allen Kategorien, ausser der Präzision, die besten Werte auf. Die grösste Präzision erreicht PSP 13, was wohl auf das Backbone-Modell ResNet 101 zurückzuführen ist, da es das alleinstellungsmerkmal dieses Setups ist. Mit der grösseren Anzahl Layern nimmt auch die Komplexität zu, die das Modell verarbeiten kann. Die Nuancen zwischen Wetland und nicht-Wetland können so von PSP 13 wohl besonders genau wahrgenommen werden. Die beiden Modelle PSP 1 und PSP 5 teilen sich das gesamte Setup, bis auf die Learn Rate, die bei PSP 1 von ArcGIS berechnet und angepasst wird, während sie für PSP 5 konstant bei 0.00002 angesetzt ist. Da PSP 5 das Default-Modell in allen Werten, ausser der Präzision übertrifft, könnte darauf geschlossen werden, dass dieses in dieser Auswertungstabelle als bestes Modell heraussticht. Dem ist aber nicht so. Das Modell klassifiziert zwar teilweise korrekt, wie es auch in Abbildung 45: Ausschnitte zweier Klassifizierungen durch PSP 5 mit der Groundtruth-Outline gezeigt wird, aber es macht auch gravierende Fehlklassifizierungen, wie sie in Abbildung 46: Fehlerhafte Klassifizierungen von PSP 5 präsentiert werden. Hier würde man wohl erwarten, dass sich solche Schwächen deutlicher in den Gütemasse zeigen, aber falls solche Fehler nur selten auf dem Validierungsset auftreten oder das Klassenungleichgewicht die Einflüsse solcher Fehler gering hält, ist dies nicht so leicht möglich.

Für PSP 1 gilt dies aber nicht. Dort ist die Präzision auch klar besser, es werden klare Abgrenzungen zwischen Wetlands und Nicht-Wetlands gezogen, wie man sie in Abbildung 47: Ausschnitte zweier Klassifizierungen durch PSP 1 mit der Groundtruth-Outline sieht.

Bei den Resultaten des korrigierten Datensatzes fallen die beiden Modelle nur noch durch ihre hohen Recalls auf, die von der liberalen Pixelklassifizierungen herrühren. Die Leitlinien der restlichen Gütemasse werden von PSP 10 und PSP 7 gesetzt. PSP 10, das Setup mit Mixup und Unfreeze, wird bei dem Parametersetting davon profitieren, dass sich das Modell durch Unfreeze besser an die Trainingsdaten anpassen kann und Mixup die Daten so augmentiert, dass das Modell flexibel anwendbar bleibt. PSP 7, die Standardausführung mit 50 Layern, bringt wohl solch gute Resultate hervor, da PSPNet dafür ausgelegt ist, auf tiefen neurale Netzwerken aufzubauen. Anscheinend ist im Kontext dieser Anwendungsaufgabe, der Wetland-Extraktion, die Erhaltung der räumlichen Auflösung aber nicht gut genug, um bei einem noch tieferen Netzwerk eine Steigerung zu erwirken, wie sich mit dem Leistungsabfall von PSP 8 mit 101 Layern zu sehen ist.

Die Klassifizierungen dieser beiden Modelle, wie sie in Abbildung 48 und Abbildung 49 zu sehen sind, spiegeln die Werte der Tabelle wieder, wobei man für PSP 7 eine leicht weniger präzise Klassifizierung sieht, dafür einen grösseren Recall hat, als bei PSP 10.

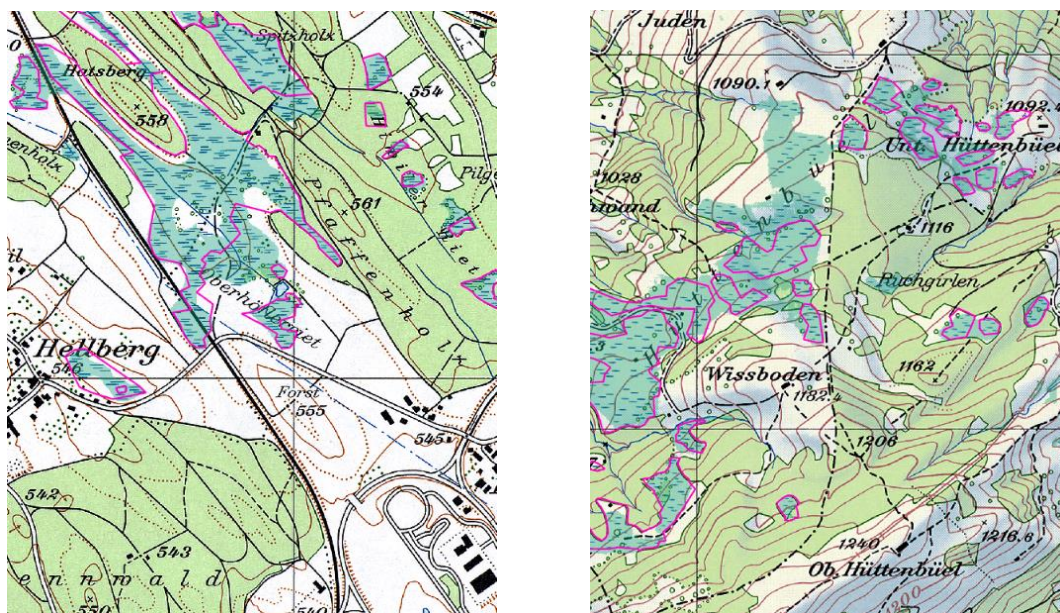


Abbildung 45: Ausschnitte zweier Klassifizierungen durch PSP 5 mit der Groundtruth-Outline



Abbildung 46: Fehlerhafte Klassifizierungen von PSP 5

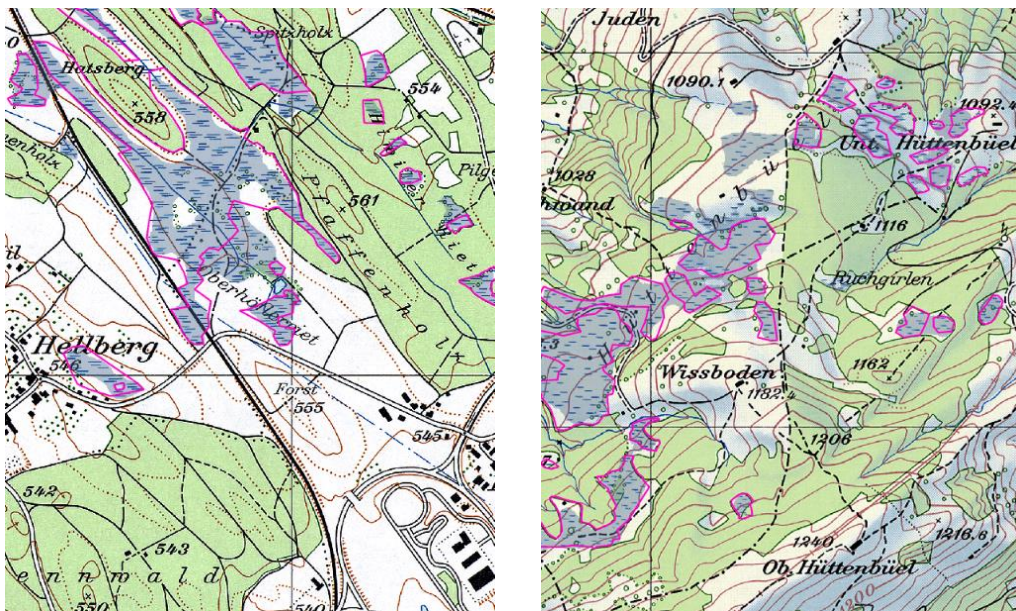


Abbildung 47: Ausschnitte zweier Klassifizierungen durch PSP 1 mit der Groundtruth-Outline

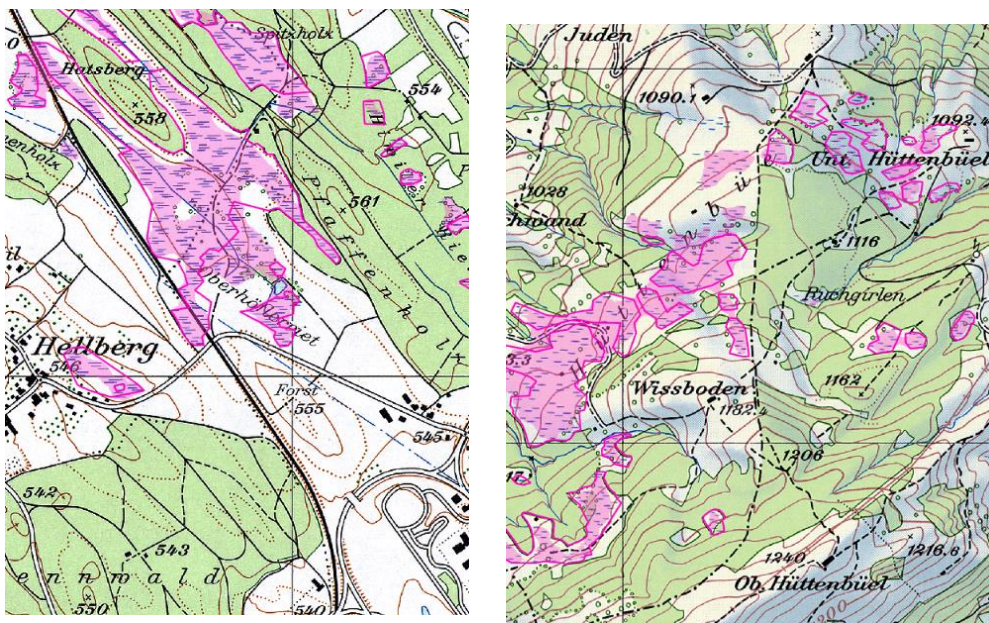


Abbildung 48: Ausschnitte zweier Klassifizierungen durch PSP 10 mit der Groundtruth-Outline

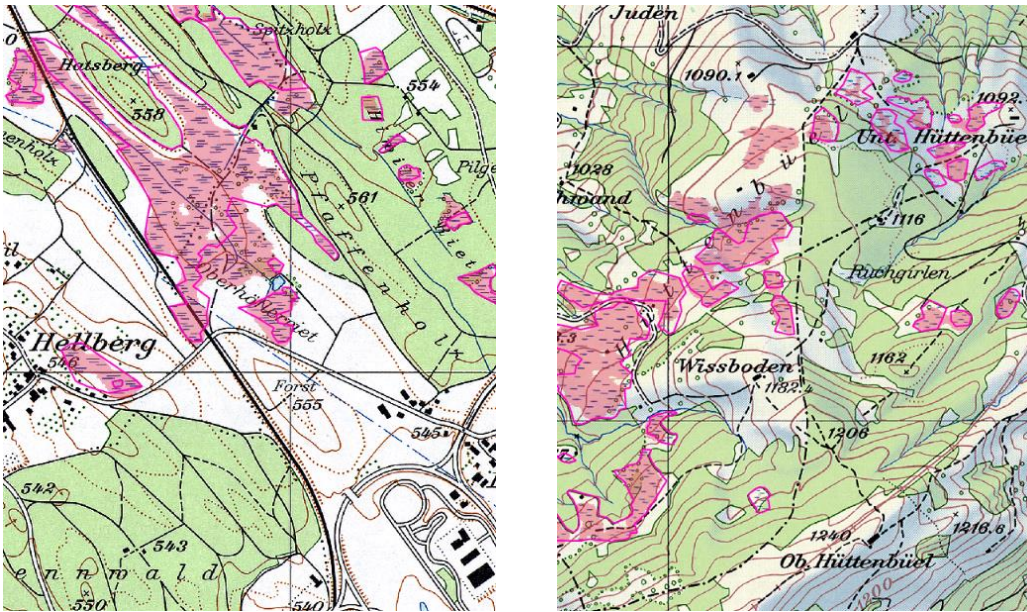


Abbildung 49: Ausschnitte zweier Klassifizierungen durch PSP 7 mit der Groundtruth-Outline

PSPNet: Metriken der Pixelklassifizierung

Tabelle 13 PSPNet Klassifizierung originaler Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
PSP 1	0.55116917	0.74330073	0.44609603	0.69434423	0.60872754
PSP 2	0.50674578	0.64995515	0.45790668	0.66847023	0.57076946
PSP 3	0.34590861	0.48000175	0.39668627	0.51085234	0.43336224
PSP 4	0.52868325	0.68905518	0.44936314	0.68534514	0.58811168
PSP 5	0.42954706	0.52993592	0.46827274	0.59699691	0.50618816
PSP 6	0.5835787	0.78709166	0.44210279	0.72803426	0.63520185
PSP 7	0.57298963	0.79673885	0.43093871	0.71810979	0.62969425
PSP 8	0.53139242	0.79739484	0.3930475	0.6844112	0.60156149
PSP 9	0.5262887	0.74697503	0.40405503	0.66746612	0.58619622
PSP 10	0.59032124	0.80116789	0.44157487	0.73116951	0.64105838
PSP 11	0.53806722	0.81958052	0.38836983	0.69114314	0.60929018
PSP 12	0.57118433	0.82740594	0.41178652	0.71650126	0.63171951
PSP 13	0.55279023	0.7973284	0.40804346	0.70170652	0.61496715

Tabelle 14 PSPNet Klassifizierung korrigierter Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
PSP 1	0.80220645	0.87521396	0.57829542	0.88906285	0.78619467
PSP 2	0.76251734	0.83185907	0.57393317	0.8622082	0.75762944
PSP 3	0.51564755	0.60796924	0.50746112	0.67194883	0.57575668
PSP 4	0.76079082	0.84253869	0.5659499	0.86241699	0.7579241
PSP 5	0.7075256	0.7685739	0.57270641	0.82051111	0.71732926
PSP 6	0.80414975	0.91564321	0.55503513	0.89055403	0.79134553
PSP 7	0.81142371	0.92839472	0.55360746	0.89519698	0.79715572
PSP 8	0.77361503	0.90980729	0.53503142	0.87047451	0.77223206
PSP 9	0.80076189	0.90938738	0.55571087	0.8882105	0.78851766
PSP 10	0.82382654	0.94065724	0.55544553	0.90265031	0.8056449
PSP 11	0.74647446	0.93942076	0.50272583	0.85399625	0.76065432
PSP 12	0.78836683	0.94371638	0.52947471	0.88108891	0.78566171
PSP 13	0.76388483	0.90513358	0.52958237	0.86420916	0.76570249

Hier gibt sich zwischen den Werten in Tabelle 13 und Tabelle 14 schon ein stimmigeres Bild, da jeweils das Setup von PSP 10 gute Werte liefert, das ja bereits anhand der Gütemasse von ArcGIS Pro als gutes Modell etabliert ist. Das Modell überzeugt also auch bei der eigentlichen Anwendung von Pixelklassifizierungen. Weiter zeigen PSP 6 bei dem Abgleich mit den regulären und PSP 7 bei dem Abgleich mit den korrigierten Daten. PSP 6, dass bei den Tabellen Tabelle 11 und Tabelle 12 noch nicht mit speziell guten Werten auffällt, ist das Setting mit weniger Layern, mit dem ResNet 18 Backbone-Modell. So kann auch mithilfe der Abbildung 50: veranschaulicht werden, dass sich der PSPNet-Modelltyp sich auch für weniger tiefe Architekturen von neuronalen Netzwerken eignen kann.

Was bei Anbetracht dieser Tabellen ebenfalls auffällt, ist die deutliche Leistungsschwäche von PSP 3, die sich bei der Auswertung durch ArcGIS nicht gezeigt hat. Während die Klassifizierung der Wetlandpixel dem Modell keine Probleme bereitet, wie in Abbildung 51: illustriert ist, liegt der Ursprung der schlechten Werte in vielen, scheinbar wahllosen Fehlklassifizierungen, wie sie in der Abbildung 52 veranschaulicht werden. Es erstaunt somit, dass das Default-Modell, welches die Parameter des Backbone-Modells nicht anpassen kann, nicht mit solchen Problemen konfrontiert ist, und damit besser abschneidet. Auch gerade dadurch, dass der Einsatz des Unfreeze-Parameters bei den U-Net-Modellen, deren Architektur ja auch für PSPNet-Modelle verwendet wird, zu Leistungssteigerungen führt, wie sie auch von ArcGIS vorhergesagt werden. Die Kombination von Unfreeze und PSPNet ist aber trotzdem legitim, wie sich ja auch durch PSP 10 zeigt. Im Verlauf des Trainingsprozesses wurde hier wohl ein ungünstiges Validierungsset gewählt, dass den Loss dieser Epoche enorm ansteigen liess und die Modellparameter dadurch unpassend verändert wurden. Auch die Loss-Funktion von PSP 10 zeigt keinen idealen Verlauf, aber im Vergleich, mit der von PSP 3 (siehe Abbildungen Abbildung 54 und Abbildung 54) zeigt sich schnell, dass der Trainingsverlauf sich hier nicht ideal entwickelt hat.

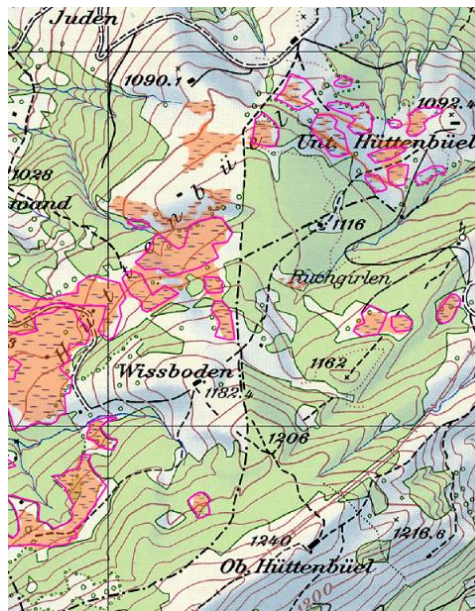
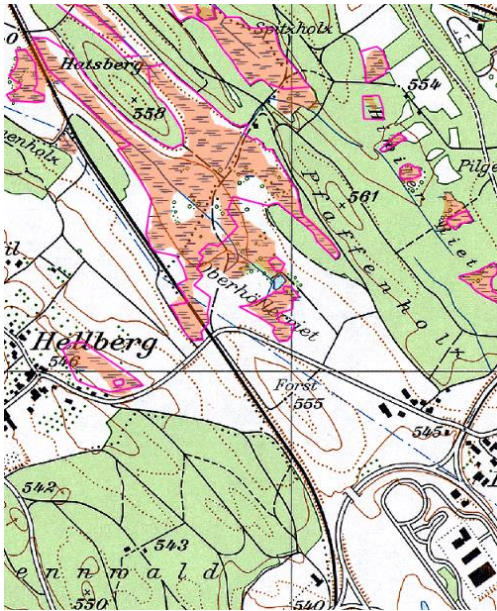


Abbildung 50: Ausschnitte zweier Klassifizierungen durch PSP 6 mit der Groundtruth-Outline

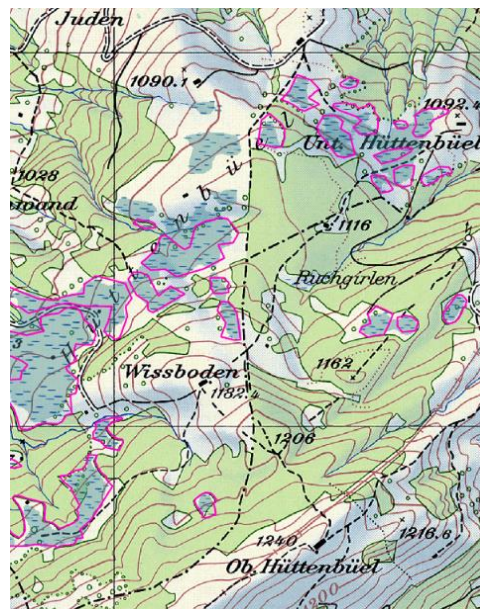
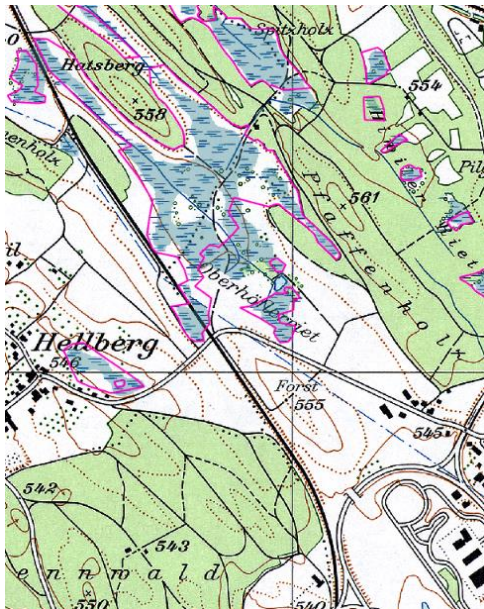


Abbildung 51: Ausschnitte zweier Klassifizierungen durch PSP 3 mit der Groundtruth-Outline

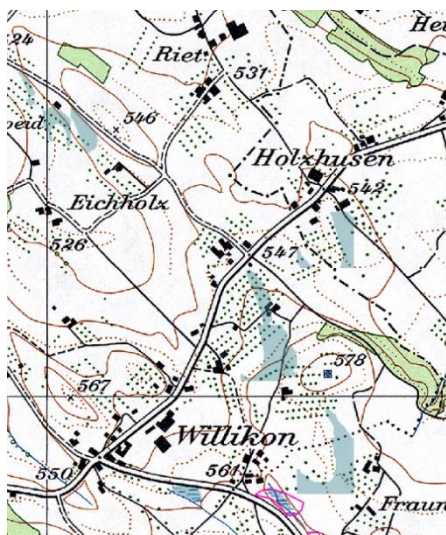


Abbildung 52: Fehlerhafte Klassifizierungen von PSP 3

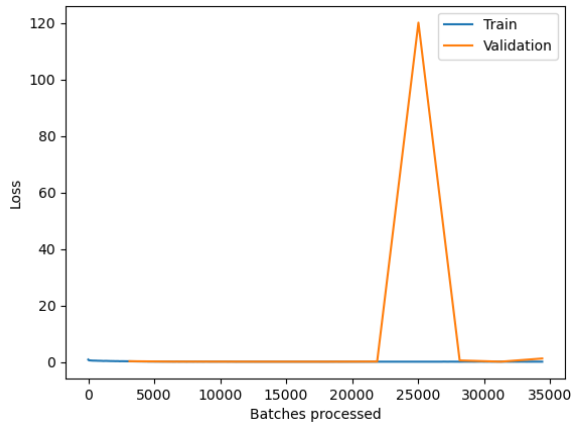


Abbildung 54: Loss-Funktion von PSP 3

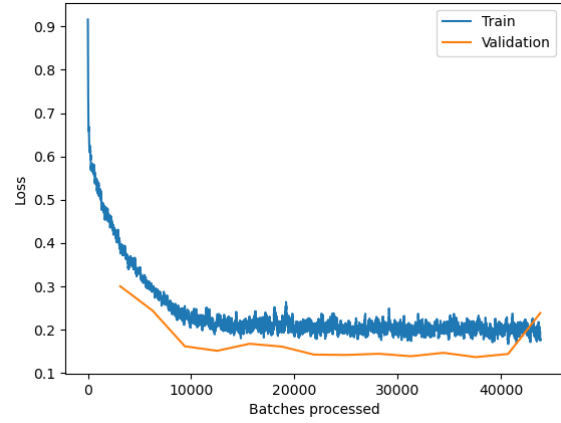


Abbildung 53: Loss-Funktion von PSP 10

4.3.1.3 DeepLabV3

DL: Metriken der Trainingsphase

Tabelle 15 DeepLab-Modell Metriken der Trainingsresultate mit den regulären Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
DL 1	0.693732584	0.862293	0.824917	0.843191	0.889268527
DL 2	0.680644048	0.873709	0.799788	0.835116	0.884536018
DL 3	0.706982138	0.882117	0.824386	0.852275	0.895408166
DL 4	0.681729863	0.856696	0.816615	0.836175	0.884585524
DL 5	0.69846816	0.836062	0.851081	0.843505	0.889597252
DL 6	0.659921807	0.864107	0.785511	0.822937	0.876461598
DL 7	0.61147922	0.814290	0.752345	0.782093	0.851488311
DL 8	0.619323454	0.852070	0.746258	0.795662	0.859487374
DL 9	0.618767016	0.851383	0.737847	0.790560	0.85748423
DL 10	0.635244936	0.851973	0.754839	0.800470	0.863592049
DL 11	0.63107165	0.837731	0.758718	0.796270	0.860695094
DL 12	0.616299298	0.860069	0.736886	0.793726	0.858722407
DL 13	0.624470854	0.862305	0.725147	0.787800	0.857826078

Tabelle 16 DeepLab-Modell Metriken der Trainingsresultate mit den korrigierten Daten

	mIOU	Präzision	Recall	F1-Score	Durchschnitt
DL 1	0.856901024	0.92758	0.928693	0.928136	0.946851905
DL 2	0.861150779	0.917386	0.942266	0.92966	0.948081978
DL 3	0.846739677	0.93565	0.912706	0.924035	0.943685523
DL 4	0.824565019	0.924651	0.926979	0.925814	0.941149575
DL 5	0.85472327	0.917664	0.910735	0.914187	0.940304103
DL 6	0.845646747	0.927935	0.918143	0.923013	0.942941443
DL 7	0.824500959	0.909197	0.912503	0.910847	0.934326119
DL 8	0.853465269	0.922304	0.926900	0.924596	0.944725867
DL 9	0.831482441	0.912394	0.916634	0.914509	0.937073544
DL 10	0.849345007	0.918365	0.928356	0.923333	0.943562128
DL 11	0.849685982	0.920326	0.923225	0.921773	0.942877724
DL 12	0.846795453	0.921785	0.921465	0.921625	0.942422693
DL 13	0.850313935	0.923844	0.923249	0.923546	0.943765808

Auch im Vergleich der beiden Tabellen Tabelle 15 und Tabelle 16 stimmen die besten Modelle der regulären und korrigierten Daten nicht überein. Es fällt aber auf, dass die Höchstwerte der einzelnen Metriken alle bei den Modell-Setups auftreten, an denen nur eine einzige Anpassung vorgenommen wird. Auch der Einsatz tieferen neuronalen Netzwerken scheint sich nicht positiv auf die DL-Modelle auszuwirken, obwohl auch diese dafür gedacht sind, unter Verwendung vieler Layer gut zu performen. Bei der regulären Groundtruth liefert DL 3 mit dem Unfreeze-Setup in allen Metriken sehr gute Werte. Nur der Recall ist etwas tiefer als beim zweiten relevanten Modell, DL 5, aber auch dieser Wert ist bei DL 3 überdurchschnittlich hoch. DL 5 überzeugt anhand der Gütemasse ebenfalls, weist aber eine tiefere Präzision auf, wodurch Vorsicht geboten werden muss. Bei den Klassifizierungen dieser beider Modelle, die Abbildungen Abbildung 55 und Abbildung 56 zeigen jeweils das Potenzial der Modelle auf, aber bei DL 5 zeigen sich bereits die fälschlicherweise als Wetland klassifizierten Pixelcluster, die in der Abbildung 57 noch näher aufgezeigt werden. Wiederum spiegelt sich dies hier in den Metriken nicht klar wider.

Bei den Validierungsdaten der korrigierten Groundtruth schneiden die DL1 und DL 2 Modelle am besten ab. Es soll aber erwähnt werden, dass alle DL-Modelle hohe Werte der Gütemasse aufweisen und die Unterschiede zwischen den Metrik-Werten sehr fein eingestuft sind. Für DL 2 mit den meisten Höchstwerten wird das Mixup-Setup angewandt. Die Augmentation der Trainingsdaten äussert sich hier also positiv. Mit DL 1 hat man wiederum das Default-Setup ohne weitere Anpassungen. Wiederum wäre es zu erwarten gewesen, dass Anstelle des Default-Modells mit ResNet 34 sich die Default-Modelle mit den komplexeren Backbone-Modells in den Vordergrund heben, oder dass die Variabilität der Parameter bei dem Unfreeze-Setup bei der Auswertung bessere Resultate liefert. Da aber alle Werte relativ nahe beieinander liegen, so lässt sich hier auch wieder das Argument machen, dass Feinheiten bei der Klassifizierung mit der Vorstellung der Wetland-Ausdehnung des Korrektors übereinstimmen müssen, und so Modelle, die das Label «Wetland» grosszügiger verteilen in der Metrik bevorzugt werden. Die Klassifizierungen sind in Abbildung 58 und Abbildung 59 dargelegt.

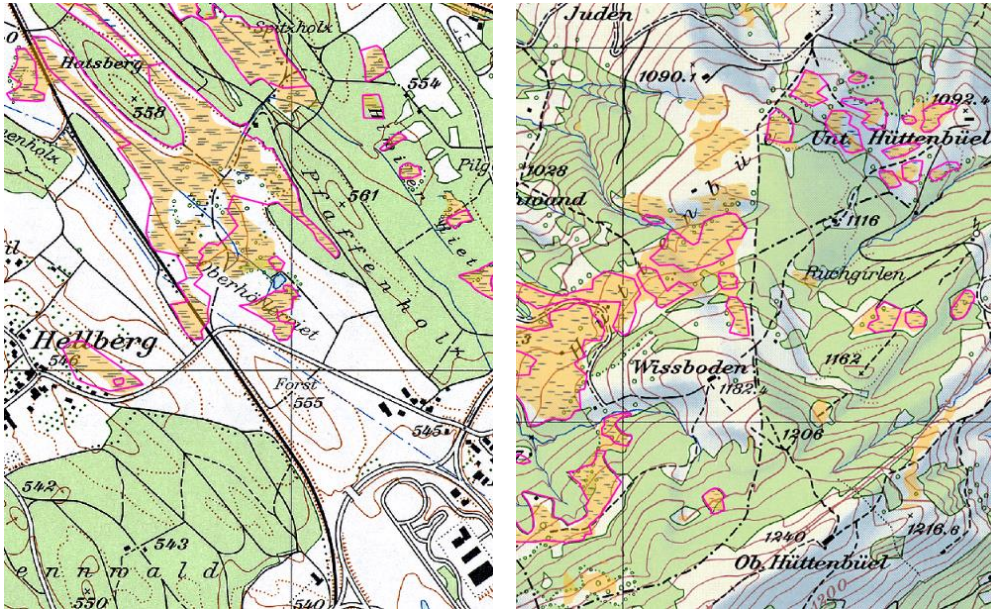


Abbildung 55: Ausschnitte zweier Klassifizierungen durch DL 3 mit der Groundtruth-Outline

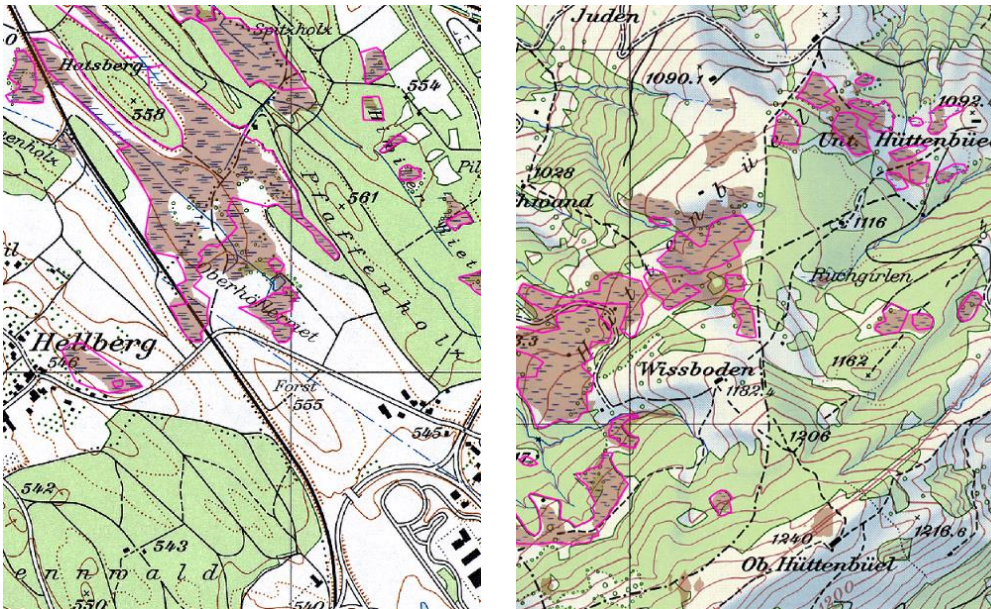


Abbildung 56: Ausschnitte zweier Klassifizierungen durch DL 5 mit der Groundtruth-Outline



Abbildung 57: Fehlerhafte Klassifizierungen von DL 5

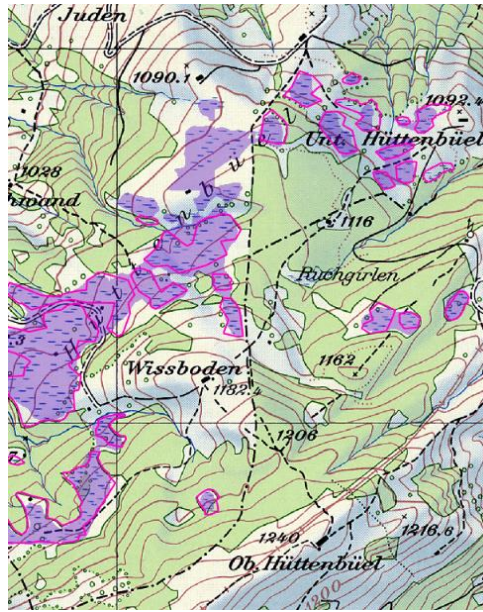
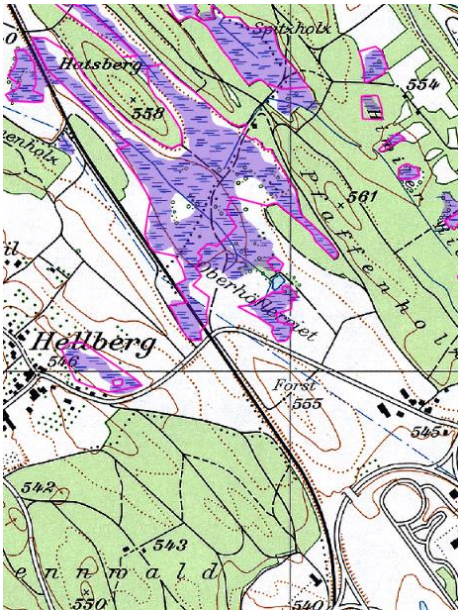


Abbildung 58: Ausschnitte zweier Klassifizierungen durch DL 2 mit der Grundtruth-Outline

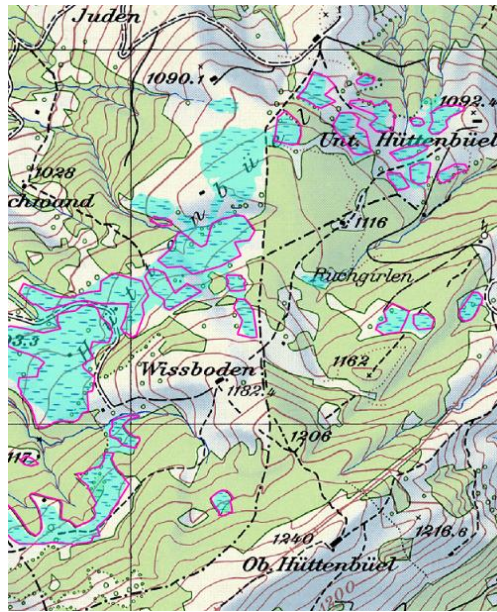
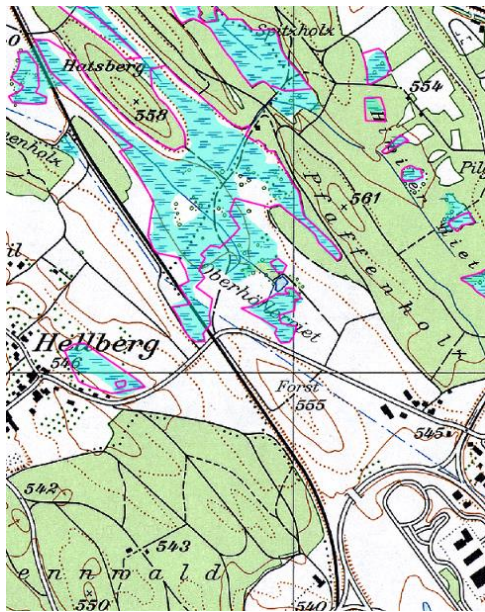


Abbildung 59: Ausschnitte zweier Klassifizierungen durch DL 1 mit der Grundtruth-Outline

DL: Metriken der Pixelklassifizierung

Tabelle 17 DeepLab Klassifizierung originaler Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
DL 1	0.58506378	0.76024878	0.45429535	0.7282907	0.63197465
DL 2	0.59591314	0.79471166	0.44536348	0.73642809	0.64310409
DL 3	0.49194987	0.62011729	0.43743506	0.64873596	0.54955954
DL 4	0.4980332	0.65038151	0.44961005	0.65559899	0.56340594
DL 5	0.5321294	0.67251635	0.45589487	0.68769139	0.587058
DL 6	0.58956416	0.7917327	0.44243047	0.73060587	0.6385833
DL 7	0.56466923	0.7574179	0.44238987	0.71193857	0.61910389
DL 8	0.54557554	0.82316201	0.39814597	0.69408629	0.61524245
DL 9	0.57208137	0.78755879	0.43076129	0.71864369	0.62726129
DL 10	0.59689982	0.80063589	0.44802568	0.73807938	0.64591019
DL 11	0.59361733	0.77842969	0.45607881	0.73358271	0.64042713
DL 12	0.57885528	0.80468759	0.42855277	0.72243769	0.63363333
DL 13	0.58094901	0.78756603	0.43637981	0.724339	0.63230846

Tabelle 18 DeepLab Klassifizierung korrigierter Kartenblätter

	IOU	Präzision	Recall	F1-Score	Durchschnitt
DL 1	0.78319336	0.86472416	0.56820263	0.87595787	0.77301951
DL 2	0.78994893	0.89110997	0.55741875	0.88084996	0.7798319
DL 3	0.72493942	0.81635896	0.54870982	0.83577052	0.73144468
DL 4	0.70296991	0.77193884	0.56137903	0.81496346	0.71281281
DL 5	0.71573558	0.78582982	0.56564366	0.83075169	0.72449019
DL 6	0.82007737	0.9319017	0.55752199	0.90020838	0.80242736
DL 7	0.79640927	0.88377041	0.5683372	0.88582625	0.78358578
DL 8	0.79059975	0.95266179	0.52673301	0.88145325	0.78786195
DL 9	0.7821329	0.91819849	0.53715128	0.87684242	0.77858127
DL 10	0.82184283	0.93518747	0.55757239	0.90179728	0.80409999
DL 11	0.80879956	0.92334852	0.55461805	0.89368932	0.79511386
DL 12	0.80248	0.93290992	0.54429027	0.88943061	0.7922777
DL 13	0.79936451	0.92993409	0.54423813	0.88789721	0.79035849

Beim Betrachten der Tabellen Tabelle 17 DeepLab Klassifizierung originaler Kartenblätter und Tabelle 18 DeepLab Klassifizierung korrigierter Kartenblätter sind die guten Metriken etwas breiter über die Setup-Variationen verteilt als bei den Gütemassen des Trainingsprozesses. DL 10 kann sich aber,

abgesehen vom Recall-Wert, bei den regulären als auch bei den korrigierten Kartenblättern profilieren. Wiederum scheint hier die Kombination von Mixup und Unfreeze erfolgreich zu sein. Die Klassifizierung der Pixel, wie sie in Abbildung 60 zu sehen sind, hinterlassen auch ein sehr stimmiges Bild von Wetland-Features. Neben DL 2, welches sich bereits bei der vorigen Abhandlung der DL-Modelle als geeignet herausstellt, fällt hier DL 6, zumindest bei den korrigierten Daten, positiv auf. Entgegen der Eignung der PSPNet-Modelle für tiefe Netzwerke handelt es sich hier um das Setup mit dem einfacheren ResNet 18 Backbone-Modell. Wiederum kann es aber gut sein, dass genau diese simplen gehaltenen Parameter und der Hang zur Generalisierung sich in der Auswertung der Kartenblätter positiv äußert. Die resultierende Klassifizierung wird in Abbildung 61 gezeigt.

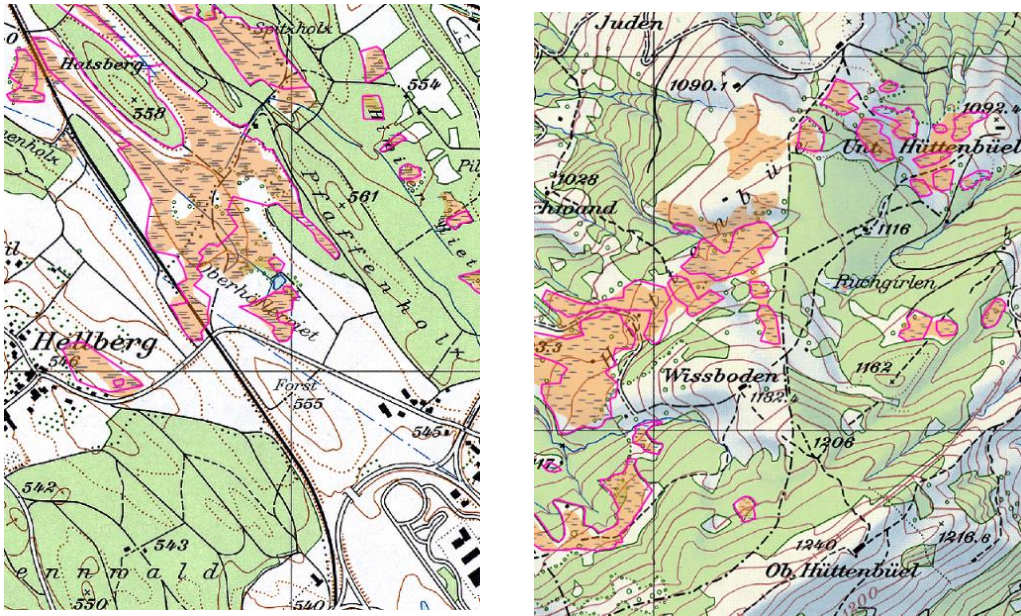


Abbildung 60: Ausschnitte zweier Klassifizierungen durch DL 10 mit der Groundtruth-Outline

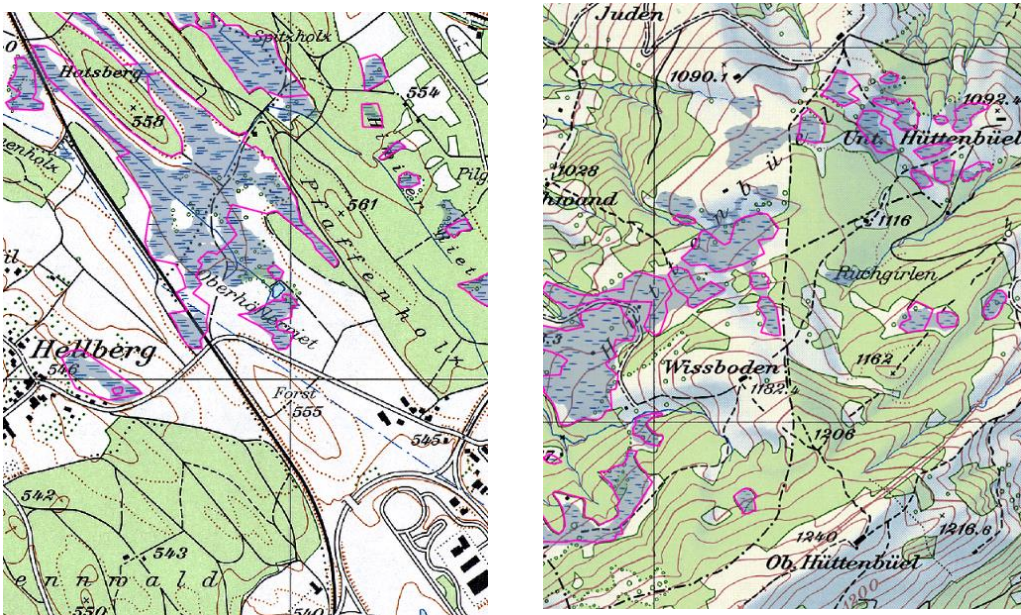


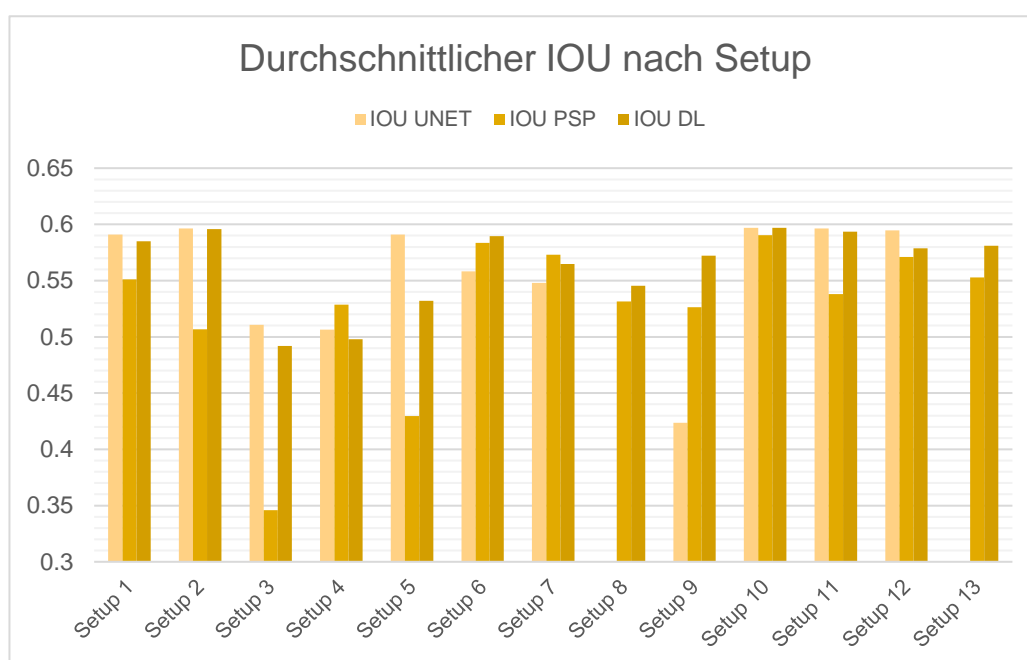
Abbildung 61: Ausschnitte zweier Klassifizierungen durch DL 6 mit der Groundtruth-Outline

4.3.2 Vergleich nach Setup

Nach der Analyse der individuellen Modelltypen folgt nun die verallgemeinerte Untersuchung mit dem Vergleich der erreichten Metriken zwischen den drei verschiedenen Modellen, eingeteilt nach dem Setup, auf welchem sie trainiert wurden. Während in 4.3.1 alle Kombinationen der Auswertungen betrachtet wurden, also einerseits die Unterscheidung nach Datensatz (regulär oder korrigiert) und die Unterscheidung nach Ursprung der Metrik (Trainingsphase oder Pixelklassifizierung) vorgenommen wird, werden die Setups hier anhand der Pixelklassifizierung und der darauffolgenden Auswertung mit den regulären Daten verglichen. So wird vermieden, dass fehleranfälligen Setups, die willkürliche Pixel als Wetland klassifizieren, trotzdem gute Werte erhalten, wie dies bei den ArcGIS-Metriken oft der Fall ist, wie es auch bereits in 4.3.1 aufgezeigt wurde. Auf die Auswertung nach den korrigierten Daten wird verzichtet, da dieser Datensatz nur drei Kartenblätter umfasst. Die Auswertung der sieben Kartenblätter mit den regulären Daten bieten hier eine grössere Spannweite. Auch im Sinne der vorliegenden Landschaften, die klassifiziert werden sollen. Durch das Vorkommen von Widersprüchlichkeiten in den Datensätzen sollen die folgenden Vergleiche also vor allem relativ betrachtet werden, nicht die Absolutwerte, die damit tief ausfallen, und die eigentliche Leistungsfähigkeit der Modelle nicht eins-zu-eins wiedergeben

IOU

Die durchschnittliche Intersection over Union erlaubt es, einen Eindruck darüber zu gewinnen, wie das Verhältnis zwischen gefundenen Wetland-Pixel und der Gesamtmenge aller wahren Wetland-Pixel, inklusive fälschlicherweise als Wetland klassifizierten Pixel, aussieht. Es gibt einen Einblick über die Genauigkeit, ohne zu stark vom Klassenungleichgewicht beeinflusst zu werden.



Plot 1: Balkendiagramm zum Vergleich der IOUs der drei Modelltypen nach Setup

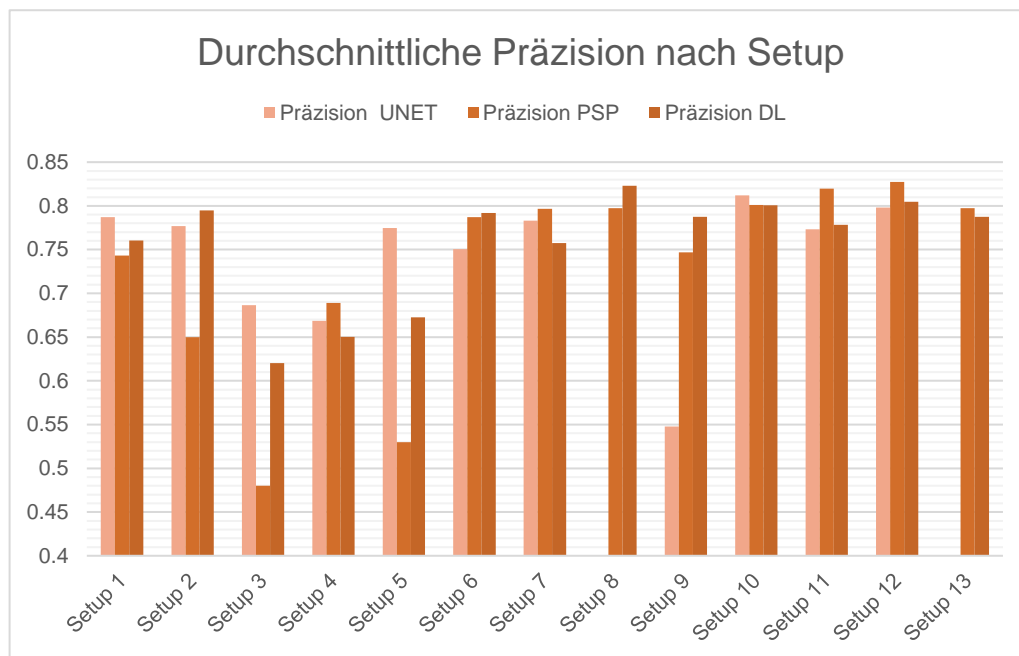
Wie dem Plot 1 zu entnehmen ist, erzielen die verschiedenen Modell-Setups nicht bei allen Modelltypen dasselbe Resultat. Generell entsteht der Eindruck, dass die IOUs der PSP-Modelle meist schwächer ausfallen als die der anderen beiden Modellarten. Es erstaunt durchaus, dass das Setup 3, welches mit dem Unfreeze-Prinzip eine bessere Anpassung an die Trainingsdaten erreichen soll, im Vergleich mit den anderen Setups so deutlich abfällt. Im Speziellen scheint PSP mit diesem Setup Probleme zu haben. Auf das Klassenungleichgewicht kann bei dieser Metrik auch nicht verwiesen werden, da die Anzahl True Negatives, also den vielen Nicht-Wetland-Pixeln, nicht in die Berechnung dieser Metrik miteinfliesst.

Weiter zeigt sich, dass die Performance über alle drei Modelle betrachtet, im Setup 10 am besten ist. Dabei handelt es sich um die Kombination von Mixup und Unfreeze. Es wirkt sehr paradox, dass die selbe Einstellung, die für die tiefsten Gütewerte verantwortlich ist, in Kombination mit der richtigen Datenaugmentation zu den besten Ergebnissen führen kann, oder diese zumindest nicht negativ beeinflusst. Die UNET- und DL-Modelle zeigen bereits bei der reinen Datenaugmentation in Setup 2 sehr starke Werte, aber die Funktionalität von PSP gewinnt durch dieses kombinierte Setup klar an Wert. Dass das Setup 5 ebenfalls sehr diverse Wertspannen liefert, kann darauf zurückgeführt werden, dass diese Learning Rate von 0.00002 sich zwar für UNET eignen mag, für DL und PSP aber eher unpassend ist.

Weiter soll noch hervorgehoben werden, dass trotz Architektur für tiefe Modelle, die Default-Settings von DL und PSP mit dem einfachsten Backbone-Modell, dem Setup 6 besser abschneiden als alle anderen mit den tieferen Architekturen.

Eine weitere spannende Eigenschaft, die sich feststellen lässt, ist dass das UNET-Modell mit Unfreeze und Focal Loss als einzelne Setups in 3 und 4 besser zurechtkommt, als wenn sie in Kombination in Setup 9 angewendet werden. In den anderen beiden Modellen resultiert dies in einer Steigerung der Leistung, nicht in einem Abfall.

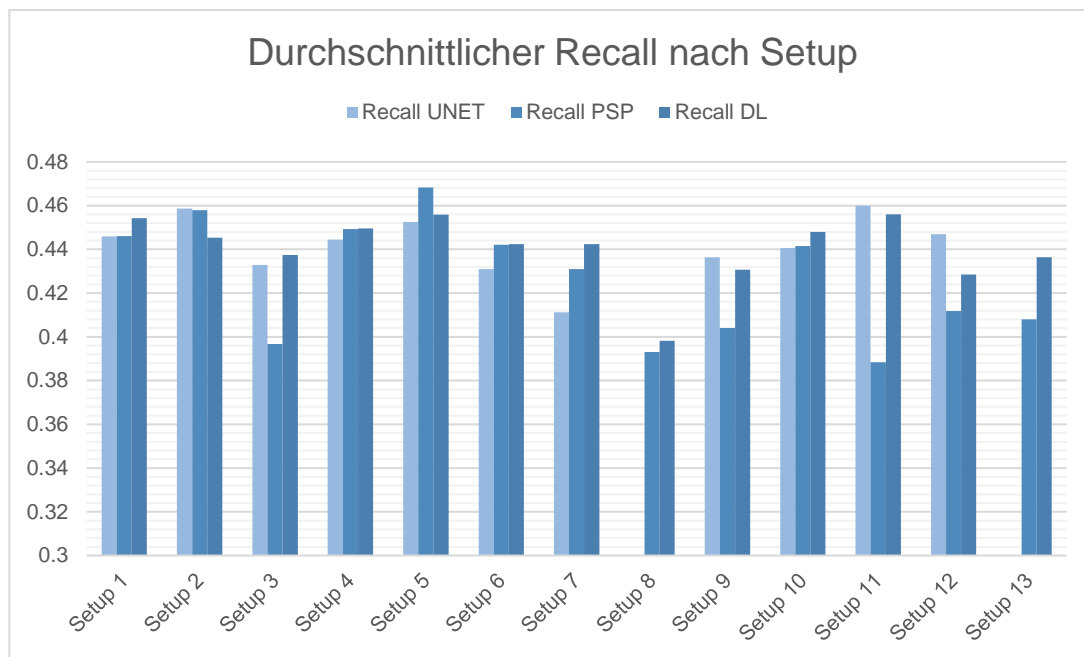
Präzision



Plot 2: Balkendiagramm zum Vergleich der Präzisionen der drei Modelltypen nach Setup

Die grobe Struktur des Plot 2 gibt eine ähnliche Silhouette wieder, wie es bei dem Plot 1, der IOU Auswertung gesehen wurde. Anders als beim IOU ist hier aber oftmals der Balken des PSP-Modells der längste. Hier findet wohl auch der Abtausch in den Metriken statt. Aufgrund der höheren Präzision, die PSP aufweist, werden weniger Pixel als Wetland klassifiziert, was zu Einbußen beim IOU führt. Es wiederholen sich die schwachen Leistungen im Setup 3 und 5. Die besten Ergebnisse verteilen sich hier auf die hinteren Setups wie 10 oder 12. Auch hier findet also die Kombination von Mixup und Unfreeze Anklang. Für den zusätzlichen Parameter des Focal Loss, wie man ihn im Setup 12 findet, scheint das Resultat wiederum besser zu werden, wenn das Backbone-Model simpler gehalten wird. Das Setup 12 mit ResNet 18 bringt allen Modellen eine bessere Präzision, als diejenigen mit ResNet 34 oder 50.

Recall



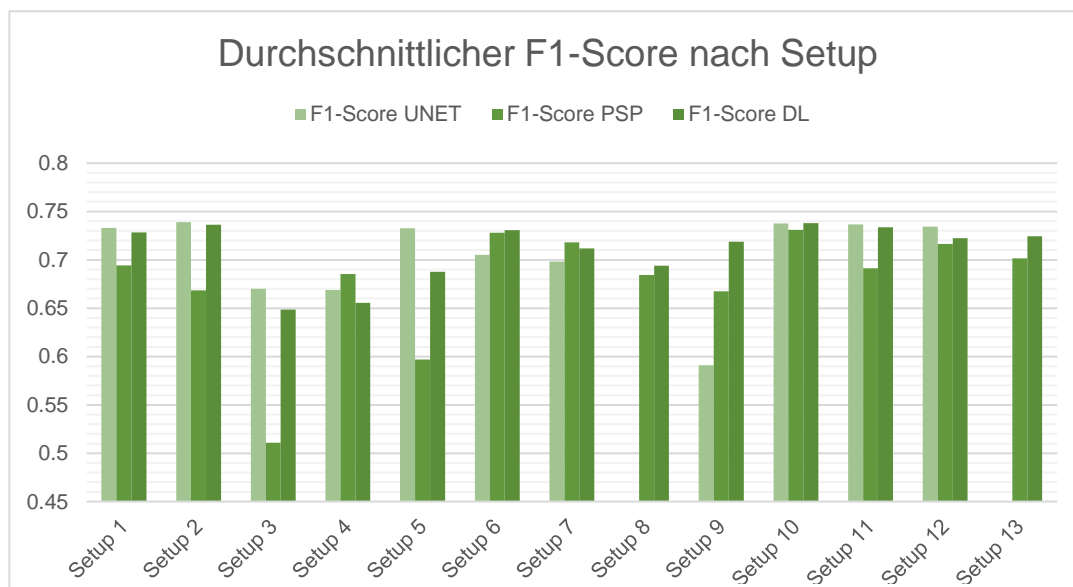
Plot 3: Balkendiagramm zum Vergleich der Recalls der drei Modelltypen nach Setup

Der Recall, der beschreibt, wie viele der wahren Wetland-Pixel vom Modell gefunden werden, liefert nun im Plot 3 ein anderes Abbild. Dadurch, dass Fehlklassifizierungen von Pixeln als False Positives keine Rolle spielen, schneiden hier nun zum Beispiel die Modelle in dem Setup 5 am besten ab. Da die Modelle, die mit der fixen Learning Rate trainiert werden, die Pixel oft wahllos und übermässig über die eigentliche Ausdehnung hinaus klassifizieren, zeigen sie gute Recall-Werte. Wenn ein Modell optimal sein soll, reicht es in der Regel nicht, wenn es nur einen hohen Recall aufweist.

Gleichzeitig sieht man, wie die tiefen neuronalen Netzwerke in Setup 7 und 8 in der Tendenz einen niedrigeren Recall-Wert aufweisen, da sie sich durch ihre Komplexität schneller spezialisieren, die Ausdehnung der Wetlands damit meist auf die Symbolisierung minimieren, und damit nicht für das generelle Klassifizieren geeignet sind.

Das Setup 10 zeigt auch hier für alle drei Modelltypen starke Resultate, wobei auch der Default im Setup 1 und die Datenaugmentation durch Mixup im Setup 2 gute Werte aufweisen.

F1-Score

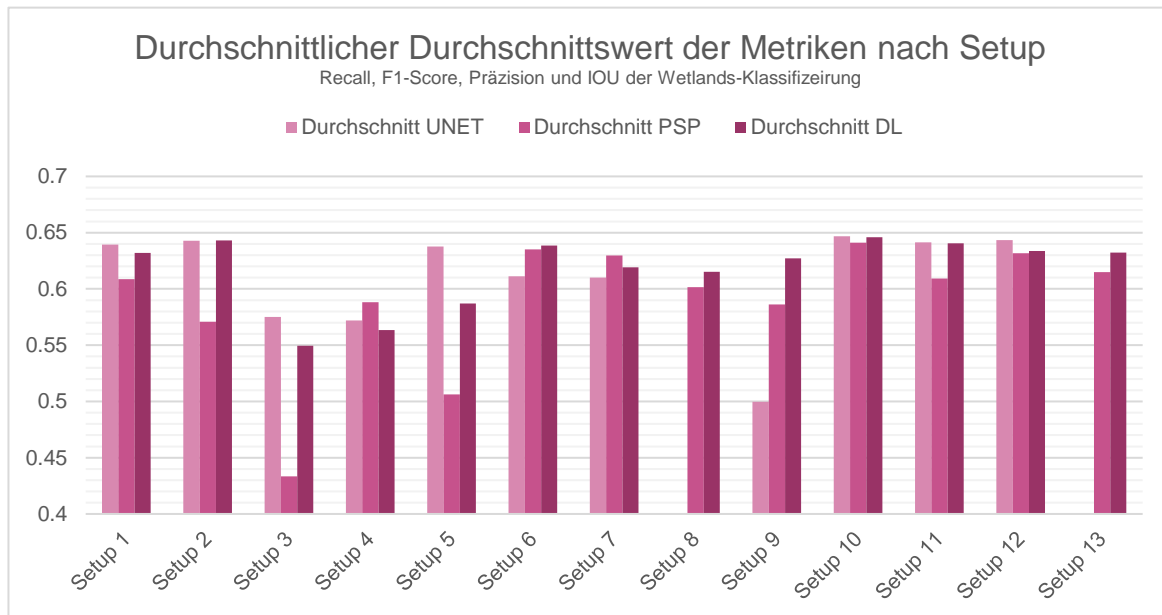


Plot 4: Balkendiagramm zum Vergleich der F1-Scores der drei Modelltypen nach Setup

Der F1-Score stellt im Plot 4 das harmonische Mittel von Präzision und Recall dar. Somit lassen sich diese beiden in einer einzigen Metriken gleichzeitig beurteilen.

So gleicht sich auch das Balkendiagramm etwas aneinander an, die Diskrepanzen zwischen den Setups fallen hier nicht mehr ganz so gross aus. Ausreisser bleiben nur Das Setup 3, im Speziellen das PSP-Modell, dass sich ohne Augmentation mit Mixup nicht für die Anwendung von Mixup eignet. Auch das Defizit des UNET-Modells im Setup 9 wird im Verhältnis von Präzision und Recall nicht ausgeglichen. Bemerkenswert ähnlich schneiden die drei Modelle erneut im Setup 10 ab. Auch das Setup 2 zeigt sich für UNET und DL sehr vielversprechend. Auch die beiden Setups 6 und 12 mit der Basis ResNet 18 sorgen bei allen Modelltypen für einen guten F1-Score.

Durchschnitt



Plot 5: Balkendiagramm zum Vergleich der Durchschnitte der drei Modelltypen nach Setup

Mit dem Durchschnitt aller vorigen behandelten Metriken wird im Plot 5 ein Bild wiedergegeben, wie es anhand der Analyse zu erwarten ist.

Das Setup 10 führt hier bei allen Modelltypen zu den generell besten Ergebnissen. In den meisten Setups wird UNET als das Modell mit den höchsten Metriken gewertet, wobei sich DL oft um die ähnlichen Wertespannen bewegt. PSP oft leicht ab, kann sich aber auch profilieren, wie zum Beispiel im Setup 4 oft dem tiefen Setup 7.

4.3.3 Vergleich nach Kartenblatt

Während dem Training des Modells in ArcGIS Pro werden die Validierungen der Modelle nur an kleinen Image-Chips vorgenommen, die sich auch aus den zur Verfügung gestellten Kartenblätter extrahieren lassen. Die zu klassifizierende Fläche, anhand welcher die Gütemasse gewonnen werden, fällt mit 256x256 Pixel sehr klein aus. Gerade auch in Anbetracht, dass die für diese Arbeit trainierten Modelle die Extraktion der Wetlands gesamter Kartenblätter vornehmen soll. Zusätzlich liegt die Annahme nahe, dass die Validierung sich auf Landschaftstypen beschränkt, die dem Modell durch das Training bekannt sind, da Trainings- und Validationsdaten von denselben Kartenblättern gewonnen werden und direkt nebeneinander verortet sein könnten. So kann es sich als interessant herausstellen, zu Vergleichen, ob und wie sich die Qualität der Klassifizierungen anhand der gewählten Geländetypen unterscheidet.

Bei den hier klassifizierten Kartenblättern werden die Auswertungen anhand der regulären Daten vorgenommen. So sind schlechte Gütemasse eines Kartenblattes, speziell wenn alle drei Modelltypen davon betroffen sind, nicht zwangsläufig Indiz für eine unzulängliche Qualität der Klassifizierung, sondern können in den meisten Fällen auf grossflächige Diskrepanzen zwischen dem Groundtruth-Layer und den eingezeichneten Wetlands auf dem historischen Kartenblatt zurückzuführen. Beispielhaft illustriert ist dies in den Ausschnitten der Abbildung 62, die das Tile 5, also das Kartenblatt 1113 zeigen.

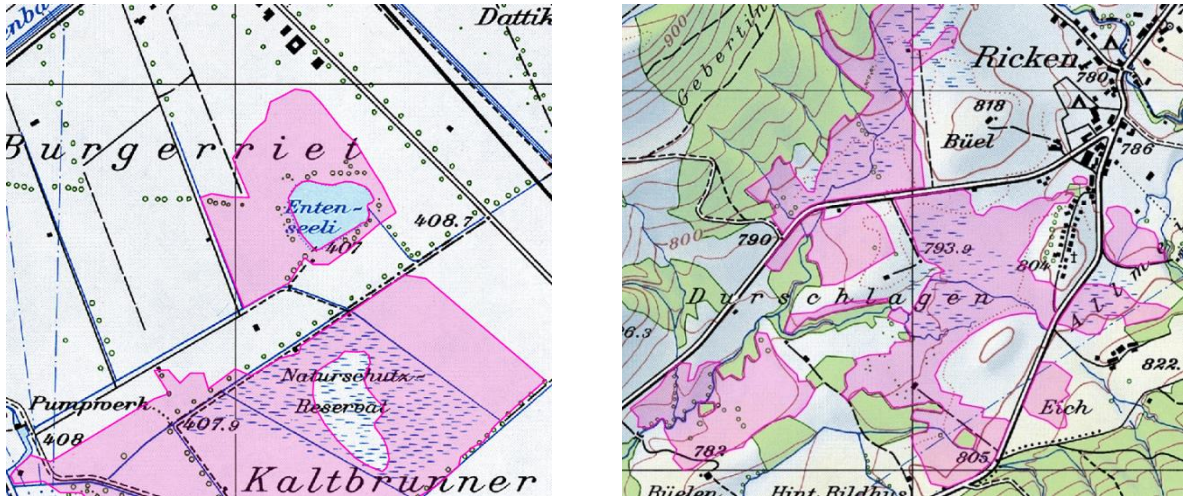
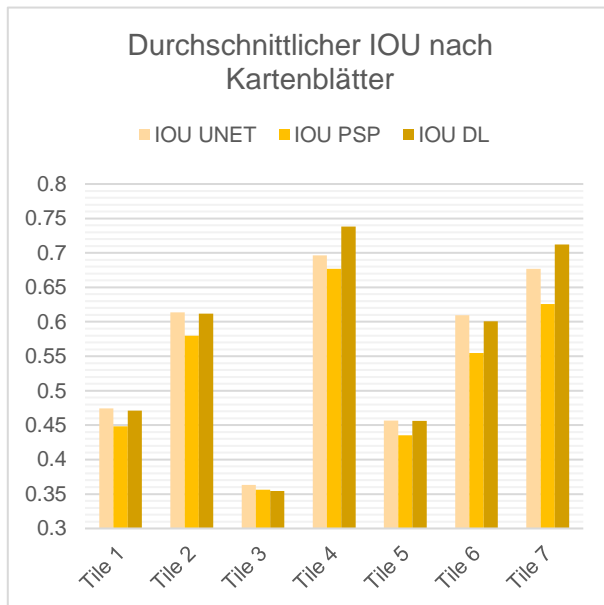


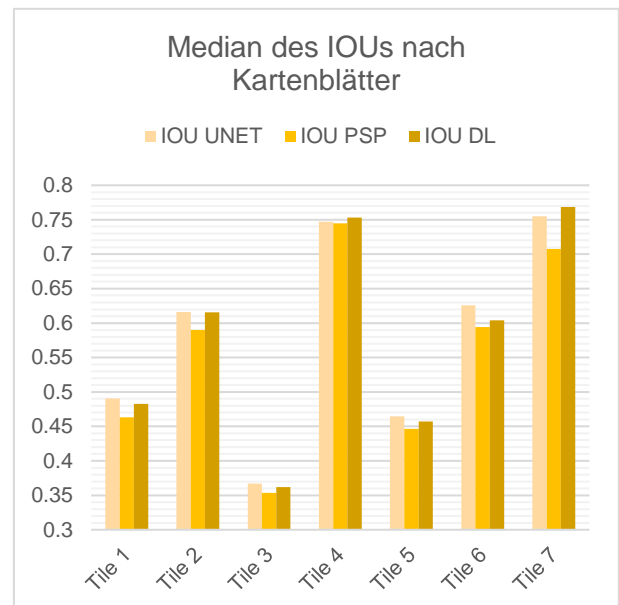
Abbildung 62: Ausschnitte fehlerbehaftete Groundtruth-Daten auf dem Tile 5

Für eine übersichtliche Analyse werden hier die durchschnittlichen Gütemasse aller vorhandenen Setups der jeweiligen Modelltypen verwendet. Um den Einfluss sehr starker und sehr schwacher Modellperformances zu unterbinden, wird ergänzend auch der Median illustriert.

IOU



Plot 6: Durchschnittlicher IOU nach Kartenblatt



Plot 7: Median IOU nach Kartenblatt

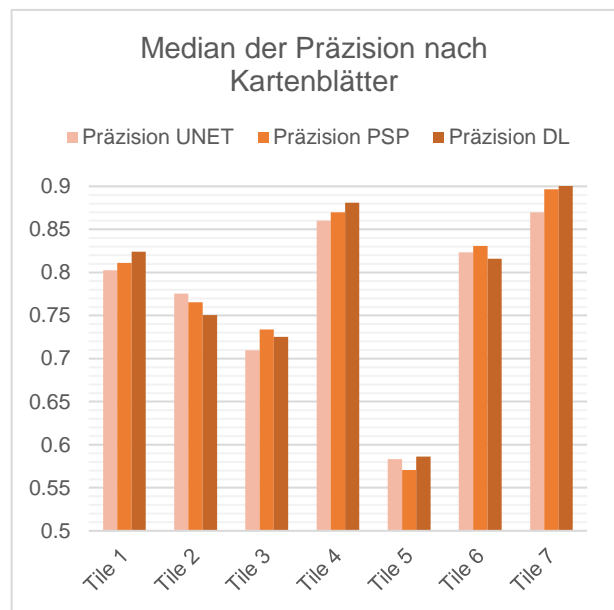
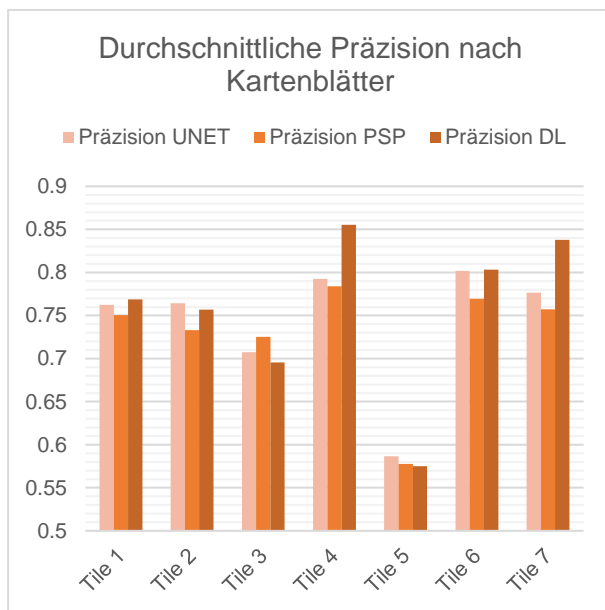
Es zeigt sich schnell, dass der erbrachte IOU von Kartenblatt zu Kartenblatt stark variiert. Da die Leistung der drei hier verglichenen Modelle aber immer in einem ähnlichen Ausmass bleibt, wird hier davon ausgegangen, dass dies der Datenqualität zugrunde liegt. Abgekoppelt von den absoluten Werten lässt sich der Trend erkennen, dass UNET und DL in dieser Metrik bessere Werte erzielt, unabhängig davon, auf welchem Kartenausschnitt das Modell angewendet wird.

Der beste arithmetische Durchschnitt liegt meist bei den DL-Modellen, während die Betrachtung der Mediane die Resultate angleicht und das UNET-Modell meist einen besseren IOU vorweisen kann.

Die beiden Kartenblätter mit den besten Resultaten, Tile 4 und Tile 7 sind diejenigen, deren Wetlands sich grossflächig über das Seeufer erstrecken. Was den Modellen hier sicher entgegenkommt, ist die dadurch klar entstehende Grenze, die zwischen Land und See gezogen werden kann. Durch die einfache Definition wo das Wetland aufhören muss, ist in der Groundtruth, als auch bei der Pixelklassifizierung klarer, welche Extent das Wetland einnimmt. Auch durch die zusammenhängenden, grossflächigen Polygone gibt es weniger Grenzen, die das Modell ziehen muss. Auch die Tiles 2 und 6 teilen diese Eigenschaft, dass viele Wetland-Features sich am Gewässerrand befinden, wenn auch nicht im selben Ausmass.

Die schlechtere Wertung des Tile 5 ist nicht auf unstimme Ausgangsdaten zurückzuführen, sondern auf die Polygonstruktur der Groundtruth-Feature. Während die Referenzdaten eher parzelliert und klein sind, neigen die trainierten Modelle dazu, die Pixel zwischen zwei Wetland-Symbolen auch als solche zu labeln, um möglichst zusammenhängende Features zu generieren. So werden die Modelle bei solchen Kartenblättern überdurchschnittliche viele False Positives generieren, was sich beim IOU negativ auswirkt.

Präzision

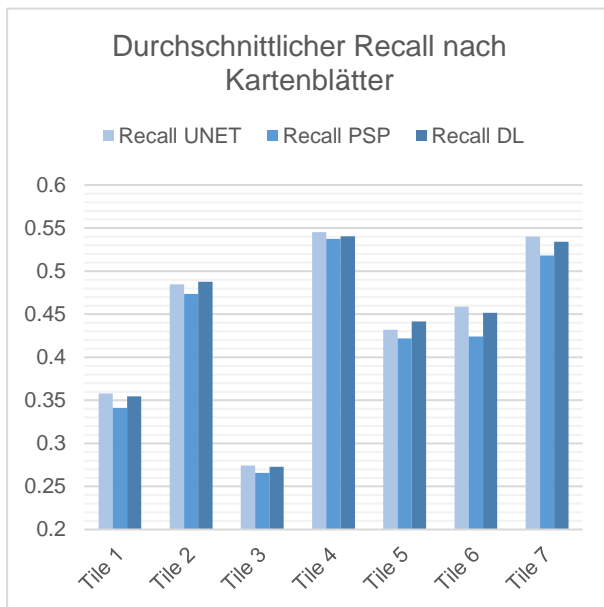


Plot 9: Durchschnittlicher Präzision nach Kartenblatt

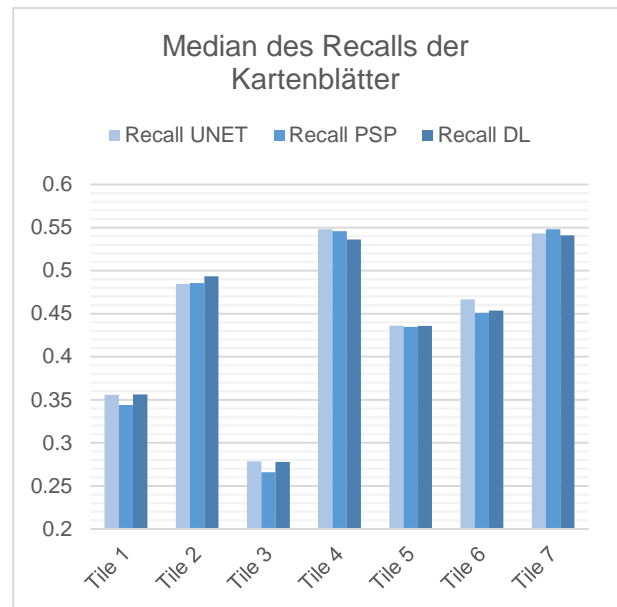
Plot 8: Median der Präzision nach Kartenblatt

In dem Vergleich der Präzisionen lässt sich kein so eindeutiges Schema wie beim IOU identifizieren. Die erreichten Werte der Präzisionsmasse fallen hier im Vergleich zwischen den Kartenblättern mit Ausnahme von Tile 5 einigermassen ähnlich aus. Die beiden Tiles 4 und 7 können nach wie vor modellübergreifend die besten Werte vorweisen. Die Präzision entspricht eigentlich dem IOU, ohne Einbezug der False Negatives, also der nicht gefundenen Wetland-Pixeln. Somit haben Groundtruth-Daten, wie sie in Abbildung 62 zu sehen keinen negativen Einfluss auf die Metrik. Für Kartenausschnitte wie Tile 5, mit dem gegenteiligen Problem, den gehäuften False Positives wird die Präzision klar abgewertet. Im Vergleich der Modelltypen untereinander gibt es zwar keinen klaren Spitzenreiter, aber die DL-Modelle im arithmetischen Mittel schlagen teilweise deutlich oben aus, können aber auch im Vergleich der Mediane am meisten überzeugen.

Recall



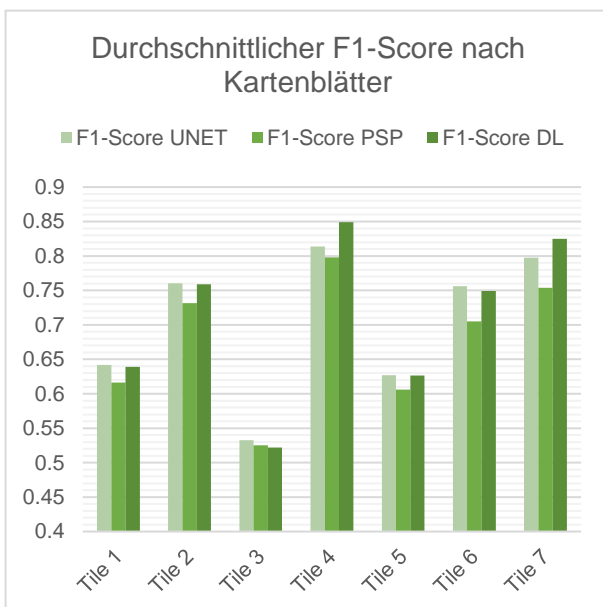
Plot 11: Durchschnittlicher Recall nach Kartenblatt



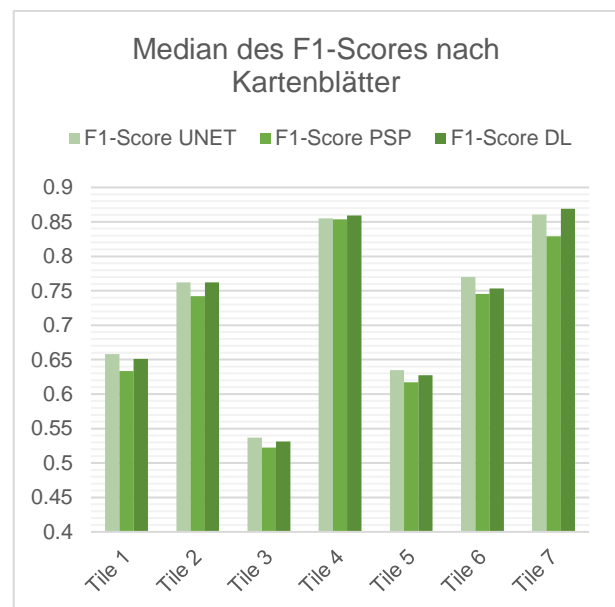
Plot 10: Median des Recalls nach Kartenblatt

Der den Metriken übergeordnete Trend der guten Resultate von Tile 4 und 7 zeigt sich auch im Recall. Auch Tile 2 kann hier sehr gute Werte vorweisen. Als Gegenstück zur Präzision wird mit dem Recall abgebildet, in welchem Verhältnis die identifizierten Wetland-Pixel zu der Gesamtmenge aller Wetland-Gebiete stehen. So spiegeln sich die Resultate von Tile 3 und Tile 5 im Bezug zur vorigen Metrik. Die Leistungen der Modelltypen bewegen sich kartenblattübergreifend auf einem ähnlichen Niveau, wobei wiederum UNET und DL meist bessere Resultate liefern als PSP. Im Schnitt, als auch beim Median sind die UNET-Modelle leicht zu bevorzugen, falls der Fokus auf dem Erfassen möglichst aller Wetland-Areale liegt.

F1-Score



Plot 12: Durchschnittlicher F1-Score nach Kartenblatt



Plot 13: Median des F1-Scores nach Kartenblatt

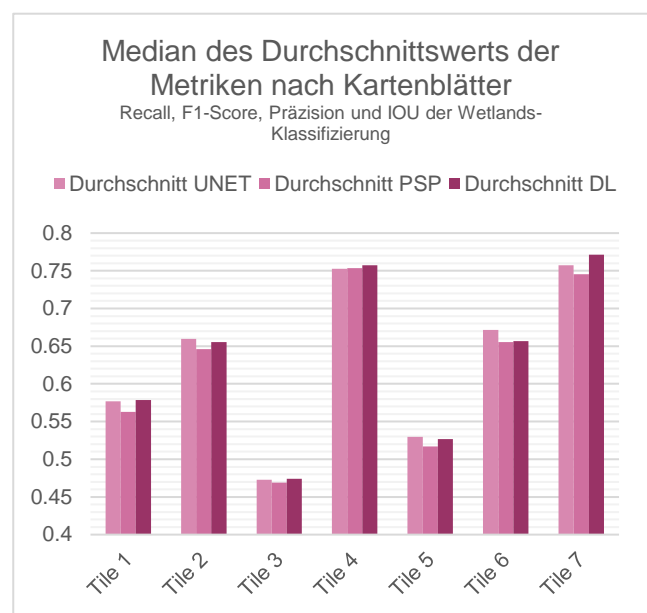
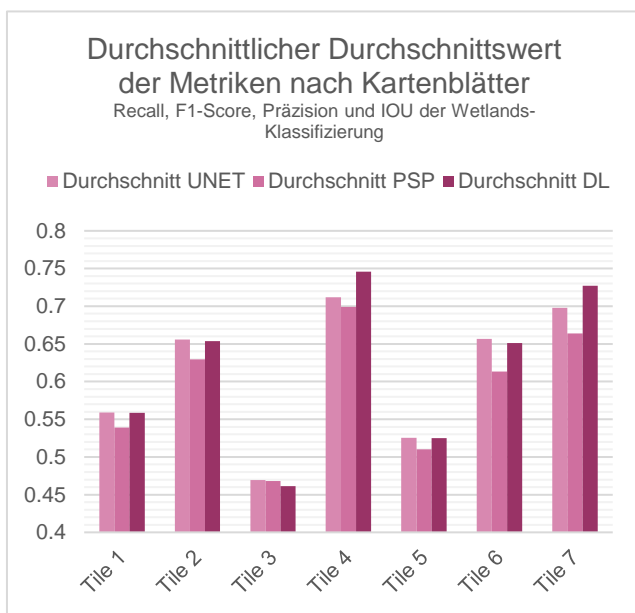
Als Harmonischer Durchschnitt von Präzision und Recall spiegelt der F1-Score diese Resultate wider. Folglich sieht man die guten Metriken der Tiles 2,4, 6 und 7. Es lässt auch darauf schliessen, dass sich Abweichungen von der Groundtruth zu den vorliegenden historischen Kartenblättern stärker auf dieses Gütemass auswirkt als die voreilige Klassifikation von Zwischenräumen zwischen Wetland-Symbolen. Dies folgt aber auch durch das Ungleichgewicht der Klassen. Für ausgeglichene Verhältnisse wären die Einflüsse ähnlicher.

Der reguläre Durchschnitt und der Median zeigen sich hier in den Plots Plot 13 und Plot 12 mit sehr ähnlichen Werten und zeigen einen ausgeglichenen Vergleich der Modelle. UNET und DL heben sich aber von PSP ab, während UNET über alle Setups betrachtet den besten F1-Score erzielt.

Durchschnitt

Zusammenfassend kann noch über alle Kartenblätter der Durchschnitt der Metriken betrachten, welche das erwartete Bild wiedergeben. Tile 3 und 5 stechen durch ihre tiefen Werte hervor, die anhand der Datenlage nachvollzogen werden kann. Tile 4 und 7 mit ihren wenigen, aber grossen Wetland-Polygonen an den Seeufern führen klar zu den besten Resultaten in den betrachteten Metriken. Und die Tiles 1,2 und 6 zeigen auf, dass bei einer regulären Datenlage auch alle Modelle robuste Werte erzielen können.

Für die Kartenblätter mit grossen Polygonen, ohne gröbere Fehlklassifikationen zeigt sich DL im Vorteil, während UNET bei den schwierigeren Kacheln mit mehr einzelnen verstreuten Polygonen die besseren Wert liefert.



4.4 Empfehlung

Auf Basis der Vergleiche der Modelle untereinander, der Setups und den Vergleichen fällt es schwer, ein finales Fazit und eine universale Empfehlung für das Extrahieren von Wetland-Features zu formulieren. Es gilt sicher zu Bedenken, was im Rahmen der Extrahierung erreicht werden soll und unter welchen Umständen diese ausgeführt wird. Ebenfalls fließt auch die Subjektivität in die Extrahierung dieser spezifischen Features mit hinein. So erschwert es sich weiter, eine Empfehlung auszusprechen, ohne die konkreten Anforderungen zu kennen, die ein potenzieller Anwender haben könnte.

Ganz grundsätzlich wird mit den drei Modelltypen, die im Fokus dieser Arbeit abgehandelt werden, schon mal ein gewisser Empfehlungsrahmen gebildet. Die Voruntersuchung aller Pixel- und Objektklassifizierungsmodelle, die im Rahmen dieser Analyse betrachtet werden, erlaubt bereits die erste Verfeinerung der möglichen Optionen auf die drei Modelltypen U-Net, PSPNet und DeepLabV3.

Wie in den Vergleichen in 4.3 gezeigt wird, lässt jedes Setup, jedes Kartenblatt je nach verwendeten Auswertungsmethoden ganz unterschiedliche Konklusionen zu.

Eine generelle Empfehlung kann bereits vorneweg für den vorgegebenen Workflow in ArcGIS Pro ausgesprochen werden. So ist zwar der Arbeitsablauf vorbestimmt und nur punktuell beeinflussbar, aber dies erlaubt auch weniger versierten Fachpersonen einen einfachen Zugang zu Deep Learning. Hier wird nun davon ausgegangen, dass die Extrahierung dieser Wetland-Features grossflächig von ganzen, historischen Kartenblättern der Schweizerischen Landeskarten erfolgen soll und die klassifizierten Pixel die Groundtruth so ähnlich wie möglich abbilden sollen.

Unter diesen angenommenen Umständen kann für den Arbeitsrahmen mit der Image Analyst Toolbox und dem Deep Learning Toolset empfohlen werden, dass man sich für das Setup 10 entscheidet. Die Kombination aus Datenaugmentation durch das Mixup-Prinzip und die ermöglichte Anpassungsfähigkeit der Backbone-Parameter durch das «Unfreeze» liefert in der praktischen Anwendung Modelle die durchwegs stabilsten Resultate. Und dies für alle drei Modelltypen, jeweils. Trotzdem lässt sich auch hier eine Empfehlung aussprechen, die aber je nach vorhandener Datenmenge oder Rechenkraft der Maschine angepasst werden kann. PSPNet liefert durchaus respektable Resultate, liegt aber im Vergleich zu den anderen beiden Modellen doch meist im Hintertreffen. So sollte die Modelltypwahl entweder auf U-Net oder DeepLabV3 fallen.

Während U-Net mit 27 der erlaubten 80 Epochen deutlich mehr Durchgänge braucht als DeepLabV3, welches bereits nach nur 14 Epochen keine Verbesserungen der Modellparameter mehr verzeichnet, so dauert es doch rund zwei Stunden länger, um ein DeepLab-Modell zu trainieren. Für Vorhaben mit sehr grossen Mengen an Trainingsdaten und einem beschränkten Zeitbudget kann also U-Net empfohlen werden. Für kleinere Anwendungsbereiche oder für allfällige Testversuche kann es sich auch lohnen, beide Modelltypen zu trainieren und Situationsabhängig zu entscheiden, welche Vorteile durch die jeweilige Modellwahl erbracht werden kann.

5 Diskussion

5.1 Interpretation

Da die Resultate, wie sie im 4.3 abgehandelt werden, grösstenteils aus Tabellen und Zahlenwerten bestehen, die für den Kontext der Interpretation von grosser Bedeutung sind, wird dies Interpretation der Resultate direkt in den Unterkapiteln von 4.3 aufgearbeitet. Durch die von der Interpretation losgelösten Präsentation der erhaltenen Gütemasse sollte zu differenzieren sein, welche Werte durch die angewandte Methodik erlangt werden, und wobei es sich um die Auslegungen dieser erhaltenen Resultate handelt.

5.2 Einschränkungen

5.2.1 Datengrundlage

Eine grosse Herausforderung der pixelbasierten Klassifikation mit den vorliegenden Daten ist, dass sich falsch klassifizierte Groundtruth-Daten je nach ihrem Ausmass so stark auf die Auswertung der Klassifizierung auswirken kann, dass diese ihre komplette Aussagekräftigkeit verliert.

Dadurch, dass die korrekten Groundtruth-Daten erst nach dem eigentlichen Trainingsprozess in den Auswertungsprozess miteinbezogen wurde, konnte nicht der volle Nutzen aus diesem zusätzlichen Datensatz gezogen werden. Da aber auch bei zukünftigen praktischen Anwendungen solcher Modelle nicht garantiert sein kann, dass die Trainingsdaten fehlerfrei sind, so ist es auch eine wertvolle Information zu wissen, dass das Modell auch mit fehlerbehafteten Daten umgehen kann.

Eine zusätzliche Einschränkung, die durch die Eigenheit der Symbolisierung der Wetland-Features gegeben ist, ist das Fehlen klarer Grenzen auf den historischen Kartenblättern. So reicht es nicht, wenn Modelle Wetland-Gebiete richtig identifizieren, sondern die Ausdehnung muss so exakt wie möglich mit der Groundtruth übereinstimmen. Durch die hohe Sensibilität der Pixelklassifizierung können so fehlende Pixel schnell ins Gewicht fallen und die Performance des Modells gemäss der Metriken schmälern.

5.2.2 Modellaufbau

Eine weitere Einschränkung wird durch ArcGIS Pro auferlegt. Während zwar offengelegt wird, auf welchen Architekturen die Modelle basieren, deren genaue Umsetzung, oder wie das trainierte Modell schlussendlich aufgebaut ist, lässt sich nicht nachvollziehen.

6 Zusammenfassung

6.1 Zielerreichung

Die Grundaufgabe, die im Zusammenhang mit dieser Arbeit als Hauptziel formuliert wurden, kann als erreicht betrachtet werden. Die Antwort auf die Frage, die welches Modell die besten Ergebnisse in der Extrahierung von Wetland-Features liefert, findet sich im Abschnitt 4.4.

6.2 Schlussfolgerung

Schlussfolgernd kann festgehalten werden, dass Deep Learning ganz klar das Potential hat, wichtige Prozesse in der Geoinformatik und Kartografie zu verbessern und zu revolutionieren. Mit dem einfachen Zugang, der durch ein weltweitverbreitetes Geoinformationssystem wie ArcGIS Pro nun gewährt wird, kann sehr bedeutend sein. Durch die einfach erklärten Parameter, der klar vorgegebene Workflow und die grösstenteils bedingungslos funktionierenden Modelle wird der Einstieg in dieses komplexe Fachgebiet so anwenderfreundlich wie möglich gestaltet. Mit laufender Anwendung des Tools und dem Vertraut machen mit dem API steigt das Verständnis für die Funktionsweise der «Deep Learning»-Modelle stetig. Was wiederum zu besseren Trainings- und Setupeinstellungen führt und die Modelle kontinuierlich verbessert werden können.

6.3 Ausblick

Da das Forschungsgebiet rund um das Thema «Deep Learning» viele wegweisende Neuerungen bringen kann, gibt es wohl auch viele Forscher, die sich mit der Entwicklung neuer revolutionärer

Technologien und Prozessen auseinandersetzen. Wenn diese Fortschritte auch weiterhin von den ArcGIS Entwicklern in ihr Programm integriert werden, so wird die jetzige Empfehlung, wie sie in dieser Arbeit ausgesprochen wurde, schnell an Relevanz verlieren, da es Modelle geben wird, die sich besser für die Wetland-Extrahierung eignen werden. Bereits im zeitlichen Rahmen dieser Arbeiten kam es zu Ergänzungen von Modellen und Modellparametern. Untersuchungen allfälliger neuer Modelle zur Pixelklassifizierung wären sicher interessant.

Im Bezug auf die nun hier vorliegenden trainierten Modelle wäre es interessant zu untersuchen, wie variabel die Rasterdaten sein können, aus denen die Features extrahiert werden sollen. Es liesse sich sicher erkunden, ob die Modelle auch fähig wären, ältere Kartenblätter oder Karten anderen Ursprungs korrekt zu klassifizieren.

7 Literaturverzeichnis

- Alberico, I., Cavuoto, G., Di Fiore, V., Punzo, M., Tarallo, D., Pelosi, N., . . . Marsella, E. (2017). *Historical maps and satellite images as tools for shoreline variations and territorial changes assessment: the case study of Volturno Coastal Plain (Southern Italy)*. doi:10.1007/s11852-017-0573-x
- Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C., & Paluri, M. (2019). *Improved Road Connectivity by Joint Learning of Orientation and Segmentation*. doi:10.1109/CVPR.2019.01063
- Bundesamt für Landestopografie swisstopo. (2022). *Hintergrundinformationen zu den Landeskarten*. Abgerufen am 3. Juni 2022 von <https://www.swisstopo.admin.ch/de/wissen-fakten/geschichte-sammlungen/historische-kartenwerke/landeskarte.html>
- Bundesamt für Landestopografie swisstopo. (2022). *Historische Kartenwerke*. Abgerufen am 3. Juni 2022 von <https://www.swisstopo.admin.ch/de/wissen-fakten/geschichte-sammlungen/historische-kartenwerke.html>
- Bundesamt für Landestopografie swisstopo. (2022). Zeichenerklärung: Landeskarten 1:10 000 bis 1:1 Million. Wabern. Von Bundesamt für Topografie swisstopo: https://www.swisstopo.admin.ch/content/swisstopo-internet/de/swisstopo/publications.download/swisstopo-internet/de/publications/kartopublications/shop/symbols_de.pdf abgerufen
- Chen, H., & Shi, Z. (2020). *A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection*. doi:10.3390/rs12101662
- Chen, L.-C., Papandreou, G., Schroff, F., & Hartwig, A. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation*. doi:10.48550/ARXIV.1706.05587
- Chiang, Y.-Y., Duan, W., Leyk, S., Uhl, J., & Knoblock, C. (2020). *Using Historical Maps in Scientific Studies*. Cham: Springer. doi:10.1007/978-3-319-66908-3
- Esri. (2022a). *How feature classifier works?* Abgerufen am 6. Juni 2022 von ArcGIS API for Python: <https://developers.arcgis.com/python/guide/how-feature-categorization-works/>
- Esri. (2022b). *Überblick über das Toolset "Deep Learning"*. Abgerufen am 6. Juni 2022 von ArcGIS Pro: <https://pro.arcgis.com/de/pro-app/latest/tool-reference/image-analyst/an-overview-of-the-deep-learning-toolset-in-image-analyst.htm>
- Esri. (2022c). *Deep-Learning-Modell trainieren (Image Analyst)*. Abgerufen am 4. Juni 2022 von ArcGIS Pro: <https://pro.arcgis.com/de/pro-app/latest/tool-reference/image-analyst/train-deep-learning-model.htm>
- Esri. (2022e). *How PSPNet works?* Abgerufen am 5. Juni 2022 von ArcGIS API for Python: <https://developers.arcgis.com/python/guide/how-pspnet-works/>
- Esri. (2022f). *How DeepLabV3 Works*. Abgerufen am 3. Juni 2022 von ArcGIS API for Python: <https://developers.arcgis.com/python/guide/how-deeplabv3-works/>

- Esri. (2022g). *Edge Detection with arcgis.learn*. Abgerufen am 3. Juni 2022 von ArcGIS API for Python: <https://developers.arcgis.com/python/guide/edge-detection-with-arcgis-learn/>
- Esri. (2022h). ArcGIS Pro. *Userinterface des "Train Deep Learning Model"-Werkzeugs*.
- Esri. (2022i). *How ChangeDetection Works?* Abgerufen am 3. Juni 2022 von ArcGIS API for Python: <https://developers.arcgis.com/python/guide/how-change-detection-works/>
- Esri. (2022m). *Introduction to the ArcGIS Pro Image Analyst extension*. Abgerufen am 3. Juni 2022 von ArcGIS Pro: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/what-is-the-arcgis-pro-image-analyst-extension-.htm>
- Garcia-Molsosa, A., Orengo, H. A., Lawrence, D., Philip, G., Hopper, K., & Petrie, C. A. (2021). *Potential of deep learning segmentation for the extraction of archaeological features from historical map series*. doi:10.1002/arp.1807
- He, J., Zhang, S., Yang, M., Shan, Y., & Huang, T. (2019). *Bi-Directional Cascade Network for Perceptual Edge Detection*. Peking University, Computer Vision and Pattern Recognition, Peking. doi:10.1109/TPAMI.2020.3007074
- Hurni, L., Heitzler, M., Jiao, C., & Xia, X. (2021). *Raumbezogene Ingenieurwissenschaften – Bachelor-Arbeiten FS 2022*. Zürich.
- Kaim, D., Kozak, J., Kolecka, N., Ziolkowska, E., Ostafin, K., Ostapowicz, K., . . . Radeloff, V. (2016). *Broad scale forest cover reconstruction from historical topographic maps*. doi:10.1016/j.apgeog.2015.12.003
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). *Focal Loss for Dense Object Detection*. doi:10.48550/ARXIV.1708.02002
- Liu, D., Toman, E., Fuller, Z., Chen, G., Londo, A., Zhang, X., & Zhao, K. (2018). *Integration of historical map and aerial imagery to characterize long-term land-use change and landscape dynamics: An object-based analysis via Random Forest*. doi:<https://doi.org/10.1016/j.ecolind.2018.08.004>
- Minar, M. R., & Naher, J. (2018). *Recent Advances in Deep Learning: An Overview*. doi:10.13140/RG.2.2.24831.10403
- Powers, D. (2020). *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. doi:10.48550/arXiv.2010.16061
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. doi:10.48550/ARXIV.1505.04597
- Schindler, K. (22. April 2021a). *Multivariate Statistik und Machine Learning. Vorlesung zum Thema Clustering*. ETH Zürich.
- Schindler, K. (6. Mai 2021b). *Multivariate Statistik und Machine Learning. Vorlesung zum Thema Deep Learning*. ETH Zürich.
- scikit-learn. (2022). *Metrics and scoring: quantifying the quality of predictions*. Abgerufen am 7. Juni 2022 von scikit-learn: https://scikit-learn.org/stable/modules/model_evaluation.html

- Stuber, M., & Bürgi, M. (2018). *Vom «eroberten Land» zum Renaturierungsprojekt. Geschichte der Feuchtgebiete in der Schweiz seit 1700.* (Bristol-Stiftung, Hrsg.) Bern: Haupt.
- Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T. F., & Achard, F. (2018). *Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar.* doi:10.1016/j.biocon.2018.04.008
- Xie, S., & Tu, Z. (2015). *Holistically-Nested Edge Detection.* doi:10.48550/ARXIV.1504.06375
- Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). *mixup: Beyond Empirical Risk Minimization.* doi:10.48550/ARXIV.1710.09412
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016). *Pyramid Scene Parsing Network.* doi:10.48550/ARXIV.1612.01105

Anhang

Eigenständigkeitserklärung



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eigenständigkeitserklärung

Die unterzeichnete Eigenständigkeitserklärung ist Bestandteil jeder während des Studiums verfassten Semester-, Bachelor- und Master-Arbeit oder anderen Abschlussarbeit (auch der jeweils elektronischen Version).

Die Dozentinnen und Dozenten können auch für andere bei ihnen verfasste schriftliche Arbeiten eine Eigenständigkeitserklärung verlangen.

Ich bestätige, die vorliegende Arbeit selbständig und in eigenen Worten verfasst zu haben. Davon ausgenommen sind sprachliche und inhaltliche Korrekturvorschläge durch die Betreuer und Betreuerinnen der Arbeit.

Titel der Arbeit (in Druckschrift):

Untersuchung von «Deep Learning»-Modellen in ArcGIS Pro zur Extraktion von Merkmalen aus historischen Karten: Objekt- und Pixelklassifizierung

Verfasst von (in Druckschrift):

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich.

Name(n):

Pironato

Vorname(n):

Aline

Ich bestätige mit meiner Unterschrift:

- Ich habe keine im Merkblatt „Zitier-Knigge“ beschriebene Form des Plagiats begangen.
- Ich habe alle Methoden, Daten und Arbeitsabläufe wahrheitsgetreu dokumentiert.
- Ich habe keine Daten manipuliert.
- Ich habe alle Personen erwähnt, welche die Arbeit wesentlich unterstützt haben.

Ich nehme zur Kenntnis, dass die Arbeit mit elektronischen Hilfsmitteln auf Plagiate überprüft werden kann.

Ort, Datum

10. Juni 2022, Dübendorf

Unterschrift(en)

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich. Durch die Unterschriften bürgen sie gemeinsam für den gesamten Inhalt dieser schriftlichen Arbeit.