

Bathymetrieableitung aus multispektralen Luftbildern über Convolutional Neural Networks

HANNES NÜBEL¹, GOTTFRIED MANDLBURGER¹, MICHAEL KÖLLE¹ & UWE SÖRGEL¹

Zusammenfassung: Zur Ableitung von Gewässertiefen werden aufgrund der einfachen Datenerfassung in den letzten Jahren vermehrt optische Ansätze basierend auf der Bildradiometrie eingesetzt. In flachen Bereichen kann die Tiefe dabei durch die Modellierung der Signaldämpfung in verschiedenen Bändern ermittelt werden. Im Rahmen des im Beitrag vorgestellten Ansatzes soll untersucht werden, inwieweit ein Convolutional Neural Network in der Lage ist, Wassertiefen aus multispektralen Luftbildern zu schätzen. Um auf die in den Bildpixel abgebildeten schrägen Wasserdistanzen zu trainieren, werden als Basis nicht das Orthophoto, sondern die originalen weitwinkeligen Luftbilder herangezogen. Der verwendete Datensatz enthält neben RGB-Bildern auch synchron erfasste panchromatische Bilder mit einem Coastal Blue Filter, dessen Nutzen in der CNN-basierten Regression analysiert wird. Die Auswertung zeigt, dass Trends für unterschiedliche Tiefen und Bodenbedeckungen erkannt werden können und eine Standardabweichung von 0.3-0.4 m erreicht werden kann. Limitiert wird der Ansatz durch direkte Reflexion von Sonnenlicht (sun glint) und durch Schattenbereiche. Die Einbeziehung des Coastal Blue Bandes bringt einen Mehrwert in Bezug auf die Verteilung der Tiefen des Testgebietes.

1 Einführung

1.1 Motivation

Die Vermessung der Erdoberfläche mittels Photogrammetrie ist ein etabliertes Verfahren. Koordinaten von Objektpunkten können bei bekannter Orientierung der Kameras mittels Vorwärtschnitt berechnet werden, wenn der jeweilige Punkt in zwei oder mehr Bildern beobachtet wird. Die Anwendung dieses Verfahrens in Wasserbereichen ist jedoch komplexer (MAAS 2015). Dennoch ist es insbesondere in Flachwassergebieten notwendig, Wassertiefen zu kartieren, zum Beispiel für sichere Schiffsnavigation oder bei der Bestimmung des Volumens eines Sees, welcher für Brandbekämpfung benötigt wird.

Die Komplexität bei Mehrmedienphotogrammetrie besteht darin, dass die Messung identischer Punkte aufgrund der spiegelnden und dynamischen Beschaffenheit der Wasseroberfläche erschwert wird. Darüber hinaus kommt es beim Übergang des Bildstrahls zwischen zwei Medien zur Brechung an der Wasseroberfläche. Für die Erzeugung eines Orthophotos bedeutet dies, dass jedes Pixel in jedem Bild einen individuellen gebrochenen Strahl aufweist, der der Wasseroberfläche entsprechend auch lokale Dynamiken aufweist.

Um die Bodenpunkte jedes Pixels zu detektieren, müssen somit die Strahlen mit der durch die Orientierung des Bildes gegebenen Richtung, unter Berücksichtigung der Brechung an der Wasseroberfläche, verfolgt werden. Des Weiteren gilt, dass selbst wenn die Richtung jedes Strahls

¹ Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl Str. 24D, D-70174 Stuttgart, E-Mail: st148625@stud.uni-stuttgart.de, [gottfried.mandlbürger, michael.koelle, uwe.soergel]@ifp.uni-stuttgart.de

bekannt ist, genügend identische Punkte gefunden werden müssen, um deren Koordinaten mit Hilfe der sich schneidenden Strahlen aus den Bildern zu berechnen. Dies ist ebenso anspruchsvoll, da Gewässerböden typischerweise homogen sind und darüber hinaus Signale im Wasser gedämpft werden. Dieser Effekt wird zudem durch Reflexion und andere Faktoren verstärkt, da dieselben Punkte unterschiedlich erscheinen, wenn sie aus verschiedenen Perspektiven betrachtet werden. Aufgrund der unterschiedlichen Stärke der Lichtabsorption in der Wassersäule für verschiedene Spektralbänder ist es auch möglich, ein lineares oder höher dimensionales Regressionsmodell aufzustellen, welches das Verhältnis von Radiometrie zu Tiefe approximiert. Sobald jedoch der Bereich verschiedene Arten von Vegetation auf dem Grund des Gewässers enthält, ist ein komplexeres Regressionsmodell erforderlich. Darüber hinaus wird die spektrale Tiefenschätzung üblicherweise auf der Grundlage von Orthophotos oder sehr schmalwinkeligen Satellitenbildern durchgeführt. Die Vernachlässigung der Strahlbrechung an der Wasseroberfläche führt dabei zu geometrischen Fehlern. Zusätzlich beschreiben nur Pixelwerte aus dem Bildmittelpunkt (Nadirrichtung) direkt die Wassertiefe, während Pixel vom Bildrand die schräge Wasserstrecke anzeigen. Jedes Pixel eines Luftbildes wiederum speichert radiometrische Informationen, die sich auf die potenziell schräge Unterwasserdistanz des jeweiligen Bildstrahls beziehen. Insbesondere bei Weitwinkelobjektiven ist es daher sinnvoll, die bathymetrische Schätzung nicht auf dem Orthophoto, sondern auf den (orientierten) Bildern durchzuführen.

Um den linearen Regressionsansatz zu erweitern, kann ein Convolutional Neural Network (CNN) verwendet werden, um Variationen des Bodenreflexionsgrades miteinzubeziehen. Die pixelweise Tiefenschätzung auf Basis der orientierten Luftbilder erfordert die schrägen Wasserdistanzen für die Bildpunkte zum Training. Diese Informationen können zum Beispiel aus bathymetrischen LiDAR (Light Detection And Ranging) Daten abgeleitet werden, insbesondere wenn diese gleichzeitig mit den Bildern erfasst wurden.

Des Weiteren bringen CNNs den Vorteil, dass räumliche Kontextinformationen berücksichtigt werden. Die Zuverlässigkeit des Netzes wird dadurch erhöht, da Nähe oft ähnliche Tiefen impliziert.

In dieser Arbeit wird der Ansatz des Trainings eines CNN zur Prädiktion der schrägen Distanzen von Bildstrahlen in einem Gewässer untersucht. Neben der Qualitätsbewertung wird auch diskutiert, inwieweit der Coastal Blue Kanal Einfluss auf das Netz hat.

Im Abschnitt 2 wird der zugrunde liegende Datensatz und seine Aufbereitung vorgestellt. Die Methoden, einschließlich der Vorverarbeitung der Referenzdaten und des Trainings des Netzes, werden dann im Abschnitt 3 beschrieben. Der darauffolgende Abschnitt 4 präsentiert und diskutiert die Ergebnisse des trainierten Netzes. Abschnitt 5 fasst schließlich die Arbeit zusammen und befasst sich zudem mit den möglichen Themen für weitergehende Forschungen.

1.2 Verwandte Arbeiten

Ein etabliertes Verfahren zur Unterwasserkartierung ist SoNAR (Sound Navigation and Ranging) (MASNADI-SHIRAZI et al. 1992). Da die Instrumente unter einem Schiff im Wasser montiert sind, gibt es im Vergleich zu Sensoren oberhalb der Wasseroberfläche keinen Übergang zwischen unterschiedlichen Medien. Die Dynamik der Wasseroberfläche beeinflusst bei der SoNAR Technik daher lediglich die Sensororientierung. Gegenüber optischen Verfahren stellt SoNAR geringere

Ansprüche an die Klarheit des Wassers und ist für die Vermessung größerer Messtiefen jenseits der Sichttiefen geeignet.

Eine weitere zunehmend angewandte Technik für die Ableitung von Wassertiefen im Flachwasserbereich ist Flugzug-getragene Laserbathymetrie (IRISH & WHITE 1998). Anstelle von Schallwellen wie bei SoNAR werden Laserpulse ausgesendet, deren Reflexion (Echos) gemessen werden. Durch den Übergang des Laserpulses von der Atmosphäre in das Wasser muss die Änderung der Ausbreitungsgeschwindigkeit und der Strahlrichtung berücksichtigt werden (Refraktionskorrektur). Ein Vergleich mit dem SoNAR-Ansatz wurde u.a. von COSTA et al. (2009) durchgeführt. Der Hauptvorteil ist die erhöhte Effizienz vor allem für große Flachwassergebiete. Darüber hinaus ist es auch möglich, Daten in abgelegenen Gebieten sowie in Randgebieten eines Gewässers zu erfassen, die mit SoNAR nicht zugänglich sind.

Für topografische Anwendungen ist die Stereophotogrammetrie eine gängige Strategie. Identische Punkte werden in mehreren Bildern erfasst und ihre Position durch Bündelausgleichung berechnet. Die Herausforderungen, die sich für diesen Prozess in bathymetrischen Anwendungen ergeben, werden von MULSOW et al. (2019) im Vergleich zur Datenerfassung mit einem luftgestützten Laserscanner diskutiert. Die Kosten der Datenerfassung können bei annähernd gleicher Genauigkeit reduziert werden. Grundlage ist jedoch das Auffinden identischer Punkte, welche durch Textur bestimmt werden. Insbesondere in tiefen oder homogenen Bereichen ist dies nicht gegeben und es treten daher größere Abweichungen auf.

Beim Ansatz der spektralen Bathymetrie dagegen, wird ein mathematisches oder physikalisches Modell erstellt, das eine Verbindung zwischen Reflexionsgrad und Wassertiefe herstellt. Somit ist es also möglich, die Wassertiefe für jedes Pixel zu schätzen. Eine häufige Annahme ist, dass sich die Reflexion vom Grund negativ exponentiell verhält, bezogen auf die Wassertiefe (LYZENGA et al. 2006). Darüber hinaus muss ein Offset einbezogen werden, der den Reflexionsgrad für eine unendliche Wassertiefe modelliert.

Spektrale Tiefenbestimmung kann, abhängig von der Lichtdämpfung in der Wassersäule, für seichte Gewässer angewendet werden. Da die Reflexion für verschiedene Bodendeckungen stark variiert, kann es sinnvoll sein, mehrere Spektralbänder (LEGLEITER et al. 2009) einzubeziehen, da die Rückstrahlung für bestimmte Bodentypen in angrenzenden Bändern ähnlich ist. Insbesondere logarithmische Verhältnisse zwischen den Bändern eignen sich für die Tiefenschätzung. Auch eine Kombination aus dem roten Kanal, der am stärksten von der Signalabsorption beeinflusst wird, und einem besser durchdringenden Spektralband weist einen hohen Informationsgehalt auf (MANDLBURGER et al. 2018).

Von WANG et al. (2019) wurde ein bathymetrisches Modell entwickelt, welches sowohl spektrale als auch räumliche Informationen mit Hilfe eines neuronalen Netzwerks einbezieht. Bei diesem Ansatz wird die räumliche Information durch die Koordinaten X und Y gegeben. Zusammen mit vier multispektralen Bandverhältnissen bilden sie einen Merkmalsvektor, welcher den ersten Layer des Netzes repräsentiert. Das neuronale Netz besteht nach dem ersten Layer aus drei versteckten (hidden) Layern und einem Output Layer, die zusammen ein Multi Layer Perceptron (MLP) bilden.

Der hier angewandte Ansatz unterscheidet sich vor allem in zwei Punkten. Erstens werden die räumlichen Informationen mit Hilfe eines CNN aufgenommen, das Gewichte für kleine Faltungskerne lernt, die einen bestimmten Bereich um ein Pixel herum abdecken und somit nicht abhängig

von der tatsächlichen X- und Y-Koordinate eines Punktes sind. Aufgrund dieser allgemeineren Annahme kann das Netz auch auf andere Bereiche übertragen werden, da eine semantische Segmentierung eines gesamten Bildes durchgeführt wird, anstatt isolierte Tiefen von Einzelpunkten anhand bestimmter Merkmale abzuleiten. Zweitens ist es dem Netz durch die größere Anzahl an Layern möglich, komplexere Features zu extrahieren.

2 Datensatz

Die Erprobung des Ansatzes erfolgte auf Basis einer Befliegung des Autobahnsees (Augsburg) im April 2018. Dabei wurden zeitgleich Luftbilder und Laserbathymetriedaten erfasst. Der Autobahnsee ist ein 5 m tiefer Baggersee mit einer Ausdehnung von ca. 750 m × 500 m und beinhaltet eine kleinere Insel. Der Untergrund des Sees besteht aus offenen Schotterflächen sowie unterschiedlichen Vegetationsfeldern (Abb. 1).

2.1 Bilddaten

Für die Bilddatenerfassung kamen zwei IGI DigiCAM 100 Kameras zum Einsatz, die auf Phase One iXU-RS 1000 Kameras mit je 11608 × 8708 Pixel basieren. Während der Sensor der RGB Kamera mit einem Bayer Pattern ausgestattet ist, wurde bei der Coastal Blue Kamera ein panchromatischen Sensor und vorgesetztem Filter für die Coastal Blue Wellenlänge verwendet (MANDLBURGER et al. 2018).

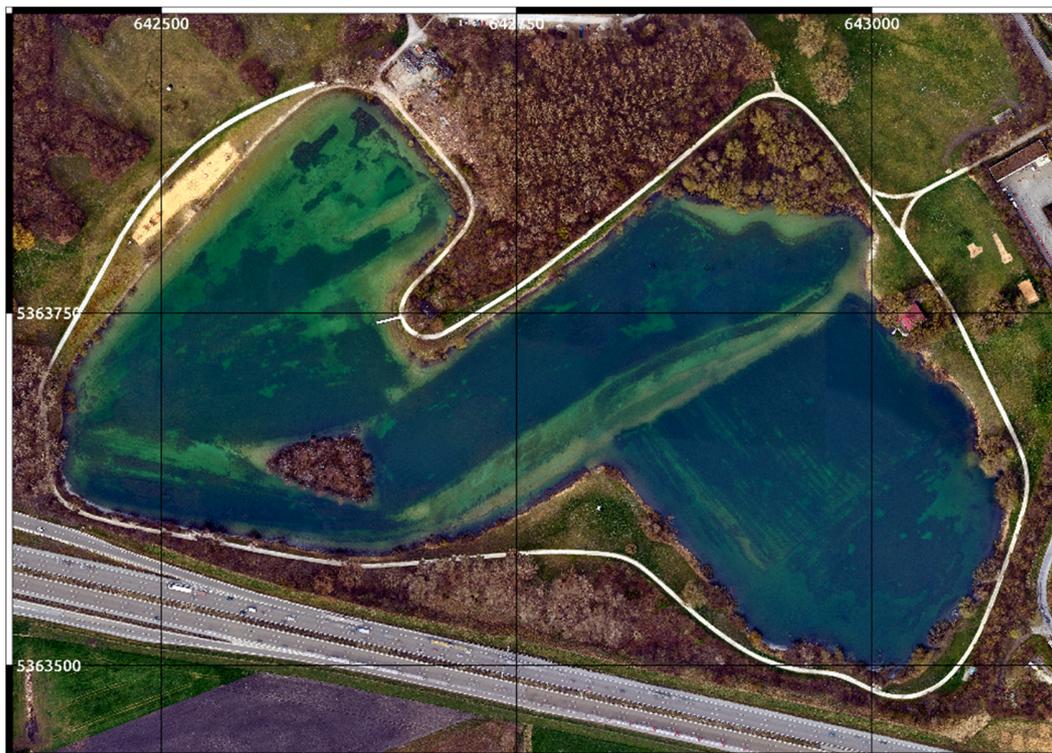


Abb. 1: Orthophoto Autobahnsee mit verschiedenen Grundbedeckungen

Um die Informationen aus beiden Bildern nutzen zu können, sollten diese von derselben Position mit derselben Orientierung aufgenommen werden. Aus praktischen Gründen mussten die Kameras jedoch nebeneinander montiert werden. Die Transformation des zeitgleich aufgenommenen Coastal Blue Bildes in den Bezugsrahmen des korrespondierenden RGB-Bildes erfolgte mittels Homographie mit der Software MATLAB (2018). Da die Höhenvariation des Geländes im Bezug zur Flughöhe vergleichsweise gering ist, sind die Restfehler vernachlässigbar. Durch die Transformation entstehen im neuen Coastal Blue-Bild leere Pixel im Überlappungsbereich mit dem RGB-Bild. Um sicherzustellen, dass ein Bild in jedem Pixel für alle Bänder Werte besitzt, wird für die zusammengeführten Bilder im RGB-Bilddatum ein kleiner Rahmen an den Grenzen jedes Bildes definiert. Alle Pixel in diesem Randbereich werden später ausmaskiert.

2.2 LIDAR

Abgesehen von den beiden Kameras integriert das eingesetzte hybride Sensorsystem auch einen RIEGL VQ-880-G topo-bathymetrischen Laserscanner (RIEGL 2019). Aus der damit synchron aufgenommenen 3D Punktwolke, konnte das Wasserflächenmodell und das Geländemodell extrahiert werden. In Abb. 2 ist das Geländemodell für das beobachtete Gebiet dargestellt. Es wird deutlich, dass es am Boden des Sees komplexe Strukturen gibt, die einerseits auf topographische Details und andererseits auf nicht durchdringbare Unterwasser-Vegetation zurückzuführen sind. Oberflächen- und Geländemodell bilden die Basis für die Extraktion der Referenzdaten, welche in diesem Fall die schrägen Distanzen der Bildstrahlen im Wasser sind. Der Scanner operiert mit grüner Laserstrahlung (Wellenlänge: 532 nm) und ist für die Flachwasserkartierung konzipiert. An Land beträgt die mittlere Punktdichte beträgt etwa 40 Punkte/m². In den tieferen Wasserbereichen ist die Punktdichte aufgrund der Dämpfung des Lasers geringer, allerdings immer noch ausreichend, um ein DGM mit 0.5 m Rasterweite zu berechnen.

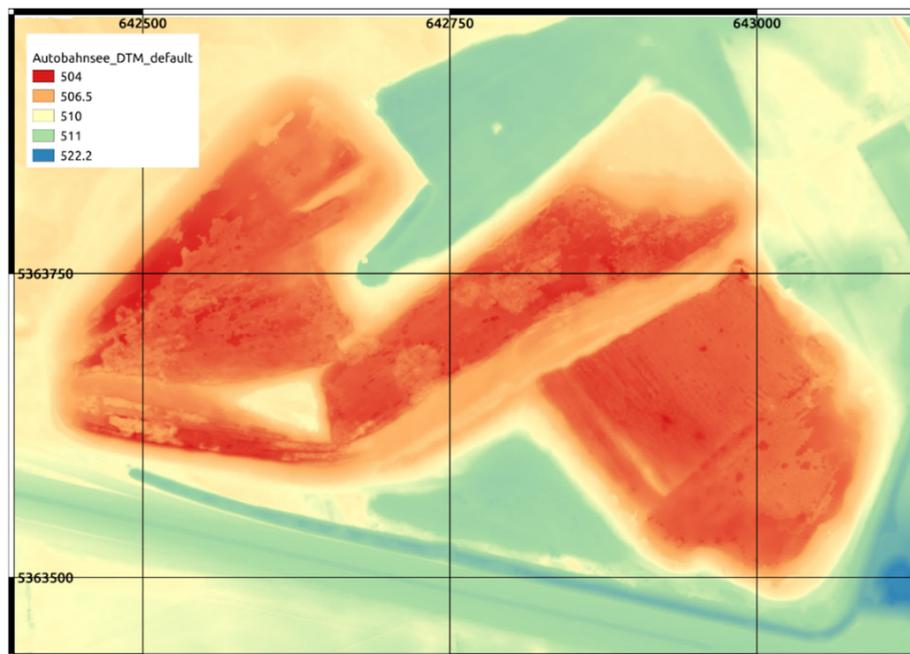


Abb. 2: Digitales Geländemodell des Autobahnsees aus topo-bathymetrischem LiDAR.

3 Methodik

3.1 Vorverarbeitung von Referenzdaten

Im folgenden Abschnitt wird die angewandte Methodik zur Ableitung der Referenzdaten diskutiert, mit denen das angewandte CNN trainiert wurde. Als Referenzdaten fungieren dabei die jeweiligen Schrägdistanzen der Bildstrahlen jedes Pixels im Wasser. Um den Strahlengang von der Kamera zum entsprechenden Bodenpunkt unter Berücksichtigung der Brechung an der Wasseroberfläche zu verfolgen, wird die Orientierungen der Kamera, ein digitales Wasseroberflächenmodell (DWM), sowie ein Geländemodell (DGM) benötigt. Das DWM wurde aus den ersten Echos des Laserscanners geschätzt, während die letzten Echos die Grundlage für die Filterung der Bodenpunkte und schließlich der DGM-Berechnung sind. Abb. 3 veranschaulicht die Zusammenhänge in Form einer Prinzipskizze.

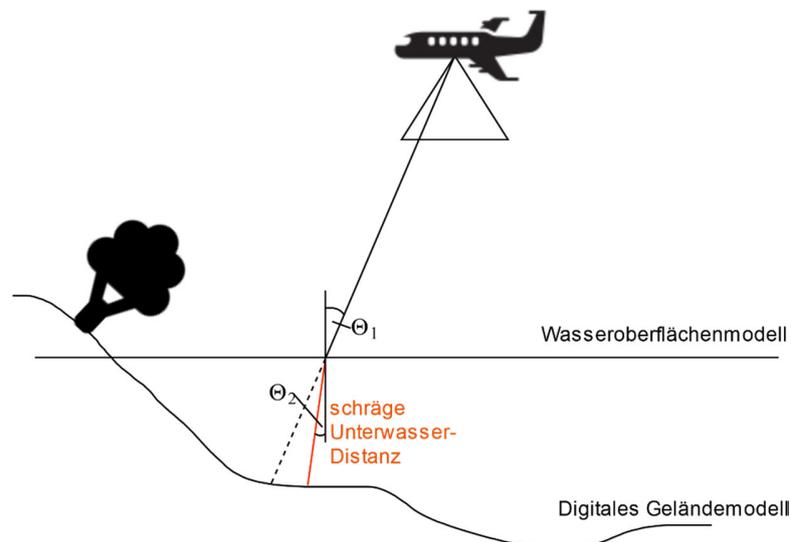


Abb. 3: Refraktion des Bildstrahls an der Wasseroberfläche.

Um die schrägen Distanzen zu erhalten, können die den einzelnen Pixel entsprechenden Bildstrahlen im lokalen Kamerakoordinatensystem aus der inneren Orientierung der Kamera berechnet werden. Über die Elemente der äußeren Orientierung wird der Bildstrahl anschließend in ein globales Koordinatensystem transformiert (KRAUS 1996). Jeder Bildstrahl wird anschließend mit dem DWM verschnitten, der Strahlbrechung gemäß dem Brechungsgesetz von Snellius unterworfen und der gebrochene Bildstrahl anschließend mit dem DGM verschnitten. Die Distanz zwischen den Strahlschnittpunkten mit dem DWM und dem DGM entspricht der im jeweiligen Pixel abgebildeten schrägen Wasserdistanz. Für die Implementierung des beschriebenen Workflows kam die Laserscanning Software OPALS (PFEIFER et al. 2014) zum Einsatz. Die Berechnung der Bildstrahlen im Objektkoordinatensystem erfolgte in Python basierend auf den Pixelkoordinaten und den Kameraorientierungen. Weitere Details zur Methodik sind in NÜBEL (2019) beschrieben.

3.2 Deep Learning

3.2.1 Convolutional Neural Networks (CNNs)

CNNs sind eine spezifischere Art von neuronalen Netzwerken, mit denen insbesondere Bilder präzise und effizient verarbeitet werden können, indem sie von ihrer räumlichen Struktur profitieren (NIELSEN 2015). Dies wird realisiert, indem man in den einzelnen Layern Faltungen durchführt. Abhängig von der Größe des Kernels liefern kleine Bereiche von z. B. 3×3 Pixel aus einem Layer Informationen für ein korrespondierendes Neuron in einem tieferen Layer. Effektive Feature extrahierende Faltungskerne können nicht nur an einer Position, sondern überall im Bild nützlich sein. Zur Reduktion wird deshalb nur ein Kernel für das gesamte Layer trainiert, anstatt verschiedene Kernel zu trainieren, die sich auf bestimmte Regionen im Bild beziehen. Anstatt also für jedes Neuron in jedem Layer Gewichte zu trainieren, müssen nur die Gewichte für ein Kernel trainiert werden. Dadurch lassen sich tiefere Architekturen mit mehr Layern realisieren, um High-Level-Features einzubeziehen, aber dennoch effizient in Bezug auf Trainingsdauer und den erforderlichen Trainingsdaten zu sein.

3.2.2 U-Net

Das U-Net (RONNEBERGER et al. 2015), welches als Grundlage für das angewandte Netz dient, ist ein CNN, welches für die biomedizinische Bildsegmentierung konzipiert wurde. Es handelt sich um ein Fully Convolutional Network, was bedeutet, dass die Größe des Ausgabebildes gleich der Eingabegröße ist. Dies ist im vorliegenden Kontext insofern erforderlich, da für jedes Pixel des Wasserbereiches ein Tiefenwert prädiziert werden soll.

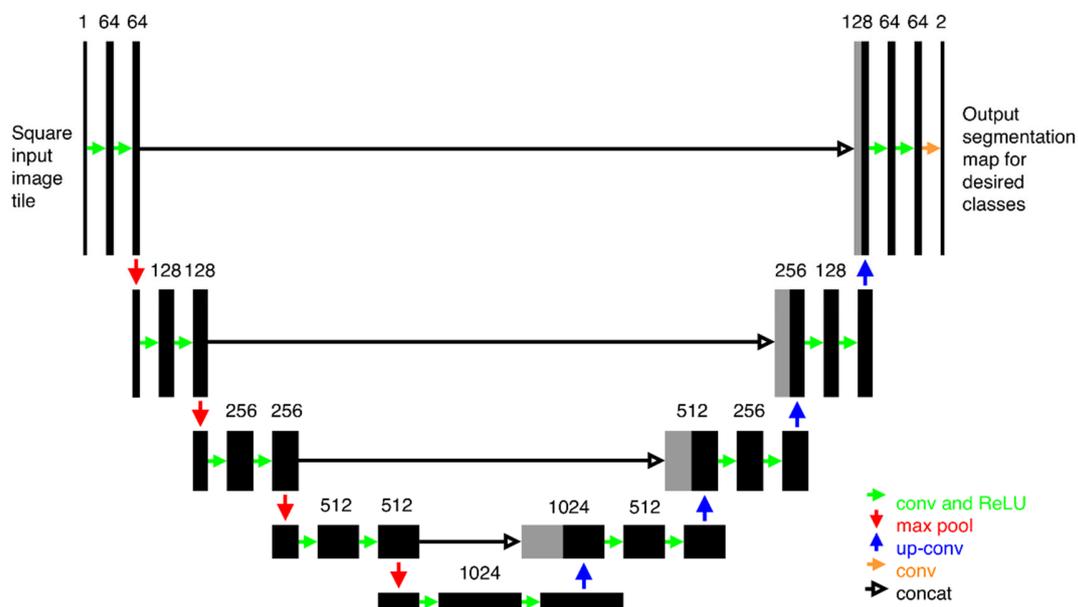


Abb. 4: U-Net-Architektur. Jede schwarze Box entspricht einer Multi-Kanal Feature Map. Die Anzahl der Kanäle ist auf der Oberseite der Box angegeben. Die grauen Boxen stellen kopierte Feature Maps dar. Die Pfeile kennzeichnen die verschiedenen Operationen (RONNEBERGER et al. 2015)

Der Name ergibt sich aus der Architektur des Netzes, da es einen komprimierenden und einen symmetrischen expandierenden Pfad aufweist (siehe Abb. 4). Es besteht aus Convolutional Layers, gefolgt von einer Rectified Linear Unit (ReLU) und Max-Pooling-Layern in den Kompressions- sowie Up-Convolution-Layern in den Expansionspfaden. Die ReLU-Aktivierungsfunktion ist definiert durch $f(x) = \max(0, x)$ und wird häufig in modernen neuronalen Netzen verwendet. Im Komprimierungspfad wird bei jedem Pooling-Schritt die Dimension der jeweiligen Bildkachel halbiert. Nach jedem Schritt wird die Anzahl der mit den Convolutional Layers berechneten Feature-Maps verdoppelt. Die Verkleinerung der Bildkacheln erfolgt durch Max-Pooling, d.h. für die jeweiligen Werte des verschiebenden Kernels wird der Maximalwert ausgegeben (GOODFELLOW et al. 2016). Um die Eingabedimension wiederherzustellen, muss für die Kacheln anschließend ein Upsampling durchgeführt werden, was durch eine inverse Faltung geschieht (DUMOULIN & VISIN 2016). Anstatt ein interpolierendes Kernel für das Upsampling zu definieren, wird dieses ebenfalls erlernt und optimiert, während das Netz trainiert wird.

Die Ergebnisse im expandierenden Pfad des U-Net werden mit den gegenüberliegenden Layern aus dem Kompressionspfad verknüpft, um die räumlichen Informationen zu erhalten. Schließlich erfolgt im letzten Layer eine 1×1 Faltung, die jeden Feature-Vektor auf die benötigte Anzahl von Klassen abbildet.

3.2.3 Anpassung der U-Net-Architektur

Die Implementierung des CNN erfolgt in Python unter Verwendung der Deep Learning Bibliothek Keras (CHOLLET 2018) mit Tensorflow Backend (ABADI et al. 2015) und das Netz wird auf einer Graphics Processing Unit (GPU) trainiert, da dies effizienter im Vergleich zu einer Central Processing Unit (CPU) ist.

Wie im Abschnitt 3.2.2 erwähnt, dient das U-Net (RONNEBERGER et al. 2015) als Grundlage für das CNN. Der Hauptunterschied besteht darin, dass das Netz nicht wie bei den meisten Ansätzen zur Segmentierung verwendet wird, sondern zur Anpassung eines Regressionsmodells. Anstatt also auf mehrere Klassen mit einer Ausgabe der jeweiligen Wahrscheinlichkeit für jedes Pixel zu trainieren, wird nur auf die schräge Wasserdistanz als Gleitkommazahl trainiert. Daher ist die Ausgabe des letzten Faltungslayers nur ein Pixel tief und anstelle einer Aktivierungsfunktion wie Sigmoid, mit der die Klassenwerte bestimmt werden können, wird erneut ReLU verwendet. Als Verlustfunktion wird der Root Mean Square Error (RMSE) zwischen der Referenz und den vorhergesagten Werten berechnet.

Um die Größe des Eingabebildes bei der Durchführung einer Faltung beizubehalten, wird Padding verwendet. Folglich hat das Ausgabebild auf der letzten Ebene die gleiche Größe wie die Eingabedaten. Wie in der ursprünglichen Architektur des U-Net werden auch hier Breite und Höhe in jedem Pooling-Schritt durch 2 dividiert, um die doppelte Anzahl an Filtern zu ermöglichen. Um Overfitting zu vermeiden, wird nach den Pooling-Schritten ein zusätzliches Dropout-Layer hinzugefügt, das während des Trainings zufällig einen Prozentsatz der Neuronen ignoriert.

Um einerseits die notwendige Speichermenge bezüglich des Arbeitsspeichers der GPU möglichst gering zu halten, andererseits aber auch nach der letzten Pooling-Schicht eine natürliche Anzahl von Pixeln in Breite und Höhe zu haben, werden Patches von 480×480 Pixel genutzt. Diese Wahl wurde getroffen, da das Netz bei kleineren Patches Probleme haben könnte, die Eigenschaften in

Uferbereichen zu erlernen, während größere Patches dazu führen könnten, dass es stärker die Struktur des Sees erlernt, anstatt das Verhalten der Spektralbänder in Bezug auf die Tiefe. Die zufällig in den Bildern gewählten Patches werden angenommen, wenn sie einen bestimmten Prozentsatz an gültigen Pixeln (Wasserpixel) aufweisen. Dadurch wird sichergestellt, dass jeder Ausschnitt genügend Daten enthält und auch Uferbereiche berücksichtigt werden, sodass nicht nur das Zentrum des Sees trainiert wird.

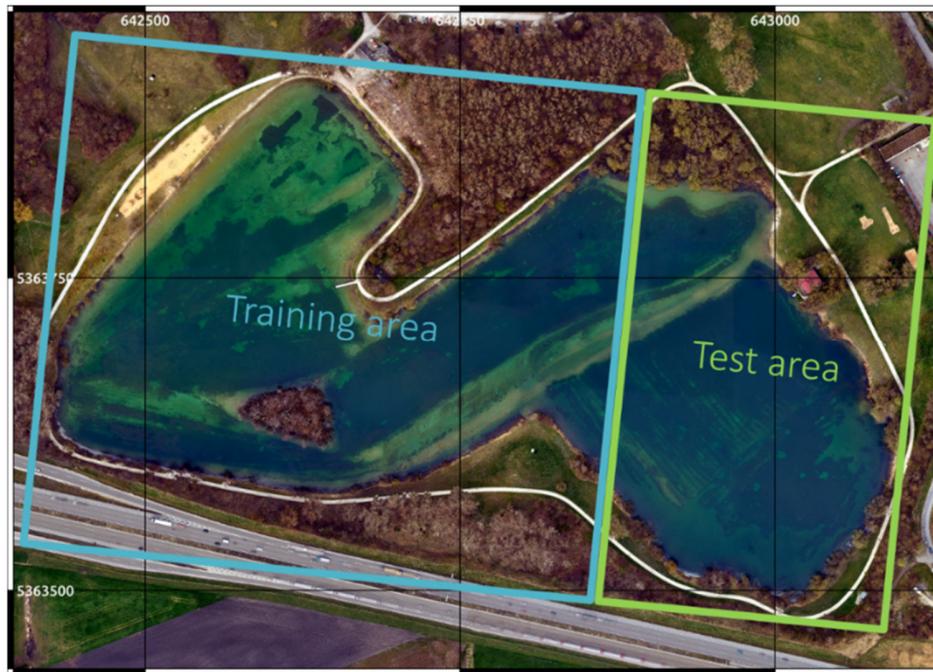


Abb. 5: Verteilung des in den Bildern beobachteten Trainings- und Testbereichs.

Zuletzt müssen die Bilder in Trainings- und Testbilder unterteilt werden. Die Trainingsbilder enthalten dabei auch einen Prozentsatz an Validierungsdaten, wogegen die Testbilder ausschließlich aus Bereichen stammen, die nicht zum Trainieren des Netzes verwendet wurden. Dazu wird die Seefläche in zwei Teile geteilt, die in Abb. 5 markiert sind. Um sicherzustellen, dass keine Teilbereiche bereits im Training enthalten waren, wurden die Bilder im Überlappungsbereich weder im Training noch zum Testen verwendet. Insgesamt wurden 41 Bilder in der Trainingsphase und 24 in der Testphase eingesetzt. Es ist jedoch zu beachten, dass die Datenmenge in jedem Bild von der Anzahl der Pixel abhängig ist, die das Gewässer abbilden. Die Strukturen in den ausgewählten Bereichen unterscheiden sich sehr stark, so dass beurteilt werden kann, ob das Netzwerk für den Trainingsbereich zum Overfitting neigt, oder ob es Eigenschaften lernt, die auch auf andere Gewässer übertragbar sind.

4 Ergebnisse und Diskussion

4.1 Angewandtes CNN für kombiniertes RGB- und Coastal Blue Band

Ein Indikator für die Qualität beim Training des Netzes ist der Loss-Plot, welcher sich hier auf den RMSE von Prädiktionen mit dem aktuellen Netz für jede Trainingsepoche bezieht (Abb. 6). Für

die Trainingsphase werden die verwendeten Bilder wiederum in zwei Teile gegliedert. Die eigentlichen Trainingsbilder werden dann zum Trainieren und Anpassen der Gewichte verwendet, während ein kleinerer Satz von Bildern zur Validierung dieser Anpassungen dient. Unter Berücksichtigung der Tatsache, dass der Loss-Plot für die Trainingsdaten ständig abnehmen sollte, gefolgt von den Validierungsdaten mit einem kleinen Offset, zeigt der Plot das erwartete Verhalten.

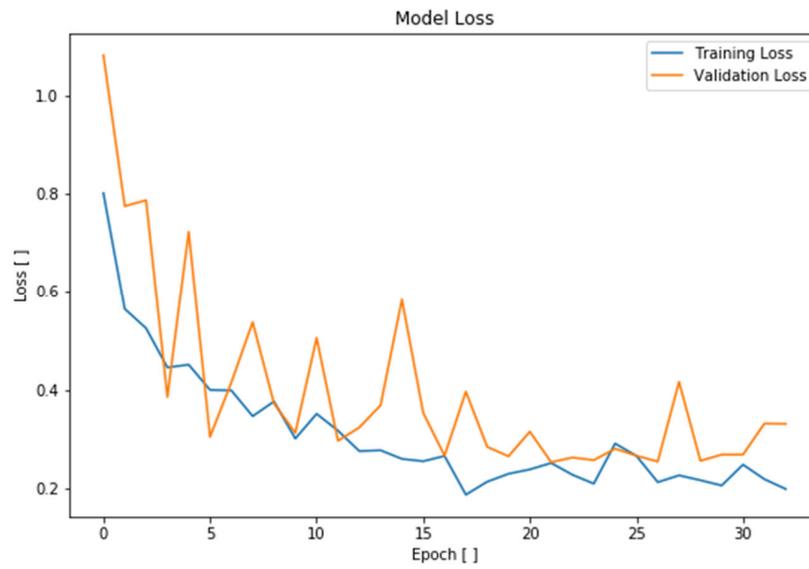


Abb. 6: Loss-Plot mit RMSE während der Epochen des Trainings.

Die Anwendung des trainierten Netzes auf bisher unbekannte Daten ermöglicht einen unabhängigen Leistungstest des Netzes. Diese Daten bestehen aus einer Teilmenge aller Bilder, die als Testbilder markiert sind. So kann überprüft werden, wie gut das Netz wirklich bestimmte Eigenschaften gelernt hat, anstatt sich nur die Trainingsdaten zu merken. Ein Beispiel für die Vorhersage eines Testbildes im Vergleich zu den Referenzdaten ist in Abb. 7 zu sehen. Abgesehen von dem Bereich oben rechts im Bild, in dem die schrägen Unterwasserdistanzen als zu groß prädiziert werden, stimmen die vorhergesagten Werte mit der Referenz generell gut überein. Außerdem gibt es keine größeren Unterschiede in Bezug auf den Trend der Wasserdistanzen. Zu beobachten ist jedoch ein Rauschen in den Prädiktionen, das durch den Kamerasensor oder durch die Dynamik der Wasseroberfläche verursacht sein kann.

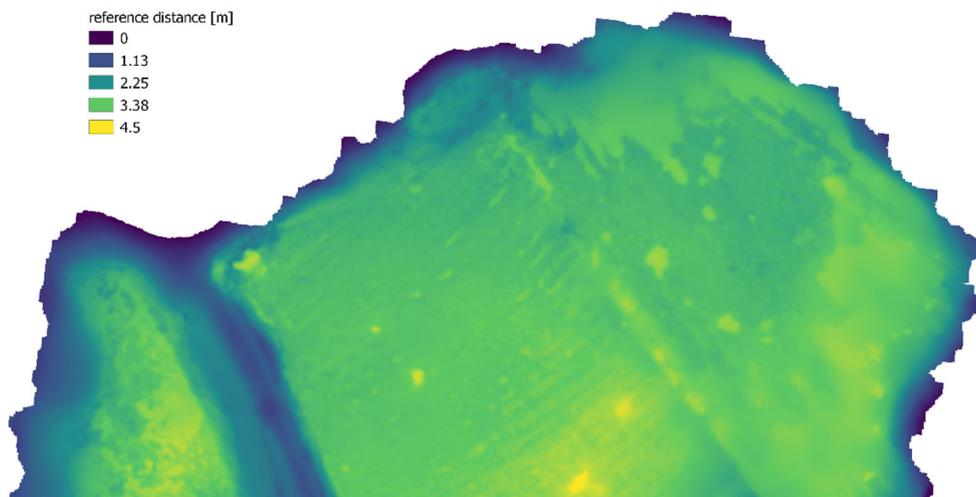


Abb. 7.1: Referenztiefenkarte eines Testbildes.

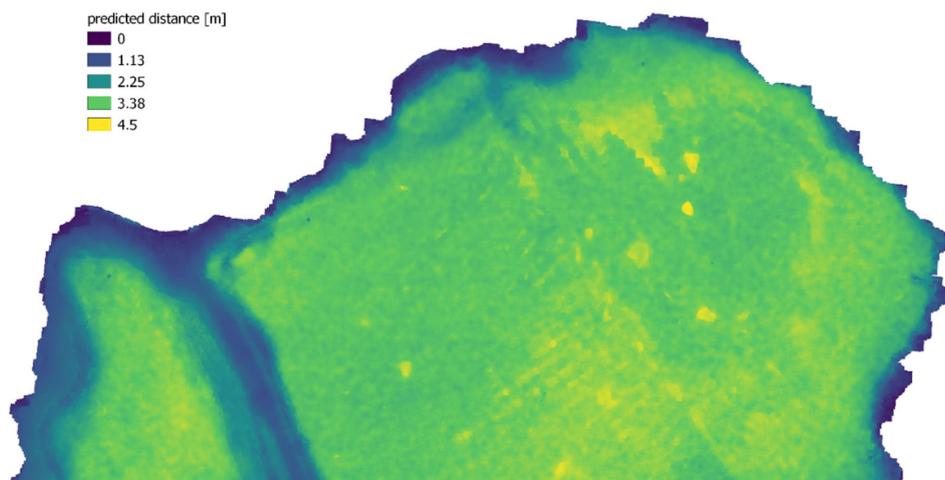


Abb. 7.2: Prädizierte Tiefenkarte eines Testbildes

Nachdem die Distanzen für alle Testbilder prädiziert wurden, kann die Genauigkeit evaluiert werden, indem die prädizierten Distanzen von der Referenz subtrahiert werden. Durch Zusammenführen der Abweichungen für alle Pixel aller Testbilder kann ein Histogramm über alle Tiefenabweichungen (Abb. 8) erhalten werden. Es wird darauf hingewiesen, dass nur Wasserpixel berücksichtigt werden, während alle Pixel in der Vegetation und auf dem Festland maskiert sind. Das Histogramm zeigt eine Verschiebung des Mittelwertes von 4,3 cm in negativer Richtung und weist eine leichte Schiefe in positiver Richtung auf. Es ist nahezu normalverteilt mit einer mittleren absoluten Abweichung von 32,2 cm.

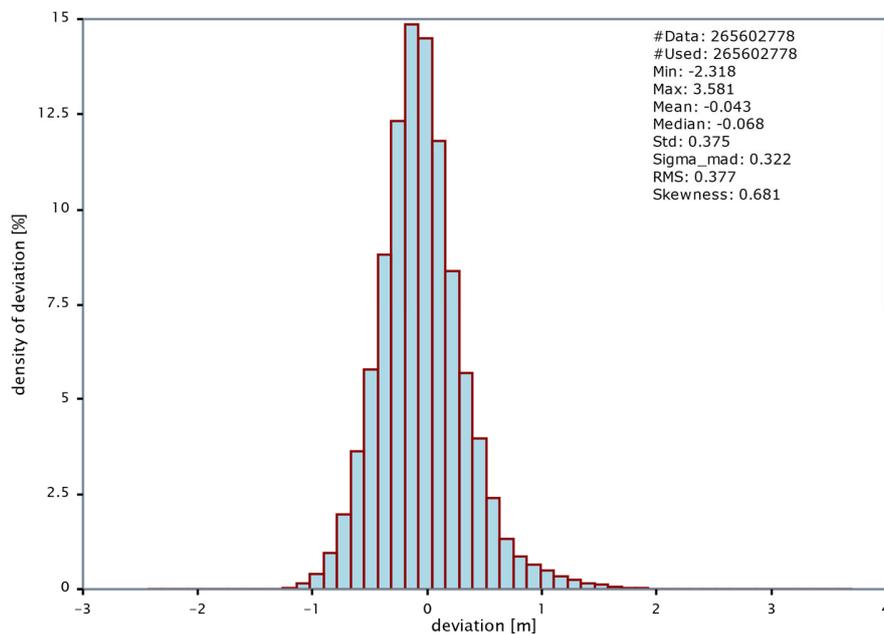


Abb. 8: Histogramm der Abweichungen der prädizierten Unterwasserdistanzen im Vergleich zur den Referenzdistanzen für 24 Testbilder.

Abb. 9 stellt mittels einer Heatmap die Verteilung der vorhergesagten Unterwasserdistanzen bezogen auf die Referenzdistanzen dar. Die Abweichungen nehmen mit zunehmender Tiefe zu. Darüber hinaus neigt das Netz bei Referenzabständen gegen Null dazu, größere Werte zu prädizieren. Dies könnte zum Teil auf schattige Bereiche an der Uferlinie zurückgeführt werden, bei welchen Dunkelheit größere Tiefen impliziert.

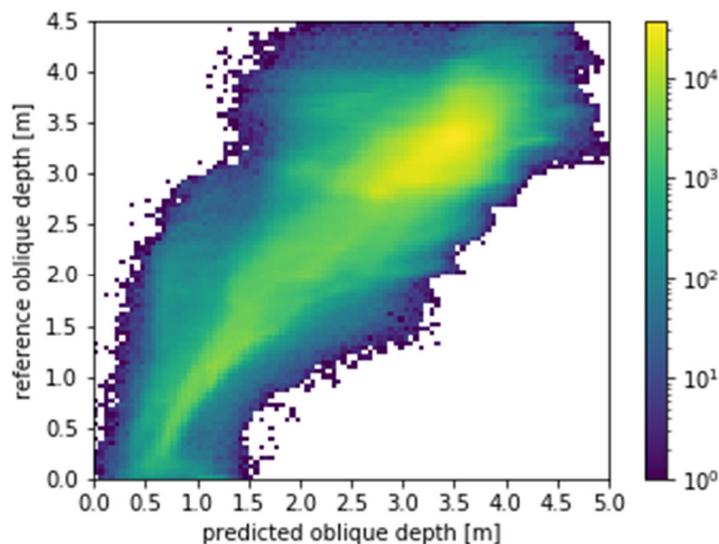


Abb. 9: Heatmap der prädizierten und Referenz-Unterwasserdistanzen.

Ein häufiges Problem bei Luftbildern von Gewässern sind Glanzlichter von der Wasseroberfläche (engl.: sun glint), welche durch direkte Reflexionen des Sonnenlichts in das Sichtfeld des Sensors hervorgerufen werden. Dies führt zu hellen Flecken in den Bildern, in denen es keine Möglichkeit gibt, Merkmale am Grund des Gewässers zu extrahieren. Für Ansätze, die von spektralen Informationen profitieren, kann dies zu großen Fehlern führen (LYZENGA et al. 2006). Ein Ausschnitt aus einem Testbild mit Glanzlichtern, sowie das Differenzbild für die vorhergesagten schrägen Unterwasserdistanzen ist in Abb. 10 dargestellt, in der vor allem in Glanzlicht-Bereichen große Abweichungen bis zu 3 m zu sehen sind. Die schrägen Unterwasserdistanzen in diesem Bereich werden aufgrund der erlernten umgekehrten Proportionalität von Helligkeit und Tiefe als nahezu Null vorhergesagt. Daher sind die Abweichungen von der Referenzdistanz für die jeweiligen Pixel abhängig, wodurch auch die hohe maximale Abweichung und die positive Schiefe in Abb. 10 erklärt werden kann. Dieses Verhalten spiegelt sich auch in der Heatmap als größere Abweichungen von der Diagonale zu kleineren prädizierten Abständen wider.

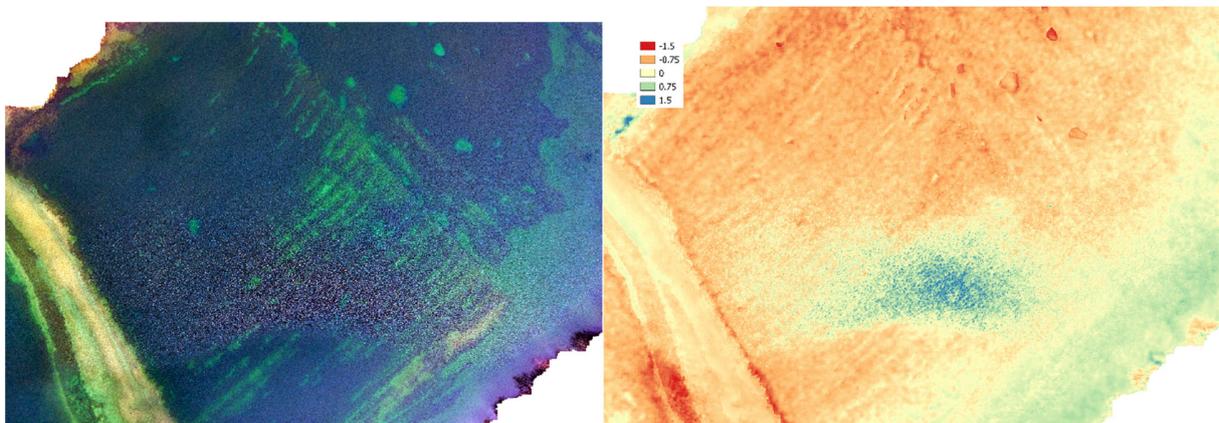


Abb. 20: Glanzlicht-Beispiel in einem Testbild sowie Auswirkung in der Prädiktion.

Darüber hinaus ist unten rechts im Bild in Abb. 10 ein kleiner Bereich mit unmaskierter Vegetation über der Wasseroberfläche, sowie daraus resultierende Schatten zu sehen, die ebenfalls Fehler verursachen. Bei der Untersuchung der kleineren hellen Bereiche oben rechts zeigt sich eine Schwäche der CNNs. Während Faltungskerne Informationen aus der Umgebung berücksichtigen, neigen sie dazu, starke Kanten zu verwischen. So finden sich an den Übergängen von Vegetation (dunkel) zu nacktem Boden (hell) große Abweichungen.

4.2 Angewandtes CNN für RGB ohne Coastal Blue Band

Um den Mehrwert des zusätzlichen Coastal Blue Bandes für die CNN-basierte Tiefenschätzung zu bewerten, wurde das Netz zusätzlich nur auf die RGB-Bilder trainiert. Beim Vergleich der Verteilung der Referenzdistanzen mit beiden Ansätzen in Abb. 11 ist zu erkennen, dass der Ansatz mit Einbeziehung des Coastal Blue Bandes eine wesentlich bessere Übereinstimmung zeigt gegenüber der Variante, welcher lediglich RGB zugrunde liegt.

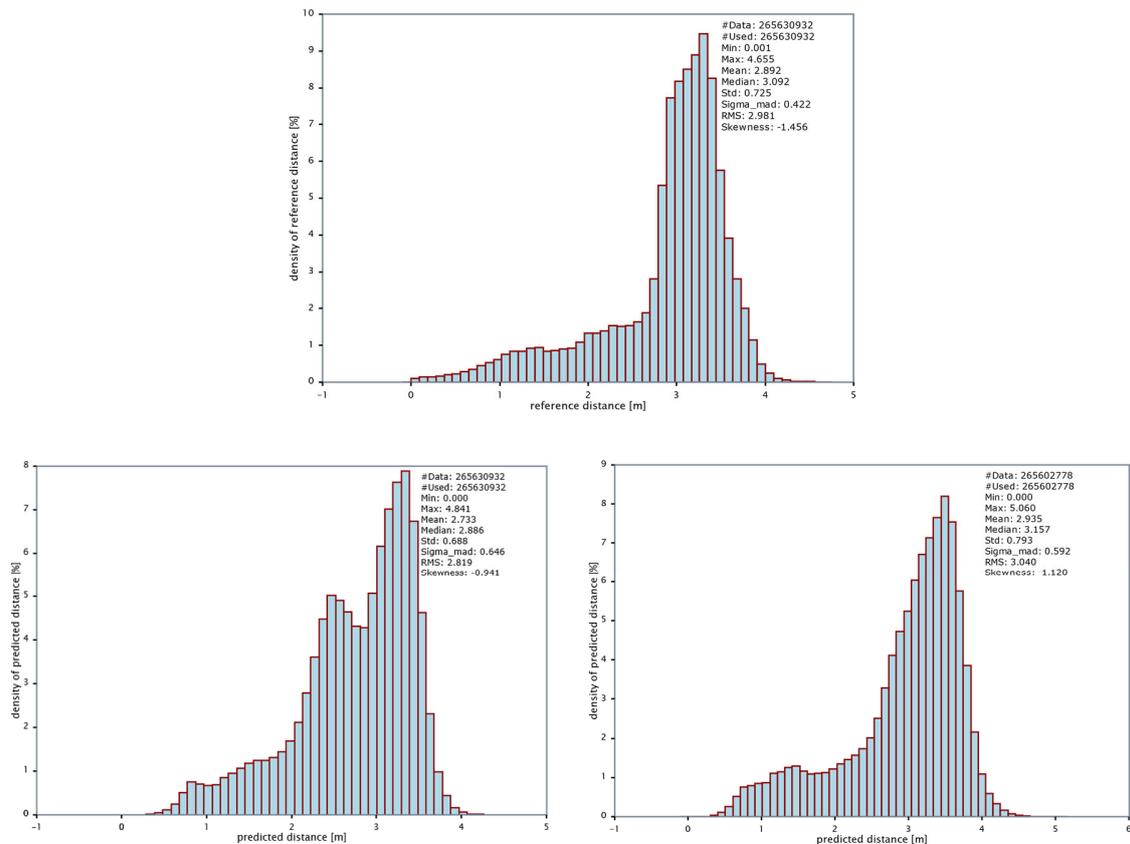


Abb. 31: Histogramme der prädierten Abstände (links: ohne Coastal Blue Kanal, rechts: mit Coastal Blue Kanal) der Testbilder im Vergleich zu dem Histogramm der Referenzabstände (oben).

5 Fazit und Ausblick

Im Rahmen des Beitrags wurde ein Ansatz vorgestellt, der die Gewässertiefenschätzung aus Luftbildern mittels faltender neuronaler Netze ermöglicht. Da der Bereich in den Testbildern dem Modell unbekannt ist, sind die Prädiktionen zuverlässig. In Bezug auf die Genauigkeit sind SoNAR oder LiDAR dem beschriebenen Ansatz überlegen. Anderenfalls sind in dieser Arbeit die Vorteile der CNN-basierten Bathymetrie-Schätzung gegenüber dem stereo-photogrammetrischen und linearen Regressionsverfahren aufgezeigt. Aufgrund der unterschiedlichen Bodenbedeckungen des Sees ist ein komplexeres Modell erforderlich als die lineare Regression. Außerdem ist das Ergebnis im Vergleich zu dem photogrammetrischen Ansatz von MULSOW et al. (2019) für homogene Bereiche wesentlich glatter. Darüber hinaus gibt es mehrere Verbesserungsmöglichkeiten. So verursachen beispielsweise Vegetation über der Wasseroberfläche sowie Glanzlichter Abweichungen, die sich sowohl auf das Training als auch auf den Test auswirken. Wenn dies berücksichtigt wird, könnte eine weitere Verbesserung der Ergebnisse erzielt werden.

Ein häufiges Problem bei der Anwendung von CNNs, ist der Mangel an Daten. Der große Vorteil der Verarbeitung im Bildraum, ist die Möglichkeit, den gesamten Datensatz mit allen überlappenden Bereichen zu nutzen, ohne ihn zu reduzieren. Dies ist auch der Grund, warum es möglich war,

die Bilder im Überlappungsbereich von Trainings- und Testbereich abzulehnen, so dass es keine Verbindung gab.

Um die Robustheit des Verfahrens zu evaluieren, kann das Netz auf ähnliche Datensätze (andere Wasserkörper) angewendet werden. Im Idealfall kann das Netz ohne Anpassungen genutzt werden. Andernfalls könnte zur Reduktion der Trainingsdaten „Transfer Learning“ angewendet werden. Der nächste logische Schritt bei diesem Verfahren wäre die Ableitung einer 3D-Punktwolke aus den vorhergesagten Unterwasserdistancen in den Bildern. Zu diesem Zweck ist ein geschätztes Wasserflächenmodell und die Orientierung der Kamera für jedes Bild erforderlich. Dadurch wäre es möglich, die überlappenden Bereiche aufeinanderfolgender Bilder zu analysieren und die unabhängig prädierten Wassertiefen auf Konsistenz zu prüfen. Darüber hinaus könnten Ausreißer in Einzelbildern, z.B. durch Glanzlichter, durch robuste Schätzverfahren für das Endergebnis kompensiert werden.

Eine weitere Überlegung ist die Einführung einer Klassifizierung in „Wasser“ und „Kein Wasser“ während des Trainings, um den derzeit notwendigen Vorverarbeitungsschritt des Maskierens von Trockengebieten zu vermeiden. Dies wurde hier nicht umgesetzt, da die Verfügbarkeit von Referenzdaten für jedes Bild automatisch eine Kein-Wasser-Klassifizierung als Nebenprodukt ermöglichte. Aber gerade bei der Übertragung auf ein anderes Gewässer ohne Referenzdaten sollte dies umgesetzt werden, zumal sonst die Wasserfläche für jedes einzelne Bild segmentiert werden müsste.

Da zumindest für Seen oft eine maximale Tiefe bekannt ist, wäre ein weiterer Ansatz, das vortrainierte Netz zu verwenden, ohne weitere Anpassungen vorzunehmen. Die Abstände könnten dann relativ bestimmt und anschließend auf die maximale Tiefe skaliert werden. Wenn eine gute Genauigkeit erreicht werden kann, wäre es möglich, Tiefenmodelle für Seen abzuleiten, ohne opulente Methoden wie Laserscanning oder Sonar einzusetzen. Der in diesem Beitrag beschriebene Ansatz hat jedenfalls gezeigt, dass moderne CNN-basierte Auswerteverfahren bei transparenten Wasserbedingungen im Flachwasserbereich eine robuste Methode für eine rein bildbasierten Tiefenschätzung darstellen.

6 Literaturverzeichnis

- ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., GOODFELLOW, I., HARP, A., IRVING, G., ISARD, M., JIA, Y., JOZEFOWICZ, R., KAISER, L., KUDLUR, M., LEVENBERG, J., MANE, D., MONGA, R., MOORE, S., MURRAY, D., OLAH, C., SCHUSTER, M., SHLENS, J., STEINER, B., SUTSKEVER, I., TALWAR, K., TUCKER, P., VANHOUCHE, V., VASUDEVAN, V., VIEGAS, F., VINYALS, O., WARDEN, P., WATTENBERG, M., WICKE, M., YU, Y. & GHEMAWAT, S., 2015: TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. 2015. URL <https://www.tensorflow.org>.
- CHOLLET, F., 2018: Keras: The python deep learning library. Astrophysics Source Code Library.
- COSTA, B.M., BATTISTA, T.A. & PITTMAN, S.J., 2009: Comparative evaluation of airborne LiDAR and ship-based multibeam SoNAR bathymetry and intensity for mapping coral reef ecosystems. Remote Sensing of Environment **113**(5), 1082-1100.

- DUMOULIN, V. & VISIN, F., 2016: A guide to convolution arithmetic for deep learning. arXiv pre-print arXiv:1603.07285.
- GOODFELLOW, I., BENGIO, Y. & COURVILLE, A., 2016: Deep learning. MIT press.
- IRISH, J.L. & WHITE, T.E., 1998: Coastal engineering applications of high-resolution lidar bathymetry. *Coastal engineering* **35**(1-2), 47-71.
- KRAUS, K., 1996: Photogrammetrie Band 2, Verfeinerte Methoden und Anwendungen. Auflage, Dümmler, Bonn.
- LEGLEITER, C.J., ROBERTS, D.A. & LAWRENCE, R. L., 2009: Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms* **34**(8), 1039-1059.
- LYZENGA, D.R., MALINAS, N.P. & TANIS, F. J., 2006: Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing* **44**(8), 2251-2259.
- MAAS, H.G., 2015: On the accuracy potential in underwater/multimedia photogrammetry. *Sensors* **15**(8), 18140-18152.
- MANDLBURGER, G., KREMER, J., STEINBACHER, F. & BARAN, R., 2018: Investigating the use of Coastal Blue imagery for bathymetric mapping of inland water bodies. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, **42**(1), 275-282, <https://doi.org/10.5194/isprs-archives-XLII-1-275-2018>.
- MASNADI-SHIRAZI, M. A., DE MOUSTIER, C., CERVENKA, P. & ZISK, S. H., 1992: Differential phase estimation with the SeaMARCII bathymetric sidescan sonar system. *IEEE Journal of Oceanic Engineering* **17**(3), 239-251.
- MATLAB, 2018: version 9.5.0 (R2018b). Natick, Massachusetts: The MathWorks Inc.
- MULSOW, C., MANDLBURGER, G., RESSL, C. & MAAS, H.G., 2019: Vergleich von Bathymetriedaten aus luftgestützter Laserscanner- und Kameraerfassung. *Publikationen der Deut. Gesellschaft für Photogrammetrie, Fernerkundung u. Geoinformation e.V.*, Band **28**, Kersten, T.P. (Hrsg.), 39. Wissenschaftlich-Technische Jahrestagung der DGPF, 20.-22. Februar 2019 in Wien, 318-333.
- NIELSEN, M. A., 2015: *Neural networks and deep learning* (Vol. 25). San Francisco, CA, USA: Determination press.
- NÜBEL, H., 2019: Bathymetry from multispectral aerial images via Convolutional Neural Networks. Bachelorarbeit im Studiengang Geodäsie und Geoinformatik, Universität Stuttgart.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS—A framework for Airborne Laser Scanning data analysis. *Computers, Environment and Urban Systems* **45**, 125-136.
- RIEGL, 2019: RIEGL VQ-880-G: Fully Integrated Topo-Hydrographic Airborne Laser Scanning System. http://www.riegl.com/uploads/tx_pxriegl/downloads/Infosheet_VQ-880-G_2016-05-23.pdf.
- RONNEBERGER, O., FISCHER, P. & BROX, T., 2015: U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234-241.
- WANG, Y., ZHOU, X., LI, C., CHEN, Y. & YANG, L., 2019: Bathymetry Model Based on Spectral and Spatial Multifeatures of Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*.