

Erzeugung von synthetischen Seitensichtsonar-Bildern mittels Generative Adversarial Networks

Ein Beitrag von YANNIK STEINIGER, JANNIS STOPPE, DIETER KRAUS und TOBIAS MEISEN

Für die Anwendung von Deep-Learning-Methoden zur automatischen Auswertung von Daten bildgebender Sonare stellt die nicht vorhandene Verfügbarkeit größerer Trainingsdatensammlungen nach wie vor ein Problem dar. In den letzten Jahren wurden jedoch sogenannte Generative Adversarial Networks (GAN) als ein Werkzeug aus dem Bereich Deep Learning für die Erzeugung synthetischer Daten entwickelt. In dieser Arbeit wird untersucht, inwieweit sich GANs zur Erzeugung künstlicher Sonarbilder eignen. Um das GAN auch mit wenigen Beispielen trainieren zu können, wird eine Art des Transfer-Lernens mit Hilfe von einfachen simulierten Bildern entwickelt. Es zeigt sich, dass die Performance eines Klassifikators durch Hinzunahme der künstlichen Bilder gesteigert werden kann.

Seitensichtsonar | Autonomes Unter-Wasser-Fahrzeug | Deep Learning | generative Netzwerke | Transfer-Lernen
side-scan sonar | autonomous underwater vehicle | deeplearning | generative adversarial network | transfer-learning

A remaining problem is the lack of large-scale sonar image data sets when applying deep learning algorithms for the automatic analysis of these data. However, over the past few years, generative adversarial networks (GAN) were developed as a tool for generating synthetic data. This work investigates how GANs can be used to generate synthetic sonar images. In order to train the GAN with only a few available samples, a transfer-learning approach is applied which uses simple simulated images. Using the additional synthetic sonar images, the performance of a classifier can be increased.

Autoren

Yannik Steiniger und Dr. Jannis Stoppe arbeiten am Institut für den Schutz maritimer Infrastrukturen des DLR in Bremerhaven.

Prof. Dr. Dieter Kraus leitet das Institut für Wasserschall, Sonartechnik und Signaltheorie an der Hochschule Bremen.

Prof. Dr. Tobias Meisen hat den Lehrstuhl für Technologien und Management der Digitalen Transformation an der Universität Wuppertal inne.

yannik.steiniger@dlr.de

1 Einleitung

Seitensichtsonare sind spezielle unterwasser-akustische Sensoren, die unter anderem an autonomen Unter-Wasser-Fahrzeugen (englisch: autonomous underwater vehicle, AUV) montiert werden, um den Meeresboden zu inspizieren und beispielsweise nach Objekten zu suchen. In den letzten Jahren wurde die Erkennungsrate von Objekten in Bildaufnahmen stetig verbessert. Grundlage hierfür sind maschinelle Lernverfahren, genauer Deep Convolutional Neural Networks (CNN), mittels derer heute Erkennungsraten erreicht werden, die der des Menschen in diesen Anwendungsbereichen entspricht (Wang et al. 2020). Sollen CNNs für die automatische Auswertung von Seitensichtsonar-Bildern herangezogen werden, stellt sich jedoch häufig das Problem einer zu geringen Datenlage für ein verlässliches Training. Das Sammeln von Seitensichtsonar-Bildern von Objekten ist besonders aufwendig, da im Allgemeinen die Position von Objekten unter Wasser im Vorfeld unbekannt ist. Auch das manuelle Auslegen von Objekten und die anschließende Aufnahme von Sonardaten ist, aufgrund der Zeit- und Kostenintensität, nur eine äußerst aufwendige Möglichkeit, einen umfangreichen und variablen Datensatz zu erstellen.

Während also die Technologie zur Unterstützung der Auswertung aufgenommener Sonardaten immer weiter voranschreitet, ist deren Anwendung auf Sonardaten als recht schwierig einzustufen, da nicht genügend Daten für das Training moderner, selbstlernender Algorithmen vorhanden sind.

Andererseits haben sich in den letzten Jahren generative neuronale Netze, sogenannte Generative Adversarial Networks (GAN) (Goodfellow et al. 2014), aus dem Bereich des Deep Learnings als eine gute Methode zum Generieren künstlicher Bilder erwiesen (Isola et al. 2017; Karras et al. 2020). GANs wurden bereits in anderen Arbeiten zur Generierung von Seitensichtsonar-Bildern (Steiniger et al. 2020) oder Synthetisches-Aperture-Sonar-Bildern (Reed et al. 2019) verwendet. Die Herausforderung, die sich jedoch ergibt, ist, dass auch GANs auf künstlichen neuronalen Netzen beruhen und entsprechend einen umfangreichen Trainingsdatensatz benötigen. Als Lösungsansatz greifen die beiden genannten Verfahren daher zunächst auf Simulationsdaten zurück. Für die Simulation kann beispielsweise ein Raytracer verwendet werden. [Abb. 1](#) zeigt zum einen ein reales Sonarbild eines Reifens und zum anderen ein mit dem Raytracer POV-Ray simuliertes Bild. Die simulierten Bilder

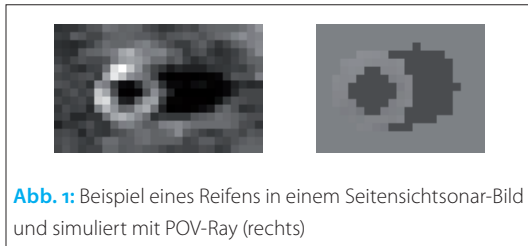


Abb. 1: Beispiel eines Reifens in einem Seitensichtsonar-Bild und simuliert mit POV-Ray (rechts)

weisen zwar die in einem echten Seitensichtsonar-Bild zu erwartenden geometrischen Formen für die Highlight- und Schattenregionen auf, enthalten aber weder weitere Details noch anwendungstypisches Rauschen oder Ähnliches.

Reed et al. verwenden in ihrem Ansatz ein GAN, um die simulierten Bilder in Sonarbilder zu »übersetzen«. Im hier vorgestellten Ansatz werden die simulierten Bilder stattdessen in einem ersten Schritt für ein Vortrainieren des GANs verwendet. Dadurch lernt das künstliche neuronale Netz, die geometrischen Formen der Highlight- und Schattenregionen zu erzeugen. Erst in einem zweiten Schritt, dem Transfer-Lernen, werden die echten Sonarbilder zum Trainieren verwendet, sodass das GAN die Verteilung der Pixelintensitäten in den jeweiligen Regionen im Bild lernt.

2 Aufnahme der Seitensichtsonar-Bilder

Für die Datenerzeugung wird das AUV SeaCat Mk1 der Atlas Elektronik GmbH verwendet (Kalwa 2019). Das AUV bietet die Möglichkeit, zuvor geplante Missionen weitgehend autonom abzufahren und dabei Messdaten mit Hilfe unterschiedlicher Sensoren zu erzeugen. Das AUV ist mit einem Doppler Velocity Log für die Messung der Geschwindigkeit über Grund, einem Drucksensor für die Messung der Tiefe und einem GNSS-System zur Positionierung über Wasser ausgestattet. In Kombination mit einem hochgenauen Inertial Navigation System (INS) ermöglichen diese Sensoren eine Positioniergenauigkeit des AUVs von bis zu 0,1 % der unter Wasser zurückgelegten Wegstrecke, wodurch eine präzise Lokalisierung von Zielobjekten unter Wasser möglich ist. Außer dem Seitensichtsonar Edge-tech 2205 am Rumpf führt das AUV in seinem modularen Sensorkopf eine Kamera mit künstlicher Beleuchtung, ein hochauflösendes Fächerecholot und einen parametrischen Sub-Bottom-Profiler mit. Das Seitensichtsonar erlaubt grundsätzlich das gleichzeitige Scannen mit zwei unterschiedlichen Frequenzen. Die Daten dieser Messung wurden mit einer Frequenz von 850 kHz erhoben, die eine Auflösung im Zentimeterbereich und eine Reichweite von 75 m zu jeder Seite erlaubt.

In der Regel werden parallel zur Aufnahme der Daten mit dem Seitensichtsonar Bathymetriedaten mit einem Fächerecholot aufgenommen. Je nach Auflösung der Daten besteht die Möglichkeit, über die einfache Bathymetrie hinaus 3D-Punktwolken von Objekten zu generieren (Heuskin 2020). Mit

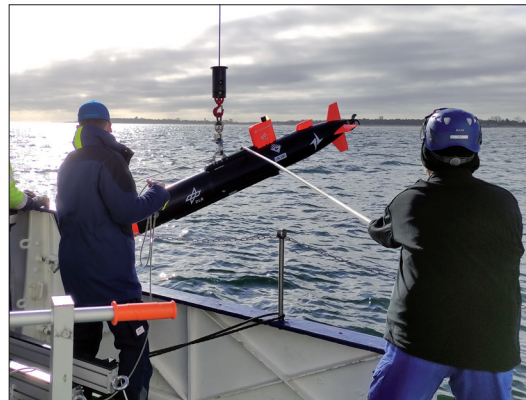


Abb. 2: AUV SeaCat Mk1 der Atlas Elektronik beim Zu-Wasser-Lassen

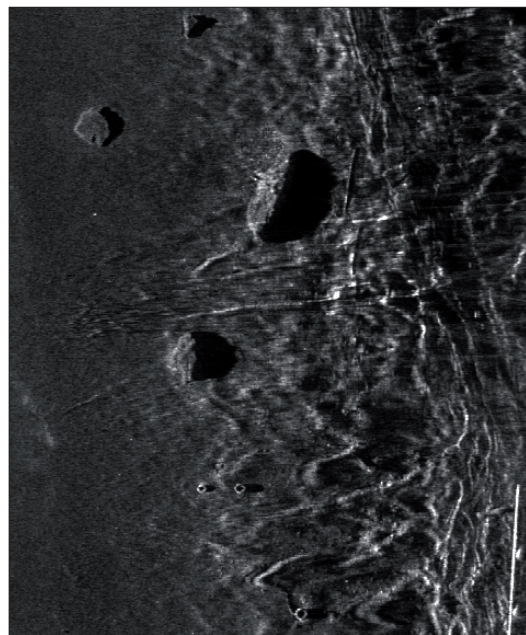


Abb. 3: Beispiel eines aufgenommenen Seitensichtsonar-Bildes

einer experimentellen Signalverarbeitungskette werden die während einer Mission aufgenommenen Rohdaten des Seitensichtsonars in Sonarbilder des Meeresbodens, wie in Abb. 3 dargestellt, umgewandelt.

Die aus den erstellten Sonarbildern extrahierten Objekte sind in Tabelle 1 aufgeführt. Zusätzlich ist die für die spätere Anwendung der neuronalen Netze relevante Aufteilung in Trainings- und Testbeispiele angegeben. Es ist zu erkennen, dass vor allem für die Klasse Reifen besonders wenig Bei-

Objekt	Anzahl der Trainingsbeispiele	Anzahl der Testbeispiele
Reifen	18	17
Stein	36	210
Hintergrund	36	210

Tabelle 1: Annotierte Objekte und Anzahl dieser im Trainings- bzw. Testdatensatz

spiele zur Verfügung stehen. Synthetische Bilder dieser Klasse können genutzt werden, um dieses Ungleichgewicht auszugleichen.

3 Künstliche Sonarbilder für die Klassifikation

Wie zuvor bereits geschrieben, sind insbesondere Objektklassen, die sehr selten auftreten, eine zusätzliche Herausforderung für Lernverfahren. In dieser Arbeit widmen wir uns daher der Synthese von Bildern derartiger Klasse, in diesem Fall also der Klasse Reifen. Obwohl sich diese Arbeit speziell mit dem Objekt Reifen befasst, lässt sich das Verfahren auch auf beliebige andere Objekte übertragen. Für komplexere Objekte wird jedoch gegebenenfalls die Simulation zum Erzeugen der Daten für das Vortrainieren aufwendiger.

Der entwickelte Trainingsprozess des GANs mit Vortrainieren und Transfer-Lernen ist in Abb. 4 schematisch dargestellt. Das GAN ist aus zwei neuronalen Netzen aufgebaut: dem Generator und dem Discriminator. Der Generator erhält als Eingangsdaten ein Bild mit zufällig verteilten Pixelintensitäten. Das formulierte Lernziel besteht

darin, diesen Eingang in ein Sonarbild zu transformieren. Der Discriminator hingegen sieht entweder ein echtes Bild aus dem Trainingsdatensatz oder ein vom Generator erzeugtes Bild und weist dem jeweiligen Bild eine Wahrscheinlichkeit P zu, die angibt ob er das Bild für ein echtes Bild ($P = 1$) oder generiertes Bild ($P = 0$) hält. Während des Trainings ist das Ziel des Generators, den Discriminator glauben zu lassen, dass das generierte Bild echt ist. Somit nähert sich die Verteilung der Pixelintensitäten der generierten Bilder im Laufe des Trainings immer mehr der der Trainingsdaten an.

Beim hier betrachteten Transfer-Lernen besteht der Trainingsdatensatz zunächst aus 10 000 simulierten Bildern. In der Simulation werden Parameter wie Abstand zum Sonar oder Größe des Objekts variiert, um einen differenzierten Datensatz zu erhalten. Wie in Abb. 4 (b) zu erkennen, weisen die generierten Bilder nach diesem Vortrainieren eine deutliche Highlight- und Schattenregion auf, wie sie für Seitensichtsonar-Bilder zu erwarten ist. Für das Transfer-Lernen wird dieses GAN mit den echten Seitensichtsonar-Bildern trainiert, wobei zuvor die Gewichte der letzten Schicht im neuro-

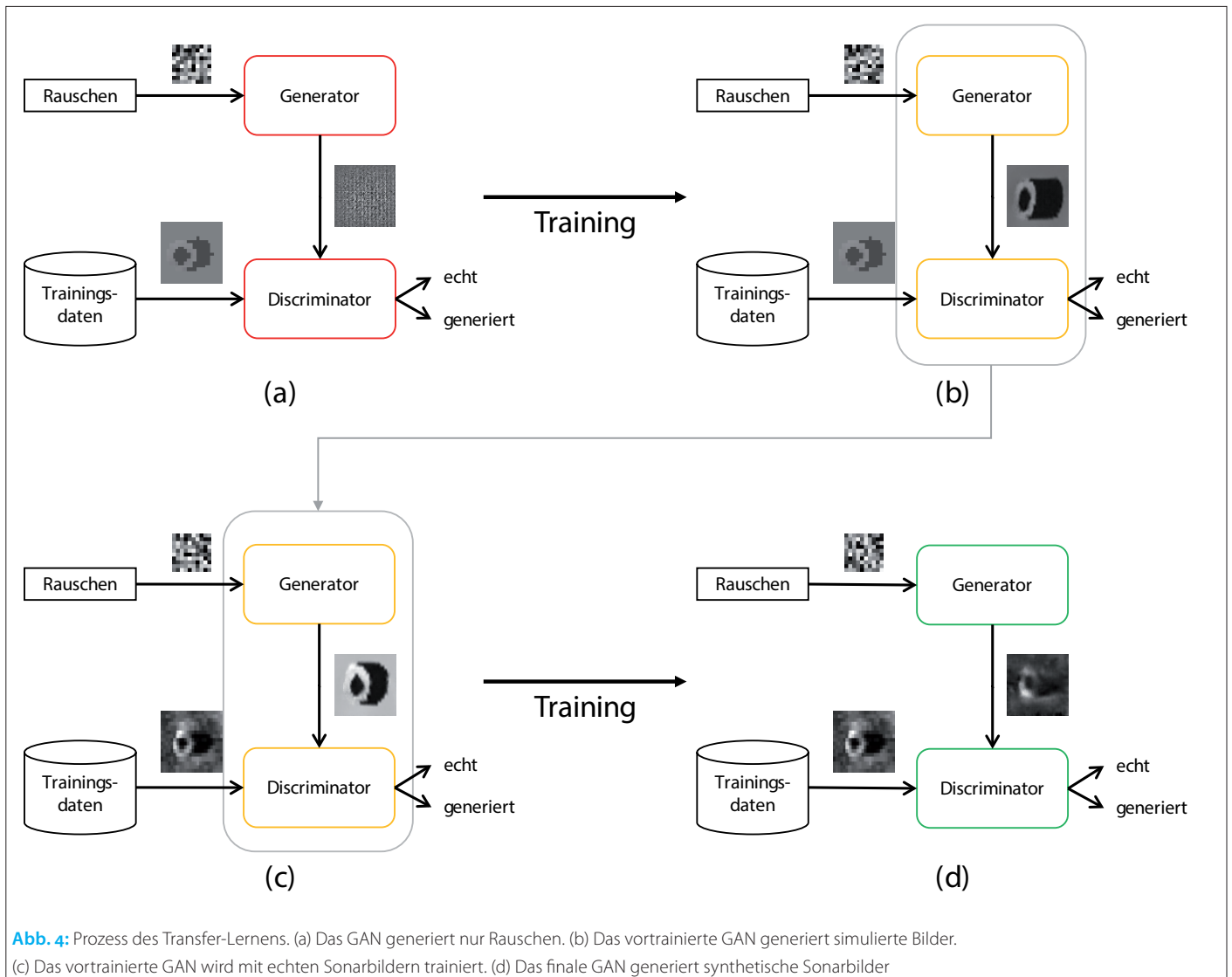


Abb. 4: Prozess des Transfer-Lernens. (a) Das GAN generiert nur Rauschen. (b) Das vortrainierte GAN generiert simulierte Bilder. (c) Das vortrainierte GAN wird mit echten Sonarbildern trainiert. (d) Das finale GAN generiert synthetische Sonarbilder

nalen Netz des Discriminators neu initialisiert werden, um eine Überanpassung und so ein Verschwinden des Gradienten während des Trainings zu verhindern. Die [Abb. 4 \(d\)](#) zeigt, dass das finale GAN nicht nur die geometrische Form, sondern auch die Pixelintensitäten in dem Bild realistisch generiert.

Das beschriebene Verfahren, basierend auf der Methodik des Transfer-Lernens, verringert das für GANs typische Problem des Mode-Collapse. Da nur wenige reale Trainingsbeispiele zur Verfügung stehen, würde ein direktes Lernen der Transformation von Eingangsruschen zu Sonarbild dazu führen, dass das GAN in einen Zustand kollabiert, in dem lediglich ein (nicht zwingend realistisches) Bild unabhängig vom Rauschen generiert wird. Der Discriminator lernt nur noch dieses eine Bild zurückzuweisen. Daraufhin muss der Generator seine Transformation nur minimal ändern, um den Discriminator zu täuschen. Somit führt ein Fortsetzen des Trainingsprozesses nicht mehr aus diesem Zustand des Mode-Collapse heraus. Die einzige Lösung ist es, das gesamte Training neu zu starten. Um dies zu umgehen, muss das Kollabieren des GANs verhindert werden.

Aber: Obwohl das Transfer-Lernen zwar das Problem des Mode-Collapse verhindert, kann die Variabilität in den generierten Daten dennoch gering sein. Wie verschieden die Daten, die von dem GAN generiert werden, letztlich sind, ist dabei von vielen Faktoren abhängig und muss noch genauer untersucht werden. Das Verhindern des Kollabierens des GANs hin zu einer einzigen Konfiguration der Ergebnisse stimmt jedoch zuversichtlich, dass durch eine Anpassung der Parameter auch die gewünschte Variabilität erreicht werden kann.

4 Performance des Klassifikators

Ziel der Erzeugung künstlicher Sonarbilder ist es letztendlich, die Performance eines Klassifikators durch die zusätzlichen Daten zu verbessern. Für

Anzahl synthetischer Bilder	ACC _{bal}	F1 _{macro}
0	0,6298	0,6302
9	0,6388	0,6379
18	0,6376	0,6364
36	0,6379	0,6368

Tabelle 2: Klassifikationsperformance des CNN für verschieden starke Augmentierungen

diese Klassifikation wird im Folgenden ein CNN verwendet. Die Augmentierung des Trainingsdatensatzes des CNN ist in [Abb. 5](#) dargestellt. Mit Hilfe der synthetischen Sonarbilder von Reifen wird der Datensatz aus [Tabelle 1](#) ausgeglichen.

Die Performance des CNN wird durch die ausbalancierte Genauigkeit (englisch: balanced accuracy, ACC_{bal}) und den Makro-F1-Wert (englisch: macro F1 score, F1_{macro}) gemessen. Die [Tabelle 2](#) gibt die Performance für eine unterschiedliche Anzahl an synthetischen Bildern im augmentierten Trainingsdatensatz an. Mit 18 synthetischen Bildern ergibt sich ein ausgeglichener Trainingsdatensatz. Alle vier Experimente wurden mit zehn verschiedenen Initialisierungen des CNN durchgeführt. In [Tabelle 2](#) sind die Mittelwerte der Metriken aufgelistet.

Das Hinzunehmen der synthetischen Daten zeigt eine leichte Steigerung in beiden Metriken von circa einem Perzentil. Es ist jedoch auch zu erkennen, dass mehr künstliche Bilder nicht zu einer weiteren Steigerung führen. Dies lässt darauf schließen, dass die Variabilität in den generierten Daten nicht ausreichend ist, um einen großen synthetischen Datensatz zu erzeugen. Es ist hierbei jedoch zu beachten, dass zum Trainieren des GANs – entsprechend der Ausgangsproblematik von nur spärlich verfügbaren Trainingsdaten – lediglich 18 reale Seitensichtsonar-Bilder verwendet wurden.

Die Performancesteigerung durch das Ver-

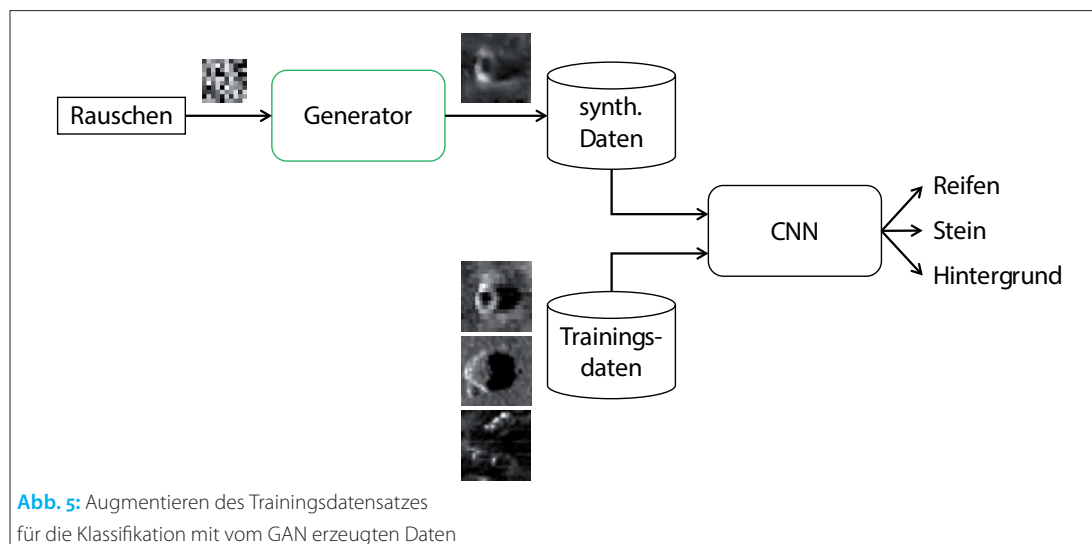


Abb. 5: Augmentieren des Trainingsdatensatzes für die Klassifikation mit vom GAN erzeugten Daten

wenden der synthetischen Daten ist messbar, verwandelt dabei jedoch – umgangssprachlich formuliert – einen passabel arbeitenden Detektor keineswegs in ein Patentrezept. Während für das Trainieren von CNNs zur automatischen Auswertung von Seitensichtsonar-Bildern die ausgiebige Erfassung von Daten nach wie vor das Mittel der Wahl darstellt, lassen sich mit synthetischen Daten gegebenenfalls noch ein paar Register ziehen, falls die Datenlage nicht verbessert werden kann.

5 Zusammenfassung und Ausblick

In diesem Artikel wurde ein Verfahren vorgestellt, um mittels Methoden des maschinellen Lernens synthetische Seitensichtsonar-Bilder zu erzeugen. Speziell wurde ein GAN mit Bildern eines Raytracers vortrainiert und anschließend ein Transfer-Lernen mit echten Sonarbildern durchgeführt. Die synthetischen Bilder weisen die zu erwartenden geometrischen Formen und Pixelintensitäten auf. Es konnte gezeigt werden, dass die Verwendung dieser synthetischen Daten zusätzlich zum Trainingsdatensatz eines Klassifikators die Performance steigern kann.

Um die Methodik weiter zu verbessern, ist ein primäres Ziel, die Variabilität in den generierten Daten zu erhöhen. Dies kann zum einen dadurch erreicht werden, dass auch der Hintergrund, sprich der Meeresboden, in einem weiteren Vortrainingschritt generiert wird. Zum anderen können bekannte Eigenschaften der Verteilungen der Pixelintensitäten in den Highlight- und Schattenregionen genutzt werden (Lehmann 2013). Damit lassen sich die Bilder nach dem Vortrainieren bereits realistischer gestalten und so etwas an Komplexität aus dem Schritt des Transfer-Lernens nehmen.

Generell sind GANs eine vielversprechende Methode, nicht um die Erfassung echter Trainingsdaten zu substituieren, sondern um fehlende Daten, die nicht ohne weiteres gewonnen werden können, zu synthetisieren. Echte Daten bleiben das Mittel der Wahl – aber sollten insbesondere für bestimmte Objekttypen eines Datensatzes mehr Beispiele benötigt werden, stellen GANs eine Möglichkeit zur Verbesserung der Datenlage als Alternative zu eventuell kostspieligen weiteren Kampagnen dar. //

Literatur

- Goodfellow, Ian; Jean Pouget-Abadie et al. (2014): Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27, MIT Press, DOI: 10.1145/3422622
- Heuskin, David; Frank Lehmann (2020): Drohnengestützte Erfassung von maritimen Infrastrukturen. *Hydrographische Nachrichten*, DOI: 10.23784/HN116-05
- Isola, Philipp; Jun-Yan Zhu et al. (2017): Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2017.632
- Kalwa, Jörg (2019): Unter-Wasser-Drohnen für Hydrographie und Seebodenerkundung. *Hydrographische Nachrichten*, DOI: 10.23784/HN114-02
- Karras, Tero; Samuli Laine et al. (2020): Analyzing and Improving the Image Quality of StyleGAN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR42600.2020.00813
- Lehmann, Benjamin (2013): Beiträge zur automatisierten Objekterkennung und insbesondere zur Bild-Segmentierung in hochaufgelösten Sonarbildern. Logos Verlag Berlin
- Reed, Albert; Isaac D. Gerg et al. (2019): Coupling Rendering and Generative Adversarial Networks for Artificial SAS Image Generation. *OCEANS 2019 MTS/IEEE SEATTLE*, DOI: 10.23919/OCEANS40490.2019.8962733
- Steiniger, Yannik; Jannis Stoppe et al. (2020): Dealing With Highly Unbalanced Sidescan Sonar Image Datasets for Deep Learning Classification Tasks. *Global OCEANS 2020 MTS/IEEE Singapore – U.S. Gulf Coast*, DOI: 10.1109/IEEECONF38699.2020.9389373
- Wang, Chien-Yao; Alexey Bochkovskiy; Hong-yuan Mark Liao (2020): Scaled-YOLOv4: Scaling Cross Stage Partial Network. <https://arxiv.org/abs/2011.08036v2>