

Bachelor-Teamprojekt

**Analyse, Design und Implementierung von
unterschiedlichen Generative Adversarial Network
Architekturen im Bereich der Bildverarbeitung**

**Analysis, design and implementation of different
Generative Adversarial Network architectures in
the field of image processing**

Elisa Du, Marcel Hoffmann

Mat.Nr.: 976090, 973043

Betreuer:

Prof. Dr. rer. nat. E.-G. Haffner

Datum:

01. Dezember 2023

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht. Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellenangabe versehen. Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Ort, Datum

Unterschrift

Abstract

Groundbreaking advances in image processing have been achieved through the continuous development of Generative Adversarial Networks (GANs). This thesis deals with the analysis, design and implementation of different GAN architectures. The focus is on the fundamentals of GANs, Pix2Pix and CycleGAN. The emphasis is on the detailed description of the generators and discriminators for Pix2Pix and CycleGAN. This is followed by an in-depth examination of the implementation and evaluation using various evaluation criteria, including generator and discriminator loss and the Structural Similarity Index (SSIM). Finally, a comparison is made between Pix2Pix and CycleGAN to show the advantages and disadvantages of each. The aim of this thesis is to provide a comprehensive overview of the diversity of GAN architectures and identify potential challenges by analysing two representative models.

Zusammenfassung

Bahnbrechende Fortschritte in der Bildverarbeitung wurden durch die kontinuierliche Entwicklung von Generative Adversarial Networks (GANs) erzielt. Diese Arbeit beschäftigt sich mit der Analyse, dem Entwurf und der Implementierung verschiedener GAN-Architekturen. Der Fokus liegt dabei auf den Grundlagen von GANs, Pix2Pix und CycleGAN. Der Schwerpunkt liegt auf der detaillierten Beschreibung der Generatoren und Diskriminatoren für Pix2Pix und CycleGAN. Es folgt eine eingehende Untersuchung der Implementierung und Evaluierung anhand verschiedener Evaluierungskriterien, einschließlich Generator- und Diskriminatorverlust und dem Structural Similarity Index (SSIM). Abschließend wird ein Vergleich zwischen Pix2Pix und CycleGAN durchgeführt, um die jeweiligen Vor- und Nachteile aufzuzeigen. Das Ziel dieser Arbeit ist es, durch die Analyse von zwei repräsentativen Modellen einen umfassenden Überblick über die Vielfalt von GAN-Architekturen zu geben und mögliche Herausforderungen zu identifizieren.

Abkürzungsverzeichnis

Adam	Adaptive Moment Estimation
CGAN	Conditional Generative Adversarial Netowrk
CT	Computertomographie
CNN	Convolutional Neural Network
DL	Deep Learning
GAN	Generative Adversarial Network
IS	Interception Score
ML	Machine Learning
MRT	Magnetresonanztomographie
MS	Mode Score
ReLU	Rectified Linear Units
ResNET	Residual Neural Network
SSIM	Structural Similarity Index Measure
cGAN	Conditional Generative Adversarial

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen	3
2.1. Generative Adversarial Networks	3
2.1.1. Funktionsweise	3
2.1.2. Training	4
2.1.3. Anwendungen	5
2.1.4. Limitationen	5
2.2. Pix2Pix	6
2.2.1. Pix2Pix-Kernkonzepte	6
2.2.2. Anwendungen von Pix2Pix	10
2.3. CycleGAN	11
2.3.1. CycleGAN - Kernkonzepte	11
2.3.2. Anwendungen von CycleGAN	14
2.4. Convolutional Layers	15
2.5. Bibliotheken	16
2.5.1. Tensorflow	16
2.5.2. Keras	16
2.5.3. Matplotlib	17
3. Literaturreview	19
4. Problembeschreibung	21
5. Lösungsbeschreibung	23
5.1. Training- und Testdaten	23
5.1.1. Datenladung für GAN-Training	23
5.1.2. Vorverarbeitung des Datensatzes	24
5.2. Implementierung der Pix2PixGAN-Architektur	26
5.2.1. Generator	26
5.2.2. Diskriminator	27
5.2.3. Verlustfunktion	28
5.3. Implementierung der CycleGAN-Architektur	29
5.3.1. Generator und Diskriminator	29
5.3.2. Verlustfunktion	30
5.4. Training und Hyperparameter	32
5.4.1. Optimierungstechnik und Optimizers	32
5.4.2. Fortschrittsüberwachung und Visualisierung	33
6. Evaluation	43
6.1. Bewertungskriterien	43
6.1.1. Diskriminator-Verlust	43
6.1.2. Generator-Verlust	43
6.1.3. Diskriminator Genauigkeit	43

6.1.4. SSIM-Score (Structural Similarity Index)	44
6.2. Ergebnisse und objektive Bewertung	44
6.2.1. Pix2Pix	44
6.2.2. CycleGAN	50
6.3. Vergleich von Pix2Pix und CycleGAN	55
7. Fazit und Ausblick	59
7.1. Fazit	59
7.2. Ausblick	59
A. Anhang - Code	61
B. Anhang - Modelle	71
Verzeichnisse	77
Literaturverzeichnis	77
Abbildungsverzeichnis	81
Tabellenverzeichnis	83
Code-Auszugs-Verzeichnis	86
Glossar	87
Arbeitsverteilung	87

T

Einleitung

In den letzten Jahren haben sich Generative Adversarial Networks (GANs), eine bahnbrechende Entwicklung in der Welt der künstlichen Intelligenz, die maßgeblich auf die Arbeit von Ian Goodfellow und seinem Team zurückgeht, als revolutionäre Methode etabliert. Diese technologische Entwicklung, die erstmals 2014 in einem wegweisenden Paper vorgestellt wurde, hat es ermöglicht, dass Maschinen in der Bildverarbeitung Leistungen erbringen, die früher undenkbar waren. GANs, ursprünglich konzipiert als Instrumente zur Erstellung von realistischen Bildern, haben weitreichende Anwendungen in verschiedenen Bereichen gefunden, insbesondere in der Bildverarbeitung, wo sie durch ihre Fähigkeit, Bilder zu generieren und zu transformieren, eine neue Ära eingeleitet haben.

Die innovative Struktur von GANs besteht aus zwei gegeneinander antretenden Netzwerken: dem Generator und dem Diskriminator. Der Generator hat die Aufgabe, Daten zu erzeugen, die von realen Daten ununterscheidbar sind, während der Diskriminator die Echtheit dieser Daten beurteilt. Durch diesen Wettbewerb lernt der Generator, immer bessere Fälschungen zu erzeugen und der Diskriminator wird effizienter im Erkennen dieser Fälschungen. Dieses Zusammenspiel führt zu einer stetigen Verbesserung beider Netzwerke und ermöglicht die Erstellung hochqualitativer generierter Daten. Diese Methodik bildet den Kern der Funktionsweise von GANs und ist entscheidend für ihre Anwendung in der Bildverarbeitung, wo sie genutzt werden, um realistische Bilder zu generieren und existierende Bilder in vielfältiger Weise zu transformieren ([Ian14](#)).

Aufbauend auf dieser grundlegenden Architektur von GANs, hat sich das spezialisierte Framework Pix2Pix als bedeutende Weiterentwicklung in der Welt der Bild-zu-Bild-Übersetzung etabliert. Entwickelt von Phillip Isola und seinem Forschungsteam, adaptiert Pix2Pix das GAN-Prinzip für spezifische Anforderungen der Bildtransformation. Während der klassische Ansatz von GANs auf die Erzeugung neuer, realistisch wirkender Daten abzielt, fokussiert Pix2Pix auf die präzise Übersetzung von Eingangsbildern in gewünschte Ausgangsbilder. Hierbei übernimmt der Generator die Rolle der Bildtransformation, indem er relevante Merkmale aus dem Eingangsbild extrahiert und diese in ein neues, zielgerichtetes Bild umwandelt. Der Diskriminator bewertet anschließend diese Übersetzung, wodurch die Genauigkeit und Realitätsnähe der Bildtransformation weiter optimiert wird. Diese spezialisierte Anwendung von GANs demonstriert die Vielseitigkeit und Anpassungsfähigkeit der Technologie und öffnet neue Horizonte in der Bildverarbei-

tung ([PJTA](#)).

CycleGAN, entwickelt von Jun-Yan Zhu und Kollegen, ist eine Erweiterung der Generative Adversarial Networks (GANs) und speziell für die Aufgabe der Bild-zu-Bild-Übersetzung in Fällen konzipiert, wo keine paarigen Trainingsdaten vorhanden sind. Es unterscheidet sich von herkömmlichen GANs und Pix2Pix durch seinen innovativen Ansatz der "Zykluskonsistenz". Im Gegensatz zu Pix2Pix, das paarige Trainingsdaten benötigt (wo jedes Eingangsbild einem spezifischen Ausgangsbild zugeordnet ist), ermöglicht CycleGAN die Übersetzung zwischen zwei Bildsammlungen, ohne, dass eine Eins-zu-eins-Beziehung zwischen den Bildern in den Sammlungen besteht. Diese Fähigkeit ist besonders wertvoll für Aufgaben wie das Umwandeln von Sommerbildern in Winterbilder oder das Übertragen von Stilen zwischen verschiedenen Künstlern, wo paarige Trainingsdaten schwierig oder unmöglich zu sammeln sind. Die Schlüsselinnovation von CycleGAN ist die Einführung einer Zykluskonsistenzverlust-Funktion. Diese Funktion sorgt dafür, dass ein Eingangsbild, das in ein Bild einer anderen Domain übersetzt und dann zurück in die ursprüngliche Domain übersetzt wird, dem Originalbild ähnlich bleibt. Zum Beispiel, wenn ein Foto eines Pferdes in ein Zebra umgewandelt wird und dann wieder zurück in ein Pferd, sollte das endgültige Bild dem ursprünglichen Pferdefoto ähnlich sein. Diese Zykluskonsistenz hilft, bedeutungsvolle und kohärente Übersetzungen zwischen unverbundenen Bildsammlungen zu gewährleisten ([ZPIE](#)).

Diese Arbeit gibt zunächst einen Überblick über verschiedene GAN-Architekturen, einschließlich Pix2Pix und CycleGAN und deren Anwendungen in der Bildverarbeitung. Darauf folgend wird eine detaillierte Beschreibung der Implementierung verschiedener GAN-Architekturen präsentiert, einschließlich der Behandlung von Trainingsdaten, der Architektur und der Optimierungstechniken.

Der Einsatz von GANs in der Bildverarbeitung bietet vielversprechende Ergebnisse, jedoch unterscheidet sich der Stil und die Herangehensweise dieser Technologie von traditionellen Bildverarbeitungsmethoden. Diese Arbeit zielt darauf ab, die Möglichkeiten von Pix2Pix- und CycleGAN in der Bildverarbeitung zu ergründen indem diese Modelle selbst Implementiert und Trainiert werden.

Abschließend werden die Ergebnisse der Durchführung des GAN-Trainings präsentiert und evaluiert. Zudem wird untersucht, welchen Einfluss die Hyperparameter auf das GAN-Training haben, um die Resultate weiter zu verbessern.

Im Anschluss wird diese Arbeit mit einer Diskussion über die erzielten Ergebnisse und einer Bewertung der verschiedenen GAN-Architekturen abgeschlossen. Ein besonderer Fokus liegt auf der praktischen Anwendung dieser Technologien in der Bildverarbeitung und den Möglichkeiten, die sich daraus für zukünftige Forschung und Entwicklung ergeben.

2

Grundlagen

2.1. Generative Adversarial Networks

Generative Adversarial Networks, kurz GANs, sind eine aufstrebende Technologie im Bereich des maschinellen Lernens und der künstlichen Intelligenz. Inspiriert von Ian Goodfellow und seinen Kollegen im Jahr 2014, bieten GANs eine effiziente Möglichkeit, tiefe Repräsentationen von Daten zu erlernen, ohne dass große Mengen an annotierten Trainingsdaten benötigt werden (CWD⁺18). Dies wird durch die Verwendung von Backpropagation und den Wettbewerb zwischen zwei neuronalen Netzen - dem Generator und dem Diskriminator - erreicht. Daraus ergeben sich zahlreiche neue Ansätze zur Generierung realistischer Inhalte. Die Anwendungen reichen von der Bildgenerierung bis hin zur Superauflösung und Textgenerierung (AMB21).

2.1.1. Funktionsweise

Der Generator und der Diskriminator sind die Hauptkomponenten eines GAN. Die beiden neuronalen Netze werden gleichzeitig trainiert und konkurrieren miteinander, wobei der Generator versucht, den Diskriminator zu täuschen, indem er synthetische Inhalte erzeugt. Um die Glaubwürdigkeit des Generators zu erhöhen, so dass der Diskriminator nicht mehr zwischen den Eingaben unterscheiden kann, wird das gesamte Netz trainiert. Die Netze werden in der Regel als mehrschichtige Netzwerke implementiert, die aus Convolutional und Fully-Connected Schichten bestehen (CWD⁺18).

Generator

Der Generator dient zur Erzeugung künstlicher Daten wie Bilder und Texte. Dieser ist nicht mit dem realen Datensatz verbunden und lernt daher nur durch die Interaktion mit dem Diskriminator. Wenn der Diskriminator nur noch 50% der Eingaben richtig vorhersagt, gilt der Generator als optimal (CWD⁺18).

Diskriminator

Die Unterscheidung zwischen echten und unechten Eingaben ist Aufgabe des Diskriminators. Der Diskriminator kann sowohl künstliche als auch reale Daten ver-

wenden. Wenn der Diskriminatior nicht mehr richtig unterscheiden kann, wird er als konvergierend bezeichnet (AMB21). Andernfalls wird er als optimal bezeichnet, wenn seine Klassifizierungsgenauigkeit maximiert ist. Im Falle eines optimalen Diskriminators wird das Training des Diskriminators gestoppt und der Generator trainiert alleine weiter, um die Genauigkeit des Diskriminators wieder zu verbessern (CWD⁺18).

2.1.2. Training

Das Training besteht in der Optimierung der Parameter für sowohl den Generator als auch den Diskriminatior durch die Anwendung von Backpropagation zur Verbesserung dieser Parameter. Dieses Verfahren wird häufig als anspruchsvoll und instabil beschrieben. Einerseits gestaltet sich die Herausforderung, beide Modelle konvergieren zu lassen. Andererseits besteht die Problematik darin, dass der Generator Muster erzeugen kann, die für verschiedene Eingaben äußerst ähnlich sind, was als "Modekollaps-Problem" bekannt ist. Der Diskriminatiorverlust kann sich rasch gegen Null konvergieren, wodurch ein zuverlässiger Gradientenfluss für die Aktualisierung des Generators verhindert wird. Zur Bewältigung dieser Herausforderungen wurden verschiedene Lösungsansätze vorgeschlagen. Ein Beispiel ist die Verwendung heuristischer Verlustfunktionen. Eine alternative Strategie, die von Sonderby et al. vorgeschlagen wurde, besteht darin, den Datensatz vor der Verwendung zu verrauschen (CWD⁺18).

Adversieller Verlust

Der adversielle Verlust, auch als GAN-Verlust bekannt, spielt eine zentrale Rolle im Trainingsprozess. Dieser Verlust basiert auf dem Konzept des MinMax-Spiels zwischen dem Generator und dem Diskriminatior. Der Generator strebt danach, den Diskriminatior zu überlisten und Daten zu erzeugen, die von echten Daten nicht zu unterscheiden sind. Gleichzeitig ist es das Ziel des Diskriminators, zwischen echten und generierten Daten zu differenzieren. Der adversielle Verlust wird in der Gleichung 2.1 repräsentiert.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2.1)$$

Hierbei bezeichnet G den Generator, D den Diskriminatior, x echte Daten, z das Rauschen und p_{data} sowie p_z die Wahrscheinlichkeitsverteilungen von echten Daten und Rauschen. Der MinMax-Ansatz impliziert, dass der Generator versucht, den Verlust zu minimieren, während der Diskriminatior versucht, ihn zu maximieren. Eine gezielte Optimierung und Anpassung des adversiellen Verlusts ist entscheidend, um Herausforderungen wie dem Modekollaps Problem und den Konvergenzproblemen zu begegnen (HHYY20).

2.1. Generative Adversarial Networks

2.1.3. Anwendungen

GANs wurden ursprünglich im Kontext des unüberwachten maschinellen Lernens konzipiert, haben jedoch herausragende Leistungen sowohl im halbüberwachten Lernen als auch im Reinforcement Learning gezeigt (AMB21). Diese Vielseitigkeit hat dazu geführt, dass GANs in verschiedenen Domänen wie dem Gesundheitswesen und dem Bankwesen Anwendung finden.

Im Bereich der medizinischen Bildgebung bieten GANs innovative Lösungsansätze, um den Herausforderungen von Datenknappheit und Patientenprivatsphäre zu begegnen. Von der Erkennung und Behandlung chronischer Krankheiten über die Segmentierung bis hin zur Bildrekonstruktion können GANs vielfältige Anwendungen haben. Zahlreiche GAN-Ansätze wurden bereits entwickelt, um die Rauschunterdrückung in medizinischen Bildverfahren zu verbessern, was wiederum die Qualität von Diagnosen steigern kann (YWB19). Darüber hinaus werden GANs nicht nur im Gesundheitswesen, sondern auch in anderen Bereichen eingesetzt, insbesondere in der Bildsynthese.

In der Finanzindustrie helfen GANs, das Handels- und Risikomanagement zu verbessern, indem sie synthetische Zeitreihen erzeugen, die wichtige Finanzdaten widerspiegeln (EO). Zheng et al. (2018) schlugen beispielsweise eine GAN vor, die auf dem Telekommunikations-Betrugsfall in China im Jahr 2017 basiert und die Wahrscheinlichkeit berechnet, wann eine Überweisung betrügerisch sein könnte (ZZS⁺18).

Des Weiteren wurden Forschungsanstrengungen unternommen, um mittels GANs menschliche Bewegungen vorherzusagen, insbesondere anhand von 3D-Skelettensequenzen (JZJ⁺20). Zudem ermöglichen GANs die Identifikation von 3D-Objekten sowie die Generierung realistischer Bilder und Texte in verschiedenen Anwendungsbereichen (AMB21). Die Vielseitigkeit von GANs eröffnet somit ein breites Spektrum an Potenzialen für verschiedene Anwendungsbereiche.

2.1.4. Limitationen

Ein kritisches Problem von GANs ist die Instabilität des Trainings aufgrund von Modekollaps, was die Weiterentwicklung des generativen Lernens und potentielle Anwendungen einschränkt (LLW⁺). Der Generator lernt nur Bilder bestimmter Arten der Datenverteilung. Andere Arten, die ebenfalls in der Verteilung vorkommen, werden hingegen vernachlässigt (SVR⁺). Ansätze wie das Hinzufügen von Rauschen zum Netzwerk, eine Manifold Entropy Estimation (LLW⁺) und implizites Variationslernen (SVR⁺) wurden bereits vorgeschlagen, um dieses Problem zu lösen. Des Weiteren birgt die Fähigkeit eines GANs zur Generierung von Inhalten, die nahezu identisch mit authentischen Inhalten sind, potenzielle Herausforderungen in realen Szenarien, insbesondere im Zusammenhang mit der menschlichen Bildsynthese. Diese Fähigkeit ermöglicht es Betrügern, gefälschte Profile in sozialen Medien zu erstellen. Gezielte Anwendungen von GANs, die darauf ausgelegt sind, einzigartige und realistische Bilder von Personen zu erzeugen, die in der Realität nicht existieren, könnten die Erstellung falscher Profile erschweren (AMB21).

2.2. Pix2Pix

Pix2Pix hat sich als zentrales Framework für Bild-zu-Bild-Übersetzungen auf der Basis von bedingten generativen adversariellen Netzwerken (cGANs) etabliert. Es ermöglicht die Erstellung einer abstrakten Abbildung von einem Eingangsbild zu einem korrespondierenden Ausgangsbild und bewältigt dabei eine vielfältige Palette an Bildübersetzungsaufgaben, wie die Transformation von Skizzen in realistische Bilder oder die Konvertierung von Tages- zu Nachtaufnahmen.

2.2.1. Pix2Pix-Kernkonzepte

Architektur des Generators

Die Bildverarbeitung hat in den letzten Jahren durch den Einsatz tiefer neuronaler Netzwerke erhebliche Fortschritte gemacht. Im Mittelpunkt vieler dieser Fortschritte steht die U-Net-Architektur, die speziell für die Bildsegmentierung entwickelt wurde. Diese Architektur zeichnet sich durch ihre ausgeklügelte Kombination aus Encoder- und Decoder- Strukturen sowie durch den Einsatz von Skip-Verbindungen aus ([PJTA](#)).

Bei der Encoder-Decoder-Struktur handelt es sich um einen Ansatz, bei dem das Eingangsbild zunächst durch den Encoder schrittweise reduziert wird. Dieser Prozess dient dazu, wesentliche Merkmale des Bildes zu erfassen. Anschließend wird das Bild durch den Decoder wiederhergestellt, indem die zuvor extrahierten Merkmale verwendet werden. Während dieser Prozesse besteht jedoch das Risiko des Informationsverlustes, insbesondere in den tieferen Schichten des Netzwerks. Um dieses Problem zu adressieren, führt die U-Net-Architektur Skip-Verbindungen ein. Diese direkten Verbindungen zwischen korrespondierenden Schichten des Encoders und Decoders sorgen dafür, dass Detailinformationen nicht verloren gehen. Genauer gesagt, ermöglichen diese Verbindungen den direkten Informationsfluss zwischen jeweils äquivalenten Schichten, wodurch die Rekonstruktion des Bildes im Decoder mit einer höheren Genauigkeit erfolgt([PJTA](#)).

Die Bedeutung von Skip-Verbindungen zeigt sich insbesondere in Anwendungen wie der Bild-zu-Bild-Übersetzung. Hier muss oft ein Bild mit niedriger Auflösung in ein Bild mit hoher Auflösung überführt werden, ohne dass Details verloren gehen. Die U-Net-Architektur, die angereichert mit diesen Verbindungen ist, ermöglicht daher eine feinere Rekonstruktion, die sowohl globale als auch lokale Informationen berücksichtigt ([PJTA](#)).

Somit kann die U-Net-Architektur durch ihre Kombination aus Encoder-Decoder-Struktur und Skip-Verbindungen ein effektives Werkzeug für die Bildsegmentierung darstellen. Ihre Fähigkeit, sowohl globale Muster als auch feine Details zu berücksichtigen, macht sie zu einer bevorzugten Wahl für viele Bildverarbeitungsaufgaben ([PJTA](#)).

In Abbildung 2.1 ist die typische U-Net-Architektur dargestellt. Die linke Seite des "U" repräsentiert den Encoder-Teil, der das Eingangsbild schrittweise reduziert und wesentliche Merkmale extrahiert. Die rechte Seite repräsentiert den Decoder-Teil,

2.2. Pix2Pix

der das Bild mithilfe der extrahierten Merkmale rekonstruiert. Die horizontalen Linien repräsentieren die Skip-Verbindungen, die sicherstellen, dass Detailinformationen zwischen den korrespondierenden Schichten des Encoder und Decoders direkt übertragen werden (**PJTA**).

In der Pix2Pix Technologie fungiert die U-Net-Architektur als zentraler Bestandteil des Generators, verantwortlich für die Bild-zu-Bild-Übersetzung. Die Wahl von U-Net für diesen Zweck basiert auf seiner Fähigkeit, feinere Details und Kontextinformationen aus dem Eingangsbild zu bewahren, was für die Qualität der Bild-zu-Bild-Übersetzung ausschlaggebend ist. Der Generator nutzt die Encoder-Decoder-Struktur von U-Net, um den globalen Kontext des Bildes zu erfassen, während die Skip-Verbindungen sicherstellen, dass auch lokale Details im resultierenden Bild erhalten bleiben. Diese Eigenschaften machen U-Net zu einem effektiven Werkzeug innerhalb des Generators, indem sie eine hohe Detailtreue und Kontextsensitivität in der Bild-zu-Bild-Übersetzung ermöglichen(**PJTA**).

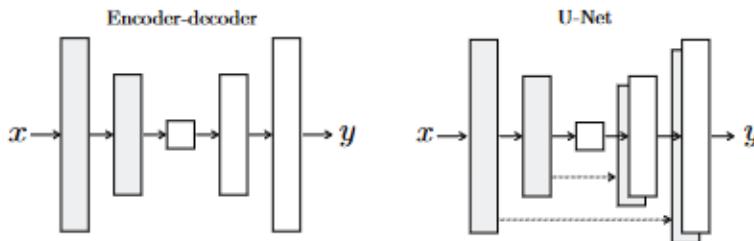


Abbildung 2.1.: Schematische Darstellung der U-Net-Architektur. Die Architektur besteht aus einem Encoder-Teil (links), einem Decoder-Teil (rechts) und Skip-Verbindungen zwischen korrespondierenden Schichten (**PJTA**).

Architektur des Diskriminators

Im Kontext von Pix2Pix GANs, spielt der PatchGAN-Diskriminator eine besonders wichtige Rolle. Der zentrale Unterschied dieses Diskriminators zu allgemeinen Diskriminatoren liegt in der Art und Weise, wie er Bilder bewertet. Statt das gesamte Bild zu beurteilen, zerlegt der PatchGAN-Diskriminator das Bild in mehrere kleinere Bildabschnitte oder Patches und bewertet jeden Patch einzeln auf seine Echtheit (**PJTA**).

Ein solches Vorgehen hat den klaren Vorteil, dass feinere Strukturen und Details in den Bildern erkannt und beurteilt werden können. Durch diese segmentierte Beurteilung kann der Diskriminator besser einschätzen, ob die Struktur und Beschaffenheit eines bestimmten Bildteils realistisch ist. Dies ist besonders nützlich, da kleinere Unstimmigkeiten in den Bildern, die ein allgemeiner Diskriminator möglicherweise übersieht, vom PatchGAN erfasst werden können.

Ein weiterer Vorteil des PatchGAN-Diskriminators ist seine Skalierbarkeit. Da er auf festen Patchgrößen basiert, kann er flexibel auf Bilder unterschiedlicher Größen angewendet werden, ohne dass das zugrunde liegende Modell geändert werden muss. Dies führt nicht nur zu einer schnelleren Bildverarbeitung, sondern ermöglicht auch eine effiziente Ausführung auf großen Bildern. Darüber hinaus reduziert

es potenzielle Kachelartefakte, die bei allgemeinen Diskriminatoren auftreten können ([PJTA](#)).

Der PatchGAN-Diskriminator kann wenn er effektiv eingesetzt wird, zu besseren und realistischeren Bildern im adverseriellen Lernprozess beitragen. Seine Fähigkeit, lokale Bildinformationen zu bewerten, ermöglicht es auch subtile Unterschiede in den Bildern zu erkennen, was zu einer verbesserten Qualität der generierten Bilder führt ([PJTA](#)).

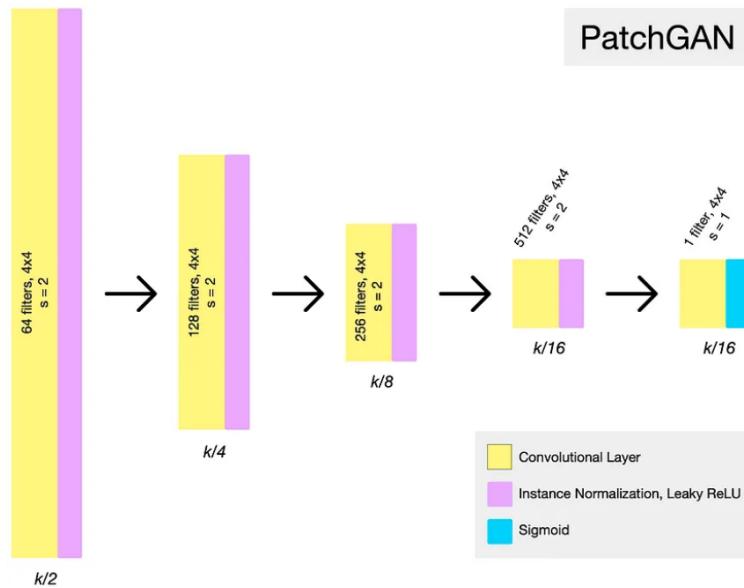


Abbildung 2.2.: Eine mögliche Architektur eines PatchGAN Diskriminator¹

L1-Verlustfunktion

Die L1-Verlustfunktion, auch als Mean Absolute Error (MAE) bekannt, ist im Pix2Pix-Modell, einem bedingten Generative Adversarial Network (cGAN) für Bild-zu-Bild-Übersetzungen, von zentraler Bedeutung. Diese Funktion misst den durchschnittlichen absoluten Unterschied zwischen den vorhergesagten und den tatsächlichen Werten, was besonders bei der Genauigkeit der niedrigen Frequenzen im Bild wichtig ist. Daher trägt die L1-Verlustfunktion maßgeblich zur Bewahrung der strukturellen Integrität und des Kontextes des Bildes bei.

Die Kombination des L1-Verlusts mit dem adversariellen Verlust im Pix2Pix-Modell ist entscheidend. Während der adversarielle Verlust darauf abzielt, die generierten Bilder realistischer wirken zu lassen, fokussiert sich der L1-Verlust auf die niedrigen Frequenzen, um die strukturelle Genauigkeit zu verbessern. Diese Kombination erlaubt es dem Modell, sowohl die niedrigen als auch die hohen Frequenzen effektiv zu erfassen, was zu generierten Bildern führt, die sowohl strukturell korrekt als auch visuell ansprechend sind (PJTA).

Die L1-Verlustfunktion wird allgemein beschrieben als: $L1(G) = \mathbb{E}_{xyz} [|y - G(x, z)|]$ wobei G das Generator-Modell darstellt, x und z den Eingabewert und den Rauschvektor repräsentieren und y das Ground-Truth-Ergebnis ist. Der L1-Verlust berechnet den durchschnittlichen absoluten Unterschied zwischen den generierten Bildern $G(x, z)$ und den echten Bildern y , was zur Reduzierung der Unschärfe in den generierten Bildern beiträgt (PJTA).

Die L1-Verlustfunktion neigt jedoch dazu, bei den hohen Frequenzen unscharfe Ergebnisse zu liefern. Dies liegt daran, dass der L1-Verlust den Median der möglichen Werte bevorzugt, was zu einer Glättung der Bildtexturen führen kann. Um dieses Problem zu adressieren und scharfe, hochfrequente Details im Bild zu erhalten, wird der L1-Verlust im Pix2Pix-Modell mit einem adversariellen Verlust kombiniert. Diese synergetische Kombination von Verlustfunktionen ermöglicht es dem Pix2Pix-Modell, hochwertige Bild-zu-Bild-Übersetzungen durchzuführen, die sowohl visuell ansprechend als auch strukturell korrekt sind (PJTA).

Zusätzlich hat sich die Synergie aus L1- und adversariellen Verlust als vielseitig einsetzbar für diverse Anwendungsfälle im Bereich der Bild-zu-Bild-Übersetzung erwiesen, darunter die semantische Segmentierung und die Farbgebung. Indem es die Pix2Pix-Modellierung ermöglicht, sowohl niedrige als auch hohe Bildfrequenzen präzise zu erfassen, trägt diese Verlustfunktionskombination dazu bei, die visuelle Qualität und die Anwendbarkeit des Modells auf ein breites Spektrum von Herausforderungen zu steigern (PJTA).

Training

Der Trainingsprozess von Pix2Pix-Generative Adversarial Networks in der Bild-zu-Bild-Übersetzung geht über die bloße Erlernung der Abbildung von Eingabe-zu Ausgabebildern hinaus. Er umfasst auch das Entwickeln einer Verlustfunktion, die speziell auf diese Art der Bildtransformation abgestimmt ist. Pix2Pix benötigt eine spezifische Art von Trainingsdaten, um effektiv zu funktionieren. Die

Trainingsdaten bestehen aus Paaren von Bildern, wobei jedes Paar ein Eingabebild und das entsprechende Ausgabebild enthält. Diese Bilder können eins bis drei Kanäle aufweisen, was bedeutet, dass das Modell sowohl mit monochromatischen (Graustufen) als auch mit farbigen Bildern (RGB) arbeiten kann. Im Rahmen des Trainingsprozesses von Pix2Pix wird eine iterative Methode verwendet, bei der der Generator und der Diskriminator abwechselnd trainiert werden. Eine Schlüsselkomponente dieses Prozesses ist die Verwendung einer zusammengesetzten Verlustfunktion, die sowohl den adversiellen Verlust (bewertet vom Diskriminator) als auch den L1-Verlust (mittlerer absoluter Fehler zwischen generiertem Bild und Zielbild) umfasst. Dadurch wird der Generator dazu angehalten, realistische Übersetzungen der Eingabebilder zu generieren. Dieses Gleichgewicht zwischen Generator und Diskriminator ist entscheidend für die Effektivität des Pix2Pix-Modells ([Haz21](#)).

Im Pix2Pix-Modell spielt die Batch-Normalisierung eine wesentliche Rolle für die Stabilisierung des Lernprozesses. Diese Normalisierung ergänzt die iterative Trainingsmethode von Pix2Pix, bei der Generator und Diskriminator abwechselnd trainiert werden. Sie ist besonders wichtig, da der Generator mit einer Vielzahl von Eingabedaten, einschließlich monochromatischer und farbiger Bilder, arbeitet. Zusammen mit der zusammengesetzten Verlustfunktion, die aus adversiellem und L1-Verlust besteht, verbessert die Batch-Normalisierung die Trainingsstabilität und Qualität der generierten Bilder ([PJTA](#)).

2.2.2. Anwendungen von Pix2Pix

Pix2Pix ist eine fortschrittliche Methode für Bild-zu-Bild-Übersetzungsaufgaben und hat eine breite Palette von Anwendungen in der Bildverarbeitung.

Ein markantes Anwendungsbeispiel ist die Umwandlung von architektonische Entwürfe oder Zeichnungen in realistische Gebäudefotos. Besonders eindrucksvoll ist diese Anwendung beim CMP Facades-Datensatz, wo aus simplen Fassadenzeichnungen detaillierte Gebäudebilder generiert werden. ([PJTA](#))

Im Bereich der Fotografie wird Pix2Pix verwendet, um Schwarz-Weiß-Fotos in farbige Bilder zu konvertieren, was besonders bei der Restaurierung alter Fotografien von Bedeutung sein kann ([PJTA](#)).

Die Transformation von Tagesaufnahmen in Nachtbilder ist eine weitere beeindruckende Leistung von Pix2Pix sowie die Umwandlung von Thermalauflnahmen in Farbfotos ([PJTA](#)).

Schließlich wird Pix2Pix auch zur Vervollständigung von Fotos mit fehlenden Pixeln verwendet, beispielsweise um unvollständige Bilder, die aus Paris StreetView stammen, zu reparieren und zu vervollständigen ([PJTA](#)).

Jedoch kann Pix2Pix auch bei medizinischen Anwendungen, insbesondere in der Augenheilkunde, hilfreich sein. Hier wird Pix2Pix für die Post-Interventions-Prognose eingesetzt, um zu zeigen, wie sich die Netzhaut nach einer Anti-VEGF-Injektionsbehandlung verändert. Dies basiert auf Bildern, die vor der Injektion bei Patienten mit exsudativer altersbedingter Makuladegeneration aufgenommen

2.3. CycleGAN

wurden. Durch diese Technik können Ärzte und Patienten eine Vorstellung davon bekommen, welche Veränderungen im Auge nach der Behandlung zu erwarten sind ([Ara22](#)).

2.3. CycleGAN

CycleGAN, das 2017 von Jun-Yan Zhu et al. vorgestellt wurde, stellt eine neue Entwicklung im Bereich des maschinellen Lernens und insbesondere der Bildübersetzung zwischen unpaaren Domänen dar. Es erweitert die Pix2Pix-Architektur durch die Einführung einer Zykluskonsistenz-Verlustfunktion (Cycle Consistency Loss), die sicherstellt, dass das Originalbild nach einem Übersetzungs- und Rückübersetzungszyklus erhalten bleibt. Der Generator G transformiert Bilder aus der Domäne X in die Domäne Y , während der Generator F den umgekehrten Prozess durchführt. Diese Transformationen werden ohne gepaarte Trainingsdaten durchgeführt.

2.3.1. CycleGAN - Kernkonzepte

Architektur des Generators

Die Architektur der Generatoren in CycleGAN spielt eine entscheidende Rolle bei der erfolgreichen Durchführung von Bildübersetzungen zwischen verschiedenen Domänen. Typischerweise basieren die Generatoren auf dem ResNet-Ansatz, der für seine Fähigkeit bekannt ist, tiefe neuronale Netze zu trainieren([HZRS](#)).

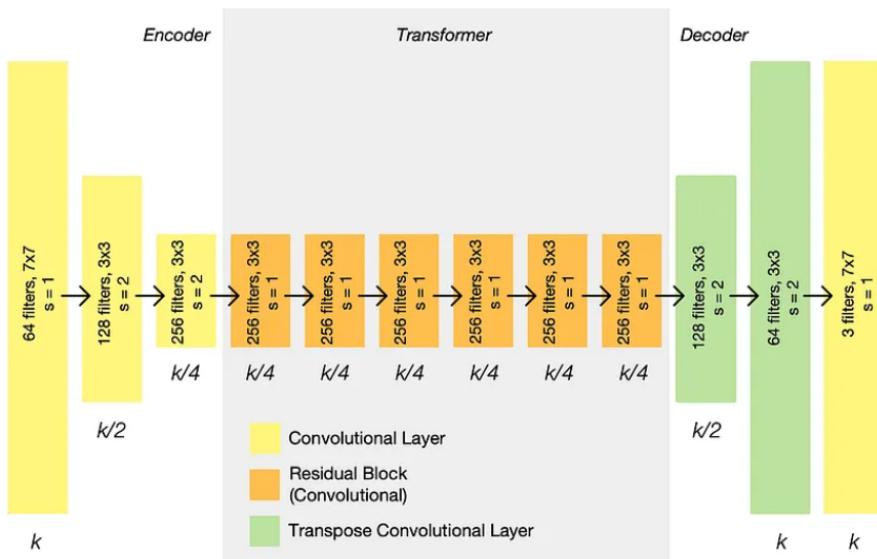


Abbildung 2.3.: Eine Architektur eines CycleGAN Generator. Instanznormalisierung und ReLU Aktivierung erfolgt nach jeder Schicht ²

Im Rahmen der Architektur von Zhu et al. manifestiert sich der Generator im CycleGAN in drei zentralen Abschnitten, wie graphisch in Abbildung 2.3 illustriert. Der Encoder besteht aus drei Convolutional-Schichten, welche unmittelbar auf das Eingabebild einwirken und dabei die Repräsentationsgröße reduzieren, sowie die Kanalanzahl erhöhen. Das resultierende Bild unterzieht sich einem Transformer, zusammengesetzt aus mehreren Residualblöcken. Die aus dieser Transformation hervorgehende Repräsentation durchläuft den Decoder, welcher aus zwei Transpose-Convolutional-Schichten besteht und somit das Bild erneut vergrößert. Die finale RGB-Ausgabe wird durch eine Ausgabeschicht generiert. Jede dieser Schichten ist mit einer Instanznormalisierung und ReLU-Aktivierung versehen, was sowohl die Trainingsstabilität fördert, als auch die Qualität der generierten Bilder optimiert([UVL, NH10](#)).

Die ResNet-Methode, von Kaiming He et al. im Jahr 2015 eingeführt, bietet eine Lösung für das Degradationsproblem. Dieses Phänomen tritt auf, wenn tiefe neuronale Netze bei Zugabe zusätzlicher Schichten schlechtere Leistungen erbringen als flachere Netze, da die Rückwärtspropagierung von Fehlern in tiefen Netzwerken erschwert wird. Die Integration von Residualblöcken ermöglicht die Überwindung dieses Problems durch die Hinzufügung einer Identitätsabbildung. Das Netzwerk lernt diese Abbildung, indem es das Residuum auf Null setzt. Residualblöcke dienen dazu, Änderungen und Fehler zu erlernen, die notwendig sind, um von der Eingabe zur gewünschten Ausgabe zu gelangen. Dies wird durch Shortcut-Verbindungen realisiert, die eine oder mehrere Ebenen überspringen und am Ende einer gestapelten Schicht hinzugefügt werden. Solche Verbindungen fügen keine zusätzlichen Parameter oder Rechenleistung hinzu, und das gesamte Netzwerk kann weiterhin mittels stochastischem Gradientenabstieg (SGD) trainiert werden([HZRS](#)).

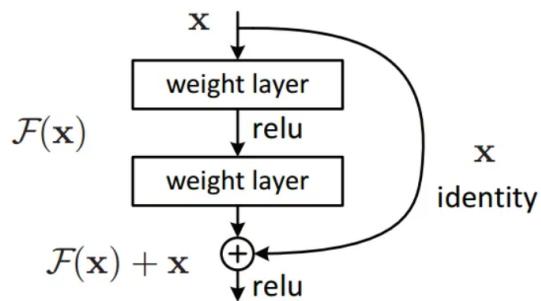


Abbildung 2.4.: Ein Aufbau eines Residualblocks

Architektur des Diskriminators

In Pix2Pix ist die gängige Architektur für Diskriminatoren ein PatchGAN, bei dem das Bild in kleine Patches unterteilt wird und jeder Patch separat klassifiziert wird (vgl. Abschnitt 2.2.1). Diese Vorgehensweise ermöglicht eine präzise Unterscheidung zwischen echten und generierten Bildern auf lokaler Ebene, was

2.3. CycleGAN

besonders in Bezug auf die feinstrukturierte Bewertung von Bildabschnitten von Vorteil ist ([ZPIE](#)).

In den Implementierungen von CycleGAN wird im Unterschied zu Pix2Pix auf die Verwendung von Batch-Normalisierung verzichtet, und stattdessen wird Instanznormalisierung bevorzugt. Bei der Instanznormalisierung wird jedes Bild individuell betrachtet, ohne Berücksichtigung über die gesamte Batch-Dimension hinweg. Dieser Ansatz bietet eine effektivere Stilübertragung im Feed-Forward-Modus und weist eine schnellere Konvergenz auf im Vergleich zur Batch Normalisierung ([HB](#)). Eine mögliche Architektur ist in Abbildung 2.2 veranschaulicht.

Training

Das Training von CycleGAN erfolgt nach einem kompetitiven Verfahren. Die Generatoren $G : X \rightarrow Y$ und $F : Y \rightarrow X$ konkurrieren mit den entsprechenden Diskriminatoren D_X und D_Y . D_X versucht, die von F erzeugten Bilder von den echten Bildern aus X zu unterscheiden, während D_Y versucht, die von G erzeugten Bilder von den echten Bildern aus der Domäne Y zu unterscheiden (Abbildung 2.5). Die adversen Verluste sind so optimiert, dass die erzeugten Bilder für die Diskriminatoren kaum von den echten Bildern zu unterscheiden sind. ([ZPIE](#)).

Cycle - Konsistenz

CycleGAN führt zusätzlich eine Cycle-Konsistenz ein. Diese stellt sicher, dass die Übersetzungen zwischen den Domänen sowohl vorwärts (X nach Y), als auch rückwärts (Y nach X) konsistent sind. Dies ist in der Abbildung 2.5 dargestellt.

Die Kernidee besteht darin, dass nach der Übersetzung von X nach Y und zurück nach X das resultierende Bild dem ursprünglichen X entsprechen sollte. Um dies zu erreichen, wird die Differenz zwischen dem Originalbild x und dem zyklisch übersetzten Bild $F(G(x))$ mit Hilfe der L1-Verlust minimiert.

Durch die Einführung dieser zyklischen Konsistenz stellt ein Ansatz zur Lösung des Problem des Modekollapses dar. Die weitere Verlustfunktion stellt sicher, dass die generierten Bilder mehr Strukturen enthalten und somit konsistenter Übersetzungen zwischen den Domänen liefern ([ZPIE](#)).

Identity - Loss

Zusätzlich zu den adversariellen und zyklischen Verlusten kann ein Identitätsverlust in die Gesamtverlustfunktion integriert werden, um sicherzustellen, dass die Farbkomposition des Eingabebildes beibehalten wird, während es ins Ausgabebild übersetzt wird. Insbesondere bei der Erstellung von Fotografien aus Gemälden hat sich diese Methode bewährt. Wenn der Generator G ein Bild aus dem Bereich Y erhält, darf es sich aufgrund seiner bereits vorhandenen Zugehörigkeit zu diesem Bereich nicht mehr verändern. Diese Unveränderlichkeit ist visuell veranschaulicht in Abbildung 2.6. Der Verlust wird dabei mittels des L1-Verlust ermittelt, bei dem

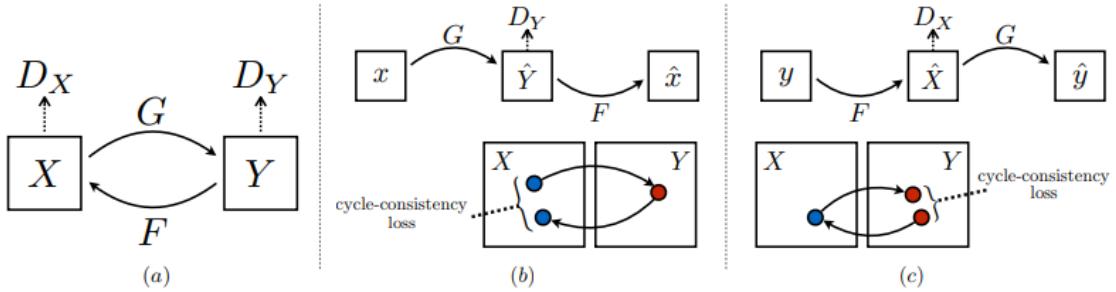


Abbildung 2.5.: (a) Modell des CycleGANs, bestehend aus zwei Generatoren $F : Y \rightarrow X$ und $G : X \rightarrow Y$ und zugehörige adversarielle Diskriminatoren D_X und D_Y , (b) Cycle-Konsistenz $F(G(x)) \approx x$, (c) Cycle-Konsistenz $G(F(y)) \approx y$

die Differenz zwischen den Pixeln des generierten Bildes $G(y)$ und dem Referenzbild $y \in Y$ erfasst wird. Das gleiche Verfahren wird für den anderen Generator F angewendet (**ZPIE**).

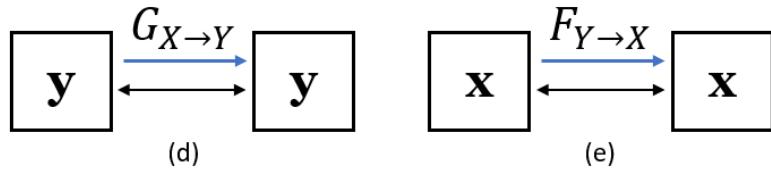


Abbildung 2.6.: Identity-Mapping für (d) Generator G und (e) Generator F

2.3.2. Anwendungen von CycleGAN

CycleGAN hat sich als äußerst vielseitiges Modell erwiesen und findet Anwendung in einer Vielzahl von Bereichen. Insbesondere seine Fähigkeit zur Bildübersetzung zwischen unpaaren Domänen hat zu zahlreichen innovativen Anwendungen geführt. Eine herausragende Nutzung von CycleGAN liegt in der Bild-zu-Bild-Übersetzung und Stilübertragung. Durch dieses Modell können Bilder zwischen verschiedenen Stilen, Szenarien oder Kunstwerken transformiert werden, wodurch die Generierung verschiedener visueller Ästhetiken in einem Bild ermöglicht wird. Diese Anwendung eröffnet kreative Ansätze in der Bildbearbeitung, wie beispielsweise die Umwandlung von Fotografien in den Stil bekannter Kunstwerke (**ZPIE**). Ein weiterer bedeutender Anwendungsbereich von CycleGAN ist die Gesichtsalterung, die in der Gesichtserkennung mit Altersprogression sowie in forensischen Untersuchungen Anwendung finden kann. Die Fähigkeit des Modells, realistische Altersprogressionen in Gesichtsbildern zu erzeugen, stellt einen innovativen Beitrag zu forensischen Methoden dar (**SSJ22**).

In der Stenografie eröffnet CycleGAN ebenfalls interessante Anwendungsmöglichkeiten. Hier kann es genutzt werden, um Satellitenbilder in Karten von Google Maps umzuwandeln und umgekehrt. Diese Anwendung zeigt die Anpassungsfähigkeit des Modells im Umgang mit unterschiedlichen Datenmodalitäten (**CZS**).

2.4. Convolutional Layers

Darüber hinaus spielt CycleGAN eine bedeutende Rolle in der medizinischen Bildverarbeitung, indem es die Möglichkeit bietet, Bilder von einer Bildgebungsmodalität in eine andere zu übersetzen, um die Diagnose zu verbessern. Erfolge wurden bereits in der Synthese von MRT-Bildern des Beckenbereichs zu CT-Bildern erzielt ([LCS⁺21](#)).

Die Flexibilität und Vielseitigkeit von CycleGAN machen es zu einem bedeutenden Werkzeug in der Bildverarbeitung, das innovative Möglichkeiten in verschiedenen Branchen eröffnet. Die fortlaufende Erforschung und Anwendung dieses Modells versprechen weiterhin bedeutende Fortschritte in der digitalen Bildtransformation und -interpretation.

2.4. Convolutional Layers

Convolutional Layers repräsentieren eine fundamentale Komponente innerhalb neuronaler Netzwerke, insbesondere im Kontext der Verarbeitung von Bildinformationen. Diese Schicht nutzt Convolution-Operationen, um durch Faltung von Eingabedaten mit Filterkernen lokale Muster zu identifizieren. Die Filterkerne, üblicherweise in Größen wie 3x3, 7x7 oder 9x9, fungieren als kleine Arrays und dienen der Extraktion spezifischer Merkmale im Bild.

Die Funktionsweise dieser Schicht basiert auf der schrittweisen Verschiebung der Filterkerne über das Eingangsbild. An den jeweiligen Pixelpositionen erfolgt eine präzise elementweise Multiplikation, gefolgt von einer anschließenden Summation aller resultierenden Werte. Diese berechneten Werte werden dann in den korrespondierenden Positionen der Feature Map eingetragen, wie in der Abbildung 2.7 veranschaulicht. Durch diesen Prozess gewinnen tiefere Schichten des Netzwerks die Fähigkeit, zunehmend komplexe und abstrakte Informationen auf höheren Ebenen der Hierarchie zu repräsentieren.

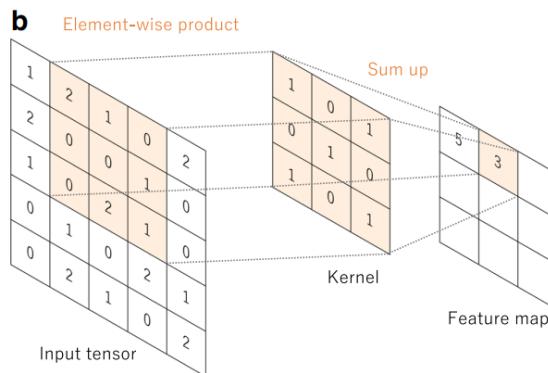


Abbildung 2.7.: Beispiel einer Convolutional-Operation mit einem 3x3 Kernel und Stride 1 ([YNDT18](#))

Die Verwendung verschiedener Kernels, sowohl in Bezug auf Größe als auch Anzahl, erlaubt die Extraktion vielfältiger Merkmale wie Kanten oder Texturen. Diese Flexibilität befähigt das Netzwerk, auf unterschiedlichste visuelle Strukturen ansprechend zu reagieren ([YNDT18](#)).

Zusätzlich fungieren die Kernels als Subsampling-Mechanismus, bedingt durch die begrenzte Ausdehnung der Convolution-Operation bis zum Bildrand. Die Wahl der Strides, als die Distanz zwischen zwei verschobenen Kernelpositionen, beeinflusst diesen Subsampling-Effekt. Die Anwendung von Padding, um das Bild vor jeder Convolution zu vergrößern, kann dazu beitragen, ungewollte Unterabtastung (Downsampling) zu minimieren.

Ein wesentlicher Vorzug von Convolutional Layers besteht in der signifikanten Reduzierung der zu trainierenden Parameter und der Komplexität des Modells. Dies führt zu einer verbesserten Effizienz, da weniger Gewichte optimiert werden müssen, und trägt zur Prävention von Overfitting bei ([YNDT18, ON](#)).

Die Ergebnisse der Convolutional Layer werden anschließend durch eine nicht-lineare Aktivierungsfunktion weitergereicht, um die Fähigkeit des Netzwerks zur Modellierung komplexer, nicht-linearer Zusammenhänge zu verbessern. Die Einbeziehung einer Aktivierungsfunktion, wie beispielsweise der ReLU (Rectified Linear Unit), ermöglicht es dem Netzwerk, nicht-linear separierbare Muster und Merkmale zu erfassen. Ohne Aktivierungsfunktionen würden die Convolutional Layers nur lineare Transformationen durchführen, was die Lernfähigkeit des Modells deutlich einschränken würde ([SSA17](#)).

2.5. Bibliotheken

2.5.1. Tensorflow

TensorFlow ist ein weit verbreitetes Open-Source-Framework, das für maschinelles Lernen und tiefe neuronale Netzwerke eingesetzt wird³. Es bietet eine umfangreiche Plattform zur Entwicklung und Umsetzung von Modellen in verschiedenen Anwendungsbereichen, darunter Bilderkennung und natürliche Sprachverarbeitung. Die Architektur von TensorFlow ermöglicht das Erstellen komplexer maschineller Lernanwendungen und stellt umfassende Tools und Bibliotheken zur Verfügung, um den gesamten Entwicklungsprozess zu unterstützen. Darüber hinaus profitiert TensorFlow von einer aktiven und engagierten Community, die kontinuierlich zur Weiterentwicklung des Frameworks beiträgt⁴.

2.5.2. Keras

Ursprünglich als eigenständiges Framework konzipiert und seit TensorFlow 2.0 in die TensorFlow Core API integriert, fungiert Keras als Open-Source-API zur Modellierung von Strukturen im Bereich des Deep Learning⁵. Diese Bibliothek, die in der Programmiersprache Python implementiert ist, umfasst sämtliche Phasen des maschinellen Lern-Workflows. Beginnend mit der Datenverarbeitung ermöglicht sie eine Fortführung bis zur präzisen Abstimmung der Hyperparameter während

³<https://www.tensorflow.org/>

⁴<https://github.com/tensorflow/tensorflow>

⁵<https://www.tensorflow.org/guide/keras>

2.5. Bibliotheken

des Trainingsprozesses. Die Keras-Prinzipien zeichnen sich durch Einfachheit, Flexibilität und Leistungsfähigkeit aus und ermöglichen es den Anwendern, die Skalierbarkeit und plattformübergreifenden Fähigkeiten der TensorFlow-Plattform zu nutzen⁶.

2.5.3. Matplotlib

Matplotlib ist eine weit verbreitete und leistungsstarke Python-Bibliothek zur Erstellung von statischen, interaktiven und animierten Visualisierungen. Entwickelt von John D. Hunter im Jahr 2003, hat sich Matplotlib zu einem Standardwerkzeug in der Datenvisualisierung und wissenschaftlichen Forschung entwickelt⁷.

⁶<https://keras.io/about/>

⁷<https://matplotlib.org/>

3

Literaturreview

Die Entstehung und Evolution der Generative Adversarial Networks (GANs) markieren für einen Wendepunkt in der Geschichte der Bildverarbeitung und der künstlichen Intelligenz. Das grundlegende Konzept der GANs wurde 2014 von Ian Goodfellow und seinen Kollegen in ihrem Papier "Generative Adversarial Nets" vorgestellt ([Ian14](#)).

In den darauffolgenden Jahren entwickelte sich die GAN-Technologie rasant weiter, weit über die bloße Bildgenerierung hinaus. GANs werden nicht nur für die Generierung von Bildern, sondern auch für die Wiederherstellung von relevanten Informationen in Trainingsdatensätzen sowie für Objekterkennung, Segmentierung und Klassifizierung eingesetzt. Unterschiedliche GAN-Modelle wurden für verschiedene Anwendungsbereiche entwickelt, die jeweils unterschiedliche Stärken aufweisen ([AMB21](#)).

Im Rahmen des Papiers "Image-to-Image Translation with Conditional Adversarial Networks" von Isola et al. (2017) wurde somit das Pix2Pix-Modell vorgestellt, das GANs für die Bild-zu-Bild-Übersetzung nutzt ([PJTA](#)). Ebenso erweiterte das von Zhu et al. (2017) präsentierte CycleGAN-Modell die Möglichkeiten der Bild-zu-Bild-Übersetzung für unpaarige Datensätze, was die Flexibilität von GANs in der Anwendung weiter erhöhte ([ZPIE](#)). Weitere bedeutende Modelle, die in der Forschung Beachtung finden, sind unter anderem das Conditional GAN (CGAN) ([MO](#)), Wasserstein GAN (WGAN) ([ACB17](#)) und Deep Convolutional GAN (DCGAN) ([KSH17](#)).

Trotz dieser Fortschritte bleibt das Training von GANs eine anspruchsvolle Aufgabe, die verschiedene Herausforderungen mit sich bringt. Um diesen zu begegnen, wurden verschiedene Methoden, Techniken und Architekturen vorgeschlagen. Schwierigkeiten wie instabiles Training und Modekollaps, bei dem Modelle dazu tendieren, eingeschränkte Vielfalt zu erzeugen, werden in der Arbeit von Salimans et al. (2016) ([SGZ⁺](#)) zu "Improved Techniques for Training GANs" adressiert, um die Stabilisierung der Netzwerke und die Optimierung des Trainingsprozesses zu unterstützen. Eine weiterführende Studie ([HHYY20](#)) präsentiert diverse GAN-Architekturen und Methoden, um das Problem des Modekollapses zu überwinden. Dabei werden die Architekturen und Methoden variiert, um die Einsatzmöglichkeiten zu erweitern ([JZJ⁺20](#), [EO](#), [SSJ22](#)) und neue Aspekte der GAN-Technologie

zu erforschen (**SVR⁺**).

Im Kontext potenzieller gesellschaftlicher Auswirkungen können GANs erhebliche Herausforderungen im Bereich der Cybersicherheit darstellen, insbesondere hinsichtlich Gesichtserkennung und Passwortentschlüsselung (**HGAPC**). Aggarwal deutete in seiner Arbeit bereits an, dass GANs in Bezug auf die menschliche Bildsynthese Besorgnis erregen könnten (**AMB21**). Diese Bedenken eröffnen neue Forschungsrichtungen für GANs, die darauf abzielen, ethische Standards in der Künstlichen Intelligenz zu wahren.

Die fortlaufende Forschung in diesem Bereich strebt danach, die Leistungsfähigkeit von GANs zu verbessern und neue Perspektiven für deren Nutzung in der Bildverarbeitung und künstliche Intelligenz zu erschließen.

4

Problembeschreibung

Generative Adversarial Networks (GANs) haben in den letzten Jahren erhebliche Aufmerksamkeit in der Forschung und Industrie erlangt. Diese neuartige Klasse von künstlichen neuronalen Netzwerken hat das Potenzial, realistische Daten zu generieren und komplexe Probleme in verschiedenen Domänen zu lösen. Im Rahmen dieser Arbeit liegt der Fokus auf zwei spezifischen GAN-Varianten: Pix2Pix und CycleGAN.

Pix2Pix konzentriert sich auf die direkte Zuordnung zwischen Eingabe- und Ausgabebildern, während CycleGAN die Fähigkeit besitzt, nicht paarweise zugeordnete Datensätze zu übersetzen. Diese Modelle sind auf die Generierung von Bildern ausgerichtet und haben das Potenzial, in verschiedenen Szenarien wie der Stilübertragung, der Bildsegmentierung und der Domänenanpassung verwendet zu werden. Trotz ihrer vielversprechenden Anwendungen gibt es jedoch verschiedene Herausforderungen im Design, in der Implementierung und in der Analyse dieser Modelle. Es stellt sich die Frage, wie die Modelle effektiv gestaltet werden können, um optimale Leistung zu erzielen und welche Strategie am besten geeignet sind, um die Modelle erfolgreich zu trainieren und zu evaluieren.

Herausforderungen im Design

Die Entwicklung von GAN, insbesondere von Modellen wie Pix2Pix und CycleGAN, ist mit zahlreichen Herausforderungen verbunden. Die Wahl der Architektur, die Anpassung der Hyperparameter und die Integration von Regularisierungstechniken sind entscheidende Aspekte, die mit besonderer Sorgfalt angegangen werden müssen. Diese Herausforderungen haben einen großen Einfluss auf die Fähigkeit der Modelle, realistische und generalisierte Ergebnisse zu liefern.

Die Wahl der Architektur spielt eine zentrale Rolle und wirkt sich direkt auf die Fähigkeit des Modells aus, komplexe Transformationen und Generierungsaufgaben durchzuführen. Die Anpassung der Hyperparameter erfordert eine Feinabstimmung, um die Konvergenz des Modells ohne Überanpassung zu gewährleisten. Die Integration von Regularisierungstechniken ist von wesentlicher Bedeutung, um das Modell vor Überanpassung zu schützen und seine Gesamtleistung zu verbessern ([SC21](#)).

Die Auswirkungen dieser Entscheidungen auf die Leistung und Konvergenz der Modelle sind daher nicht unbedeutend und erfordern eine eingehende Analyse, um

sicherzustellen, dass die GANs in der Lage sind, qualitativ hochwertige und realistische Ergebnisse zu liefern und gleichzeitig eine stabile Konvergenz während des Trainings zu gewährleisten.

Herausforderungen in der Implementierung

Die Auswahl geeigneter Datensätze, der Umgang mit Datenungleichgewichten, die Optimierung der Trainingsparameter und die Vermeidung von Overfitting sind wichtige Schritte bei der Implementierung von Pix2Pix und CycleGAN. Dies umfasst die Datenaufbereitung, das Training und die Evaluierung der Modelle.

Die Auswahl des Datensatzes hat einen großen Einfluss auf die Fähigkeit des Modells, realistische Ergebnisse zu liefern. Dabei ist nicht nur die Menge, sondern auch die Vielfalt der Daten von Bedeutung. Auch der Umgang mit Ungleichgewichten in den Daten ist von hoher Relevanz, um sicherzustellen, dass das Modell nicht in Richtung bestimmter Merkmale verzerrt wird ([YNDT18](#)).

Die Optimierung der Trainingsparameter, einschließlich der Lernraten und der Batchgrößen, ist ein Feinabstimmungsprozess, um eine stabile Konvergenz des Modells zu gewährleisten ([SC21](#)). Gleichzeitig ist es von entscheidender Notwendigkeit, Overfitting durch die Implementierung geeigneter Regularisierungstechniken zu vermeiden ([YNDT18](#)).

Das Verständnis und die zielgerichtete Bewältigung dieser Implementierungsherausforderungen sind entscheidend, um sicherzustellen, dass Pix2Pix und CycleGAN effektiv in verschiedenen Anwendungsbereichen eingesetzt werden können. Durch eine gründliche Untersuchung dieser Aspekte können Modelle entwickelt werden, die nicht nur leistungsfähig, sondern auch robust und generalisierbar sind.

Herausforderungen in der Analyse

Die Analyse von Pix2Pix und CycleGAN umfasst mehrere Schlüsselaspekte, die zum Verständnis der Leistung und Zuverlässigkeit dieser GANs beitragen. Dazu gehören die Bewertung der Generierungsfähigkeiten, die Quantifizierung von Artefakten und die Untersuchung von Konvergenzproblemen.

Die Evaluierung umfasst qualitative Bewertungen der von den Modellen erzeugten Bilder. Dazu können visuelle Inspektionen und Vergleiche mit den Originaldaten durchgeführt werden. Unerwünschte Muster und Unvollkommenheiten können ebenfalls visuell oder mit Hilfe von Metriken identifiziert werden. Dazu gehören Messungen wie der Mode Score (MS), SSIM-Score oder Interception Score (IS) ([PYY⁺19](#)).

Die Analyse kann ferner die Überwachung von Verlustkurven umfassen, um Konvergenzprobleme während des Trainings zu untersuchen.

Analyse und Bewertung sind entscheidend, um die Modelle weiter zu verbessern, ihre Anwendbarkeit zu erweitern und sicherzustellen, dass sie ihren Zweck erfüllen.

5

Lösungsbeschreibung

5.1. Training- und Testdaten

Das Training und die Evaluierung von Generative Adversarial Networks (GANs) erfordern die klare Definition von Trainings- und Testdatensätzen. Der entscheidende Unterschied zwischen diesen Datensätzen besteht darin, dass das Modell während des Trainings auf den Trainingsdaten optimiert wird, während die Testdaten verwendet werden, um die Leistung und die Generalisierungsfähigkeiten des Modells zu bewerten.

5.1.1. Datenladung für GAN-Training

Für das effektive Training von GANs ist der Zugriff auf qualitativ hochwertige Datensätze von entscheidender Bedeutung. In diesem Kontext bietet TensorFlow eine umfassende Sammlung öffentlich verfügbarer Datensätze. Die verwendeten Datensätze werden von der Quelle <https://efrosgans.eecs.berkeley.edu> heruntergeladen und lokal extrahiert.

```
1 path_to_zip = tf.keras.utils.get_file(fname=f'{dataset_name}.tar.gz',
2 origin=_URL, extract=True)
```

Code-Auszug 5.1: Laden eines Datensatzes von einer URL

Die Transformation und Vorverarbeitung der Bilddaten erfolgt durch die TensorFlow-Datensatz-API. Diese API bietet eine effiziente Datenpipeline für das Laden und Verarbeiten von Daten, insbesondere für den Einsatz in Machine-Learning-Modellen. In den folgenden Implementierungen werden die Datensätze durch eine Liste von Dateipfaden als Zeichenketten erzeugt.

```
1 train_horses = tf.data.Dataset.list_files(str(PATH / 'trainA/*.jpg'))
```

Code-Auszug 5.2: Erzeugung eines Tensorflow-Dataset aus der CycleGAN Implementierung

5.1.2. Vorverarbeitung des Datensatzes

Um die Leistung von GAN-Modellen zu optimieren, werden vor dem Training Variationen in den Trainingsdaten eingeführt. Dieser Prozess umfasst Datenjittering und Normalisierung. Durch die Integration von Variationen wird das Modell robuster, da es eine erhöhte Invarianz gegenüber unterschiedlichen Eingabedaten entwickelt. Dies trägt wesentlich dazu bei, eine verbesserte Konvergenz während des Trainings zu erreichen.

Bei Pix2Pix bestehen die Trainingsdaten aus einem Paar von Eingabe- und Zielbildern, während bei CycleGAN unpaare Daten berücksichtigt werden. Um eine konsistente Skalierung mit der Tanh-Aktivierungsfunktion sicherzustellen, werden diese Bilder im Bereich von -1 bis +1 skaliert. Diese Normalisierung ist von entscheidender Bedeutung für die Stabilisierung des Trainingsprozesses. Durch die Bereitstellung eines standardisierten Datensatzes kann das Modell effektiver mit einer verbesserten Lernrate und Konvergenzgeschwindigkeit arbeiten ([RMC](#)).

```

1 def normalize(image):
2     image = (image / 127.5) - 1
3     return image

```

Code-Auszug 5.3: Vorverarbeitung des Datensatzes: Normalisierung

Datenjittering bezieht sich auf die Einführung von zufälligen Variationen oder Veränderungen in den Trainingsdaten, was in der Implementierung durch zufälliges Zuschneiden und eine zufällige Spiegelung erreicht wird. Die heruntergeladenen Bilder mit einer Auflösung von 256x256 werden zuerst auf eine größere Größe von 286x286 skaliert, wobei die Nearest-Neighbor-Methode verwendet wird.

```

1 def random_jitter(image):
2     image = tf.image.resize(image, [286, 286], method=tf.image.
                           ResizeMethod.NEAREST_NEIGHBOR)
3     image = tf.image.random_crop(image, size=(286, 286, 3))
4     image = tf.image.random_flip_left_right(image)
5     return image

```

Code-Auszug 5.4: Vorverarbeitung des Datensatzes: Jittering

Diese Methode skaliert Bilder einfach und effizient, indem sie für jedes Pixel im skalierten Bild den Farbwert des nächstgelegenen Pixels im Originalbild übernimmt¹. Nach dem Skalieren wird das Bild zufällig auf die Originalgröße reduziert und gleichzeitig mit einer Wahrscheinlichkeit von 50% gespiegelt. Anschließend werden

¹https://en.wikipedia.org/wiki/Nearest-neighbor_interpolation

die Bildpfade mittels der *load_image*-Funktion geladen und in das resultierende JPEG-Format decodiert (Code A.1).

Die vorverarbeiteten Trainings- und Testbilder werden darauf in TensorFlow-Datasets integriert. Nachfolgend wird der Trainingsdatensatz zufällig gemischt und in Batches gruppiert, wodurch sichergestellt wird, dass das Modell nicht von der Reihenfolge der Datenpunkte beeinflusst wird.

```
1 def preprocess_image_train(image_path):
2     image = load_image(image_path)
3     image = random_jitter(image)
4     image = normalize(image) if not tf.reduce_all(tf.math.logical_and(
5         image >= 0.0, image <= 1.0)) else image
6     return image
7 train_dataset = tf.data.Dataset.list_files(str(PATH / 'trainA/*.*'))
8 train_dataset = train_dataset.map(preprocess_image_train,
9     num_parallel_calls=tf.data.AUTOTUNE)
9 train_dataset = train_dataset.shuffle(BUFFER_SIZE)
10 train_dataset = train_dataset.batch(BATCH_SIZE)
```

Code-Auszug 5.5: Integration der vorverarbeiteten Trainingsbilder in Tensorflow-Datasets (CycleGAN Implementierung)

Diese umfassende Vorverarbeitung stellt sicher, dass das GAN-Modell auf optimal vorbereiteten Daten trainiert wird, um eine maximale Leistung und Generalisierungsfähigkeit zu erreichen.

5.2. Implementierung der Pix2PixGAN-Architektur

5.2.1. Generator

Die Struktur des Generator in der Pix2Pix-Implementierung ist ein wesentlicher Aspekt, der die Leistungsfähigkeit des Modells bestimmt. Der Generator ist als U-Net-Architektur aufgebaut, die aus dem Encoder und Decoder bestehen die wiederum aufeinanderfolgende Downsampling- und Upsampling-Schritten beinhaltet. Im Encoder-Teil des Generators wird das Downsampling durch eine Reihe von Convolutional Neuronal Network (CNN) Schichten realisiert, die durch die downsample-Funktion innerhalb des downstack definiert sind (Code 5.6).

```

1      down_stack = [
2          downsample(64, 4, apply_batchnorm=False), # (batch_size, 128,
3              128, 64)
4          downsample(128, 4), # (batch_size, 64, 64, 128)
5          downsample(256, 4), # (batch_size, 32, 32, 256)
6          downsample(512, 4), # (batch_size, 16, 16, 512)
7          downsample(512, 4), # (batch_size, 8, 8, 512)
8          downsample(512, 4), # (batch_size, 4, 4, 512)
9          downsample(512, 4), # (batch_size, 2, 2, 512)
10         downsample(512, 4), # (batch_size, 1, 1, 512)
11     ]

```

Code-Auszug 5.6: Downsampling-Schritt in Pix2Pix

Die downsample-Funktion erstellt eine Downsampling-Schicht, die mittels einer Conv2D-Schicht mit spezifischen Filtern und Kernel-Größen die räumlichen Dimensionen der Eingabebilder reduziert. Zur Verbesserung der Stabilität und Leistungsfähigkeit des Modells integriert die Funktion optional eine Batch-Normalisierung. Diese Normalisierung reguliert und standardisiert die Ausgabe der Conv2D-Schicht, was dazu beiträgt, das Training effizienter zu gestalten. Darüber hinaus beinhaltet die Funktion eine LeakyReLU-Aktivierung, eine Variation der herkömmlichen ReLU-Aktivierungsfunktion. LeakyReLU ermöglicht es, dass auch für negative Eingabewerte ein kleiner Gradient erhalten bleibt, wodurch das Problem der inaktiven Neuronen, bekannt als "sterbende ReLUs", vermieden wird (Code 5.7).

Im Anschluss daran erfolgt das Upsampling im Decoder-Teil des Generators, das durch die upsample-Funktion innerhalb des upstack repräsentiert wird. Diese Schichten arbeiten daran, die Merkmale auf ein höher ausgelöstes Format zu projizieren und die Bildgröße wiederherzustellen (Code 5.8).

Die upsampling-Funktion verwendet eine spezielle Art von Convolutional Layer, die Conv2DTranspose-Schicht, um die Bildgröße zu erhöhen. Diese Schicht kehrt den Prozess einer Convolutional Layer Schicht um, indem sie die Eingabedaten

5.2. Implementierung der Pix2PixGAN-Architektur

expandiert, was für das Wiederherstellen einer größerem Bildgröße im Generator unerlässlich ist. Zusätzlich zur Conv2DTranspose-Schicht integriert die upsample-Funktion eine Batch-Normalisierung, die zur Stabilisierung des Lernprozesses beiträgt, indem sie die Ausgaben der Conv2DTranspose-Schicht normalisiert (Code 5.9).

Dies ist ein wichtiger Schritt, um die interne Kovariantenverschiebung zu reduzieren und die Leistung des Modells zu verbessern. Ein weiteres wichtiges Merkmal der Funktion ist die optional Anwendung von Dropout um Overfitting zu vermeiden. Dies trägt dazu bei dass das Modell robustere und generalisierbare Merkmale lernt. Schließlich wird eine ReLU-Aktivierungsfunktion angewendet, die dafür sorgt, dass das Modell nicht-lineare Zusammenhänge lernt.

Die Skip-Verbindungen werden im Generator durch die Speicherung und spätere Verwendung der Ausgaben der Downsampling-Schichten in der skips-Liste realisiert. Nach dem Downsampling-Prozess werden diese gespeicherten Ausgaben in umgekehrter Reihenfolge durchlaufen und mit den Ausgaben der Upsampling-Schichten mittels einer Concatenate-Operation verbunden. Diese Kombination von hoch- und niedrigstufigen Merkmalen führt zu einer detaillierteren und genaueren Bildrekonstruktion.

Schließlich wird das endgültige Bild durch die letzte Schicht des Generators erzeugt, eine Conv2DTranspose-Schicht, die die Ausgabe des Generators darstellt. Diese letzte Schicht spielt eine entscheidende Rolle bei der Erzeugung des endgültigen Bildes, das die kombinierten Merkmale aus dem gesamten Netzwerk nutzt (Code 5.10).

5.2.2. Diskriminator

Der Diskriminator im Pix2Pix-Modell ist eine tiefgehende neuronale Netzarchitektur, die zwei Eingabebilder - ein Eingabebild und ein Zielbild - verarbeitet. Diese Bilder werden entlang ihrer Farbkanäle zu einem einzigen Bild kombiniert. Anschließend durchläuft das Bild eine Reihe von Downsampling-Schritten mittels Convolutional Layern, um die räumliche Auflösung zu reduzieren und wichtige Bildmerkmale zu extrahieren. Nach mehreren Downsampling-Phasen folgt eine entscheidende Verarbeitungsschicht, die die Bildmerkmale weiter analysiert und normalisiert. Der finale Output des Diskriminators erfolgt durch einen spezialisierten Convolutional Layer, der eine Patch-basierte Bewertung des Bildes vornimmt. Diese Patch-basierte Analyse ist zentral für die Funktionsweise des Diskriminators, indem sie es ihm ermöglicht, die Echtheit jedes Bildbereichs unabhängig zu bewerten, was ein Schlüsselement der PatchGAN-Architektur ist. Insgesamt handelt es sich um ein komplexes System zur Bildanalyse, das speziell für die Aufgabe der Bild-zu-Bild-Übersetzung entwickelt wurde (Code 5.11).

5.2.3. Verlustfunktion

Generatorverlust

Der Generatorverlust im Pix2Pix-Modell besteht aus zwei wesentlichen Komponenten: dem adversariellen Verlust (*gan_loss*) und dem L1-Verlust (*l1_loss*). Der adversarielle Verlust wird durch Anwendung der *BinaryCrossentropy*-Funktion von TensorFlow ermittelt. Hierbei wird die Ausgabe des Diskriminators für generierte Bilder (*disc_generated_output*) mit einem Tensor, der ausschließlich aus Einsen besteht, verglichen. Dies dient dazu, die Effektivität des Generators bei der Erzeugung von Bildern zu bewerten, die für den Diskriminator von echten Bildern nicht unterscheidbar sind. Der L1-Verlust hingegen berechnet den mittleren absoluten Fehler zwischen dem vom Generator erzeugten Bild (*gen_output*) und dem tatsächlichen Zielbild (*target*). Dieser Verlust trägt maßgeblich dazu bei, die inhaltliche Übereinstimmung und Ähnlichkeit des generierten Bildes mit dem Zielbild zu fördern.

Der Gesamtverlust des Generators (*total_gen_loss*) ist die Summe des adversariellen Verlust und des L1-Verlusts, wobei der L1-Verlust mit einem Faktor λ gewichtet wird. Die Gewichtung des L1-Verlusts hilft dabei, die strukturelle Integrität und die Genauigkeit des generierten Bildes zu verbessern, indem sie darauf abzielt, die Pixeldifferenzen zwischen dem generierten Bild und dem Zielbild zu minimieren.

Nach der Berechnung des Gesamtverlusts werden Gradienten bezüglich der Generatorparameter berechnet (*gen_tape.gradient*) und diese Gradienten werden dann verwendet, um den Generator mittels des Adam-Optimierers

(*generator_optimizer.apply_gradients*) zu aktualisieren. Dieser Prozess ist ein integraler Bestandteil des Trainings, da er dem Generator hilft, sich schrittweise zu verbessern und immer realistischere Bilder zu erzeugen.

Diskriminatorverlust

Der Diskriminatorverlust wird durch die *discriminator_loss*-Funktion im Code bestimmt. Diese Funktion enthält zwei Eingaben: *disc_real_output*, die Diskriminatierausgabe für das echte Bild und *disc_generated_output*, die Diskriminatierausgabe für das vom Generator erzeugte Bild. Der Verlust für echte Bilder (*real_loss*) wird berechnet, indem die *BinaryCrossentropy*-Funktion zwischen *disc_real_output* und einem Tensor aus Einsen angewendet wird. Dieser Schritt bewertet, wie gut der Diskriminator echte Bilder als solche erkennen kann. Der Verlust für generierte Bilder (*generated_loss*) wird berechnet, indem die *BinaryCrossentropy*-Funktion zwischen *disc_generated_output* und einem Tensor aus Nullen angewendet wird. Dies bewertet, wie gut der Diskriminator generierte Bilder als falsch erkennen kann.

Der Gesamtverlust des Diskriminators (*total_disc_loss*) ist die Summe von *real_loss* und *generated_loss*. Diese Kombination zwingt den Diskriminator, zwischen echten und generierten Bildern besser zu unterscheiden.

Für die Optimierung des Diskriminators werden die Gradienten des Diskriminators

5.3. Implementierung der CycleGAN-Architektur

verlusts in Bezug auf die Diskriminatorparameter berechnet (*disc_tape.gradient*). Diese Gradienten werden dann verwendet, um den Diskriminator mittels des Adam-Optimierers (*discriminator_optimizer.apply_gradients*) zu aktualisieren.

5.3. Implementierung der CycleGAN-Architektur

Nach den theoretischen Grundlagen der CycleGAN-Architektur und der Datenvorverarbeitung in den vorherigen Abschnitten, wird nun die konkrete Implementierung der Architekturen unter Verwendung von TensorFlow und Keras vorgestellt.

5.3.1. Generator und Diskriminator

Die Architekturen des Generators und Diskriminators wurden gemäß den theoretischen Grundlagen umgesetzt, insbesondere den Richtlinien von Zhu et al. (2017) (ZPIE).

Der Generator besteht aus einem Encoder-Block, gefolgt von sechs Residual-Blöcken und einem Decoder-Block, welche in Abbildung 2.3 dargestellt ist. Der Diskriminator wurde als sequentielles Modell implementiert und umfasst mehrere Convolutional Schichten, konzipiert als PatchGAN.

Nach jeder Convolutional Schicht im Generator und Diskriminator, abgesehen von der letzten, wird eine Instanznormalisierung und eine ReLU-Aktivierung angewendet. Die Instanznormalisierung dient dazu, die Aktivierungen zu normalisieren und das Training zu stabilisieren, indem sie die Eingaben jedes Minibatches normalisiert. Die ReLU-Aktivierung fördert die Einführung von Nichtlinearitäten in das Modell und ermöglicht es, komplexere Merkmale zu erfassen.

Für die Output-Schicht wird die tanh-Aktivierungsfunktion angewendet. Diese Schicht begrenzt die Ausgabewerte auf den Bereich zwischen -1 und +1. Diese Begrenzung ist nicht nur wichtig, um eine konsistente Skalierung mit den Trainingsdaten sicherzustellen, sondern trägt auch zur Stabilisierung des Trainingsprozesses bei (RMC).

Die spezifischen Implementierungsdetails, inklusive der Helferfunktionen, sind im beigefügten Code zu finden.

5.3.2. Verlustfunktion

Während des Trainings werden verschiedene Verlustfunktionen verwendet, um sicherzustellen, dass der Generator qualitativ hochwertige Bilder generiert und dass die Transformationen zwischen den Domänen konsistent sind. Die zentralen Verlustfunktionen, insbesondere die Gesamtverlustfunktionen für den Generator und den Diskriminator, werden im Folgenden erläutert.

Für die Klassifizierung, ob es sich um echte oder generierte Bilder handelt, wird in der Implementierung der *BinaryCrossentropy*-Verlustfunktion aus TensorFlow/Keras verwendet. Dieser berechnet den binären Kreuzentropieverlust zwischen den Zielwerten und den Vorhersagen².

Gesamtverlust des Generators

Der Gesamtverlust des Generators setzt sich aus dem adversiellen Verlust und dem Zykluskonsistenz-Verlust zusammen. Optional kann der Identitätsverlust berücksichtigt werden, was zu einem konsistenteren Transformationsprozess führt (siehe Implementierung 5.15). Die Integration des Identitätsverlusts gewährleistet die Bewahrung der Struktur des Originalbildes.

Im adversiellen Verlust, implementiert durch die Funktion *generator_adversarial_loss*, werden die generierten Bilder mit einem Tensor aus Einsen verglichen, welcher die Zielwerte für 'echt' repräsentiert. Die Verlustberechnung erfolgt mittels einer spezifizierten Verlustfunktion.

Für den Zykluskonsistenz-Verlust (*cycle_loss*-Funktion) wird der mittlere absolute Unterschied zwischen einem echten Bild und seiner zyklisch transformierten Version berechnet. Dieser L1-Verlust stellt sicher, dass die Übersetzung zwischen den Domänen und zurück nahe an der Identitätsabbildung liegt.

Die Funktion *identity_loss* ermittelt den Identitätsverlust, wobei der L1-Verlust zwischen einem echten Bild und seiner übersetzten Version in derselben Domäne verwendet wird.

Zur Berücksichtigung ihrer Bedeutung für den Gesamtverlust wird der Verlust an Zykluskonsistenz und Identität jeweils mit einem gewichteten Faktor multipliziert. Der Code-Anhang enthält detaillierte Implementierungen der genannten Verlustfunktionen.

²https://www.tensorflow.org/api_docs/python/tf/keras/losses/BinaryCrossentropy

Gesamtverlust des Diskriminators

Die Gesamtverlustfunktion des Diskriminators setzt sich aus dem adversariellen Verlust für echte und generierte Bilder zusammen. Der Diskriminator wird dementsprechend trainiert, echte Bilder als Einsen und generierte Beispiele als Nullen zu klassifizieren.

Zu Beginn wird der Verlust berechnet, wenn der Diskriminator echte Bilder betrachtet. Hierbei kommt die BinaryCrossentropy-Verlustfunktion zum Einsatz, die den Verlust zwischen den echten Vorhersagen (*real*) und den Zielwerten berechnet. Anschließend erfolgt die Berechnung des Verlusts zwischen den generierten Vorhersagen (*generated*) und den Zielwerten.

Der Gesamtverlust ergibt sich als Summe der beiden Teilverluste. Um sicherzustellen, dass die Gradientenaktualisierung während des Trainings angemessen skaliert wird, erfolgt eine Multiplikation des Gesamtverlustes mit dem Wert 0.5, was einer Bildung des Durchschnitts des Gesamtverlustes entspricht.

5.4. Training und Hyperparameter

Während des Trainings von Generative Adversarial Networks (GANs) wird eine iterative Methode angewendet, bei der der Generator und der Diskriminator abwechselnd trainiert werden. Dieser Trainingszyklus erstreckt sich über mehrere Epochen, wobei in jeder Epoche Bilder durch einen Feed-Forward-Prozess vom Generator generiert und vom Diskriminator bewertet werden. Gleichzeitig wird der Diskriminator mit den echten Zielbildern trainiert, um seine Fähigkeit zu verbessern, zwischen echten und generierten Bildern zu unterscheiden. Der Trainingsschritt beinhaltet die Berechnung und Anwendung von Gradienten für sowohl den Generator als auch den Diskriminator, basierend auf ihren jeweiligen Verlustfunktionen (Code A.4, A.11, A.12).

Die Wahl der Hyperparameter, einschließlich der Anzahl der Epochen, Lernraten der Optimizer und der *LAMBDA*-Werte der jeweiligen Architekturen, wurde durch abgestimmte Experimente ermittelt. Dabei wurde darauf geachtet, dass die gewählten Werte zu einer stabilen und konvergenten Schulung führen.

Um Überanpassung entgegenzuwirken, wird auf das Training mit umfangreichen Datensätzen zurückgegriffen. Die Integration zusätzlicher Regularisierung in Form von Dropout trägt dazu bei, die Generalisierungsfähigkeit des Modells zu steigern und seine Robustheit zu verbessern. Beim Dropout-Mechanismus werden während des Trainingsprozesses zufällig bestimmte Neuronen deaktiviert, was dazu führt, dass das Netzwerk nicht übermäßig abhängig von spezifischen neuronalen Pfaden wird. Diese Strategie fördert die Robustheit des Modells, indem sie es dazu zwingt, mit dem Fehlen bestimmter Neuronen umzugehen und alternative Pfade zu nutzen. Durch die erzwungene Redundanz werden verschiedene Teile des Netzwerks aktiviert, was zu einer breiteren Erfassung von Merkmalen und einer vielfältigeren Repräsentation führt. Dies ist entscheidend, um sicherzustellen, dass das Modell nicht zu stark auf spezifische Trainingsdaten reagiert, sondern stattdessen Merkmale erlernt, die auf einer breiteren Palette von Eingaben generalisiert werden können ([YNDT18](#)).

5.4.1. Optimierungstechnik und Optimizers

Als Optimierungstechnik wird das Gradientenabstiegsverfahren angewendet, bei dem die Parameter in Richtung des negativen Gradienten der Verlustfunktion aktualisiert werden, um den Verlust zu minimieren. Die Parameter beider Modelle werden mittels korrespondierenden Optimizern aktualisiert. Die Lernrate der Optimizer bestimmt die Geschwindigkeit, mit der das Modell lernt, und sie wird sorgfältig abgestimmt, um eine stabile und konvergente Schulung sicherzustellen. In der Implementierung wird der Adam-Optimizer verwendet, eine gängige Wahl in der GAN-Literatur. Dieser Optimizer ist eine verbesserte Variante des stochastischen Gradientenabstiegsverfahrens und passt die Lernraten für jede Variable

dynamisch an. Dadurch wird die Konvergenz des Trainingsprozesses beschleunigt und das Modell wird robuster gegenüber unterschiedlichen Lernraten (KB).

5.4.2. Fortschrittsüberwachung und Visualisierung

Um inkrementelle Verbesserungen nach jeder Iteration zu visualisieren, wird während des Trainings ein Bild aus dem Trainingsdatensatz ausgewählt, auf das der Generator angewendet wird. Die Eingabebilder und die generierten Bilder werden mithilfe der Matplotlib- und TensorFlow-Bibliotheken gespeichert, um sie nach dem Training zugänglich und vergleichbar zu machen.

Zusätzlich werden die Metriken zur Evaluation der Modelle, wie der Verluste des Generators und des Diskriminators, sowie die Genauigkeit des Diskriminators und der SSIM-Score nach jeder Epoche in einer CSV-Datei festgehalten (Code A.13). Dieser Ansatz ermöglicht eine visuelle Überprüfung des Trainingsverlaufs und erleichtert die Identifikation von möglichen Verbesserungsbereichen. Darüber hinaus ermöglicht die Datei die spätere Erstellung von Verlaufskurven für eine eingehendere Analyse (Code A.15).

Zur Initialisierung der Metriken kommen TensorFlow-Keras-Metriken zum Einsatz:

Diese Metriken werden verwendet, um den durchschnittlichen Verlust des Generators, den durchschnittlichen Verlust des Diskriminators und die Genauigkeit des Diskriminators zu erfassen.

Während des Trainingsprozesses wird für jeden Batch von Trainingsbildern der Verlust und die Genauigkeit berechnet. Die Verluste werden durch die oben genannten Metriken erfasst und nach jeder Epoche in die CSV-Datei aufgezeichnet. Die Genauigkeit des Diskriminators wird durch die BinaryAccuracy-Metrik überwacht, wobei ein Schwellenwert von 0,5 festgelegt ist, was in binären Klassifikationsaufgaben typisch ist.

Die Berechnung und Aktualisierung der Metriken erfolgen wie folgt:

Der SSIM-Score kann mithilfe der TensorFlow-Funktion `tf.image.ssim()` berechnet werden (Code A.14).

Diese systematischen Vorgehensweisen ermöglichen eine umfassende Überprüfung der Modellleistung und bieten visuelle Einblicke in den Fortschritt während des Trainingsprozesses.

```

1 def downsample(filters, size, apply_batchnorm=True):
2     initializer = tf.random_normal_initializer(0., 0.02)
3     result = tf.keras.Sequential()
4     result.add(
5         tf.keras.layers.Conv2D(filters, size, strides=2, padding='same',
6                               kernel_initializer=initializer, use_bias=False))
7     if apply_batchnorm:
8         result.add(tf.keras.layers.BatchNormalization())
9
10    result.add(tf.keras.layers.LeakyReLU())
11
12    return result

```

Code-Auszug 5.7: Downsampling-Schicht in Pix2Pix

```

1 up_stack = [
2     upsample(512, 4, apply_dropout=True), # (batch_size, 2, 2,
3             1024)
4     upsample(512, 4, apply_dropout=True), # (batch_size, 4, 4,
5             1024)
6     upsample(512, 4, apply_dropout=True), # (batch_size, 8, 8,
7             1024)
8     upsample(512, 4), # (batch_size, 16, 16, 1024)
9     upsample(256, 4), # (batch_size, 32, 32, 512)
10    upsample(128, 4), # (batch_size, 64, 64, 256)
11    upsample(64, 4), # (batch_size, 128, 128, 128)
12 ]

```

Code-Auszug 5.8: Upsampling-Schritt in Pix2Pix

```

1 def upsample(filters, size, apply_dropout=True):
2     initializer = tf.random_normal_initializer(0., 0.02)
3     result = tf.keras.Sequential()
4     result.add(
5         tf.keras.layers.Conv2DTranspose(filters, size, strides=2,
6                                         padding='same', kernel_initializer=initializer,
7                                         use_bias=False))
8     result.add(tf.keras.layers.BatchNormalization())
9
10    if apply_dropout:
11        result.add(tf.keras.layers.Dropout(0.5))
12
13    result.add(tf.keras.layers.ReLU())
14
15    return result

```

Code-Auszug 5.9: Upsampling-Schritt in Pix2Pix

```

1     initializer = tf.random_normal_initializer(0., 0.02)
2     last = tf.keras.layers.Conv2DTranspose(
3         OUTPUT_CHANNELS, 4, strides=2, padding='same',
4         kernel_initializer=initializer, activation='tanh') # (
5             batch_size, 256, 256, 3)
6     x = inputs
7
8     # Downsampling through the model
9     skips = []
10    for down in down_stack:
11        x = down(x)
12        skips.append(x)
13    skips = reversed(skips[:-1])
14
15    # Upsampling and establishing the skip connections
16    for up, skip in zip(up_stack, skips):
17        x = up(x)
18        x = tf.keras.layers.concatenate([x, skip])
19    x = last(x)
20
21    return tf.keras.Model(inputs=inputs, outputs=x)

```

Code-Auszug 5.10: Skip Verbindungen in Pix2Pix

 5.4. Training und Hyperparameter

```

1 def Discriminator():
2     inp = tf.keras.layers.Input(shape=[256, 256, 3], name='
3         input_image')
4     tar = tf.keras.layers.Input(shape=[256, 256, 3], name='
5         target_image')
6
7     x = tf.keras.layers.concatenate([inp, tar]) # (batch_size, 256,
8                                     256, channels*2)
9
10    down1 = downsample(64, 4, False)(x) # (batch_size, 128, 128,
11                                     64)
12    down2 = downsample(128, 4)(down1) # (batch_size, 64, 64, 128)
13    down3 = downsample(256, 4)(down2) # (batch_size, 32, 32, 256)
14    down4 = downsample(512,4)(down3)
15
16    # Zweite letzte Ausgabeschicht
17    initializer = tf.random_normal_initializer(0., 0.02)
18    last = tf.keras.layers.Conv2D(512, 4, padding='same',
19                                kernel_initializer=initializer)(down4)
20    last = tf.keras.layers.BatchNormalization()(last)
21    last = tf.keras.layers.LeakyReLU()(last)
22
23    # Patch-Ausgang
24    patch_out = tf.keras.layers.Conv2D(1, 4, padding='same',
25                                    kernel_initializer=initializer)(last)
26    patch_out = tf.keras.layers.Activation('sigmoid')(patch_out)
27
28    return tf.keras.Model(inputs=[inp, tar], outputs=patch_out)

```

Code-Auszug 5.11: Diskriminator Pix2Pix

```

1 def Generator():
2     inputs = tf.keras.layers.Input(shape=[None,None,3])
3
4     # Layer 1
5     x = layers.Conv2D(64, 7, strides=1, padding='same')(inputs)
6     x = layers.BatchNormalization()(x)
7     x = layers.Activation('relu')(x)
8
9     # Layer 2 + 3
10    x = convolutional_layer(x, 128, 3,2)
11    x = convolutional_layer(x, 256, 3, 2)
12
13    # Residual Blocks
14    for _ in range(6):
15        x = residual_block(x, 256)
16
17    # Layer 10 + 11
18    x = t_convolutional_layer(x, 128, 3, 2)
19    x = t_convolutional_layer(x, 64, 3, 2)
20
21    # Dropout
22    x = tf.keras.layers.Dropout(0.5)(x)
23
24    # Layer 12
25    outputs = layers.Conv2D(3, 7, strides=1, padding='same',
26                           activation='tanh')(x)
27
28    return tf.keras.Model(inputs=inputs, outputs=outputs)

```

Code-Auszug 5.12: CycleGAN Generator in Tensorflow

```

1 def Discriminator():
2     model = tf.keras.Sequential()
3
4     # Layer 1
5     model.add(layers.Conv2D(64, 4, strides=2, padding='same',
6         input_shape=(256,256,3)))
7     model.add(layers.BatchNormalization())
8     model.add(layers.Activation('relu'))
9
10    # Layer 2
11    model.add(layers.Conv2D(128, 4, strides=2, padding='same'))
12    model.add(layers.BatchNormalization())
13    model.add(layers.Activation('relu'))
14
15    # Layer 3
16    model.add(layers.Conv2D(256, 4, strides=2, padding='same'))
17    model.add(layers.BatchNormalization())
18    model.add(layers.Activation('relu'))
19
20    # Layer 4
21    model.add(layers.Conv2D(512, 4, strides=2, padding='same'))
22    model.add(layers.BatchNormalization())
23    model.add(layers.Activation('relu'))
24
25    # Layer 5
26    model.add(layers.Conv2D(1, 4, strides=1, padding='same'))
27    model.add(layers.BatchNormalization())
28    model.add(layers.Activation('sigmoid'))
29
30    # Dropout
31    model.add(tf.keras.layers.Dropout(0.5))
32
33    return model

```

Code-Auszug 5.13: CycleGAN Diskriminator in Tensorflow

```

1 loss_obj = tf.keras.losses.BinaryCrossentropy(from_logits=True)

```

Code-Auszug 5.14: Initialisierung des BinaryCrossentropy-
Verlustfunktion

```

1 def generator_loss(real_x, real_y, cycled_x, cycled_y, disc_fake,
2                     identity):
3     gen_loss = generator_adversarial_loss(disc_fake)
4     total_cycle_loss = cycle_loss(real_x, cycled_x) + cycle_loss(real_y,
5                                   cycled_y)
6     id_loss = identity_loss(real_y, identity)
7     total_gen_loss = gen_loss + total_cycle_loss + id_loss
8     return total_gen_loss

```

Code-Auszug 5.15: Gesamtverlust des Generators in CycleGAN

```

1 def discriminator_adversarial_loss(real, generated):
2     real_loss = loss_obj(tf.ones_like(real), real)
3     generated_loss = loss_obj(tf.zeros_like(generated), generated)
4     total_disc_loss = real_loss + generated_loss
5     return total_disc_loss * 0.5

```

Code-Auszug 5.16: Gesamtverlust des Diskriminators in CycleGAN

```

1 generator_optimizer = tf.keras.optimizers.Adam(learning_rate=0.0002,
2                                               beta_1=0.5)
3 discriminator_optimizer = tf.keras.optimizers.Adam(learning_rate
4 =0.0002, beta_1=0.5)

```

Code-Auszug 5.17: Initialisierung der Adam-Optimizers aus der Pix2Pix Implementierung

```

1 def generate_images(model, test_input, tar, epoch):
2     prediction = model(test_input, training=True)
3     tf.keras.preprocessing.image.save_img(os.path.join(save_path,
4                                                 'input_image.png'), test_input[0] * 0.5 + 0.5)
4     tf.keras.preprocessing.image.save_img(os.path.join(save_path,
5                                                 'ground_truth.png'), tar[0] * 0.5 + 0.5)
5     tf.keras.preprocessing.image.save_img(os.path.join(save_path,
6                                                 'predicted_image.png'), prediction[0] * 0.5 + 0.5)

```

Code-Auszug 5.18: Speicherung der generierten Bilder

```
1 gen_loss_metric = tf.keras.metrics.Mean()  
2 disc_loss_metric = tf.keras.metrics.Mean()  
3 disc_accuracy_metric = tf.keras.metrics.BinaryAccuracy(threshold=0.5)
```

Code-Auszug 5.19: Initialisierung der Metriken

```
1 gen_loss, disc_loss, disc_real_output, disc_generated_output =  
    train_step(input_image, target, epoch, train_discriminator=True)  
2 gen_loss_metric(gen_loss)  
3 disc_loss_metric(disc_loss)  
4  
5 # Echte Bilder : Label 1, generierte Bilder :Label 0  
6 real_labels = tf.ones_like(disc_real_output)  
7 generated_labels = tf.zeros_like(disc_generated_output)  
8 disc_accuracy_metric.update_state(real_labels, disc_real_output)  
9 disc_accuracy_metric.update_state(generated_labels,  
    disc_generated_output)
```

Code-Auszug 5.20: Aktualisierung der Metriken

6

Evaluation

6.1. Bewertungskriterien

6.1.1. Diskriminator-Verlust

Der Diskriminator-Verlust ist ein wesentliches Maß für die Effektivität des Diskriminators in einem GAN-Modell. Ein niedriger Diskriminator-Verlust bedeutet, dass der Diskriminator erfolgreich zwischen echten und generierten Bildern unterscheiden kann. Im Kontext eines GANs zeigt ein abnehmender Diskriminator-Verlust über die Epochen hinweg, dass der Diskriminator zunehmend besser darin wird, die Authentizität der Bilder zu beurteilen. Ein hoher Diskriminator-Verlust weist hingegen darauf hin, dass das Modell Schwierigkeiten hat, echte von generierten Bildern zu unterscheiden, was ein Indikator für Verbesserungspotenzial im Training oder der Modellarchitektur sein könnte.

6.1.2. Generator-Verlust

Der Generator-Verlust ist ein Indikator für die Fähigkeit des Generators, Bilder zu erzeugen, die vom Diskriminator als echt eingestuft werden. Ein hoher Generator-Verlust deutet darauf hin, dass der Generator die realen Daten noch nicht effektiv nachahmen kann, während ein niedriger Generator-Verlust ein Zeichen dafür ist, dass die generierten Bilder den echten immer ähnlicher werden. Im Verlauf des Trainings sollte der Generator-Verlust tendenziell abnehmen, was darauf hindeutet, dass der Generator lernt, überzeugendere Bilder zu produzieren.

6.1.3. Diskriminator Genauigkeit

Die Diskriminator-Genauigkeit gibt an, wie oft der Diskriminator korrekt zwischen echten und gefälschten Bildern unterscheidet. Eine hohe Genauigkeit bedeutet, dass der Diskriminator effektiv arbeitet, während eine niedrige Genauigkeit auf Probleme bei der Unterscheidung hinweisen kann. Es ist wichtig, ein Gleichgewicht zu finden, denn eine zu hohe Genauigkeit kann darauf hindeuten, dass der Diskriminator die generierten Bilder zu leicht erkennt, was auf ein Problem mit dem Generator hinweisen könnte.

In einem ideal funktionierenden GAN sollte der Diskriminator eine Genauigkeit

von etwa 50% erreichen. Das bedeutet, dass der Diskriminator die echten von den generierten Bildern nur zufällig unterscheiden bzw. nicht mehr zuverlässig unterscheiden kann, was zeigt, dass das GAN gut trainiert wurde.

6.1.4. SSIM-Score (Structural Similarity Index)

Der SSIM-Score ist ein Maß für die visuelle Ähnlichkeit zwischen den generierten Bildern und den entsprechenden Originalbildern. Er bewertet die Bildqualität anhand von Faktoren wie Helligkeit, Kontrast und Struktur. Ein hoher SSIM-Wert deutet darauf hin, dass die generierten Bilder den echten Bildern sehr ähnlich sind, was ein Zeichen für eine hohe Bildqualität ist. Ein niedriger SSIM-Wert kann auf deutliche Unterschiede in der visuellen Struktur hinweisen, was Verbesserungsbedarf im Generator-Training oder in der Modellarchitektur signalisiert.

6.2. Ergebnisse und objektive Bewertung

In diesem Abschnitt liegt der Fokus auf der objektiven Bewertung der Leistung des CycleGAN-Algorithmus. Es werden verschiedene Metriken und Kriterien herangezogen, um eine umfassende Bewertung durchzuführen, einschließlich der Verluste der Generatoren und Diskriminatoren bei Variation der Hyperparameter. Darüber hinaus wird eine detaillierte Analyse der Genauigkeit der Diskriminatoren und des SSIM-Scores durchgeführt. Diese Messungen sind sowohl für die genaue Bestimmung der Qualität der generierten Ergebnisse als auch für die Bewertung der Lernfähigkeit der Modelle von Bedeutung.

Die verwendeten Datensätze für die Evaluierung stammen aus dem Berkeley-Repository. Besondere Beachtung wird dem 'maps'-Datensatz geschenkt, bei dem Transformationen zwischen Kartenansichten und Satellitenbildern durchgeführt wurden. Das Training wurde mehrmals mit verschiedenen Hyperparameter-Konfigurationen durchgeführt, wobei Parameter wie der Lambda-Wert, die Anzahl der Epochen und die Lernrate des Optimizers variiert wurden. Ziel ist es, eine Konfiguration zu finden, bei der die Qualität der erzeugten Bilder nahe an der Qualität des Originals und die Leistung des Diskriminators optimal ist.

Die subjektive Bewertung erfolgt durch die visuelle Beurteilung der generierten Bilder aus dem Testdatensatz hinsichtlich stimmiger Farben, Struktur und Rauschen im Vergleich zu den Originalbildern. Darauf aufbauend erfolgt eine objektive Bewertung mithilfe der definierten Metriken.

6.2.1. Pix2Pix

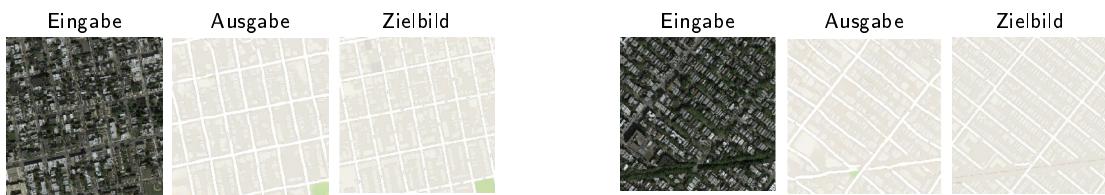
Die Entwicklung eines Pix2Pix GANs stellt eine große Herausforderung dar, insbesondere im Hinblick auf das Training des Modells. Eine der zentralen Schwierigkeiten besteht darin, ein angemessenes Gleichgewicht zwischen effektivem Lernen und der Vermeidung von Overfitting zu finden. Overfitting tritt auf, wenn das Modell zu stark an die spezifischen Muster und Merkmale des Trainingsdatensatzes

6.2. Ergebnisse und objektive Bewertung

angepasst wird, was dazu führt, dass der Generator Probleme beim Übersetzen von unbekannten Daten hat. Dieses Problem tritt oft in Szenarien auf, in denen mit einem begrenzten Datensatz trainiert wird, wie in diesem Fall.

In der praktischen Anwendung des Pix2Pix-Modells, das in diesem Fall speziell für die Transformation von Satellitenbildern in Karten entwickelt wurde, ist aufgefallen, dass der Generator besonders effizient bei der Erzeugung von Straßenlinien ist. Dies deutet auf eine starke Fähigkeit des Modells hin, lineare, kontrastreiche Strukturen zu erkennen und wiederzugeben. Jedoch stößt das Modell an seine Grenzen, sobald es um die Darstellung von Farben und komplexeren Strukturen geht, insbesondere bei Grünflächen oder Wasser. Bei der Generierung dieser Elemente zeigt der Generator deutliche Schwierigkeiten, was sich in einer weniger präzisen und oft fehlerhaften Darstellung widerspiegelt. Die Beobachtungen ergaben, dass obwohl diese Elemente oftmals nur einen geringen Teil des Gesamtbildes ausmachen, sie dennoch einen großen negativen Einfluss auf das Gesamtergebnis haben.

Für das Training des Pix2Pix-Modells wurde der Adam-Optimierer mit einer initialen Lernrate von 0.0002 und den Momentum-Parametern $\beta_1 = 0.5$ und $\beta_2 = 0.999$ eingesetzt. Diese Einstellungen erweisen sich als effektiv, um einen guten Kompromiss zwischen Lerngeschwindigkeit und Stabilität des Trainingsprozesses zu ermöglichen. Eine Batchgröße von 1 wurde gewählt, um die Trainingseffizienz zu maximieren und qualitativ hochwertigere Ergebnisse zu erzielen. Während verschiedener Testdurchläufe konnte beobachtet werden, dass eine größere Batchgröße zwar das Training beschleunigt, jedoch zu schlechteren Trainingsergebnissen führt. Der Hyperparameter λ wird genutzt, um das Gewicht des L1-Verlusts im Generatorverlust zu steuern und die Inhaltsähnlichkeit zwischen den generierten und den Zielbildern zu regulieren. Insgesamt hat das Modell eine Trainingsdauer von 100 Epochen durchlaufen, wobei die ersten 80 Epochen mit $\lambda = 100$ und die restlichen 20 mit $\lambda = 150$ absolviert wurden.



Die Evaluierung der Trainingsresultate des Pix2Pix-Modells basiert auf einer Analyse der 100 durchgeführten Trainingsepochen. Während des Durchlaufs dieser Epochen wurden spezifische Metriken wie Generator- und Diskriminator-Verluste sowie Diskriminator-Genauigkeit und der Structural Similarity Index Measure (SSIM) überwacht.

Der Verlust des Generator beginnt bei 14.44 in der ersten Epoche und sinkt dann kontinuierlich ab, was auf ein erfolgreiches Lernen hinweist. Es gibt Schwankungen im Verlauf des Trainings, aber insgesamt zeigt sich ein Abwärtstrend, was darauf hindeutet, dass der Generator zunehmend besser in der Lage ist, glaubwürdige

Bilder zu erzeugen.

Der Diskriminator-Verlust beginnt bei 0.93 und zeigt eine leichte Abnahme über die Zeit. Dies deutet darauf hin, dass der Diskriminator im Laufe des Trainings besser darin wird, echte Bilder von gefälschten zu unterscheiden. Mit der Erhöhung von λ bei Epoche 80 fokussiert sich der Generator mehr darauf, die exakten Pixelwerte des Zielbildes nachzubilden, dadurch muss sich der Diskriminator an die neuen Unterschiede zwischen realen und generierten Bildern anpassen. Aus diesem Grund kann dies anfänglich zu einem Anstieg des Diskriminatorverlusts führen, da seine Aufgabe nun schwieriger wird. In Abbildung 6.2 wird die Entwicklung der Verlustfunktionen nochmal grafisch dargestellt.

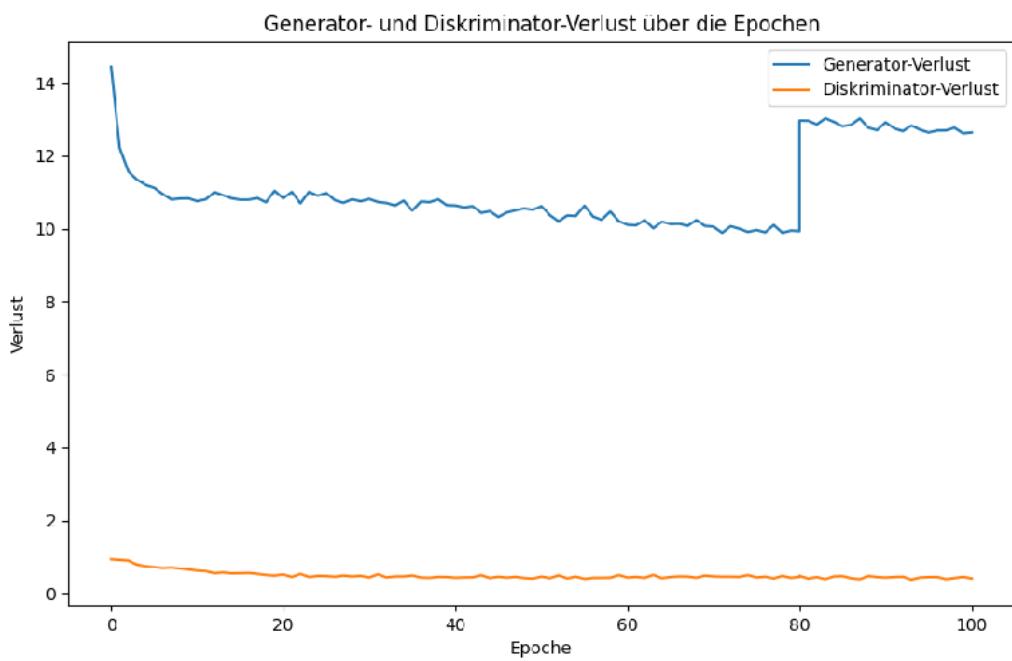


Abbildung 6.2.: Generator- und Diskriminatorenverlust des Trainingsdurchlaufs mit 100 Epochen, eigene Erstellung mittels matplotlib.pyplot-Bibliothek

6.2. Ergebnisse und objektive Bewertung

Die Analyse der Diskriminatorgenaugkeit im Trainingsverlauf zeigt einen Anstieg der Genauigkeit über 90% was weit über den Idealwert 50%. Dies ist ein Anzeichen dafür das der Diskriminator möglicherweise zu effektiv darin geworden ist, generierte Bilder von realen zu unterscheiden. Dieser Wert ist ein Indikator dafür, dass der Diskriminator fast immer richtig liegt und könnte darauf hindeuten, dass der Generator nicht genügend Raum bekommt, um zu lernen und sich zu verbessern. Zusätzlich weisen die Schwankungen, wie in Abbildung 6.3 gezeigt, auf eine gewisse Inkonsistenz im Lernprozess hin.

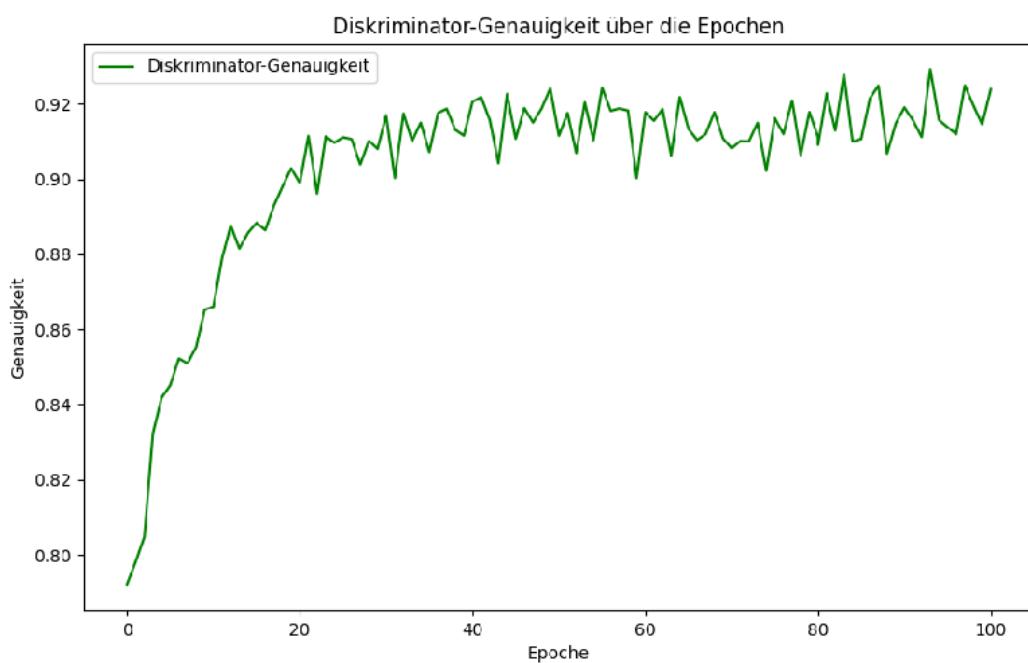


Abbildung 6.3.: Diskriminatorgenaugkeit des Trainingsdurchlaufs mit 100 Epochen,
eigene Erstellung mittels matplotlib.pyplot-Bibliothek

Der SSIM-Score hatte anfänglich einen Wert von 0.5549 der allmählich anstieg, was auf eine Verbesserung der Bildqualität hindeutet, die der Generator im Laufe des Trainings erzielt. Dieser Trend ist ein positives Zeichen dafür, dass der Generator zunehmend besser darin wird, Bilder zu erzeugen, die den Zielbildern ähnlicher sind. Interessanterweise steigt der SSIM-Wert nicht kontinuierlich, wie in Abbildung 6.4 zu sehen ist, sondern zeigt einige Schwankungen, was auf unterschiedliche Leistungen des Generators bei verschiedenen Bildern hinweist. Diese Inkonsistenzen könnten auf die Variabilität innerhalb des Trainingsdatensatzes oder auf die begrenzte Fähigkeit des Generators zurückzuführen sein bei bestimmten Strukturen wie beispielsweise Grünflächen oder Wasser Probleme zu haben.

Bemerkenswert ist der Anstieg des SSIM-Wertes nach der 80. Epoche, was zeitlich mit der Erhöhung des Lambda-Wertes von 100 auf 150 korreliert. Dies könnte darauf hindeuten, dass die Erhöhung des Lambda-Wertes dazu beigetragen hat, die Qualität der generierten Bilder in Bezug auf ihre strukturelle Ähnlichkeit zu verbessern. Der höhere SSIM-Wert in späteren Epochen könnte ein Indikator dafür sein, dass der Generator lernt, sich auf die wichtigen Aspekte der Bilder zu konzentrieren, die zur Ähnlichkeit mit den Zielbildern beitragen.

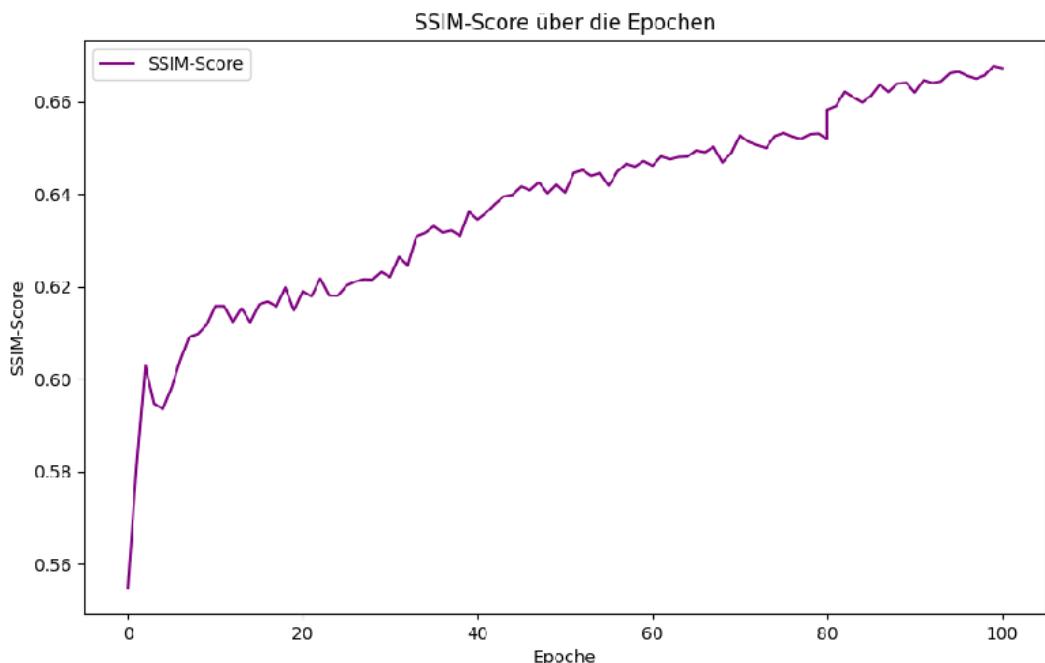


Abbildung 6.4.: SSIM-Score des Trainingsdurchlaufs mit 100 Epochen, eigene Erstellung mittels matplotlib.pyplot-Bibliothek

Das Pix2Pix GAN-Modell demonstriert seine Stärken in der Umwandlung von Satellitenbildern zu Karten, mit besonderer Kompetenz in der Darstellung von Straßenstrukturen. Trotzdem stößt es bei der Generierung komplexerer Elemente wie Grünflächen und Wasser auf Herausforderungen. Die Feinabstimmung der Hyperparameter, einschließlich einer geringen Batchgröße und einer spezifischen Anpassung des Lambda-Wertes, hat die Trainingseffizienz verbessert, was sich in

6.2. Ergebnisse und objektive Bewertung

einem erhöhten SSIM-Score niederschlägt. Allerdings weist die hohe Diskriminatoren genauigkeit von über 90% darauf hin, dass das Modell weiter optimiert werden muss, um ein Gleichgewicht zwischen Generator und Diskriminator zu erlangen. Hervorzuheben ist die signifikante Zeitersparnis durch die Verwendung einer GPU im Vergleich zur CPU: Während eine Epoche auf der CPU 320 Sekunden dauert, wird diese Zeit auf der GPU auf nur 20 Sekunden reduziert.

6.2.2. CycleGAN

Die Datensätze X (Karten) und Y (Satellitenbilder) werden durch die Generatoren $G : X \rightarrow Y$ und $F : Y \rightarrow X$ sowie die Diskriminatoren D_X und D_Y repräsentiert, die jeweils darauf abzielen, Bilder in den Domänen X und Y zu unterscheiden. Die Hyperparameter wurden initial gemäß den Angaben in der Originalarbeit von Zhu et al. ([ZPIE](#)) festgelegt und nachfolgend durch explorative Anpassungen basierend auf der Bildqualität und der Genauigkeit der Diskriminatoren verfeinert. Die Evaluierung der Leistungsfähigkeit erfolgt nach dem Training anhand einer definierten Anzahl von Testdaten.

Ein schwieriges Problem für die Generatoren ist die Wiedergabe feiner und detaillierter Strukturen. Insbesondere ist festzustellen, dass die Generatoren bei der Erzeugung von Straßenlinien effizienter sind. Dies ist auf die Überrepräsentation von Straßen und Häusern in den Trainingsdaten zurückzuführen. Dagegen stellt die Vielfalt der Bilder, wie Flüsse und Wälder, eine Herausforderung dar. Die genaue Erfassung der Struktur und Farbcodierung dieser Gebiete stellt die Generatoren vor Schwierigkeiten. Das Ergebnis ist oft die Generierung von Flächen mit ähnlichen Farben wie Straßen, wobei versucht wird, kleine Straßen auf der Basis dieser großen offenen Flächen zu generieren.

Die Anwendung eines anderen Datensatzes, 'zebra2horse', zeigt ähnliche Herausforderungen. Die Umwandlung eines Pferdes in ein Zebra bereitet dem Generator weniger Schwierigkeiten als die umgekehrte Umwandlung. Es ist einfacher, Strukturen auf eine homogene Pferdefläche zu projizieren als unterschiedliche Kontraste und Muster auf eine homogene Zebrafläche.

Die Wahl der Lernrate erwies sich als kritisch, wobei die besten Ergebnisse mit einer Lernrate von 0,0002 erzielt wurden. Diese Einstellung führte zu den höchsten SSIM-Werten und Diskriminator-Testgenauigkeiten. Insbesondere erzielte der Generator F einen SSIM-Score von 0.592, während der Generator G einen Wert von 0.198 erreichte. Die Diskriminatoren zeigten mit 50.1% für D_Y und 48.5% für D_X nahezu optimale Genauigkeiten. Trotz der auf den ersten Blick großen visuellen Ähnlichkeit der generierten Satellitenbilder zeigte eine genauere Analyse, dass kleinere Strukturen im generierten Bild fehlten. Dies spiegelte sich im vergleichsweise niedrigen SSIM-Wert wider.

Bei einem Gewichtungsfaktor von $\lambda = 120$ wurden sowohl bei der visuellen Beurteilung als auch beim SSIM-Score die besten Ergebnisse erzielt. Die Abbildung 6.5 zeigt einen Teil der Ergebnisse. Dieser spezifische Wert für λ führte zu erzeugten Bildern, deren Signal-Rausch-Verhältnis (SNR) entweder besser oder nahezu gleich dem des Originalbildes war. Der SNR-Wert gibt das Verhältnis zwischen dem Signal, repräsentiert durch relevante Bildinformationen, und dem Rauschen, repräsentiert durch unerwünschte Störungen, an. Ein höherer SNR-Wert bedeutet eine bessere Bildqualität und Klarheit gegenüber dem Rauschen.

6.2. Ergebnisse und objektive Bewertung

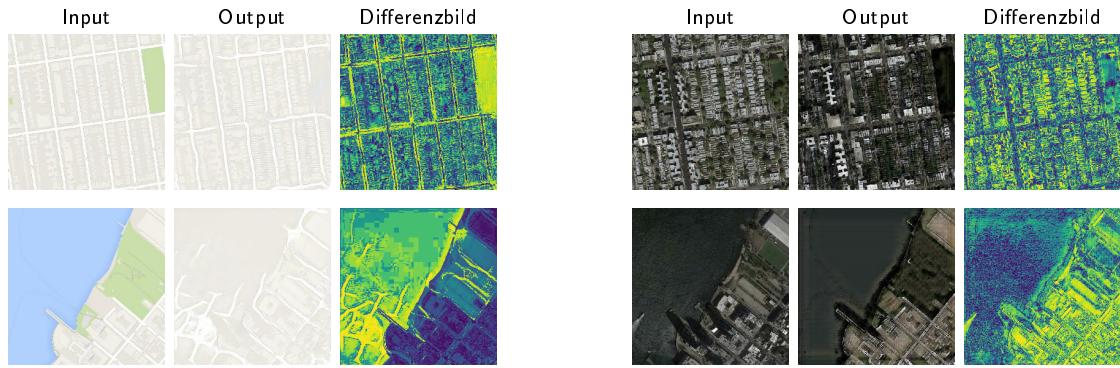


Abbildung 6.5.: Testergebnisse nach dem Training mit $\lambda = 120$, Lernrate = 0.0002 und 100 Epochen

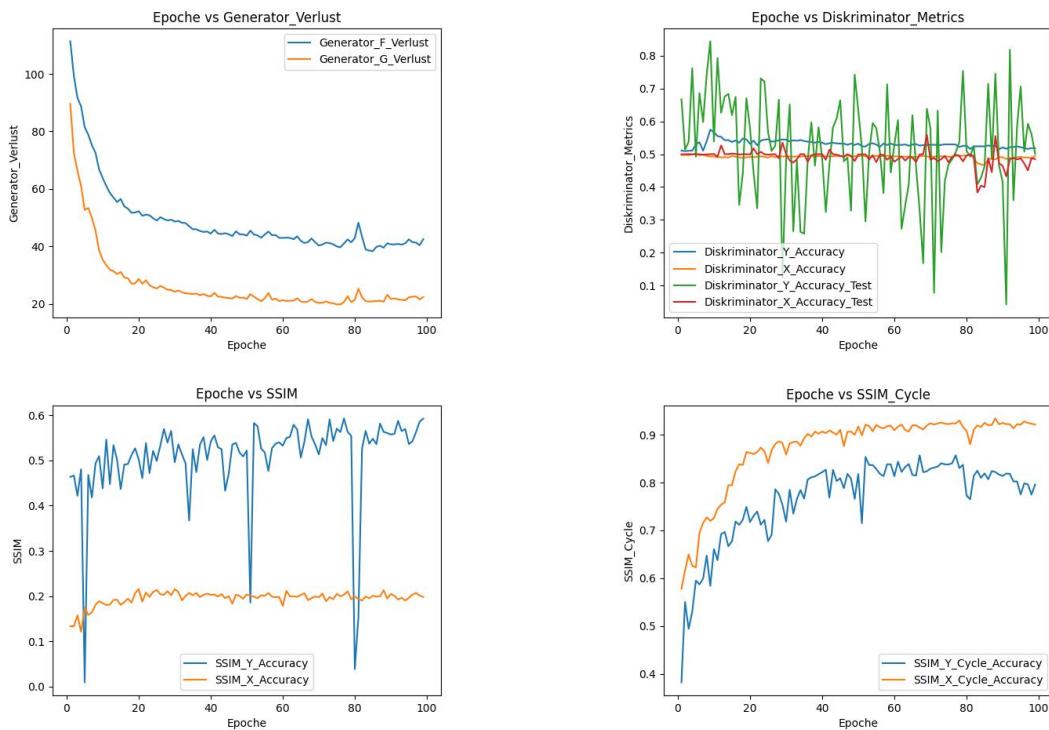


Abbildung 6.6.: Verlauf der Metriken beim Training mit $\lambda = 120$, Lernrate = 0.0002 und 100 Epochen. Die Metriken umfassen: Generator-Verlust (oben links), Diskriminator-Genauigkeiten auf den Trainings- und Testdaten (oben rechts), sowie SSIM-Score (unten links) und Zyklus-SSIM-Score (unten rechts). Die Metriken werden in Abhängigkeit von der Anzahl der Epochen auf der x-Achse untersucht.

Trotz dieser Erfolge bleibt das Training instabil, und es besteht die Möglichkeit eines Modekollapses. Die Generatoren zeigen Anzeichen eines exponentiellen Verfalls in ihren Verlustkurven, während der SSIM-Score nur langsam ansteigt (Abbildung 6.6). Dies deutet darauf hin, dass die Generatoren zwar ihre Verluste minimieren, jedoch möglicherweise nicht die gewünschte visuelle Qualität erreichen.

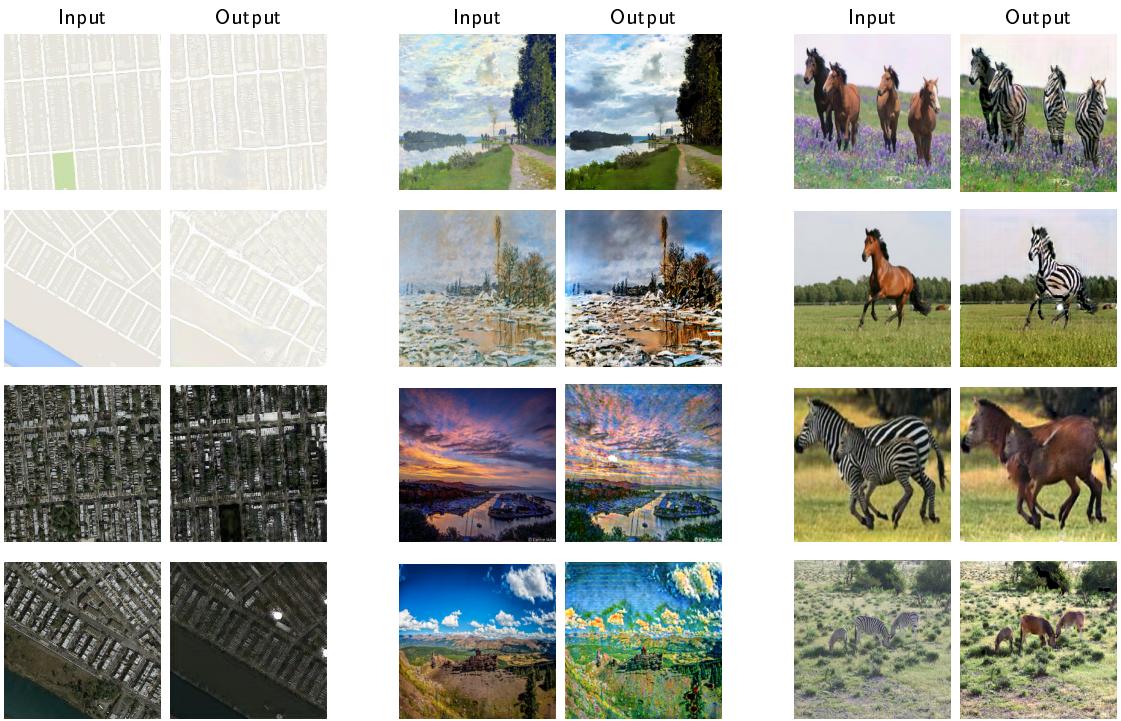


Abbildung 6.7.: CycleGAN-Ergebnisse aus verschiedenen Datensätzen aus dem Berkeley-Repository: 'maps' (links), 'monet2photo' (mitte), 'horse2zebra' (rechts)

Die generierten Bilder zeigen bereits zu diesem Zeitpunkt eine grobe visuelle Ähnlichkeit mit der zugrunde liegenden Originalbildstruktur mit minimalen Abweichungen. Im SSIM-Score werden jedoch die Schwierigkeiten der Generatoren bei der Erzielung signifikanter Verbesserungen in Bezug auf feinere Details und Farbübereinstimmungen deutlich. Dies ist auch visuell an den Differenzbildern zu erkennen, die mit Hilfe eines Histogrammabgleichs kontrastverbessert wurden. Helle Bereiche in diesen Differenzbildern zeigen größere Unterschiede. Insbesondere bei kleinen pixelweisen Details sowie bei unterschiedlichen Farbkodierungen häufen sich die hellen Bereiche. Die Abbildung 6.5 zeigt ein solches Beispiel.

Insbesondere in den frühen Epochen sind die Generatoren nicht in der Lage, Straßenlinien zuverlässig als gerade Linien zu generieren. Diese Schwächen verbessern sich mit zunehmender Trainingsdauer, so dass nach ca. 300 Epochen relativ konsistente Geraden erzeugt werden.

Die Analyse des SSIM-Wertes in Abhängigkeit von der Zyklusübersetzung zeigt einen exponentiellen Anstieg. Diese Beobachtung deutet darauf hin, dass der Verlust der Zykluskonsistenz einen signifikanten Einfluss auf das Training hat. Der exponentielle Anstieg des SSIM-Wertes deutet darauf hin, dass die Konsistenz zwischen den Originalbildern und den rückübersetzten Bildern mit der Zeit kontinuierlich zunimmt. Dies kann auf eine kontinuierliche Verbesserung der Gene-

Lambda	Epoche	Lernrate	D_Y Tr-Acc	D_X Tr-Acc	D_Y Te-Acc	D_X Te-Acc
120	100	0.000025	0.549	0.494	0.307	0.497
120	100	0.00005	0.537	0.491	0.321	0.489
120	100	0.0001	0.519	0.490	0.418	0.485
120	100	0.0002	0.518	0.495	0.501	0.485
120	100	0.0004	0.524	0.487	0.771	0.433
100	100	0.0002	0.521	0.491	0.527	0.481
140	100	0.0002	0.562	0.492	0.544	0.469
120	200	0.0002	0.514	0.500	0.563	0.489
100	300	0.0002	0.510	0.484	0.470	0.676
120	300	0.0002	0.508	0.495	0.468	0.709

Tabelle 6.1.: CycleGAN Ergebnisse unter verschiedenen Hyperparametern mit Batchgröße 1, Teil1

F Loss	G Loss	X SSIM	Y SSIM	X Z-SSIM	Y Z-SSIM	X SNR	Y SNR
53.113	22.556	0.545	0.191	0.895	0.831	-2.991	0.228
44.626	20.417	0.541	0.191	0.916	0.836	-0.589	1.071
44.376	21.943	0.580	0.194	0.935	0.842	-4.604	1.367
42.489	22.373	0.592	0.198	0.921	0.796	-5.308	0.851
47.239	29.084	0.588	0.207	0.601	0.845	-2.407	0.823
34.973	18.837	0.551	0.194	0.930	0.830	-3.292	1.261
13.048	6.842	0.510	0.179	0.513	0.409	-2.087	1.005
46.061	23.058	0.371	0.177	0.802	0.894	1.915	1.992
34.881	20.899	0.577	0.191	0.931	0.756	-2.705	1.373
46.652	26.360	0.585	0.207	0.901	0.707	-3.409	1.248

Tabelle 6.2.: CycleGAN Ergebnisse unter verschiedenen Hyperparametern mit Batchgröße 1, Teil2

ratoren hinsichtlich ihrer Fähigkeit zur Zyklusübersetzung hindeuten. Insgesamt erreichen die SSIM-Scores für die Zyklusübersetzung Werte von 80% für Y und 90% für X (Abbildung 6.6).

Im Gegensatz dazu bleibt die Genauigkeit der Diskriminatoren für die Trainingsdaten stabil bei etwa 50%. Für die Testdaten, insbesondere für D_Y , sind jedoch erhebliche Schwankungen in Abhängigkeit von der Epoche festzustellen. Dennoch liegt die Genauigkeit am Ende des Trainings in den meisten Fällen in einem nahezu optimalen Wertebereich (Abbildung 6.6). Dies deutet darauf hin, dass die Diskriminatoren möglicherweise Schwierigkeiten haben, mit neuen, untrainierten Daten umzugehen. Diese Schwankungen können auf eine Überanpassung oder eine mangelnde Generalisierungsfähigkeit des Modells hinweisen, trotz der implementierten Maßnahmen wie dem Hinzufügen von Rauschen und Dropout-Schichten.

Die Trainingszeit spiegelt die Komplexität der Modelle wider, insbesondere auf der GPU, wo jede Epoche im Durchschnitt etwa 2 Minuten dauert. Im Vergleich dazu benötigt die CPU, selbst nach Reduktion der Residualblöcke auf einen Block, etwa 6 Minuten pro Epoche. Die Verwendung der Instanznormalisierung trägt zur Beschleunigung der Konvergenz und zur Verbesserung der Laufzeit bei. Auch im Vergleich zu Generatoren mit zusätzlichen Schichten zeigt sich, dass die Verwendung von Restblöcken eine bessere Laufzeit ermöglicht.

Insgesamt zeigt die Evaluierung, dass das CycleGAN-Modell trotz einiger Herausforderungen bei der Generierung detaillierter Strukturen und der Generalisierung auf neue Daten vielversprechende Ergebnisse liefert.

6.3. Vergleich von Pix2Pix und CycleGAN

Pix2Pix und CycleGAN sind beides Modelle von Generative Adversarial Networks, die für Bild-zu-Bild-Übersetzungen eingesetzt werden, jedoch unterscheiden sie sich wesentlich in ihrer Methodik und den Anwendungsbereichen aufgrund ihrer strukturellen Unterschiede.

Die Pix2Pix-Architektur verwendet typischerweise einen U-Net-basierten Generator und einen PatchGAN-Diskriminator, der auf die einzelnen Bildabschnitte einwirkt, um festzustellen, ob sie echt oder gefälscht sind. Diese Architektur ermöglicht eine effektivere Erfassung hochauflösender Details. Der Diskriminator von CycleGAN basiert ebenfalls auf PatchGAN, verwendet jedoch Instanznormalisierung anstelle von Batchnormalisierung. Ein wesentlicher Unterschied besteht in den Generatoren. Diese basieren auf dem ResNet-Ansatz und erweitern Pix2Pix um einen Zykluskonsistenzverlust. Dieser stellt sicher, dass das Eingabebild in die Zieldomäne übersetzt und anschließend wieder in die Ursprungsdomäne zurückübersetzt wird, ohne dass der ursprüngliche Inhalt verloren geht. Diese bidirektionale Struktur soll helfen, die Mode-Kollapsen zu vermeiden, und stellt sicher, dass der Generator vielfältige Ausgaben erzeugt.

Pix2Pix benötigt paarweise Trainingsdaten. Das bedeutet, dass jedes Trainingsbeispiel aus einem korrespondierenden Bildpaar, einem Eingabe- und einem Zielbild, besteht. Daher eignet sich das Pix2Pix-Modell besonders für Aufgaben, bei denen eine strukturierte und koordinierte Beziehung zwischen Eingabe- und Ausgabebild besteht, wie bei der Umwandlung von Satellitenbildern in Karten oder beim Einfärben von Schwarzweißfotos. Im Gegensatz dazu ist CycleGAN für die Arbeit mit ungepaarten Trainingsdaten konzipiert, die das Lernen von Übersetzungen ohne eine Eins-zu-Eins-Zuordnung zwischen Quell- und Zielbildern ermöglichen. Dies ist besonders nützlich für Stiltransformationen, beispielsweise für Pferde in Zebras oder Äpfel in Orangen, für die es keine gepaarten Beispiele gibt. Des Weiteren wird eine größere Flexibilität und Unabhängigkeit spezifischer Datensätze ermöglicht.

Die Evaluierung zeigt die wesentlichen Stärken und Schwächen der beiden GAN-Architekturen. Beide haben mit einer Lernrate von 0,0002 gute Ergebnisse erzielt. Der λ -Wert variiert bei Pix2Pix, während bei CycleGAN ein konstanter λ -Wert verwendet wird.

Es stellt sich heraus, dass beide Generatoren effizient sind, wenn es darum geht, Straßenlinien zu erzeugen, aber dass es für beide schwierig ist, die Struktur und Farbkodierung von vielfältigeren Bildern, wie Flüssen und Wäldern, genau zu erfassen. Außerdem steigt der SSIM-Wert von CycleGAN exponentiell mit großen Ausreißern, während der SSIM-Wert von Pix2Pix relativ linear mit der Zeit ansteigt. Dies spiegelt sich auch in den Verlusten der Generatoren und Diskriminatoren wider. Bei CycleGAN ist die Abnahme exponentiell. Bei Pix2Pix ist sie linear. Pix2Pix zeigt insbesondere eine große Verbesserung der Genauigkeit des

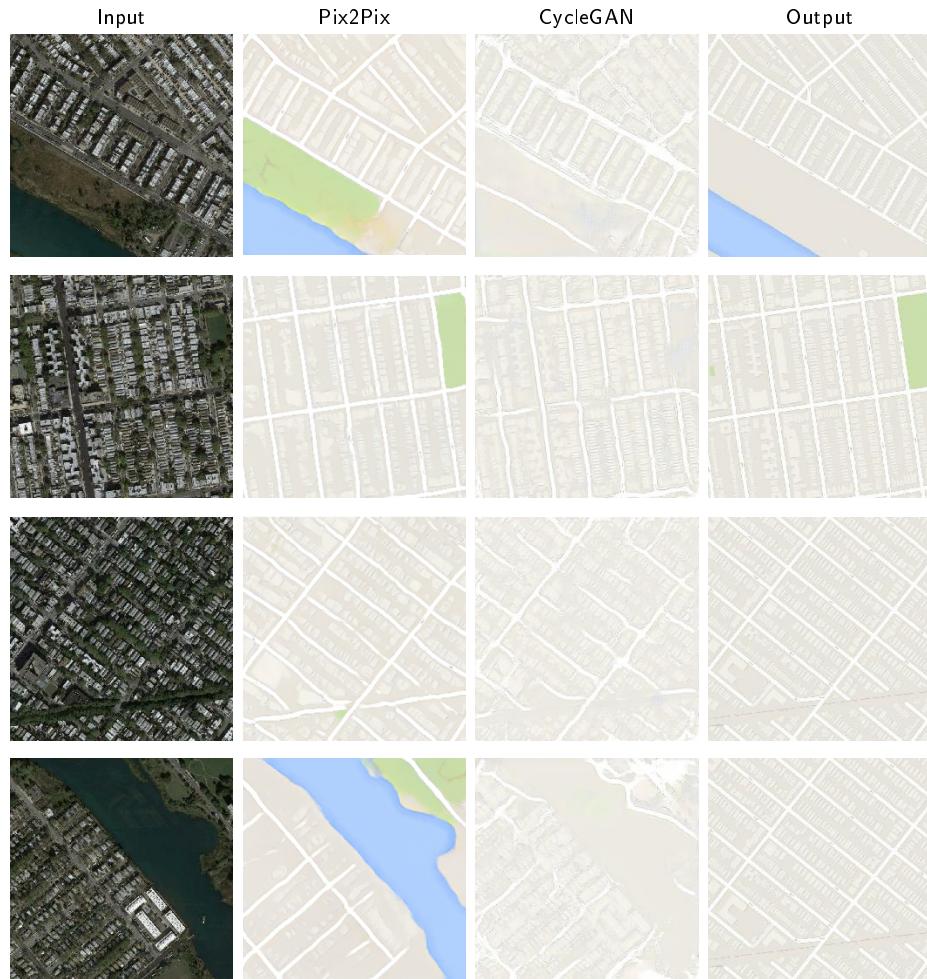


Abbildung 6.8.: Vergleich der Ergebnisse von Pix2Pix und CycleGAN

Diskriminators auf den Trainingsdaten, die nach einigen Epochen 90% übersteigt. Im Gegensatz dazu bleibt die Genauigkeit des CycleGAN Diskriminators konstant bei 50%, was als optimal angesehen wird. Die Genauigkeit der Testdaten spiegelt dies teilweise wider. Die Testgenauigkeit bei Pix2Pix ist mit 37.5% dagegen niedrig, obwohl die Trainingsgenauigkeit im Vergleich sehr hoch ist. Dies und die großen Schwankungen bei CycleGAN können auf eine Überanpassung der Diskriminatoren zurückgeführt werden.

Des Weiteren ist die Komplexität der Pix2Pix-Architektur aufgrund der 6-mal schnelleren Laufzeit geringer. Darüber hinaus wurde festgestellt, dass die Anpassung der Hyperparameter und die Stabilität gegenüber Mode-Kollapse bei CycleGAN schwieriger sind als bei Pix2Pix.

Aufgrund der direkten Korrespondenz zwischen Eingangs- und Ausgangsbildern zeigt Pix2Pix bessere Tendenzen bei der Projektion von Farbkodierungen als CycleGAN. Letzteres hat bereits große Schwierigkeiten, die richtige Stelle für die Farb-

6.3. Vergleich von Pix2Pix und CycleGAN

kodierung zu finden. Andererseits ist es mit den beiden Generatoren von CycleGAN möglich, nach dem Training beide Bilddomänen abzubilden, während Pix2Pix auf die Abbildung nur einer Bilddomäne beschränkt ist. Die Wahl zwischen Pix2Pix und CycleGAN hängt daher stark von den verfügbaren Daten und den spezifischen Anforderungen der jeweiligen Aufgabe ab.

7

Fazit und Ausblick

Hier wird ein Fazit und ein Ausblick gegeben.

7.1. Fazit

Im Zuge dieses Teamprojekts wurden zentrale Forschungsfragen im Bereich der Bildverarbeitung mit Generative Adversarial Networks (GANs) untersucht und beantwortet. Dabei standen insbesondere die Frameworks Pix2Pix und CycleGAN im Fokus, um die vielfältigen Anwendungsmöglichkeiten und Herausforderungen von GANs zu beleuchten. Durch die praktische Implementierung und das Training der Modelle Pix2Pix und CycleGAN auf einem spezifischen Datensatz konnten wir die Leistungsfähigkeit und Grenzen dieser Technologien demonstrieren. Die Ergebnisse zeigen, dass beide Modelle effektiv in der Bild-zu-Bild-Übersetzung sind, sich jedoch in ihren Stärken und Herausforderungen unterscheiden.

Während Pix2Pix für seine Effizienz bei gepaarten Trainingsdaten bekannt ist, demonstriert CycleGAN seine Stärke in der Zykluskonsistenz bei ungepaarten Daten. Die Analyse der SSIM-Scores und Diskriminatorgenauigkeiten und weiterer Metriken beider Modelle hat einen tiefgreifenden Einblick in ihre Fähigkeit zur Erzeugung qualitativ hochwertiger Bilder ermöglicht. Besonders auffällig war die kontinuierliche Verbesserung der Bildqualität bei CycleGAN im Laufe des Trainings sowie die Notwendigkeit der sorgfältigen Anpassung der Trainingsparameter bei Pix2Pix und CycleGAN. Diese Erkenntnisse unterstreichen die Bedeutung einer gezielten und durchdachten Konfiguration der Trainingsparameter für die erfolgreiche Anwendung von GANs in der Bildverarbeitung.

7.2. Ausblick

Der Ausblick für die weitere Forschung und Anwendung von Generative Adversarial Networks (GANs) im Bereich der Bildverarbeitung ist vielversprechend und bietet zahlreiche Möglichkeiten für Innovationen. Die Ergebnisse dieses Projekts legen nahe, dass zukünftige Arbeiten sich darauf konzentrieren könnten, die Effizienz und Genauigkeit von GAN-Modellen weiter zu verbessern, insbesondere in Bereichen, in denen hochpräzise Bild-zu-Bild-Übersetzungen erforderlich sind.

Des Weiteren könnte die Erforschung neuer Anwendungsbereiche, in denen GANs bisher noch nicht eingesetzt wurden, neue Horizonte eröffnen. Dies könnte die Entwicklung maßgeschneiderter Lösungen für spezifische Aufgabe oder kreative Prozesse beinhalten. Auch die Auswirkungen unterschiedlicher Datensätze und die Entwicklung von GANs, die effektiv mit weniger oder ungepaarten Daten arbeiten können, sind vielversprechende Forschungsfelder. Letztlich könnte die Erweiterung der ethischen und rechtlichen Rahmenbedingungen, insbesondere in Bezug auf die Verwendung generierter Bilder, eine Schlüsselrolle spielen, um das volle Potenzial von GANs in einer verantwortungsbewussten und nachhaltigen Weise zu nutzen.

A

Anhang - Code

Dieser Abschnitt enthält die wichtigsten Funktionen, die in der Implementierung verwendet werden. Hier finden sich auch die fehlenden Funktionen aus dem Kapitel 'Lösungsbeschreibung'.

Datenvorverarbeitung

```
1 def load_image(image_path):
2     image = tf.io.read_file(image_path)
3     image = tf.io.decode_jpeg(image, channels=3)
4     image = tf.cast(image, tf.float32)
5     return image
```

Code-Auszug A.1: Lesen eines Bildes (CycleGAN Implementierung)

Pix2Pix

```
1 def generator_loss(disc_generated_output, gen_output, target):  
2     gan_loss = loss_object(tf.ones_like(disc_generated_output),  
3                             disc_generated_output)  
4     # Mean absolute error  
5     l1_loss = tf.reduce_mean(tf.abs(target - gen_output))  
6     total_gen_loss = gan_loss + (LAMBDA * l1_loss)  
7     return total_gen_loss, gan_loss, l1_loss
```

Code-Auszug A.2: Pix2Pix Generatorverlust

```
1 def generator_loss(disc_generated_output, gen_output, target):  
2     gan_loss = loss_object(tf.ones_like(disc_generated_output),  
3                             disc_generated_output)  
4     # Mean absolute error  
5     l1_loss = tf.reduce_mean(tf.abs(target - gen_output))  
6     total_gen_loss = gan_loss + (LAMBDA * l1_loss)  
7     return total_gen_loss, gan_loss, l1_loss
```

Code-Auszug A.3: Residualblock Implementierung

```
1 def train_step(input_image, target, epoch, train_discriminator=True):
2     with tf.GradientTape() as gen_tape, tf.GradientTape() as disc_tape:
3         gen_output = generator(input_image, training=True)
4
5         disc_real_output = discriminator([input_image, target], training=True)
6         disc_generated_output = discriminator([input_image,
7             gen_output], training=True)
8
9         gen_total_loss, gen_gan_loss, gen_l1_loss = generator_loss(
10            disc_generated_output, gen_output, target)
11        disc_loss = discriminator_loss(disc_real_output,
12            disc_generated_output)
13
14        generator_gradients = gen_tape.gradient(gen_total_loss, generator.
15            trainable_variables)
16        generator_optimizer.apply_gradients(zip(generator_gradients,
17            generator.trainable_variables))
18
19        if train_discriminator:
20            discriminator_gradients = disc_tape.gradient(disc_loss,
21                discriminator.trainable_variables)
22            discriminator_optimizer.apply_gradients(zip(
23                discriminator_gradients, discriminator.trainable_variables))
24
25    return gen_total_loss, disc_real_output, disc_real_output,
26        disc_generated_output
```

Code-Auszug A.4: Pix2Pix Trainingsfunktion

CycleGAN

```
1 def residual_block(x, filters ):  
2     shortcut = x  
3     x = tf.keras.layers.Conv2D(filters, kernel_size=3, strides=1,  
4         padding='same')(x)  
5     x = InstanceNormalization(axis=-1)(x)  
6     x = tf.keras.layers.LeakyReLU()(x)  
7     x = tf.keras.layers.Conv2D(filters, kernel_size=3, strides=1,  
8         padding='same')(x)  
9     x = InstanceNormalization(axis=-1)(x)  
10    x = tf.keras.layers.concatenate([x, shortcut])  
11    return x
```

Code-Auszug A.5: Residualblock Implementierung

```
1 def convolutional_layer(x, filters , kernel_size=3, strides=2):  
2     x = tf.keras.layers.Conv2D(filters, kernel_size=kernel_size, strides  
3         =strides, padding='same')(x)  
4     x = InstanceNormalization(axis=-1)(x)  
5     x = tf.keras.layers.LeakyReLU()(x)  
6     return x
```

Code-Auszug A.6: Convolutional Block in CycleGAN

```
1 def t_convolutional_layer(x, filters, kernel_size=3, strides=2):
2     x = tf.keras.layers.Conv2DTranspose(filters, kernel_size=kernel_size
3                                         , strides=strides, padding='same')(x)
4     InstanceNormalization(axis=-1)(x)
5     x = tf.keras.layers.LeakyReLU()(x)
6     return x
```

Code-Auszug A.7: Transpose Convolutional Block in CycleGAN

```
1 def generator_adversarial_loss(generated):
2     return loss_obj(tf.ones_like(generated), generated)
```

Code-Auszug A.8: Adversieller Verlust des Generators in CycleGAN

```
1 def cycle_loss(real_image, cycled_image):
2     loss1 = tf.reduce_mean(tf.abs(real_image - cycled_image))
3     return LAMBDA * loss1
```

Code-Auszug A.9: Zykluskonsistenz Verlust in CycleGAN

Training und Hyperparameter

```
1 def identity_loss(real_image, same_image):
2     loss = tf.reduce_mean(tf.abs(real_image - same_image))
3     return LAMBDA * 0.5 * loss
```

Code-Auszug A.10: Identitätsverlust in CycleGAN

```
1 def train_step(real_x, real_y):
2     with tf.GradientTape(persistent=True) as tape:
3         fake_y = generator_g(real_x, training=True)
4         cycled_x = generator_f(fake_y, training=True)
5
6         fake_x = generator_f(real_y, training=True)
7         cycled_y = generator_g(fake_x, training=True)
8
9         same_x = generator_f(real_x, training=True)
10        same_y = generator_g(real_y, training=True)
11
12        disc_real_x = discriminator_x(real_x, training=True)
13        disc_real_y = discriminator_y(real_y, training=True)
14        disc_fake_x = discriminator_x(fake_x, training=True)
15        disc_fake_y = discriminator_y(fake_y, training=True)
16
17        total_gen_g_loss = generator_loss(real_x, real_y, cycled_x,
18                                           cycled_y, disc_fake_y, same_y)
19        total_gen_f_loss = generator_loss(real_x, real_y, cycled_x,
20                                           cycled_y, disc_fake_x, same_x)
21
22        disc_x_loss = discriminator_loss(disc_real_x, disc_fake_x)
23        disc_y_loss = discriminator_loss(disc_real_y, disc_fake_y)
```

Code-Auszug A.11: CycleGAN Trainingsfunktion, Teil 1

```

1 # Calculate the gradients for generator and discriminator
2 generator_g_gradients = tape.gradient(total_gen_g_loss,
3                                         generator_g.trainable_variables)
4 generator_f_gradients = tape.gradient(total_gen_f_loss,
5                                         generator_f.trainable_variables)
6 discriminator_x_gradients = tape.gradient(disc_x_loss,
7                                         discriminator_x.
8                                         trainable_variables)
9 discriminator_y_gradients = tape.gradient(disc_y_loss,
10                                         discriminator_y.
11                                         trainable_variables)
12
13 # Apply the gradients to the optimizer
14 generator_g_optimizer.apply_gradients(zip(generator_g_gradients,
15                                         generator_g.
16                                         trainable_variables))
17 generator_f_optimizer.apply_gradients(zip(generator_f_gradients,
18                                         generator_f.
19                                         trainable_variables))
20 discriminator_optimizer_x.apply_gradients(zip(
21                                         discriminator_x_gradients,
22                                         discriminator_x.
23                                         trainable_variables))
24 discriminator_optimizer_y.apply_gradients(zip(
25                                         discriminator_y_gradients,
26                                         discriminator_y.
27                                         trainable_variables))
28
29 return total_gen_f_loss, total_gen_g_loss, disc_y_loss, disc_x_loss,
30       fake_x, fake_y, disc_real_x, disc_real_y, disc_fake_x,
31       disc_fake_y

```

Code-Auszug A.12: CycleGAN Trainingsfunktion, Teil 2

```

1 def save_losses(epoch, gen_loss, disc_loss, disc_accuracy, ssim):
2     with open(losses_path, 'a') as f:
3         writer = csv.writer(f)
4         writer.writerow([epoch, gen_loss.result().numpy(), disc_loss.result()
5                         .numpy(), disc_accuracy, ssim])

```

Code-Auszug A.13: Speicherung der Metriken in CSV-Datei

```

1 def calculate_ssime(generator, dataset, num_samples=10):
2     ssim_values = []
3     for input_image, target_image in dataset.take(num_samples):
4         prediction = generator(input_image, training=False)
5         ssim_value = tf.image.ssim(prediction, target_image, max_val
6                                     =2.0)
6         ssim_values.append(ssim_value)
7     return tf.reduce_mean(ssim_values)

```

Code-Auszug A.14: Berechnung des SSIM-Score

```

1 def plot_curve(name, epoch, col2, col_name, num_curves=1):
2     path = f"./satelite/try1/Curves/{name}.jpg"
3     if not os.path.exists("./satelite/try1/Curves/"):
4         os.makedirs("./satelite/try1/Curves/")
5     for i in range(num_curves):
6         plt.plot(epoch, col2.iloc[:, i], label=col_name[i])
7     plt.xlabel('Epoch')
8     plt.ylabel(name)
9     plt.title(f'Epoch vs {name}')
10    plt.legend()
11    plt.savefig(path)
12    plt.show()
13
14 def save_curves():
15     df = pd.read_csv('./satelite/try1/losses.csv')
16     epoch = df.iloc[:, 0]
17     gen_loss = df.iloc[:, [1, 2]]
18     plot_curve("Generator_Verlust", epoch, gen_loss, [
19                 "Generator_F_Verlust", "Generator_G_Verlust"], 2)

```

Code-Auszug A.15: Ausschnitt zur Erstellung einer Verlaufskurve
(CycleGAN Implementierung)

B

Anhang - Modelle

In diesem Abschnitt werden die Modelle der Generatoren und Diskriminatoren aufgelistet.

```

1 def Generator():
2     inputs = tf.keras.layers.Input(shape=[256, 256, 3])
3
4     down_stack = [
5         downsample(64, 4, apply_batchnorm=True), # (batch_size, 128, 128,
6         64)
6         downsample(128, 4), # (batch_size, 64, 64, 128)
7         downsample(256, 4), # (batch_size, 32, 32, 256)
8         downsample(512, 4), # (batch_size, 16, 16, 512)
9         downsample(512, 4), # (batch_size, 8, 8, 512)
10        downsample(512, 4), # (batch_size, 4, 4, 512)
11        downsample(512, 4), # (batch_size, 2, 2, 512)
12        downsample(512, 4), # (batch_size, 1, 1, 512)
13    ]
14
15    up_stack = [
16        upsample(512, 4, apply_dropout=True), # (batch_size, 2, 2, 1024)
17        upsample(512, 4, apply_dropout=True), # (batch_size, 4, 4, 1024)
18        upsample(512, 4, apply_dropout=True), # (batch_size, 8, 8, 1024)
19        upsample(512, 4), # (batch_size, 16, 16, 1024)
20        upsample(256, 4), # (batch_size, 32, 32, 512)
21        upsample(128, 4), # (batch_size, 64, 64, 256)
22        upsample(64, 4), # (batch_size, 128, 128, 128)
23    ]

```

Code-Auszug B.1: Pix2Pix Generator in Tensorflow Teil 1

```
1   initializer = tf.random_normal_initializer(0., 0.02)
2   last = tf.keras.layers.Conv2DTranspose(OUTPUT_CHANNELS, 4,
3                                         strides=2,
4                                         padding='same',
5                                         kernel_initializer=initializer ,
6                                         activation='tanh') # (
7                                         batch_size, 256, 256, 3)
8
9   x = inputs
10
11  # Downsampling through the model
12  skips = []
13  for down in down_stack:
14      x = down(x)
15      skips.append(x)
16
17  skips = reversed(skips[:-1])
18
19  # Upsampling and establishing the skip connections
20  for up, skip in zip(up_stack, skips):
21      x = up(x)
22      x = tf.keras.layers.concatenate([x, skip])
23
24  x = last(x)
25
26  return tf.keras.Model(inputs=inputs, outputs=x)
```

Code-Auszug B.2: Pix2Pix Generator in Tensorflow Teil 2

```

1 def Discriminator():
2     inp = tf.keras.layers.Input(shape=[256, 256, 3], name='input_image')
3     tar = tf.keras.layers.Input(shape=[256, 256, 3], name='target_image')
4
5     x = tf.keras.layers.concatenate([inp, tar])  # (batch_size, 256, 256,
6                                         channels*2)
7
8     down1 = downsample(64, 4, False)(x) # (batch_size, 128, 128, 64)
9     down2 = downsample(128, 4)(down1) # (batch_size, 64, 64, 128)
10    down3 = downsample(256, 4)(down2) # (batch_size, 32, 32, 256)
11    down4 = downsample(512,4)(down3)
12
13    # Zweite letzte Ausgabeschicht
14    initializer = tf.random_normal_initializer(0., 0.02)
15    last = tf.keras.layers.Conv2D(512, 4, padding='same',
16                                kernel_initializer=initializer)(down4)
17    last = tf.keras.layers.BatchNormalization()(last)
18    last = tf.keras.layers.LeakyReLU()(last)
19
20    # Patch-Ausgang
21    patch_out = tf.keras.layers.Conv2D(1, 4, padding='same',
22                                     kernel_initializer=initializer)(last)
23    patch_out = tf.keras.layers.Activation('sigmoid')(patch_out)
24
25    return tf.keras.Model(inputs=[inp, tar], outputs=patch_out)

```

Code-Auszug B.3: Pix2Pix Diskriminator in Tensorflow

```
1 def Generator():
2     inputs = tf.keras.layers.Input(shape=[None,None,3])
3
4     # Layer 1
5     x = layers.Conv2D(64, 7, strides=1, padding='same')(inputs)
6     x = layers.BatchNormalization()(x)
7     x = layers.Activation('relu')(x)
8
9     # Layer 2 + 3
10    x = convolutional_layer(x, 128, 3,2)
11    x = convolutional_layer(x, 256, 3, 2)
12
13    # Residual Blocks
14    for _ in range(6):
15        x = residual_block(x, 256)
16
17    # Layer 10 + 11
18    x = t_convolutional_layer(x, 128, 3, 2)
19    x = t_convolutional_layer(x, 64, 3, 2)
20
21    # Dropout
22    x = tf.keras.layers.Dropout(0.5)(x)
23
24    # Layer 12
25    outputs = layers.Conv2D(3, 7, strides=1, padding='same',
26                           activation='tanh')(x)
27
28    return tf.keras.Model(inputs=inputs, outputs=outputs)
```

Code-Auszug B.4: CycleGAN Generator in Tensorflow

```
1 def Discriminator():
2     model = tf.keras.Sequential()
3
4     # Layer 1
5     model.add(layers.Conv2D(64, 4, strides=2, padding='same',
6         input_shape=(256,256,3)))
7     model.add(layers.BatchNormalization())
8     model.add(layers.Activation('relu'))
9
10    # Layer 2
11    model.add(layers.Conv2D(128, 4, strides=2, padding='same'))
12    model.add(layers.BatchNormalization())
13    model.add(layers.Activation('relu'))
14
15    # Layer 3
16    model.add(layers.Conv2D(256, 4, strides=2, padding='same'))
17    model.add(layers.BatchNormalization())
18    model.add(layers.Activation('relu'))
19
20    # Layer 4
21    model.add(layers.Conv2D(512, 4, strides=2, padding='same'))
22    model.add(layers.BatchNormalization())
23    model.add(layers.Activation('relu'))
24
25    # Layer 5
26    model.add(layers.Conv2D(1, 4, strides=1, padding='same'))
27    model.add(layers.BatchNormalization())
28    model.add(layers.Activation('sigmoid'))
29
30    # Dropout
31    model.add(tf.keras.layers.Dropout(0.5))
32
33    return model
```

Code-Auszug B.5: CycleGAN Diskriminator in Tensorflow

Literaturverzeichnis

- [.2010] *Proceedings of the 27th International Conference on International Conference on Machine Learning.* Madison, WI, USA : Omnipress, 2010 (ICML'10). – ISBN 9781605589077
- [.2017] *Proceedings of the 34th International Conference on Machine Learning - Volume 70.* JMLR.org, 2017 (ICML'17)
- [.2019] *Proceedings of The 7th International Conference on Intelligent Systems and Image Processing 2019.* The Institute of Industrial Application Engineers, 2019 . – ISBN 9784907220198
- [ACB17] ARJOVSKY, Martin ; CHINTALA, Soumith ; BOTTOU, Léon: Wasserstein Generative Adversarial Networks. In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, JMLR.org, 2017 (ICML'17), S. 214–223
- [AMB21] AGGARWAL, Alankrita ; MITTAL, Mamta ; BATTINENI, Gopi: Generative adversarial network: An overview of theory and applications. In: *International Journal of Information Management Data Insights* 1 (2021), Nr. 1, S. 100004. <http://dx.doi.org/10.1016/j.jjimei.2020.100004>. – DOI 10.1016/j.jjimei.2020.100004. – ISSN 26670968
- [Ara22] ARAM YOU, JIN KUK KIM, IK HEE RYU, TAE KEUN YOO: Application of generative adversarial networks (GAN) for ophthalmology imagedomains: a survey. (2022)
- [CWD⁺18] CRESWELL, Antonia ; WHITE, Tom ; DUMOULIN, Vincent ; ARULKUMARAN, Kai ; SENGUPTA, Biswa ; BHARATH, Anil A.: Generative Adversarial Networks: An Overview. In: *IEEE Signal Processing Magazine* 35 (2018), Nr. 1, S. 53–65. <http://dx.doi.org/10.1109/MSP.2017.2765202>. – DOI 10.1109/MSP.2017.2765202. – ISSN 1053–5888
- [CZS] CHU, Casey ; ZHMOGINOV, Andrey ; SANDLER, Mark: *CycleGAN, a Master of Steganography*
- [EO] ECKERLI, Florian ; OSTERRIEDER, Joerg: *Generative Adversarial Networks in finance: an overview*
- [Haz21] HAZEM ABDELMOTAAL, AHMED A. ABDOU, AHMED F. OMAR, DALIA MOHAMED EL-SEBAITY, KHALED ABDELAZEEM: Pix2pix Conditional Generative Adversarial Networks for Scheimpflug Camera Color-Coded Corneal Tomography Image Generation. (2021)
- [HB] HUANG, Xun ; BELONGIE, Serge: *Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization*

- [HGAPC] HITAJ, Briland ; GASTI, Paolo ; ATENIESE, Giuseppe ; PEREZ-CRUZ, Fernando: *PassGAN: A Deep Learning Approach for Password Guessing*. <http://arxiv.org/pdf/1709.00440v3>
- [HHYY20] HONG, Yongjun ; HWANG, Uiwon ; YOO, Jaeyoon ; YOON, Sungroh: How Generative Adversarial Networks and Their Variants Work. In: *ACM Computing Surveys* 52 (2020), Nr. 1, S. 1–43. <http://dx.doi.org/10.1145/3301282>. – DOI 10.1145/3301282. – ISSN 0360–0300
- [HZRS] HE, Kaiming ; ZHANG, Xiangyu ; REN, Shaoqing ; SUN, Jian: *Deep Residual Learning for Image Recognition*
- [Ian14] IAN J. GOODFELLOW, JEAN POUGET-ABADIE, MEHDI MIRZA, BING XU, DAVID WARDE-FARLEY, SHERJIL OZAIR, AARON COURVILLE, YOSHUA BENGIO: Generative Adversarial Nets. (2014)
- [JZJ⁺20] JAIN, Deepak K. ; ZAREAPOOR, Masoumeh ; JAIN, Rachna ; KATHURIA, Abhishek ; BACHHETY, Shivam: GAN-Poser: an improvised bidirectional GAN model for human motion prediction. In: *Neural Computing and Applications* 32 (2020), Nr. 18, S. 14579–14591
- [KB] KINGMA, Diederik P. ; BA, Jimmy: *Adam: A Method for Stochastic Optimization*
- [KSH17] KRIZHEVSKY, Alex ; SUTSKEVER, Ilya ; HINTON, Geoffrey E.: ImageNet classification with deep convolutional neural networks. In: *Communications of the ACM* 60 (2017), Nr. 6, S. 84–90. <http://dx.doi.org/10.1145/3065386>. – DOI 10.1145/3065386. – ISSN 0001–0782
- [LCS⁺21] LIU, Yanxia ; CHEN, Anni ; SHI, Hongyu ; HUANG, Sijuan ; ZHENG, Wanjia ; LIU, Zhiqiang ; ZHANG, Qin ; YANG, Xin: CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 91 (2021), S. 101953. <http://dx.doi.org/10.1016/j.compmedimag.2021.101953>. – DOI 10.1016/j.compmedimag.2021.101953
- [LLW⁺] LIU, Haozhe ; LI, Bing ; WU, Haoqian ; LIANG, Hanbang ; HUANG, Yawen ; LI, Yuexiang ; GHANEM, Bernard ; ZHENG, Yefeng: *Combating Mode Collapse in GANs via Manifold Entropy Estimation*
- [MO] MIRZA, Mehdi ; OSINDERO, Simon: *Conditional Generative Adversarial Nets*
- [NH10] NAIR, Vinod ; HINTON, Geoffrey E.: Rectified Linear Units Improve Restricted Boltzmann Machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Madison, WI, USA : Omnipress, 2010 (ICML'10). – ISBN 9781605589077, S. 807–814

- [ON] O'SHEA, Keiron ; NASH, Ryan: *An Introduction to Convolutional Neural Networks*
- [PJTA] PHILLIP ISOLA ; JUN-YAN ZHU ; TINGHUI ZHOU ; ALEXEI A. EFROS: Image-to-Image Translation with Conditional Adversarial Networks
- [PYY⁺19] PAN, Zhaoqing ; YU, Weijie ; YI, Xiaokai ; KHAN, Asifullah ; YUAN, Feng ; ZHENG, Yuhui: Recent Progress on Generative Adversarial Networks (GANs): A Survey. In: *IEEE Access* 7 (2019), S. 36322–36333. <http://dx.doi.org/10.1109/ACCESS.2019.2905015>. – DOI 10.1109/ACCESS.2019.2905015
- [RMC] RADFORD, Alec ; METZ, Luke ; CHINTALA, Soumith: *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*
- [SC21] SAXENA, Divya ; CAO, Jiannong: Generative adversarial networks (GANs) challenges, solutions, and future directions. In: *ACM Computing Surveys (CSUR)* 54 (2021), Nr. 3, S. 1–42
- [SGZ⁺] SALIMANS, Tim ; GOODFELLOW, Ian ; ZAREMBA, Wojciech ; CHEUNG, Vicki ; RADFORD, Alec ; CHEN, Xi: *Improved Techniques for Training GANs*
- [SSA17] SHARMA, Sagar ; SHARMA, Simone ; ATHAIYA, Anidhya: Activation functions in neural networks. In: *Towards Data Sci* 6 (2017), Nr. 12, S. 310–316
- [SSJ22] SHARMA, Neha ; SHARMA, Reecha ; JINDAL, Neeru: Comparative analysis of CycleGAN and AttentionGAN on face aging application. In: *Sādhanā* 47 (2022), Nr. 1. <http://dx.doi.org/10.1007/s12046-022-01807-4>. – DOI 10.1007/s12046-022-01807-4. – ISSN 0256–2499
- [SVR⁺] SRIVASTAVA, Akash ; VALKOV, Lazar ; RUSSELL, Chris ; GUTMANN, Michael U. ; SUTTON, Charles: *VEEGAN: Reducing Mode Collapse in GANs using Implicit Variational Learning*
- [UVL] ULYANOV, Dmitry ; VEDALDI, Andrea ; LEMPITSKY, Victor: *Instance Normalization: The Missing Ingredient for Fast Stylization*
- [YNDT18] YAMASHITA, Rikiya ; NISHIO, Mizuho ; Do, Richard Kinh G. ; TOGASHI, Kaori: Convolutional neural networks: an overview and application in radiology. In: *Insights into imaging* 9 (2018), Nr. 4, S. 611–629. <http://dx.doi.org/10.1007/s13244-018-0639-9>. – DOI 10.1007/s13244-018-0639-9. – ISSN 1869–4101

Literaturverzeichnis

- [YWB19] YI, Xin ; WALIA, Ekta ; BABYN, Paul: Generative adversarial network in medical imaging: A review. In: *Medical image analysis* 58 (2019), S. 101552. <http://dx.doi.org/10.1016/j.media.2019.101552>. – DOI 10.1016/j.media.2019.101552
- [ZGQZ19] ZHU, Miao M. ; GONG, Shengrong ; QIAN, Zhenjiang ; ZHANG, Lifeng: A Brief Review on Cycle Generative Adversarial Networks. In: *Proceedings of The 7th International Conference on Intelligent Systems and Image Processing 2019*, The Institute of Industrial Application Engineers, 2019. – ISBN 9784907220198, S. 235–242
- [ZPIE] ZHU, Jun-Yan ; PARK, Taesung ; ISOLA, Phillip ; EFROS, Alexei A.: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*
- [ZZS⁺18] ZHENG, Yu-Jun ; ZHOU, Xiao-Han ; SHENG, Wei-Guo ; XUE, Yu ; CHEN, Sheng-Yong: Generative adversarial network based telecom fraud detection at the receiving bank. In: *Neural networks : the official journal of the International Neural Network Society* 102 (2018), S. 78–86. <http://dx.doi.org/10.1016/j.neunet.2018.02.015>. – DOI 10.1016/j.neunet.2018.02.015

Abbildungsverzeichnis

2.1.	Schematische Darstellung der U-Net-Architektur. Die Architektur besteht aus einem Encoder-Teil (links), einem Decoder-Teil (rechts) und Skip-Verbindungen zwischen korrespondierenden Schichten (PJTA).	7
2.2.	Eine mögliche Architektur eines PatchGAN Diskriminators ¹	8
2.3.	Eine Architektur eines CycleGAN Generator. Instanznormalisierung und ReLU Aktivierung erfolgt nach jeder Schicht ²	11
2.4.	Ein Aufbau eines Residualblocks	12
2.5.	(a) Modell des CycleGANs, bestehend aus zwei Generatoren $F : Y \rightarrow X$ und $G : X \rightarrow Y$ und zugehörige adversarielle Diskriminatoren D_X und D_Y , (b) Cycle-Konsistenz $F(G(x)) \approx x$, (c) Cycle-Konsistenz $G(F(y)) \approx y$ (ZPIE)	14
2.6.	Identity-Mapping für (d) Generator G und (e) Generator F	14
2.7.	Beispiel einer Convolutional-Operation mit einem 3x3 Kernel und Stride 1 (YNDT18)	15
6.2.	Generator- und Diskriminatorenverlust des Trainingsdurchlaufs mit 100 Epochen, eigene Erstellung mittels matplotlib.pyplot-Bibliothek	46
6.3.	Diskriminatoren Genauigkeit des Trainingsdurchlaufs mit 100 Epochen, eigene Erstellung mittels matplotlib.pyplot-Bibliothek	47
6.4.	SSIM-Score des Trainingsdurchlaufs mit 100 Epochen, eigene Erstellung mittels matplotlib.pyplot-Bibliothek	48
6.5.	Testergebnisse nach dem Training mit $\lambda = 120$, Lernrate = 0.0002 und 100 Epochen	51
6.6.	Verlauf der Metriken beim Training mit $\lambda = 120$, Lernrate = 0.0002 und 100 Epochen. Die Metriken umfassen: Generator-Verlust (oben links), Diskriminatoren-Genauigkeiten auf den Trainings- und Testdaten (oben rechts), sowie SSIM-Score (unten links) und Zyklus-SSIM-Score (unten rechts). Die Metriken werden in Abhängigkeit von der Anzahl der Epochen auf der x-Achse untersucht.	51
6.7.	CycleGAN-Ergebnisse aus verschiedenen Datensätzen aus dem Berkeley-Repository: 'maps' (links), 'monet2photo' (mitte), 'horse2zebra' (rechts)	52
6.8.	Vergleich der Ergebnisse von Pix2Pix und CycleGAN	56

¹<https://towardsdatascience.com/cyclegan-learning-to-translate-images-without-paired-training-data-1f3a2a2a23>

²<https://towardsdatascience.com/cyclegan-learning-to-translate-images-without-paired-training-data-1f3a2a2a23>

Tabellenverzeichnis

6.1. CycleGAN Ergebnisse unter verschiedenen Hyperparametern mit Batchgröße 1, Teil1	53
6.2. CycleGAN Ergebnisse unter verschiedenen Hyperparametern mit Batchgröße 1, Teil2	53

Code-Auszugs-Verzeichnis

5.1.	Laden eines Datensatzes von einer URL	23
5.2.	Erzeugung eines Tensorflow-Dataset aus der CycleGAN Implementierung	23
5.3.	Vorverarbeitung des Datensatzes: Normalisierung	24
5.4.	Vorverarbeitung des Datensatzes: Jittering	24
5.5.	Integration der vorverarbeiteten Trainingsbilder in Tensorflow-Datasets (CycleGAN Implementierung)	25
5.6.	Downsampling-Schritt in Pix2Pix	26
5.7.	Downsampling-Schicht in Pix2Pix	35
5.8.	Upsampling-Schritt in Pix2Pix	35
5.9.	Upsampling-Schritt in Pix2Pix	36
5.10.	Skip Verbindungen in Pix2Pix	36
5.11.	Diskriminator Pix2Pix	37
5.12.	CycleGAN Generator in Tensorflow	38
5.13.	CycleGAN Diskriminator in Tensorflow	39
5.14.	Initialisierung des BinaryCrossentropy- Verlustfunktion	39
5.15.	Gesamtverlust des Generators in CycleGAN	40
5.16.	Gesamtverlust des Diskriminators in CycleGAN	40
5.17.	Initialisierung der Adam-Optimizers aus der Pix2Pix Implementierung	40
5.18.	Speicherung der generierten Bilder	40
5.19.	Initialisierung der Metriken	41
5.20.	Aktualisierung der Metriken	41
A.1.	Lesen eines Bildes (CycleGAN Implementierung)	61
A.2.	Pix2Pix Generatorverlust	62
A.3.	Residualblock Implementierung	62
A.4.	Pix2Pix Trainingsfunktion	63
A.5.	Residualblock Implementierung	64
A.6.	Convolutional Block in CycleGAN	64
A.7.	Transpose Convolutional Block in CycleGAN	65
A.8.	Adversieller Verlust des Generators in CycleGAN	65
A.9.	Zykluskonsistenz Verlust in CycleGAN	65
A.10.	Identitätsverlust in CycleGAN	66
A.11.	CycleGAN Trainingsfunktion, Teil 1	67
A.12.	CycleGAN Trainingsfunktion, Teil 2	68
A.13.	Speicherung der Metriken in CSV-Datei	68
A.14.	Berechnung des SSIM-Score	69
A.15.	Ausschnitt zur Erstellung einer Verlaufskurve (CycleGAN Implementierung)	69
B.1.	Pix2Pix Generator in Tensorflow Teil 1	71
B.2.	Pix2Pix Generator in Tensorflow Teil 2	72
B.3.	Pix2Pix Diskriminator in Tensorflow	73

Code-Auszugs-Verzeichnis

B.4. CycleGAN Generator in Tensorflow	74
B.5. CycleGAN Diskriminator in Tensorflow	75

Glossar

- **Keras:**

Open-Source-API für Deep Learning, seit TensorFlow 2.0 integraler Bestandteil der TensorFlow Core API.

- **Matplotlib:**

Python-Bibliothek zur Erstellung von statischen, interaktiven und animierten Visualisierungen.

- **Python:**

Skript- und Programmiersprache, die unter Anderem objektorientiertes Programmieren ermöglicht.

- **TensorFlow:**

Open-Source-Framework für maschinelles Lernen und tiefe neuronale Netzwerke, bekannt für seine Skalierbarkeit und umfangreiche Plattformunterstützung.

Arbeitsverteilung

Teilnehmer 1: Elisa Du

Inhalte:

Abstract/Zusammenfassung
Grundlagen - Kapitel (exklusiv Pix2Pix)
Literaturreview
Problembeschreibung - Kapitel
Lösungsbeschreibung - Kapitel (exklusiv Implementierung der Pix2PixGAN-Architektur)
Evaluation - Ergebnisse und objektive Bewertung (exklusiv Pix2Pix)
Evaluation - Vergleich von Pix2Pix und CycleGAN
Anhänge

Teilnehmer 2: Marcel Hoffmann

Inhalte:

Einleitung
Pix2Pix in Grundlagen - Kapitel
Implementierung der Pix2PixGAN-Architektur in Lösungsbeschreibung - Kapitel
Evaluation - Bewertungskriterien
Evaluation - Pix2Pix Ergebnisse und objektive Bewertung
Evaluation - Vergleich von Pix2Pix und CycleGAN (Theoretischer Vergleich)
Fazit und Ausblick