

---

Bachelor-Teamprojekt

Analyse, Design und Implementierung von  
unterschiedlichen Generative Adversarial Network  
(GAN) Architekturen im Bereich der  
Bildverarbeitung

Analysis, design and implementation of different  
Generative Adversarial Network (GAN)  
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 Quellennachweis versehen. Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

---

Ort, Datum

---

Unterschrift



---

# Abstract

---

This is a summary of all the important points and achievements of this work.

# Zusammenfassung

---

Hierbei handelt es sich um eine Zusammenfassung aller wichtige Punkte und Errungenschaften dieser Arbeit.



# Abkürzungsverzeichnis

**GAN** Generative Adversarial Network





# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Grundlagen</b>	<b>3</b>
2.1. Generative Adversarial Networks . . . . .	3
2.2. Pix2Pix . . . . .	5
2.2.1. Pix2Pix-Kernkonzepte . . . . .	5
2.2.2. Anwendungen von Pix2Pix . . . . .	9
2.3. CycleGAN . . . . .	9
2.3.1. CycleGAN - Kernkonzepte . . . . .	9
2.3.2. Anwendungen von CycleGAN . . . . .	11
<b>3. Literaturreview</b>	<b>13</b>
<b>4. Problembeschreibung</b>	<b>15</b>
4.1. Bibliotheken . . . . .	15
4.2. Datenbeschaffung und -vorverarbeitung . . . . .	15
4.3. Architekturentwurf . . . . .	15
<b>5. Lösungsbeschreibung</b>	<b>17</b>
5.1. Bewertungskriterien . . . . .	17
5.2. Pix2Pix: Ergebnisse und objektive Bewertung . . . . .	17
5.3. CycleGAN: Ergebnisse und objektive Bewertung . . . . .	17
5.4. Vergleich und Bewertung von Pix2Pix und CycleGAN . . . . .	17
<b>6. Evaluation</b>	<b>19</b>
<b>7. Fazit und Ausblick</b>	<b>21</b>
7.1. Fazit . . . . .	21
7.2. Ausblick . . . . .	21
<b>A. Anhang - Code</b>	<b>23</b>
<b>B. Anhang - Dokumentationen</b>	<b>25</b>
<b>Verzeichnisse</b>	<b>27</b>
Literaturverzeichnis . . . . .	27
Abbildungsverzeichnis . . . . .	29
Tabellenverzeichnis . . . . .	31
Code-Auszugs-Verzeichnis . . . . .	33
Glossar . . . . .	35



## 1

## Einleitung

Hier wird in die Arbeit eingeleitet.



## 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. 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.

### 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 Mehrschichtnetz implementiert, die aus Faltungsschichten und vollständig verbundenen Schichten bestehen.

### Generator

Der Generator dient zur Erzeugung künstlicher Daten wie Bilder und Texte. Der Generator 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.

### Diskriminator

Die Unterscheidung zwischen echten und unechten Eingaben ist Aufgabe des Diskriminators. Der Diskriminator kann sowohl künstliche als auch reale Daten verwenden. Wenn der Diskriminator nicht mehr richtig unterscheiden kann, wird er als

konvergierend bezeichnet. 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.

## Training

Durch das Finden von Parametern für beide Netze wird das Training durchgeführt. Ziel ist die Optimierung beider Netze durch Anwendung von Backpropagation zur Verbesserung der Parameter. Das Training wird oft als schwierig und instabil beschrieben, da es einerseits schwierig erscheint, beide Modelle konvergieren zu lassen. Auf der anderen Seite kann der Generator sehr ähnliche Muster für verschiedene Eingaben erzeugen, und der Diskriminatorverlust kann schnell gegen Null konvergieren, so dass es keinen zuverlässigen Weg für die Gradientenaktualisierung zum Generator gibt. Zur Lösung dieser Probleme wurden verschiedene Ansätze vorgeschlagen, wie z.B. die Verwendung heuristischer Verlustfunktionen. Eine weitere Möglichkeit, die von Sonderby et al. vorgeschlagen wurde, besteht darin, den Datensatz vor der Verwendung zu verrauschen.

## Adversarial Loss

Der Erfolg von GAN liegt zum Einen an dem verwendeten Adversarial Verlustfunktion. Dieser

// TODO: PAPER ABOUT THAT

## Anwendungen

GAN wurde ursprünglich für unüberwachtes maschinelles Lernen entwickelt. Die Architektur liefert jedoch ebenso gute Ergebnisse beim halbüberwachten Lernen und beim Reinforcement Learning. Aus diesem Grund wird sie in einer Vielzahl von Bereichen wie dem Gesundheitswesen, dem Maschinenbau und dem Bankwesen eingesetzt. Beispielsweise wird GAN in der Medizin zur Erkennung und Behandlung chronischer Krankheiten eingesetzt. Aber auch die Identifikation von 3D-Objekten und die Generierung von realen Bildern und Texten ist durch den Einsatz von GANs möglich.

## Limitationen

Die Tatsache, dass ein Generative Adversarial Network in der Lage ist, Inhalte zu erzeugen, die nahezu identisch mit realen Inhalten sind, kann in der realen Welt zu Problemen führen, insbesondere bei der menschlichen Bildsynthese. Diese Bilder können von Betrügern verwendet werden, um falsche Profile in sozialen Medien zu erstellen. Auch dies kann durch den Einsatz von GANs verhindert werden, in-

dem einzigartige und pragmatische Bilder von nicht existierenden Personen erzeugt werden.

## 2.2. Pix2Pix

Pix2Pix, initiiert von Isola et al., 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.

Pix2Pix fungiert hier als Generative Adversarial Network (GAN), spezialisiert auf diverse Formen der Bildübersetzung. Darunter fallen die Umwandlung von Schwarz-Weiß-Fotos in Farbbilder, die Transformation von Skizzen in realistische Bilder, und relevant für diese Arbeit, die Konvertierung von Satellitenbildern in kartographische Darstellungen, ähnlich den Visualisierungen von Google Maps.

Die Architektur von Pix2Pix besteht aus einem Generator und einem Diskriminator. Der Generator, der eine U-Net-Architektur verwendet, besteht aus einem Encoder und einem Decoder. Der Encoder komprimiert das Eingangsbild schrittweise zu einer niedrigdimensionalen Repräsentation, während der Decoder diese dazu nutzt, das Ausgangsbild zu rekonstruieren. Skip-Verbindungen zwischen Encoder und Decoder helfen dabei, sowohl globale als auch lokale Informationen im generierten Bild zu bewahren.

Der Diskriminator nimmt die Form eines PatchGAN-Modells an und bewertet Patches des Bildes, indem er die Wahrscheinlichkeit für die Echtheit jedes Patches ausgibt. Dies ermöglicht die Anwendung des Diskriminators auf Bilder unterschiedlicher Größen. Im Zuge des adversariellen Trainingsprozesses passen sowohl der Generator als auch der Diskriminator ihre Fähigkeiten fortlaufend an. Während der Generator lernt, immer realistischere Übersetzungen zu erzeugen, wird der Diskriminator stetig besser darin, zwischen echten und generierten Bildern zu unterscheiden.

### 2.2.1. Pix2Pix-Kernkonzepte

#### Generator

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 angeklügelte Kombination aus Encoder- und Decoder-Strukturen sowie durch den Einsatz von Skip-Verbindungen aus (IZZE).

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 (IZZE).

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 (IZZE).

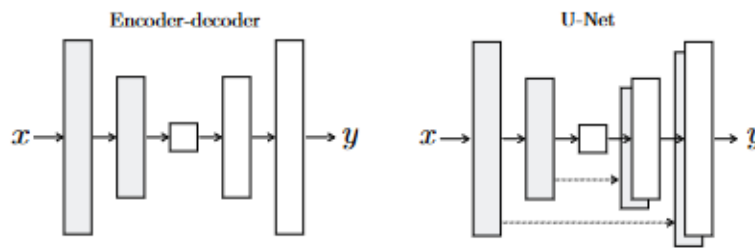
Somit kann die U-Net-Architektur durch ihre Kombination aus Encoder-Decoder-Struktur und Skip-Verbindungen ein effektives Werkzeug für die Bildsegmentierung darstellen. Ihre Fähigkeit, sowohl globale Muster als auch feine Details zu berücksichtigen, macht sie zu einer bevorzugten Wahl für viele Bildverarbeitungsaufgaben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

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, 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 (IZZE).

In der Pix2Pix Technologie dient diese U-Net-Architektur als Generator. Er ist das zentrale Element, das für die Bild-zu-Bild-Übersetzung verantwortlich ist. Die Wahl der U-Net-Struktur für den Generator liegt in ihrer Fähigkeit, feinere Details und Kontextinformationen aus dem Eingangsbild beizubehalten, was für die Bild-zu-Bild-Übersetzung von entscheidender Bedeutung ist. Die Encoder-Decoder-Struktur des U-Net ermöglicht es dem Generator, den globalen Kontext des Bildes zu erfassen, während die Skip-Verbindungen sicherstellen, dass auch lokale Details im resultierenden Bild berücksichtigt werden (IZZE).

### Diskriminator

Im adversariellen Lernprozess spielen Generatoren und Diskriminatoren eine entscheidende Rolle. Während der Generator versucht, Daten zu erzeugen, die von echten Daten kaum zu unterscheiden sind, evaluiert der Diskriminator die vom Generator erzeugten Daten und versucht, zwischen echten und gefälschten Daten zu unterscheiden (IZZE).



**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.

Im Kontext von Generative Adversarial Networks (GANs), insbesondere im speziellen Fall des Pix2Pix GANs, spielt der PatchGAN-Diskriminator eine besonders wichtige Rolle. Der zentrale Unterschied dieses Diskriminators zu traditionellen 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 (IZZE).

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 traditioneller 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 traditionellen Diskriminatoren auftreten können (IZZE).

Eine Metrik, die oft verwendet wird, um die Leistung des Diskriminators zu beurteilen, ist der FCN-Score. Dieser bewertet die Qualität der vom Generator erzeugten Bilder. Ein hoher FCN-Score zeigt, dass der Diskriminator erfolgreich echte von gefälschten Bildern unterscheiden kann.

Der PatchGAN-Diskriminator kann wenn er effektiv eingesetzt wird, zu besseren und realistischeren Bildern im adversariellen 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 (IZZE).

## L1-Verlustfunktion

Die L1-Verlustfunktion, auch bekannt als Mean Absolute Error (MAE), spielt eine entscheidende Rolle im Pix2Pix-Modell, einem bedingten Generative Adversari-

al Network (cGAN) für Bild-zu-Bild-Übersetzungen. Diese Verlustfunktion misst den durchschnittlichen absoluten Unterschied zwischen den vorhergesagten und den tatsächlichen Werten, wodurch sie die Genauigkeit der generierten Bilder verbessert, insbesondere im Hinblick auf die niedrigen Frequenzen im Bild. Die L1-Verlustfunktion trägt somit maßgeblich zur Bewahrung der strukturellen Integrität und des Kontexts des Bildes bei (IZZE).

Die Verwendung des L1-Verlusts zusätzlich zum adversariellen Verlust im Pix2Pix-Modell ist von entscheidender Bedeutung. Während der adversarielle Verlust darauf abzielt, die generierten Bilder realistisch erscheinen zu lassen, konzentriert sich der L1-Verlust auf die Genauigkeit der niedrigen Frequenzen, um die strukturelle Integrität und den Kontext des Bildes zu bewahren. Diese Kombination ermöglicht es, sowohl die niedrigen als auch die hohen Frequenzen im Bild effektiv zu erfassen, was zu generierten Bildern führt, die sowohl strukturell korrekt als auch visuell ansprechend sind (IZZE).

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 (IZZE).

Darüber hinaus hat sich die Kombination von L1-Verlust und adversariellen Verlust im Pix2Pix-Modell als nützlich für eine Vielzahl von Bild-zu-Bild-Übersetzungsproblemen erwiesen, einschließlich semantischer Segmentierung und Farbgebung. Durch die effektive Erfassung sowohl der niedrigen als auch der hohen Frequenzen im Bild trägt das Pix2Pix-Modell dazu bei, die Qualität der generierten Bilder zu verbessern und ihre Anwendbarkeit auf verschiedene Probleme zu erweitern (IZZE).

## Training

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 1-3 Kanäle haben und müssen nicht in einer bestimmten Weise vorverarbeitet werden, da das Modell direkt auf den Rohpixeln arbeitet. Die Trainingsdaten können aus verschiedenen Quellen stammen und unterschiedliche Größen haben. Beispielsweise besteht der Satellit-To-Image-Datensatz aus 1092 Bildern.

Das Training von Pix2Pix-Modellen erfolgt iterativ, wobei Generator und Diskriminator abwechselnd trainiert werden. Der Generator versucht, Bilder zu erzeugen, die nicht von echten Bildern unterschieden werden können, während der Diskriminator versucht, echte Bilder von generierten Bildern zu unterscheiden. Diese dynamische Interaktion führt dazu, dass beide Netzwerke im Laufe der Zeit besser werden und letztendlich dazu beitragen, dass der Generator hochwertige Bilder erzeugt (IZZE).

### 2.2.2. Anwendungen von Pix2Pix

Hier können Sie über die Anwendungen von Pix2Pix schreiben.

## 2.3. CycleGAN

CycleGAN, eingeführt von Jun-Yan Zhu et al. im Jahr 2017, repräsentiert eine wegweisende Entwicklung im Bereich des maschinellen Lernens, insbesondere in der Bildübersetzung zwischen unpaaren Domänen. Dieses Kapitel erläutert im Detail die Mechanismen, die CycleGAN ermöglichen, und vertieft sich in die komplexen Wechselwirkungen zwischen Generatoren und Diskriminatoren. CycleGAN erweitert das klassische GAN-Paradigma, indem es Generatoren und Diskriminatoren einführt, die speziell darauf abzielen, Bildübersetzungen zwischen zwei unterschiedlichen Domänen zu erzeugen. Der Generator  $G$  transformiert Bilder aus Domäne  $X$  in Domäne  $Y$ , während der Generator  $F$  den umgekehrten Prozess durchführt. Diese Transformationen erfolgen ohne die Notwendigkeit für gepaarte Trainingsdaten.

### 2.3.1. CycleGAN - Kernkonzepte

#### Architektur der Generatoren

Die Architektur der Generatoren in CycleGAN spielt eine entscheidende Rolle bei der erfolgreichen Durchführung von Bildübersetzungen zwischen unterschiedlichen Domänen. Typischerweise basieren die Generatoren auf dem Residual-Netzwerk (ResNet)-Ansatz, der für seine Fähigkeit, tiefe neuronale Netzwerke zu trainieren, bekannt ist.

ResNet-Blöcke, bestehend aus Convolutional-Schichten, Normalisierungsschichten und Aktivierungsfunktionen, ermöglichen den Generatoren, komplexe Transformationen durchzuführen. Die Nutzung von ResNet-Blöcken erleichtert auch das Training tiefer Netzwerke, was wichtig ist, um hochwertige Übersetzungen zu erreichen.

#### Architektur der Diskriminatoren

Ebenso wie beim Pix2Pix ist die gängige Architektur für die Diskriminatoren in CycleGAN der PatchGAN.

Die Diskriminatoren bestehen in der Regel aus Convolutional-Schichten, gefolgt von Normalisierungsschichten und Aktivierungsfunktionen. In einigen Implementierungen von CycleGAN wird statt der üblichen Batch-Normalisierung die Instance Normalisierung bevorzugt.

Die Instance Normalisierung ist eine Variante der Normalisierung, die auf Instanzebene durchgeführt wird. Im Gegensatz zur Batch-Normalisierung, bei der die Normalisierung über die gesamte Batch-Dimension erfolgt, betrachtet die Instance Normalisierung jede Instanz oder jedes Bild einzeln. Dies kann besonders vorteilhaft sein, wenn die statistischen Eigenschaften der einzelnen Instanzen variieren. Die Verwendung von Instance Normalisierung in den Diskriminatoren von CycleGAN kann dazu beitragen, eine stabilere und konsistentere Konvergenz während des Trainings zu erreichen.

### Training

Das Training von CycleGAN erfolgt durch einen kompetitiven Prozess. Die Generatoren  $G$  und  $F$  konkurrieren mit den zugehörigen Diskriminatoren  $D_X$  und  $D_Y$ .  $D_X$  versucht, zwischen den von  $F$  generierten und den echten Bildern aus  $X$  zu unterscheiden, während  $D_Y$  die von  $G$  generierten Bilder von den echten aus der Domäne  $Y$  unterscheidet. Die adversarialen Verluste werden dabei optimiert, um sicherzustellen, dass die generierten Bilder für die Diskriminatoren kaum von den echten zu unterscheiden sind.

### Cycle - Konsistenz

Der CycleGAN führt zusätzlich einen Cycle-Konsistenz ein. Diese gewährleistet, dass die Übersetzungen zwischen den Domänen sowohl vorwärts ( $X$  nach  $Y$ ) als auch rückwärts ( $Y$  nach  $X$ ) konsistent sind. Die Cycle-Konsistenz fördert die Generierung qualitativ hochwertiger und konsistenter Übersetzungen.

Die Kernidee besteht darin, dass nach der Übersetzung von  $X$  nach  $Y$  und zurück nach  $X$  das resultierende Bild dem ursprünglichen  $X$  entsprechen sollte. Dabei wird die Differenz zwischen dem ursprünglichen Bild  $x$  mit dem zyklisch übersetzten Bild  $F(G(x))$  mittels der L1-Verlustfunktion minimiert.

Diese zyklische Konsistenz stellt sicher, dass die Generatoren konsistente und realitätsnahe Übersetzungen zwischen den Domänen erzeugen.

### Identity - Loss

Zusätzlich zu adversarialen und zyklischen Verlusten spielt der Identitätsverlust eine bedeutende Rolle im Training von CycleGAN. Der Identitätsverlust zielt darauf ab, sicherzustellen, dass die Transformationen nicht zu stark sind und wichtige Merkmale der Ausgangsbilder bewahrt bleiben. Dieser Verlust wird durch die Verwendung von Bildern der jeweiligen Ausgangsdomäne als Eingabe für den Generator und die Berechnung des Verlusts zwischen den generierten und den Eingangsbildern erreicht.

Die Identitätsverluste für  $G$  und  $F$  werden als zusätzliche Terme in die Gesamtverlustfunktion integriert. Dieser Mechanismus stellt sicher, dass der Generator nicht unerwünschte Veränderungen vornimmt und hilft, feine Details und Strukturen in den generierten Bildern zu erhalten.

### 2.3.2. Anwendungen von CycleGAN

CycleGAN hat sich als äußerst vielseitiges Modell erwiesen und findet in verschiedenen Anwendungsbereichen Anwendung. Die Fähigkeit, Bildübersetzungen zwischen unpaaren Domänen durchzuführen, hat zu zahlreichen innovativen Anwendungen geführt.

Eine der prominentesten Anwendungen von CycleGAN ist die Bild-zu-Bild-Übersetzung. Dies umfasst die Transformation von Bildern zwischen verschiedenen Stilen, Szenarien oder Kunstwerken. Zum Beispiel kann CycleGAN dazu verwendet werden, Fotos in den Stil berühmter Kunstwerke zu überführen, was einen einzigartigen und kreativen Ansatz zur Bildbearbeitung ermöglicht. Der Stiltransfer ist ein weiteres aufregendes Anwendungsgebiet von CycleGAN. Hierbei können Stile von einem Bild auf ein anderes übertragen werden, ohne dass gepaarte Trainingsdaten benötigt werden. Dies ermöglicht es, den Stil eines Gemäldes auf ein Fotografiebild anzuwenden oder umgekehrt. CycleGAN kann auch für die Übersetzung zwischen verschiedenen Farbdomänen verwendet werden. Zum Beispiel könnte es genutzt werden, um Schwarz-Weiß-Bilder in farbige Versionen zu übersetzen oder den Farbton von Bildern anzupassen, ohne dass dazu paare Trainingsdaten notwendig sind.



# 3

## Literaturreview

Hier wird der Inhalt der Arbeit präsentiert.





# 4

## Problembeschreibung

4.1. Bibliotheken

4.2. Datenbeschaffung und -vorverarbeitung

4.3. Architekturentwurf

Pix2Pix

CycleGAN



# 5

## Lösungsbeschreibung

- 5.1. Bewertungskriterien
- 5.2. Pix2Pix: Ergebnisse und objektive Bewertung
- 5.3. CycleGAN: Ergebnisse und objektive Bewertung
- 5.4. Vergleich und Bewertung von Pix2Pix und CycleGAN





# Evaluation

Hier wird die Arbeit evaluiert.



# 7

## Fazit und Ausblick

Hier wird ein Fazit und ein Ausblick gegeben.

### 7.1. Fazit

Fazit.

### 7.2. Ausblick

Ausblick.







## Anhang - Code

Hier sehen Sie den gesamten Quellcode!



# B

## Anhang - Dokumentationen

Hier sehen Sie die gesamten Dokumentationen zu den erstellten Programmen.



# Literaturverzeichnis

- [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.jjime.2020.100004>. – DOI 10.1016/j.jjime.2020.100004. – ISSN 26670968
- [CWD<sup>+</sup>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
- [IZZE] ISOLA, Phillip ; ZHU, Jun-Yan ; ZHOU, Tinghui ; EFROS, Alexei A.: Image-to-Image Translation with Conditional Adversarial Networks



# 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. . . 7





# Tabellenverzeichnis



# Code-Auszugs-Verzeichnis



# Glossar

- **C++:**  
Hardwarenahe, objektorientierte Programmiersprache.
- **HTML:**  
Hypertext Markup Language - textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente.
- **HTTP:**  
Hypertext Transfer Protocol - Protokoll zur Übertragung von Daten auf der Anwendungssicht über ein Rechnernetz.
- **iARS:**  
innovative Audio Response System - System mit zwei Applikationen (iARS-master-App; iARS-student-App), dass sich zum Einsetzen von e-TR-ainer-Inhalten in Vorlesungen eignet.
- **ISO:**  
Internationale Vereinigung von Normungsorganisationen.
- **JavaScript:**  
Skriptsprache zu Auswertung von Benutzerinteraktionen.
- **Konstruktor:**  
Beim Erzeugen einer Objektinstanz aufgerufene Methode zum Initialisieren von Eigenschaften.
- **MySQL:**  
Relationales Datenbankverwaltungssystem.
- **OLAT:**  
Online Learning and Training - Lernplattform für verschiedene Formen von webbasiertem Lernen.
- **OOP:**  
Objektorientierte Programmierung - Programmierparadigma, nach dem sich die Architektur eine Software an realen Objekten orientiert.
- **Open Source:**  
Software, die öffentlich von Dritten eingesehen, geändert und genutzt werden kann.
- **PHP:**  
Skriptsprache zur Erstellung von Webanwendungen.
- **Python:**  
Skript- und Programmiersprache, die unter Anderem objektorientiertes Programmieren ermöglicht.

- **Shell:**  
Shell oder auch Unix-Shell - traditionelle Benutzerschnittstelle von Unix-Betriebssystemen
- **Spyder:**  
Entwicklungsumgebung für wissenschaftliche Programmierung in der Programmiersprache Python.
- **SymPy:**  
Python-Bibliothek für symbolische Mathematik.