

# Rekonstruktion von 3D-Modellen mit Generative Adversarial Networks

**Andreas Wiegand**

Masterthesis im Schwerpunkt künstliche Intelligenz



Gutachterin: Prof. Dr. Ute Schmid

in Zusammenarbeit mit  
Fraunhofer-Institut für Integrierte Schaltungen IIS/EZRT  
betreut von Oliver Scholz und Franz Uhrmann

Fakultät für Angewandte Informatik, insbesondere Kognitive Systeme  
Otto-Friedrich-Universität Bamberg  
Bamberg  
28.Februar.2019

# Table of Contents

Rekonstruktion von 3D-Modellen mit Generative Adversarial Networks ..	4
<i>von Andreas Wiegand</i>	
1 Einleitung .....	5
1.1 Zielsetzung .....	6
1.2 Überblick .....	6
2 Grundlagen und ähnliche Arbeiten .....	7
2.1 Künstliche Neuronale Netzwerke .....	7
2.1.1 Ziel-Funktion .....	10
2.1.2 Backpropagation Algorithmus .....	10
2.1.3 Momentum .....	12
2.1.4 Regularization .....	12
2.1.5 Batch Normalisation .....	13
2.2 Datenformat Punktwolken .....	14
2.2.1 Datenaufnahme von Tabakpflanzen .....	16
2.2.2 Die Schwierigkeit bei mit 3D-Data bei Machine Learning	
Ansätzen .....	18
2.3 Convolution Neural Networks .....	18
2.4 Autoencoder .....	21
2.5 Generative Adversarial Network .....	24
2.5.1 Probleme mit GANs .....	27
2.5.2 Lösungsansätze für die GAN Probleme .....	27
2.6 Conditional-GAN .....	30
2.7 3D-GAN .....	31
2.8 Rekonstruktion von Daten .....	32
3 Methoden .....	33
3.1 Datensatz 1. Versuchsaufbau .....	34
3.2 Datensatz 2. Versuchsaufbau .....	35
3.3 Versuchsaufbau 1 - GAN .....	37
3.4 Versuchsaufbau 2 - CGAN für Punktwolkenrekonstruktion .....	38
4 Evaluation und Ergebnisse .....	40
4.1 Ergebnisse - Versuchsaufbau 1 .....	40
4.2 Ergebnisse - Versuchsaufbau 2 .....	48
5 Zusammenfassung und Diskussion .....	58
Literatur .....	62

**Abb.** Abbildung

**C-GAN** Conditional Adversarial Networks

**CD** Chamfer Distance

**EMD** Earth Mover Distance

**KNN** Künstliches Neuronales Netzwerk

**GAN** Generativ Adversarial Network

**DC GAN** Deep Convolution Generative Adversarial Network

**CNN** Convolutional Neural Network

**KL** Kulback-Leibler Divergenz

# Rekonstruktion von 3D-Modellen mit Generative Adversarial Networks

von Andreas Wiegand

Masterthesis im Schwerpunkt künstliche Intelligenz

**Zusammenfassung.** Generative Adversarial Network(GAN) ist ein künstliches neuronales Netzwerk(KNN) aus dem Bereich der generativen Modelle. Die Aufgabe des GANs ist es, die Wahrscheinlichkeitsverteilung von Trainingsdaten zu erlernen und anschließend neue Samples aus dieser Wahrscheinlichkeitsverteilung zu generieren. Ziel der hier vorliegenden Arbeit ist es, das Konzept der GANs auf 3D-Punktwolken von Tabakblättern anzuwenden, um die Wahrscheinlichkeitsverteilung von 3D-Daten zu erlernen(Goodfellow et al., 2014). Des weiteren soll geprüft werden, ob es mit Hilfe von GANs möglich ist, Punktwolken zu ergänzen, welche beim Erfassen durch ein Scanverfahren unvollständig erfasst worden sind. Dabei werden auf Verfahren zurückgegriffen welche bei 2D-Daten, wie Bilder, für Rekonstruktion verwendet worden ist, so zum beispielsweise Conditional-GANs(C-GAN) (Dong, Loy, He, & Tang, 2016a). Diese Unvollständigkeit kann zustande kommen, wenn Artefakte beim Scannen die Sicht auf das zu scannende Objekt verdecken und dieses dann unvollständig aufgenommen wird. Da im verwendeten Trainingsdatensatz nicht genügend valide Daten vorhanden sind, wurde die Datenerhebung durch das Einfügen von Artefakte in die Tabakpunktwolken simuliert. Die Differenzmenge zwischen Tabakpunktwolke und Artefakt simuliert dann das nicht vollständig gescannte Tabakblatt. Die Ergebnisse dieser Arbeit zeigen auf, dass es möglich ist die Punktwolken von Tabakblättern zu erlernen und neue Blätter zu produzieren. Die Qualität der generierten Daten hat jedoch noch Potential zur Steigerung, vergleicht man diese mit den Ergebnissen von Achlioptas, Panos und Diamanti(Achlioptas, Diamanti, Mitliagkas, & Guibas, 2018). Die vollständige Rekonstruktion mit Hilfe von GANS konnte nicht produziert werden. Es zeichnet sich jedoch eine Qualitätssteigerung der Rekonstruktion, durch unterschiedliche Aufbauten von GANs, ab und die Möglichkeit durch eine Veränderung des GAN Aufbaus in zukünftigen Arbeiten, die vollständige Rekonstruktion zu ermöglichen. Durch einen Nebeneffekt, welcher durch die Zielfunktion von Autoencoder gegeben ist, ist es möglich Lücken in der Punktwolke eines Tabakblattes zu schließen und dieses zu rekonstruieren. Dabei handelt es sich jedoch nur um einen Nebeneffekt und es kann in dieser Arbeit nicht gezeigt werden, dass dies als generelle Lösung für die Rekonstruktion von 3D-Punktwolken heran genommen werden kann.

## 1 Einleitung

Um exaktere Prognosen über Ernteerträge zu treffen und die Früherkennung von Krankheiten an Pflanzen zu verbessern, werden 3D-Scanverfahren eingesetzt. Diese Verfahren scannen Pflanzen und nutzen die gewonnenen Daten um aussagekräftige Faktoren und Probleme zu ermitteln, die mit Hilfe der Daten behoben werden können. Um phänotypische Eigenschaften von Pflanzen heraus zu finden, können Machine Learning Ansätze verwendet werden. Diese können dazu beitragen, Korrelationen zwischen Input-Daten und Eigenschaften der Pflanze, wie Ertrag, Krankheitsresistenz oder Nährstoffverbrauch, heraus zu finden.

Jedoch kommt es bei den Scanverfahren zu Verdeckung von Pflanzen, so dass nicht die komplette Pflanze also Datenmodel generiert wird. Dies ist ein Informationsverlust, den man verhindern möchte. Um diesem Verlust von Daten entgegen zu wirken, muss die Möglichkeit geprüft werden, ob die Daten vervollständigt und rekonstruiert werden können. In der hier vorliegenden Arbeit soll ein Ansatz überprüft werden, welcher es ermöglichen soll 3D-Modelle, die als Punktwolken vorhanden sind, zu rekonstruieren. Ein Ansatz der dafür herangezogen werden kann, ist Deep Learning. Deep Learning kann dabei helfen den latenten Raum der 3D-Modelle zu erlernen und eine Rekonstruktion von nicht vollständigen Datensätzen zu erhalten.

In den letzten Jahren haben sich im Deep Learning Bereich besonders die discriminativen Modelle hervorgehoben, welche Input Daten wie Bilder, Audio oder Text bestimmten Klassen zuordnen. Der Grund für das wachsende Interesse liegt in der niedrigen Fehlerrate bei discriminativen Aufgaben, im Vergleich zu anderen Maschine Learning Ansätzen, wie Decision Trees oder Markov Chains (Goodfellow, Bengio, Courville, & Bengio, 2016). Die Modelle lernen eine Funktion, welche Input Daten  $X$  auf ein Klassen Label  $Y$  abbildet. Die Modelle werden dabei von künstlichen Neuralen Netzwerken repräsentiert. Man kann auch sagen, das Model lernt die bedingte Wahrscheinlichkeitsverteilung  $P(Y|X)$  (Ng & Jordan, 2002). Generative Modelle haben die Aufgabe die Wahrscheinlichkeitsverteilung von Trainingsdaten zu erlernen. Diese generativen Modelle lernen eine multivariate Verteilung  $P(X, Y)$ , was dank der Bayes Regel auch zu  $P(Y|X)$  umgeformt werden kann. Somit kann dieses Modell auch für discriminative Aufgaben herangezogen werden. Gleichzeitig können aber neue  $(x, y)$  Paare erzeugt werden, was zu dem Ergebnis von neuen Datensätzen führt, welche nicht Teil der Trainingssample sind (Ng & Jordan, 2002). In dieser Arbeit wird speziell auf GAN, aus der Vielzahl von generativen Modellen, eingegangen. Diese wurden von Goodfellow (Goodfellow et al., 2014) entwickelt und ebneten den Weg für Variationen, welche auf der Grundidee von GANs aufbauen. GANs wurden bereits verwendet um 2D-Daten, wie Bilder, zu rekonstruieren, wie beispielsweise im Paper (Dong, Loy, He, & Tang, 2016b) gezeigt wurde, in welchem auf super-hochauflösenden Bildern Artefakte entfernt und das Bild so wieder zu seinem Urzustand zurück geführt wurde.

## 1.1 Zielsetzung

In dieser Arbeit wird versucht die derzeit bestehende Forschungslücke anzugehen, welche in der Verarbeitung von Punktwolken durch künstliche Neuronale Netzwerke besteht. Der Bereich, zu dem noch keine Lösung gefunden werden konnte, ist die vollständige Rekonstruktion von Punktwolken. Bei der Rekonstruktion wird eine Punktwolke, welche von ihrem Urzustand abweicht, transformiert und wieder in diesen gebracht. Dieser Zustand kann durch Artefakte entstehen, welche beim Erfassen der Punktwolke im Scanverfahren den Blickwinkel verdecken. Ein Vorverarbeitungsschritt ist es zu prüfen, ob es möglich ist den latenten Raum des Urzustandes dieser Punktwolke zu erlernen. Die Forschungsfragen der hier vorliegenden Arbeit lassen sich definieren als:

**Fragestellung 1** Können durch GANs 3D-Punktwolken von Tabakblättern erlernt werden um neue Datensätze zu generieren?

**Fragestellung 2** Können durch GANs 3D-Punktwolken von Tabakblättern, die von ihrem Urzustand abgebracht wurden, rekonstruiert werden?

## 1.2 Überblick

Diese Arbeit ist folgender Maßen strukturiert. In Kapitel 2 "Grundlagen und ähnliche Arbeiten" werden theoretische Grundlagen welche für diese Arbeit benötigt werden, genauer beleuchtet. Außerdem sollen vorangeegangenen Arbeiten welche einfluß auf diese Ausüben vorgestellt werden um zu veranschaulichen aus welchen Gründen diese Funktionieren kann.

**Kapitel 1 - Grundlage und ähnliche Arbeiten** Zunächst wird auf die fundamentalen Grundlagen zu dieser Arbeit eingegangen, wie beispielsweise auf künstliche Neuronale Netzwerken, welche für das Verständnis von Generativ Adversarial Networks obligatorisch sind. Daraufhin werden GANs genauer spezifiziert, auf welchen Theorien sie aufbauen und aus welchen Modulen sie zusammen gesetzt sind. Auch werden Punktwolken genauer spezifiziert und aufgezeigt, auf welche Probleme Machine Learning Ansätze bei diesen stoßen. Zuletzt wird in diesem Kapitel auf die bisherigen Ergebnisse bei der Rekonstruktion von Daten eingegangen.

**Kapitel 2 - Methoden** In diesem Kapitel werden die einzelnen Versuchsaufbauten definiert, welche versuchen die Ziele der Datenverteilung von Tabakblättern zu erlernen, sowie die Rekonstruktion von Tabakblätter zu prüfen.

**Kapitel 3 - Evaluation und Ergebnisse** Es werden die Ergebnisse aus den Versuchsaufbauten 1 und 2 evaluiert und auf ihre Verwendbarkeit in der Praxis hingewiesen.

**Kapitel 4 - Zusammenfassung und Diskussion** Gibt eine Konklusion über die Arbeit und diskutiert die Ergebnisse kritisch. Des weiteren wird ein Ausblick drauf gegeben, inwiefern das Ergebnis für zukünftige Arbeiten von Relevanz ist.

## 2 Grundlagen und ähnliche Arbeiten

Im folgenden Kapitel wird auf theoretische Grundlagen eingegangen, welche zum Verständnis für Arbeit benötigt werden. Zunächst werden generative Modelle im Allgemeinen vorgestellt, welche den Grundgedanken der Datengeneration für GANs aufzeigen und mit deren Hilfe es möglich gemacht werden soll unkenntliche beziehungsweise beschädigte Modelle von Objekten wieder herzustellen. Diese Theorie baut zunächst auf Datenstruktur von künstlichen Neuronalen Netzwerken auf, welche durch verschiedene Modelle dargestellt werden können. Im Kapitel 2.7 und 2.6 werden die theoretischen Grundbausteine der vorherigen Themen vertieft und bilden die Grundlage für das in dieser Arbeit vorgestellte Verfahren zur 3D-Datenrekonstruktion von 3D-Modellen aus Tabakblättern.

### 2.1 Künstliche Neuronale Netzwerke

Künstliche Neuronale Netzwerke(KNN) sind Datenstrukturen, welche von biologischen neuronalen Netzen wie sie bei Lebewesen vorkommen, inspiriert sind. KNN haben das Ziel ein Funktion  $f^*$  zu approximieren. Dabei werden Parameter  $\Theta$  eines Modells angepasst, um die Abbildung von  $y = f(x; \Theta)$  zu approximieren. Die Modelle werden auch als feedforward Neuronale Netzwerke betitelt weil der Informationsfluss des Models von Input zu Output fließt und keine Rekursion von Output zu Input statt findet. Diese Approximation wird durch Machine Learning Ansätzen erlernt, man spricht auch von einem Optimierungsproblem. KNN können aus mehreren Schichten, sogenannten Hidden Layer  $n$ , bestehen, welche als  $f^{(n)}$  dargestellt werden wobei,  $n \geq 1$  sein muss. Ein 2-Layer KNN ist dann definiert durch  $f(x) = f^{(2)}(f^{(1)}(x))$ . Man spricht auch von Fully-Connected-Layer. Es gibt einen Input Layer, welcher den Input in das Netzwerk aufnimmt (Goodfellow et al., 2016). Wird ein KNN mit mehr als nur 1 Layer mit Maschine Learning trainiert spricht man auch von Deep Learning. Im vorherigen Beispiel ist dies  $f^{(1)}$  und der letzte Layer des Netzwerkes wird Output Layer genannt. Im vorherigen Beispiel ist das  $f^{(2)}$ . In Abb. 1 sind ein KNN exemplarisch dargestellt. Es soll die Konnektivität der einzelnen Layer veranschaulichen und den Informationsfluss von Input zu Output.

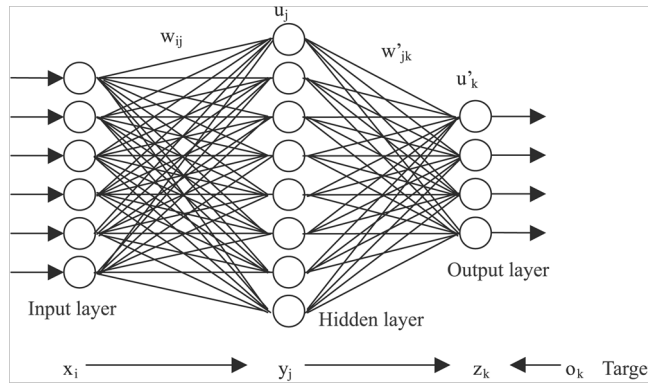


Abb. 1: Künstliches Neuronales Netzwerk(Haque et al., 2018)

Jeder Layer besteht aus künstliche Neuronen. Diese haben ihre Namensgebung von aus der Natur stammende Neuron in Gehirnen von Lebewesen. Neuronen sind die Bausteine, aus denen die Gehirne von Lebewesen wie Fischen, Vögeln und Säugetiere zusammen gesetzt sind. Neuronen, oder auch Nervenzellen haben einen Zellkern der Zentrum der Zelle ist. Um sie herum sind Dendriten welche die Verbindung zu anderen Neuronen her stellt. Neuronen sind untereinander mit dem Axonen verbunden, welche an den enden Synapsen haben die Grenze von Axon zur Nervenzelle einen Spalt bilden. Dieser Spalt kann überwunden werden indem von der Synapse Botenstoffe abgesendet werden, die sich dann an den Rezeptor der gegenüberliegenden Synapse anhaften. Diese Übertragung findet statt wenn an der Synapse ein bestimmter Schwellenwert überschritten wurde von elektrischen Reizen, welche die Zelle abfeuert lässt. Künstliche Neuronen haben diesen Schwellenwert durch sogenannte Gewichte  $w_{ij}$  diese sind auf den Verbindungen zwischen den Neuronen in den unterschiedlichen Layern im KNN (vgl. 2). Dabei steht  $i$  für die Position im Layer und  $j$  in welchen Layer sich das Neuron befindet. Jeder Layer eines KNN besteht aus mehreren Neuronen. Ein Neuron  $\theta$  kann mehrere Inputs  $X_n$  erhalten und produziert einen Output  $\omega$ . Die Berechnungen, welche von den Neuron durchgeführt werden sind zunächst jeden Input  $X_n$  mit einen Gewicht  $w_{ij}$  zu multiplizieren. Anschließend wird die Summe von  $x \cdot w$  gebildet. Das Ergebnis wird dann in eine Aktivierungsfunktion  $\lambda$  gegeben. Ein Neuron ist definiert durch:

$$y_k = \lambda(\sum_{j=0}^m (w_{kj} + x_j) + b_k)$$

Die Aufgabe der Aktivierungsfunktion  $\lambda$  ist es eine nicht lineare Transformation des Inputs zu erzeugen. Damit kann das KNN nicht lineare Funktionen abbilden und somit komplexere Aufgaben lösen. Es gibt unterschiedliche Aktivierungsfunktionen im folgenden werden einige aufgezählt:



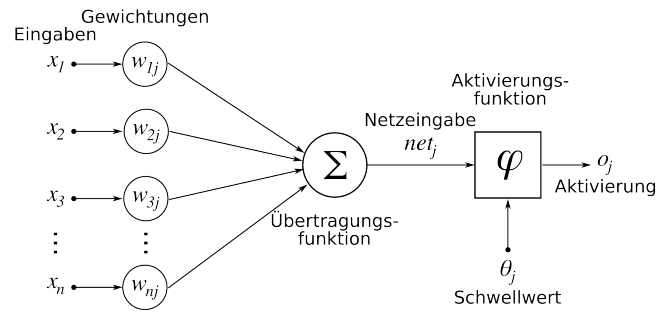


Abb. 2: künstliches Neuron

**Sigmoid-Funktion**  $\sigma(x) = \frac{1}{1+\exp(-x)}$

**Softmax-Funktion**  $\zeta(x)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}}$

**Rectified Linear Unit (ReLU):**  $f(x) = \max(0, x)$

**Leaky Rectified Linear Unit(Leaky Relu):**  $f(x) = \begin{cases} x & \text{if } x > 0 \\ 0.01x & \text{sonst} \end{cases}$

Der Funktionsplot von Sigmoid Funktion kann Abb.3 entnommen werden. Dieser veranschaulichen den Wertebereich, welcher von den Aktivierungsfunktion angenommen werden kann und ihren Verlauf.

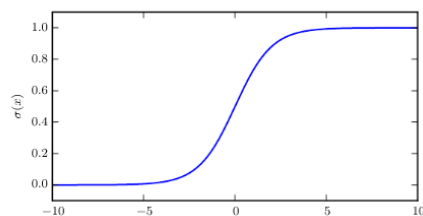


Abb. 3: Sigmoid Funktion

### 2.1.1 Ziel-Funktion

Die Ziel-Funktion oder auch Loss-Funktion genannt  $J(\theta)$ , muss differenzierbar sein wobei  $\theta$  das KNN ist und  $J$  für die Zielfunktion steht. Das heißt die Funktion  $f: D \rightarrow R$  ist differenzierbar an der Stelle  $x_0$ . Wenn nun  $f': x \rightarrow f(x)$  an jedem Punkt  $x^n$  ableitbar ist, ist  $f$  differenzierbar. Die Aufgabe der Ziel-Funktion ist es zu messen wie gut unsere Model  $\theta$   $f^*(x)$  approximiert. Es wird der Begriff Kosten verwendet, wie viel Kosten erzeugt das Model beim Lösen der zugewiesenen Aufgabe. Die Wahl, welche Ziel-Funktion gewählt wird ergibt sich aus der Aufgabe des KNN. Diese kann beispielsweise für Klassifikations oder Regressions Aufgaben hergenommen werden. Bei Regression soll eine kontinuierliche Variable von  $\theta$  als Output generiert werden, wohingegen bei Klassifikationsproblemen der Output Klassen-Labels darstellt (Goodfellow et al., 2016). Es gibt verschiedene Ziel Funktionen im folgenden wird die Cross Entropy Loss Function vorgestellt (Golik, Doetsch, & Ney, 2013). Diese kann verwendet werden um beispielsweise Klassifikationsprobleme zu lösen. Diese ist definiert als:

$$\hat{q}(c | x) = \arg \min_{q(c|x)} \left\{ - \sum_n \log q(c_n | x_n) \right\}$$

Wobei  $x_n: n=1, \dots, N$  die Trainingsdaten sind und  $c_n: n=1, \dots, N$  die möglichen Klassen. Der Output ist die Wahrscheinlichkeit zwischen 0 und 1, ob  $x_i \in$  der bestimmten Klasse enthalten ist. Auch kann sie hergenommen um zu messen wie hoch die Differenz zwischen zwei Wahrscheinlichkeitsverteilungen ist.

### 2.1.2 Backpropagation Algorithmus

Um nun KNNs zu trainieren und den gewünschten Output  $y$  zu generieren wird der Backpropagation Algorithmus benutzt. Dieser zählt zu den Optimierungs Algorithmen für KNN und arbeitet schneller und effizienter auf Neuronalen Netzwerken als andere Optimierungsalgorithmen vor ihm. Das mathematische zugrundeliegende Konzept ist ein Optimierungsproblem der die partielle Ableitung von  $\frac{\partial J}{\partial w}$ , wobei  $J$  die Zielfunktion und  $w$  die Gewichte im zu optimierenden Neuronalen Netzwerk, sind. Für eine Funktion  $f(x) = y$  ist die Ableitung definiert als  $f'(x)$  oder  $\frac{dy}{dx}$  und gibt die Steigung der Funktion an Punkt  $x$  an. Durch die Steigung am Punkt  $x$  ist man nun in der Lage eine Änderung von der Ableitung  $x$  dahin gehend zu optimieren (Goodfellow et al., 2016). Dieses Verfahren hilft dabei das KNN dahingehen zu optimieren den gewünschten Output  $y$  zu erlangen. Dieses Ziel wird durch die Ziel-Funktion eines KNN beschrieben (vgl. 2.1.1). Das Verfahren wird in Abb. 4 veranschaulicht.

Wenn nun  $f'(x) = 0$  gibt es keinerlei Information über die Steigung. Dies ist aber kein Indiz dafür, dass  $f$  ein Optimum erreicht hat. Es könnte, wie in Abb. 5 unten dargestellt ein lokales Minimum sein, was bedeutet, dass an diesen Punkt ein Minimum erreicht ist aber im Funktionsverlauf ein noch niedrigeres Minimum vorhanden ist. Oder einen Sattelpunkt welche einen Übergang zu einem Anstieg

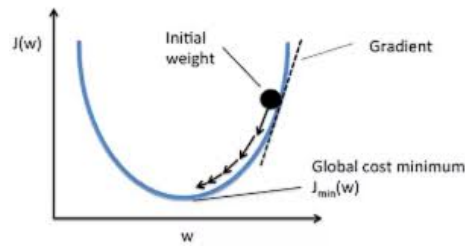


Abb. 4: Optimierung einer Zielfunktion

der Funktion bildet. Ist nun die die Funktion  $f$  definiert als  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  hat sie als Input mehrere Variablen. Die partielle Ableitung  $\frac{\partial f(x)}{\partial x_i}$  zeigt an wie sehr sich  $f(x)$  ändert wenn  $x_i$  geändert wird. Der Gradient  $\nabla_x f(x)$  ist ein Vektor, welcher alle partiellen Ableitungen von  $f$  enthält. Nun kann  $f$  optimiert werden, in dem, in die Richtung die Gewichte  $w_{ij}$  von  $\theta$  dahin gehen verändert werden in dem das Gradientenverfahren absteigend zu  $f'(x) = 0$  optimiert wird. Dieses Verfahren wird Gradient Descent genannt (Goodfellow et al., 2016). Beim Mini-batch Gradient Descent wird die Anpassung der Gewichte während des Trainings in  $n$  Batches unterteilt. Die Batches bestehen aus jeweils  $i$  Datensätze welche aus dem Trainingsdatensatz enthalten sind. Nach dem der Gradient für die  $i$  Datensätze des jeweiligen Batch berechnet wurden, werden die Gewichte  $\theta$  des ANN angepasst und der nächste Batch wird zum trainieren verwendet. Solange bis  $e$  Epochen wobei  $e$  die Durchläufe des kompletten Trainingsdatensatzes stehen, passiert sind. Dieses Verfahren führt zu stabileren Trainingsergebnisse (Ruder, 2016).

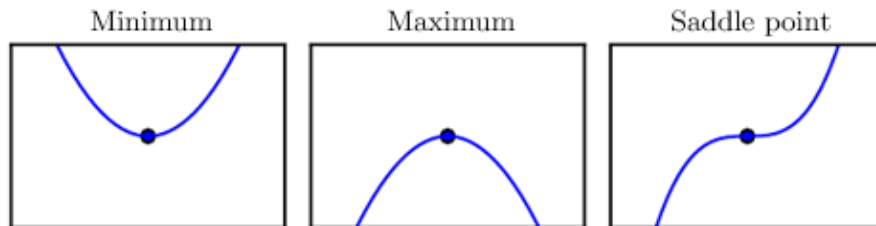


Abb. 5: Minimum, Maximum und Sattelpunkt einer Zielfunktion (Morabbi et al., 2018)

### 2.1.3 Momentum

Momentum hat das Ziel das Gradientenverfahren des Backpropagation Algorithmus 2.1.2 zu beschleunigen, um ein effizienter Ergebnis herbei zu führen und ein schnelleres Lernen erreichen. Definiert ist dies durch:

$$v_{t+1} = \mu v_t - \epsilon \nabla f(\theta_t) \theta_{t+1} = \theta_t + v_{t+1}$$

wobei  $\epsilon > 0$  die Lernrate ist,  $\mu \in [0,1]$  das Momentum und  $\nabla f(\theta_t)$  der Gradient von  $\theta_t$  ist. Je größer das Momentum, desto schneller bewegt sich der Gradient abwärts. Da dieser am Anfang einer Lernphase der üblicherweise hoch ist empfiehlt sich zunächst mit einem niedrigen Momentum zu Arbeiten da sonst die Gefahr besteht über das globale Optimum hinaus zuschießen. Wenn nun das Training stagniert, was aus Gründe der Aufbau der Zielfunktion zurückzuführen ist das zur Nähe des globalen Optimums flache Täler entstehen welche das Trainings verlangsamen und es zu keiner Verbesserung kommt, kann man durch Momentum erzwingen größere Gradienten Sprünge einzugehen. Und sich somit schneller zu einem globalen Optimum zu bewegen oder aber aus einem lokalen Optimum hinaus Richtung eines globalen Optimum zu bewegen kann Momentum benutzt werden (Sutskever, Martens, Dahl, & Hinton, 2013).

### 2.1.4 Regularization

Ein Problem welches alle Machine Learning Anwendungen teilen ist es wenn der Trainingsalgorithmus zunächst ein gutes Trainingsergebnis erzielt, aber dann auf dem Testdatensatz ein schlechtes Ergebnis erlangt, was Overfitting genannt wird. Es gibt einige Möglichkeiten die eingesetzt werden können, welche dann aber ebenfalls auch das Trainingsergebnis verschlechtern. Man spricht davon das Model zu Regularisieren.

#### Data Argumentation

Je mehr Daten beim Training verwendet werden desto mehr kann generalisiert werden. Data Argumentation heißt, dass man mehr Datensätze erstellt. Dies kann durch eine Veränderung der Trainingsdatensätze im Allgemein geschehen (Goodfellow et al., 2016), wenn man beispielsweise 3D-Punktwolken rotiert, um diese mehrfach zu nutzen. Eine weitere Möglichkeit ist es, bestimmte Datenpunkte in einen Datensatz ändern indem man ein Objekt, welches auf einem Bild dargestellt wird, verkleinert oder vergrößert (Goodfellow et al., 2016).

#### Dropout

Bei Dropout wird während des Lernprozess ein gewisser Prozentsatz der Neuronen in jedem Layer nicht verwendet. Dies zwingt das Netzwerk, welches sonst die Abhängigkeiten zwischen den Neuronen lernt, eine generalisiertere

Lösung zu finden(Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

## L2-Regularization

Dabei wird das Gewicht des Neurons während des Trainings näher zu seinem Ursprung, das heißt dem Wert seiner Initialisierung, gedrängt. Dies passiert indem ein Regularisierungswert  $\Omega = \frac{1}{2}||w||_2^2$  an die Zielfunktion gehängt wird. Dieses sorgt dafür, dass während dem Training der Gewichts Vektor welche durch das Gradienten Verfahren ermittelt wird, schrumpft, was für eine höhere Varianz während des Trainings sorgt, welches wiederum das KNN dazu zwingt mehr zu generalisieren(Goodfellow et al., 2016).

### 2.1.5 Batch Normalisation

Wie in Kapitel 2.1.2 gezeigt hat das stochastischen Gradienten Verfahren Vorteile gegenüber dem normalen Gradienten Ermittlung Verfahren beim Training von KNN. Dadurch das der Input in jeden Layer abhängig von den vorherigen Layern ist können Änderungen von Werten in frühen Layern des KNN große Auswirkungen in tieferen Layern im Netzwerk haben. Dadurch resultiert dass in Trainingsabläufen die Verteilung der Gewichte in den jeweiligen Layern verlangsamt wird. Batchnormalization soll die Werteänderung von Gewichten verringern. Dieses Problem wird auch Covariance Shift genannt. Um dies zu verhindern zeigten Sergey Ioffe und Christian Szegedy (Ioffe & Szegedy, 2015) eine Methode, welche Batch Normalisation genannt wird. Je mehr Layer das Netzwerk hat desto stärker ist der Covariance Shift. Batch Normalisation besteht aus zwei Algorithmen. Algorithmus 1 verändert den eigentlichen Input von Layer n zu einen normalisierten Input y und Algorithmus 2 verändert das eigentliche Training eines Batch normalisierten Netzwerkes(Ioffe & Szegedy, 2015).

---

**Algorithm 1:** Batch Normalisierung angewand auf x über Input bei Mini-Batch

---

1 Input: Werte von x über einen Mini-Batch  $B = \{x_1, \dots, x_n\}$

2  $\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x^i$

3  $\sigma_B \leftarrow \frac{1}{m} \sum_{i=1}^m (x^i - \mu_B)^2$

4  $\hat{x}_{iB} \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B + \epsilon}}$

5  $y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma; \beta}(x_i)$

6 Output:  $\{y_i = BN_{\gamma; \beta}(x_i)\}$

---

In Schritt 2 des Algorithmus 1 wird der Erwartungswert für alle Inputs von Mini-Batch B berechnet und in Schritt 3 die Varianz. In Schritt 3 wird nun der der normalisierte  $x_i$  berechnet welche dann mit  $\beta$  und  $\gamma$  multipliziert werden. Diese Werte sind neue Gewichte im Neuronalen Netzwerk welche während des Trainingsprozesses angepasst werden.  $\epsilon$  in der Gleichung in Zeile 4 ist nur dafür da damit nicht durch 0 geteilt werden kann. In Zeile 8 - 11 werden die Inferenz Schritte beschrieben, bei welchen der Minibatch des Trainings ersetzt wird (Ioffe & Szegedy, 2015). In Abb. 6 ist veranschaulicht wie die Batch-Normalization-Layer in das KNN eingebaut werden.

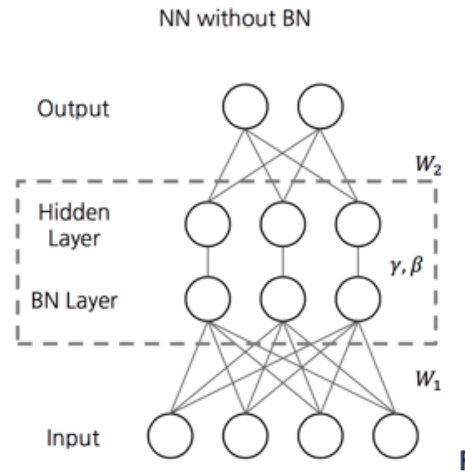


Abb. 6: Neuronales Netzwerk mit Batch-Normalization-Layer(*batchnorm* Chun, n.d.)

## 2.2 Datenformat Punktwolken

Punktwolken sind eine Menge von  $N$  Punkten, welche im Vektorraum dargestellt werden können. Jeder Punkt  $n \in N$  wird durch seine  $(x,y,z)$  Koordinaten im Euklidischen Raum dargestellt. Punkte können zusätzliche Features gegeben werden, wie Farbe oder Material. Es gibt unterschiedliche Dateiformate, welche für die Abspeicherung von Punktwolken herangezogen werden können Beispiele dafür sind PLY, STL oder OBJ. Das Polygon File Format (PLY) speichert die einzelnen Koordinaten in einer Liste, welche Vertex List genannt ist. In Abb. 7 kann eine Beispieldatei entnommen werden in der dieser Aufbau dargestellt ist.

Punktwolken können als Menge betrachtet werden. Die jeweiligen Punkte sind geordnet in dieser Liste gespeichert jedoch spielt es keine Rolle für den Punktwolkencompiler bei der Visualisierung der Liste an welcher Listenposition ein

```

1 ply
2 format binary_little_endian 1.0
3 element vertex 2800
4 property float x
5 property float y
6 property float z
7 end_header
8 0 0 0
9 0 0 1
10 0 1 1
11 0 1 0
12 1 0 0
13 1 0 1
14 1 1 1
15 1 1 0
16 1 0 0
17 1 0 1
18 1 1 1
19 1 1 0
20 1 0 0

```

Abb. 7: Polygon File Format

jeweiliger Punkt geführt wird, die Liste wird jedes mal gleich angezeigt, egal welche Permutation der einzelnen Punkte in der Liste durchgeführt wird. In Abb. 8 kann eine visualisierte Punktwolke eines Tabakblattes entnommen werden.

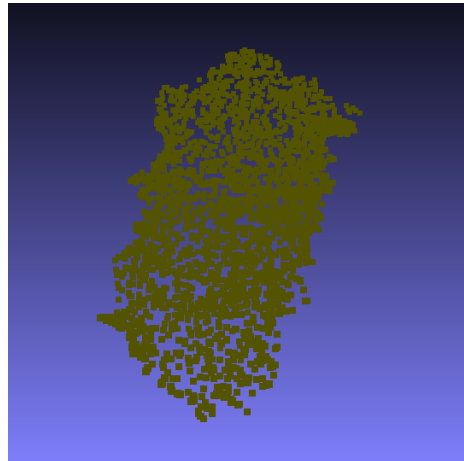


Abb. 8: Visualisierte Punktwolke eines Tabakblattes

### 2.2.1 Datenaufnahme von Tabakpflanzen

Die Punktwolken können Beispielsweise von den TERRA-REF Feld Scanner von der University von Arizona Maricopa Agricultural Center and USD Arid Land Research Station in Maricopa aufgenommen werden. In Abb. 9 ist Recht eine Skizze dargestellt. In Abb. 9 Links ist der Scankopf des TERRA-REF dar gestellt. Der 3D Laser Scanner erzeugt 3D Punktwolken. Dabei werden die Objekte

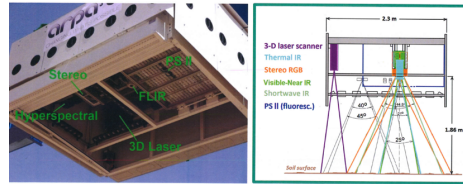


Abb. 9: Scankopf und Scanaufbau für die 3D-Punktwolken Gewinnung

durch den Scanner erfasst und eine 3D Repräsentation, welche durch Punkte in einen 3 Dimensionalen Koordinaten System erfasst werden können, dargestellt. Dabei wird ein Laser über das zu scannende Objekt gefahren durch reflektion des Laserstrahls auf der Oberfläche des Objektes können x,y,z Koordinaten des jeweiligen Punktes auf den Objekt bestimmt werden. Eine 3D-Scan einer Tabakpflanze kann in Abb.28 entnommen werden. Diese wurden in dieser Arbeit von Fraunhofer-Institut für Integrierte Schaltungen zur Verfügung gestellt. Da nun beim Scannen eines Objektes durch andere Objekte die Oberfläche verdeckt sein können. Wie beispielsweise beim Scannen von Tabakpflanzen, Blätter den Scankopf es nicht Möglich macht Objekte unterhalb dieses zu erreichen. Können Scans unvollständig sein. Dieses Problem für zu den Ziel dieser Arbeit, eine Möglichkeit zu Prüfen diese unvollständigen Punktwolken zu vervollständigen. Genauer zum diesen Thema in Kapitel 2.8.





Abb. 10: 3D Punktwolke einer Tabakpflanze

### 2.2.2 Die Schwierigkeit bei mit 3D-Data bei Machine Learning Ansätzen

Vergleicht man 3D-Data auf ihre Dimensionalität mit anderen Datenformaten wie Bild-, Audio-, und Textdateien steigt der Informationsgehalt und dadurch auch die Komplexität für das Anwenden von Maschine Learning Algorithmen enorm. Besonders bei Nicht-Euklidischen 3D-Daten wie Punktwolken, welchen keine Struktur zu Grunde liegt ist dieses gegeben (Ahmed et al., 2018).

Wie im Kapitel 2.2 gezeigt, sind Punktwolken als Menge gespeichert, in der keine Relation untereinander besteht. Was bedeutet, dass es für den Punktwolkencompiler nicht von Relevanz ist auf welchem Platz die einzelnen Punkte abgespeichert werden. Die Punktwolke wird immer gleich angezeigt, egal in welcher Permutation der einzelnen Punkte abgespeichert abgespeichert werden. Vergleicht man nun ein Bild mit 512 Pixeln und 3 RGB-Farbkanälen ist einer Dimension von 391680 erreicht. Vergleicht man dies mit einer Punktwolke in einen 125 cm<sup>3</sup> großen Bereich. Da die einzelnen Koordianten eines Punktes als Rationale Zahlen dargestellt werden und rationale Zahlen abzählbar unendlich sind ist der Suchraum unendlich groß. Dies führt zu einen erheblichen mehr Aufwand für Machine Learning Ansätzen wie, Deep Learning für Punktwolken.

Ein weitere Schwierigkeit, die für Deep Learning mit Punktwolken besteht ist, dass die aus Kapitel Convolutional Neural Network beschriebenen Convolutionel-Layer einen großen Beitrag bei dem Fortschritt von Deep Learning gebracht hat. Da sie helfen die Strukturen von strukturierten Daten zu lernen und den latenten Raum zu entdecken. Da jedoch Pointclouds unstrukturiert sind hilft es nicht unbedingt diese Tools auch bei Punktwolken einzusetzen(Achlioptas et al., 2018).

## 2.3 Convolution Neural Networks

Convolution Neural Networks(CNN) sind eine besondere Art von künstlichen neuronalen Netzwerken, sie sind dafür konzipiert auf Datensätzen zu arbeiten welche in eine Matrix Form gebracht worden sind. Der Input eines CNN können

beispielsweise Bilder sein welche durch die Matrix  $A = \begin{bmatrix} a_1 & a_1 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ \vdots & n_n & \ddots & \vdots \\ x_1 & x_2 & x_3 & x_n \end{bmatrix}$

dargestellt werden. Jedes Element  $x_{ij}$  stellt einen Pixel eines Bildes da, wobei  $x_n \in [0,255]$ . Die Matrix  $A^{w \cdot b \cdot c}$  stellt  $w \cdot b \cdot c = N$  dimensionale Matrix da. Wobei  $w$  die Länge und  $b$  Breite des Bildes entspricht.  $c$  sind die Farbspektren eines Bildes und sind in einen RGB-Farbraum 3 beziehungsweise in einen schwarz-weiß Bild 1. Nachdem der Input eines CNN definiert ist kommt nun der Aufbau. CNN setzen sich aus mehrere Schichten von Convolution Layern zusammen. Ein Netzwerk kann mehrere N-Layer haben. Wobei jeder Layer aus

mehreren Convolution oder auch Kernels genannt, zusammengesetzt ist. Ein Aufbau kann aus Abb. 11 entnommen werden (Goodfellow et al., 2016). Die Ker-

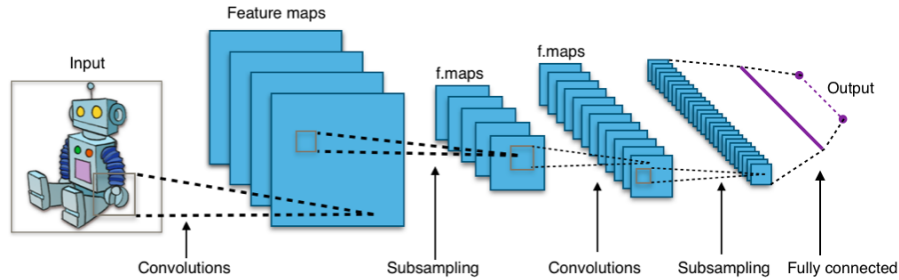


Abb. 11: Convolutional Neural Network

nels, also die einzelnen Filter, von denen jeder N-Layer  $k$  besitzt sind  $K^{n \times n}$  Matrizen jedes  $k_{ij}$  in einem Filter entspricht einen, aus den üblichen Neuronalen Netzwerk Architektur bekannten Gewichte. Diese Gewichte werden dann durch den Backpropagation-Algorithmus in der Trainingsphase des Netzwerkes angepasst um den Verlust der Ziel-Funktion durch bestimmten des Gradienten zu minimieren. Das in Abb. 11 dargestellte Subsampling ist der Output aus den Convolutional Layern (Goodfellow et al., 2016).

Da Input und Kernel unterschiedliche Größen haben und man den gesamten Input mit den Kernel abdecken möchte, bewegt sich der Filter um  $s$  Position auf den Input und führt erneut einen Berechnungsschritt durch. Dieser Vorgang wird Stride genannt. An jeder Position wird das Produkt von jedem  $x_{ij}$  des Input und  $k_{ij}$  des Kernel durchgeführt. Anschließend werden alle Produkte aufsummiert. In Abb. 12 ist dieser Vorgang verdeutlicht. Zusätzlich gibt es die Möglichkeit für das sogenannte Zero Padding  $P$ . Dabei werden mehrere 0 um die Input Feature Map, am Anfang und Ende der Axen anfügt. Dies ist notwendig wenn Kernel und Input Größe nicht kompatibel zueinander sind. Die Anzahl der möglichen Positionen ergeben sich aus der Kernel Größe und den Input des jeweiligen Kernel sowie des Strides. Die Output Größ  $W$  kann berechnet werden durch  $W = (W-F+2P)/s+1$  berechnet werden. Wobei  $F$  für die Größe des Kernel steht (Dumoulin & Visin, 2016).

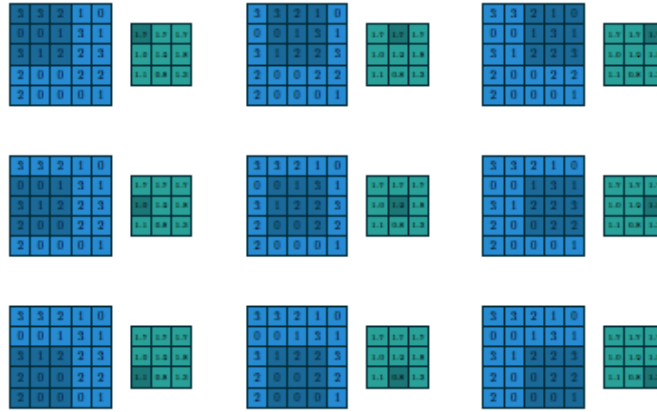


Abb. 12: Convolution Beispiel(Dumoulin & Visin, 2016)

Um besser zu verstehen welche Auswirkungen die Anzahl der Kernels in Layer  $n$  auf die Größe des Outputs von  $n$  und die Anzahl der Kernels in Layer  $n+1$  für den nächsten Layer haben, wird ein Beispiel aufgezeigt. Der erste Layer hat 20 Kernels mit der Größe  $7 \times 7$  und Stride 1. Der Input  $A$  für einen Kernel  $K$  ist ein  $28 \times 28$  Matrix. Der Output aus diesen Filter sind 20  $22 \times 22$  Feature Maps. Wäre der Input ein  $28 \times 28 \times 3$  Bild mit 3 RGB Channels sein, der Output 60  $22 \times 22$  Feature Maps. Allgemein kann Convolution Layer als Supersampling gesehen werden und Stride gibt an wieviele Dimensionen bei diesen Prozess pro Convolution Layer entfernt werden soll. Der letzte Layer ist ein Fully-Connected Layer welcher den typischen Anforderungen von ANN entspricht (Dumoulin & Visin, 2016).

Transposed Convolution, auch genannt Fractionally Strided Convolution oder Deconvolution, ist eine Umkehrfunktion von der üblichen Convolution. Es verwendet die gleichen Variablen wie Convolution. Dabei wird ein Kernel  $K$  mit der Größe  $N \times N$  definiert der Input  $I$  mit der Größe  $N \times N$  und Stride  $s = 1$ . Deconvolution kann wie Convolution angesehen werden mit Zero Padding auf dem Input. Das in Abb. 13 gezeigte Beispiel zeigt einen deconvolution Vorgang mit eine  $3 \times 3$  Kernel über einen  $4 \times 4$  Input. Dies ist gleich mit einen Convolution Schritt mit einen  $3 \times 3$  kernel auf einen  $2 \times 2$  Input und einer  $2 \times 2$  Zero Padding Grenze. Convolution ist Supersampling und mit Deconvolution wird Upsampling betrieben. Durch diesen Schritt kommt es zu einer Dimensionserhöhung des Inputs. Die Gewichte der Kernels bestimmen wie der Input transformiert wird. Durch mehrere Schichten von Deconvolution Layer kann von einer Input Größe  $N \times N$  auf eine Output Größe  $K \times K$ , wobei  $K > N$  mit Abhängigkeit von Kernel und Stride abgebildet werden(Dumoulin & Visin, 2016).

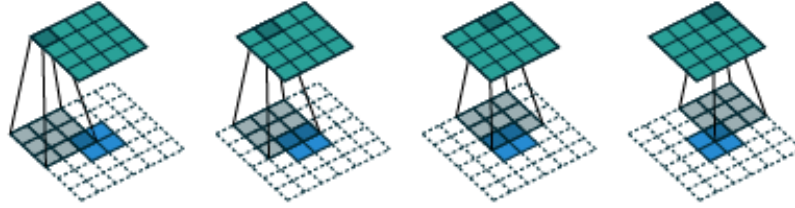


Abb. 13: Deconvolution Beispiel (Dumoulin & Visin, 2016)

## 2.4 Autoencoder

Autoencoder gehören zu den generativen Modellen im Bereich des Machine Learnings. Generative Modelle haben das Ziel eine Wahrscheinlichkeitsverteilungen zu erlernen. Anschließend kann diese als ein Modell genutzt werden um Samples aus dieser zu erzeugen. Die Modelle können dabei beispielsweise auf KNN oder Markov Chains trainiert werden. Im folgenden liegt der Fokus auf KNNs. Allgemein gehalten können jegliche Typen von Daten wie Text, Bild oder Audiodateien für generative Modelle herangezogen werden. Es gibt unterschiedliche Typen von generativen Modellen, welche sich vom Aufbau des Neuronalen Netzwerk und der Zielfunktion unterscheiden. Beispiele dafür sind Boltzmann Maschine, Autoencoder oder Deep Belief Networks(Goodfellow et al., 2016).

Autoencoder sind eine andere Art von Model aus den Bereich der generativen Modelle. Ihre Aufgabe besteht darin einen Input zu komprimieren und aus der komprimierten Information den Input wieder herzustellen. Die Technik auf welche Autoencoder zurückgreifen, nennt sich Dimensionreduktion. Dabei wird die Dimension der Daten so reduziert um Informationen bei zu behalten, welche als relevant gelten. Diese Technik findet auch in anderen Machine Learning Anwendung, wie beispielsweise der Principale component anaysis(PCA) Anwendung(Hinton & Salakhutdinov, 2006). Ein Autoencoder besteht aus 2 Bestandteilen. Erstens einen Encoder  $e$  dargestellt als  $z=f(x)$  welcher einen Input  $x \in \mathbb{R}^i$  wo  $x$  ein Vektor der Länge  $i$  ist und damit die Input Dimension bestimmt. Dieser wird durch den Encoder auf einen Vektor  $z^k$  abgebildet wobei  $k < i$  ist. Zweitens der Decoder  $d$  dargestellt als  $x'=g(z)$ , bekommt als Input  $z^k$  und bildet  $z$  auf  $\mathbb{R}^l$  ab wobei  $l = i$ . Und somit die gleiche Dimension wie der Input. Aufgabe ist es nun, dass der Encoder den Input  $x$  so gut komprimiert, dass der Decoder es schafft das  $x \approx x'$ . Eine grafische Darstellung kann aus Abb. 14 entnommen werden(Goodfellow et al., 2016).

Der Autoencoder wird durch den in Kapitel 2.1.2 vorgestellten Algorithmus trainiert und erlernt so können beispielsweise durch Fully-Connected-Layer, Convolutional-Layer oder Deconvolutional-Layer modelliert werden. Eine Metrik um

zu messen wie das Model seine Aufgabe erfüllt, könnte beispielsweise die Cross-Entropy-Funktionen sein welche schon in Kapitel 2.1.1 vorgestellt worden ist. Weitere spezifische Zielfunktionen für Autoencoder welche mit 3D Punktwolken arbeiten und für die folgende Arbeit von belangen sind, werden nun vorgestellt.

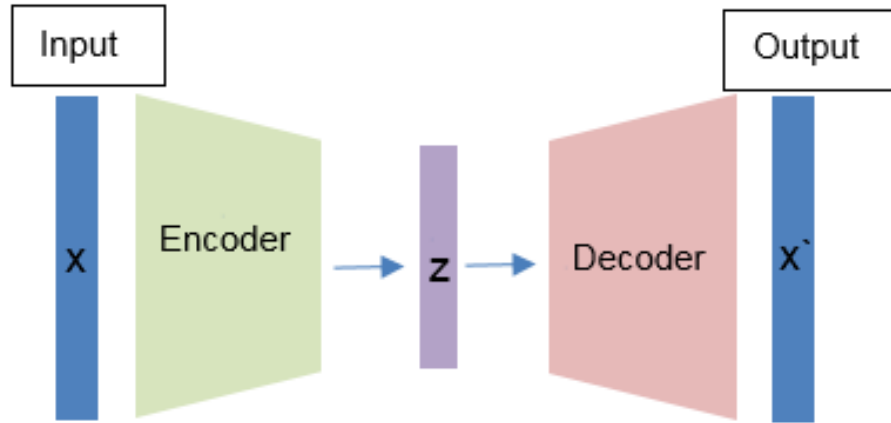


Abb. 14: Autoencoder

Bei Punktwolken als Datentyp erzielt die Cross-Entropy-Funktion keine guten Ergebnisse da diese invariant zu ihrer Permutation sind. Da dieser Arbeit der Input 3d Pointcloud sind und diese Sets invariant zu ihren Permutationen sind. Ändert sich Anordnung meiner einzelnen Punkte in mein Set bleibt das dargestellte Ergebnis unverändert. Deshalb kann nicht auf übliche Zielfunktionen welche für strukturierte Daten wie Bilder verwendet werden, zurück gegriffen. Die Herausforderung besteht darin zwischen zwei unterschiedliche Sets von Punkten heraus zu finden, wie hoch die Diskrepanz zwischen den beiden Sets ist (Ravanbakhsh, Schneider, & Poczós, 2016).

Eine Möglichkeit, speziell für Autoencoder, welche mit Punktwolken arbeiten, ist die Earth Mover Distance (EMD). Bei dieser sind  $X_1$  und  $X_2$  zwei Punktwolken mit jeweils  $x_n$  definierten Punkten (Fan, Su, & Guibas, 2017). Definiert ist sie durch die Funktion:

$$d_{\text{EMD}} = \min_{\theta: X_1 \rightarrow X_2} \sum_{x \in X_1} \|x - \theta(x)\|_2; \text{ wobei } \theta: X_1 \rightarrow X_2 \text{ bijektiv ist}$$

Grafisch kann man sich die Berechnung wie in Abb. 15 darstellen. Ein Punkt

von  $x_n \in X_1$  wird den nächsten Punkt  $y_n \in X_2$  zugewiesen. Wobei die Distanz durch die Euklidische Distanz der jeweiligen Punkte ermittelt wird. Eine weitere

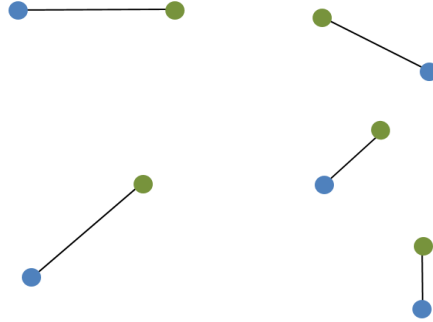


Abb. 15: Earth Mover Distance

Möglichkeit ist die Chamfer Distance(CD). Wie auch zuvor sind  $X_1$  und  $X_2$  zwei Punktwolken mit jeweils  $x_n$  definierten Punkten(Fan et al., 2017). Definiert ist sie durch die Funktion:

$$d_{CD}(X_1, X_2) = \sum_{x \in X_1} \min_{y \in X_2} \|x - y\| + \sum_{y \in X_2} \min_{x \in X_1} \|x - y\|$$

Der Unterschied ergibt sich zwischen den beiden Metriken, dass bei der EMD von der Ausgangswolke die Punkte zu der anderen jeweils optimiert werden. Wohingegen bei der CD die Distanzen von und zu der Ausgangspunktwolke berechnet werden. Dies geht aus den Abb. 16 hinaus in welche die Distanzberechnung der einzelnen Punkte dargestellt wird(Fan et al., 2017).

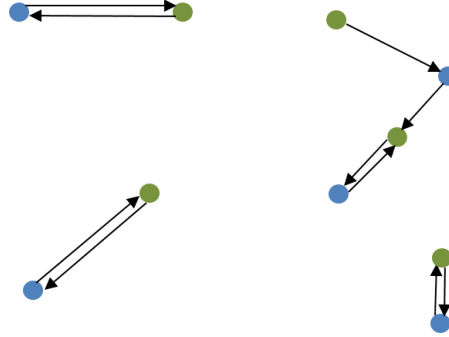


Abb. 16: Chamfer Distance

## 2.5 Generative Adversarial Network

GAN arbeiten nach dem Maximum Likelihood Schätzverfahren (ML-Schätzer) in dem die Parameter  $\theta$  dahingegen angepasst werden, dass die unsere beobachtet Daten am ehesten passen. Man kann ML-Schätzer als Kulback-Leibler(KL) Divergenz darstellen und das generative Modelle das Ziel haben die KL Divergenz zwischen den Trainingsdaten  $P_r$  und den generierten Daten  $P_g$  zu minimieren. Diese ist definiert durch:

$$KL(P_r||P_g) = \int_x P_r \log \frac{P_r}{P_g} dx$$

Ein GAN besteht aus zwei KNN, dem Discriminator D und dem Generator G. Das Ziel des G ist es, Daten  $x$  zu erzeugen, welche nicht von Trainingsdaten  $y$  unterschieden werden können. Dabei wird eine vorangegangene Input Noise Variable  $p_z(z)$  verwendet, welche eine Abbildung zum Datenraum  $G(z; \Phi_g)$  herstellt. Dabei sind  $\Phi_g$  die Gewichte des neuronalen Netzwerkes von G. Der Discriminator hat die Aufgabe zu unterscheiden, ob der jeweilige Datensatz von G erzeugt wurde und somit ein fake Datensatz ist, oder von Trainingsdaten  $y$  stammt (Goodfellow et al., 2014). Die Zusammensetzung zwischen den beiden Netzwerken kann aus Abb. 17 entnommen werden.

Der Discriminator ist definiert durch  $D(x; \Phi_d)$ . Wobei  $\Phi_d$  die Gewichte des Discriminators sind und  $D(x)$  die Wahrscheinlichkeit ist, dass  $x$  von den Trainingsdaten stammt und nicht von  $p_g$ . Die Wahrscheinlichkeitsverteilung für unsere Trainingsdaten ist  $p_r$ . Im Training werden dann  $\Phi_d$  so angepasst, dass die Wahrscheinlichkeit Trainingsbeispiele richtig zu klassifizieren maximiert wird. Und  $\Phi_g$  wird dahingegen trainiert die Wahrscheinlichkeit zu minimieren, so dass D erkennt dass Trainingsdatensatz  $x$  von G erzeugt wurde. Mathematisch ausgedrückt durch  $\log(1 - D(G(z)))$ . Die gesamte Loss-Funktion des vanilla GAN ist definiert als:



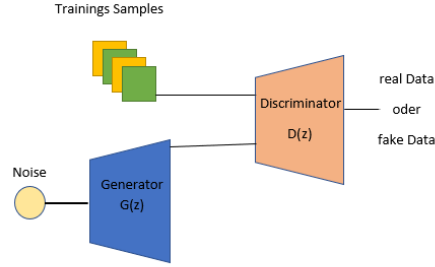


Abb. 17: Generativ Adverserial Network

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

diese beschreibt ein Minmax Spiel zwischen G und D. Welches das globale Optimum erreicht hat wenn  $p_g = p_r$ . Das heißt, wenn die Datenverteilung, welche von G erzeugt wird, gleich der unserer Trainingsdaten ist (Goodfellow et al., 2014). Das Training erfolgt durch den folgenden Algorithmus:

---

**Algorithm 2:** Minibatch stochastic gradient descent Training für Generative Adversarial Networks. Die Anzahl der Schritte welche auf den Discriminator angewendet wird ist k

---

```

1 for Anzahl von Training Iterationen do
2   for k Schritte do
3     • Sample minibatch von m noise Samples  $z^{(1)}, \dots, z^{(m)}$  von noise
       $p_g(z)$ 
4     • Sample minibatch von m Beispielen  $x^{(1)}, \dots, x^{(m)}$  von Daten
      Generationsverteilung  $p_{\text{data}}(x)$ 
5     • Update den Discriminator zum aufsteigenden stochastischen
      Gradienten:
6      $\nabla_{\phi_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$ 
7   end
8   • Sample minibatch von m noise Samples  $z^{(1)}, \dots, z^{(m)}$  von noise  $p_g(z)$ 
9   • Update den Generator mit den absteigenden stochastischen
      Gradienten:
10   $\nabla_{\phi_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$ 
11 end
```

---

Beim Training wird ein stochastischer Minibatch von mehreren Trainingsdaten gleichzeitig erstellt. Dies soll dabei helfen, dass der Generator sich nicht auf bestimmte Gewichte fest fährt und auf Trainingssätze kollabiert. So weisen die erzeugten Daten mehr Variationen auf (Salimans et al., 2016). D wird zunächst in einer inneren Schleife auf n Trainingsätzen trainiert, womit man Overfitting

von D vermeiden will, was zur Folge hätte, dass D nur den Trainingsdatensatz kopieren würde. Deshalb wird  $k$  mal D optimiert und ein mal G in der äußeren Schleife.

Ein möglicher Aufbau von GAN wird in Abb. 18 dargestellt. Dies ist das sogenannte Deep Convolutional GAN(DC GAN), welches dafür konzipiert wurde auf Bilddaten zu arbeiten. Dabei besteht der Generator aus mehreren Schichten von Deconvolution-Layern. Welche den Input Noise Variable  $p_z(z)$  auf  $y$  abbildet. D besteht aus mehreren Schichten von Convolution-Layern und bekommt als Input die Trainingsdaten, oder die von G erzeugten  $Y$ , und entscheidet über die Klassifikation(Radford, Metz, & Chintala, 2015). Das asymmetrische Ver-

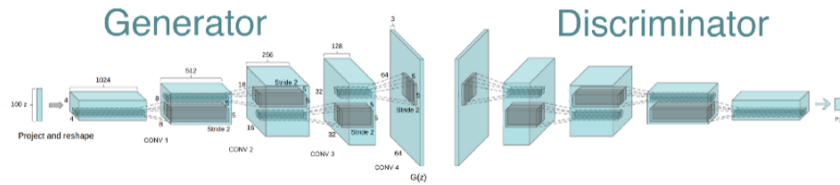


Abb. 18: Deep Convolutional GAN(*dc-gan online book*, n.d.)

halten der KV Divergenz zu schlechten Trainingsergebnissen führt, welches von üblichen generativen Modellen wie Autoencoder verwendet wird. Goodfellow (Goodfellow et al., 2014) zeigte, dass sich die MinMax Loss-Funktion des GAN auch als Jensen-Shannon Divergenz(JS Divergenz) darstellen lässt. Diese ist definiert als:

$$D_{JS}(P_r||P_g) = \frac{1}{2}D_{KL}(P_r||\frac{P_g+P_r}{2}) + D_{KL}(P_g||\frac{P_g+P_r}{2})$$

wobei  $P_r$  die Wahrscheinlichkeitsverteilung der Trainingsdaten ist und  $P_g$  die des Generators. Huzár (Huszár, 2015) zeigte, dass durch das symmetrische Verhalten der JS Divergenz ein potentiell besseres Trainingsergebnis entstehen kann, im Vergleich zu der KL Divergenz. Damit zeigte er weshalb GANs im Vorteil gegenüber anderen generativen Modellen sind.

### 2.5.1 Probleme mit GANs

Wie auch andere generative Modelle haben auch GANs noch Schwächen bezüglich der Trainingsabläufe und der Qualität ihrer generierten Daten. Im Folgenden wird auf einige Probleme eingegangen welche im darauffolgenden Kapitel Lösungsansätze aufgezeigt werden.

#### Vanishing gradient

Dies beschreibt das Problem, wenn D perfekt trainiert ist weil die beiden Datenverteilungen  $p_r$  und  $p_g$  disjunkt sind. Und es keine Überschneidung der Datenverteilungen gibt. Die Loss-Funktion würde in diesem Fall auf 0 fallen und es gäbe keinen Gradienten, für den die Gewichte von G angepasst werden können. Dies verlangsamt den Trainingsprozess bis hin zu einem kompletten Stopp des Trainings. Würde D zu schlechten trainieren mit  $D(x) = 0$  für  $p_r$  und  $D(x) = 1$  für  $p_g$ . Bekommt G kein Feedback über seine Leistung bei der Datengeneration und hat keine Möglichkeit  $p_r$  zu erlernen (Bang & Shim, 2018).

#### Equilibrium

D und G betreiben ein MinMax Spiel. Beide versuchen das Nash Equilibrium zu finden. Dies ist der bestmögliche Endpunkt in einen nicht kooperativen Spiel. Wie in dem Fall von GAN wäre das wenn  $p_g = p_r$ . Es wurde gezeigt, dass das Erreichen dieses Punktes sehr schwierig ist, da durch die Updates der Gewichte mit den Gradienten der Loss-Funktion starke Schwingungen der Funktion entstehen können. Dies kann zu einem instabilen Orbit für das laufende Training führen (Salimans et al., 2016).

#### Mode Collapse

Ein weiteres Problem kann passieren wenn der Generator während des Trainings, auf ein bestimmtes Setting seiner Gewichte festhält und er bei der Datengeneration sehr ähnlichen Samples generiert. Dieses Problem wird Mode Collapse genannt(Salimans et al., 2016).

### 2.5.2 Lösungsansätze für die GAN Probleme

Nun werden einige Techniken aufgezeigt, welche die unter Abschnitt Probleme mit GAN genannten Schwierigkeiten angehen und zu einem effizienteren Training führen, damit eine schnellere Konvergenz während des Trainings erreicht wird.

#### Feature matching

Dies soll die Instabilität von GANS verbessern und gegen das Problem des Vanishing Gradient angehen. G bekommt eine neue Loss-Funktion und ersetzt die des üblichen Vanilla GAN. Diese soll G davon abhalten, sich an

D über zu trainieren und sich zu sehr darauf zu fokussieren, D zu täuschen und gleichzeitig auch versuchen die Datenverteilung der Trainingsdaten abzudecken (Salimans et al., 2016).

### Minibatch discrimination

Um das Problem des Mode Collapse zu umgehen, so dass es nicht zu einem Festfahren der Gewichten von G kommt, wird beim Trainieren die Nähe von den Trainingsdatenpunkten gemessen. Anschließend wird die Summe über der Differenz aller Trainingspunkte genommen und dem Discriminator als zusätzlicher Input beim Training hinzugegeben (Salimans et al., 2016).

### Historical Averaging

Beim Training werden die Gewichte von G und D aufgezeichnet und je Trainingsschritt verglichen. Anschließend wird an die Lossfunktion je Trainingsschritt die Veränderung zu dem Trainingsschritt davor an die Loss-Funktion addiert. Damit soll verhindert werden dann das Model sich zu sehr von seinen vorherigen Zustand der Gewichte entfernt. (Salimans et al., 2016).

### One-sided Label Smoothing

Die üblichen Label für den Trainingsdurchlauf von 1 und 0 werden durch die Werte 0.9 und 0.1 ersetzt. Dies führt zu besseren Trainingsergebnissen. Es gibt derzeit nur empirische Belege für den Erfolg, jedoch nicht weshalb diese Technik besser funktioniert (Salimans et al., 2016).

### Wasserstein-GAN

Die Wasserstein Metrik, oder auch Earth Mover Distance(EMD) genannt misst die Minimum Kosten welche entstehen wenn man Daten von der Datenverteilung  $p_r$  zur Datenverteilung  $p_g$  überträgt. Es wird oft auch von Masseöder Fläche” gesprochen, welche von  $p_r$  zu  $p_g$  getragen wird. Definiert ist sie durch:

$$W(p_r, p_g) = \inf_{\gamma \in \Pi(p_r, p_g)} E_{(x, y) \sim \gamma} [\|x - y\|]$$

$p_r$  steht für die reale Datenverteilung zu welche uns Daten im Form von Trainingsdaten zur Verfügung stellen und  $p_g$  steht für die generierte Datenverteilung, welche von einem Model erzeugt wird. Dabei wird nun das Infimum von allen möglichen Transportplänen  $\gamma$ , welche in  $\Pi$  enthalten sind ausgewählt, was der kostengünstigste Plan ist, die Daten von  $x$  zu  $y$  zu übertragen. Die EMD sind nun die Kosten von einem optimalen Transportplan. Vorteile sind unter anderem, dass der Gradient gleichmäßiger ist und das WGAN lernt besser auch wenn der Generator schlechtere Daten erzeugt im Vergleich zum üblichen GAN (Arjovsky, Chintala, & Bottou, 2017). Durch die Kantorovich-Rubinstein Methode kann die Wasserstein-Distanz umgeformt werden in zu:

$$W(p_r, p_g) = \sup_{\|f\|_L \leq 1} E_{x \sim p_r} [f(x)] - E_{x \sim p_g} [f(x)]$$

Durch diese Umformung ist es nun möglich, dass GAN die EMD nutzt um die von G generierten Daten mit den realen Daten zu vergleichen. Dabei übernimmt der Discriminator nun die Aufgabe eines Critic, welcher nun nicht mehr 0 oder 1 für Fake oder Real ausgibt sondern, einen Score, welcher angibt wieviel Masse von der  $P_\theta$  also der Generator, umverteilt werden muss damit dieser bessere Ergebnisse liefert. Das Wasserstein GAN liefert bis dahin die erfolgreichste Verbesserung zum üblichen Vanilla-GAN von Goodfellow und erlaubt es, gegen die in Kapitel 2.5.1 aufgezeigten Probleme Vanishing gradient und Model Collapse anzugehen(Arjovsky et al., 2017).

## 2.6 Conditional-GAN

Conditional-GAN(C-GAN) ist eine Modifikation des ursprünglichen GAN von Goodfellow, welches erlaubt bedingte Wahrscheinlichkeiten in Datensätze zu erlernen. Das heißt zusätzliche Informationen in den Lernprozess einzuspeisen um den Output zu modifizieren. Im ursprünglichen GAN gibt es keine Möglichkeit auf den Output des Generators Einfluss zu nehmen. Dabei wird das Modell so verändert das eine zusätzliche Information  $y$  als Input in den Discriminator und Generator zugefügt wird. Dabei kann  $y$  jegliche Information sein wie Label, Bilddaten oder 3D-Daten. Dementsprechend muss die Zielfunktion dahin gehend angepasst werden bedingte Wahrscheinlichkeiten zu lernen

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z|y)))]$$

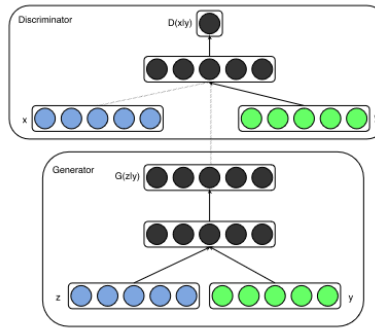


Abb. 19: Conditional Adversarial Network(Mirza & Osindero, 2014)

Im Abb. 19 kann der Informationsfluss und die Konnektivität der einzelnen Module entnommen werden. Die Module Generator und Discriminator bleiben gleich und können von ihrem Aufbau für die jeweiligen Datentyp verändert werden und Beispielsweise durch Convolutional-Layer, Deconvolutional-Layer oder Fully-Connected-Layer bestehen(Mirza & Osindero, 2014).

## 2.7 3D-GAN

Das besondere am 3D Raum im Vergleich zu normalen 2D Bildern ist die Steigerung der Dimension und zu gleich der hohe Informationsgehalt welcher in 3D Objekten steckt. Das Ziel von 3D-GAN ist es die Datenverteilung der zugrunde liegenden 3D-Modellen zu erlernen. Dabei wird der latente Objektraum erfasst und soll dadurch die Wahrscheinlichkeiten für einzelne Objektklassen enthalten.

Es wurden schon mehrere Versuche von generativen Modellen auf 3D Daten durchgeführt wie von Wu, Jiajun und Zhang (Wu, Zhang, Xue, Freeman, & Tenenbaum, 2016) welche in ihren Modell mit mit 3D-Voxel Daten arbeiten und damit ein GAN trainiert haben. Auch Achlioptas, Panos und Diamanti (Achlioptas et al., 2018) welche ein GAN auf Punktwolken trainiert. Da in folgender Arbeit die 3D-Daten durch Punktwolken dargestellt werden wird die Arbeit von Achlioptas, Panos und Diamanti näher beleuchtet.

Die Architektur des typischen 3D-GAN oder auch RAW-GAN betitelt ist dem Vanilla GAN von Goodfellow ähnlich. Der Input Layer ist ein Fully-Connected Layer welcher der Anzahl der Punkte je Punktwolke  $\cdot 3$  entspricht. Dieser bekommt als Input einen Noise-Vektor welcher aus einer Normal Verteilung entnommen wird und durch mehrere Layern gereicht bis hin zum Output Layer welche die Anzahl der gewünschten Punkte besteht. Also Zielfunktion kann mit der KL-Zielfunktion gearbeitet werden oder mit der Wasserstein Metrik welche im Versuchsaufbau von Achlioptas, Panos und Diamanti besser Ergebnisse geliefert hat. Der Aufbau unterscheidet sich nicht von Vanilla-GAN (Achlioptas et al., 2018).

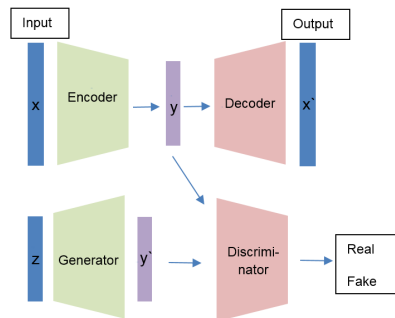


Abb. 20: Latent-GAN

Eine weitere Möglichkeit ist das Latent-GAN, dieses benutzt eine andere Aufbau gegenüber dem RAW-GAN, welches dabei helfen soll den latenten Raum

der Objekte zu erlernen. Zunächst wird ein Autoencoder mit den vorhandenen Trainingsdaten(vgl. 2.4) trainiert. Der Aufbau ist der selbe einen üblichen 3D-Autoencoder beschrieben in Kapitel 2.4. Ziel dabei ist den latenten Raum der Trainingsdaten zu erlernen und eine Kompression der Daten um den Suchraum zu welcher beim RAW-GAN durch die Inputdimension gegeben ist, zu verringern und dadurch das Training des GANs zu erleichtern(Achlioptas et al., 2018).

Bevor das Training des Latent-GAN beginnen kann werden nun die Trainingsdaten durch den vorher trainierten Encoder des 3D-Autoencoder auf die festgelegte Output-Dimension komprimiert. Anschließend werden diese komprimierten Daten verwendet um das GAN zu trainieren. Dabei erlernt das GAN komprimierten Code zu produzieren, welcher anschließend durch den Decoder des 3D-Autoencoder wieder auf die ursprüngliche Größe gebracht werden kann. Durch dieses Verfahren war es möglich Daten in einer guten Qualität zu produzieren und eine Datenverteilung des zugrundeliegenden Modells zu erlernen. Der gesamte Ablauf kann in Abb. 20 entnommen werden(Achlioptas et al., 2018).

## 2.8 Rekonstruktion von Daten

Derzeit setzt Deep Learning neue Maßstäbe bei der Rekonstruktion von Daten, wie Bildern oder Texten. Bei der Rekonstruktion geht es darum Daten, welche von ihren Urzustand verändert wurden, sei es durch Artefakte oder manuelle Bearbeitung, wieder dahin zurück geführt werden. In Abb. 21 ist eine Rekonstruktion eines Bildes dargestellt welches eine Wiederherstellung eines Hundes zeigt. Deep Learning hat in diesem Bereich besonders bei Bilder große Erfolge erzielt. Da diese als Matrizen dargestellt werden können, liefern sie eine Datenstruktur, auf welche KNN arbeiten können. Auch Bearbeitung mit Convolutionen-Layer auf Bilddaten zählt als Erfolg für die Weiterverarbeitung(Dong et al., 2016a). Durch die genannten Methoden werden bessere Strukturen für das jeweilige Ziel, in diesem Fall die Rekonstruktion, gelernt und ermöglicht es wie in verschiedenen Papern wie (Dong et al., 2016b) welche auf super-hochauflösenden Bildern Artefakte entfernt und das Bild wieder zum Urzustand zurück führt.

Auch wie die Arbeit von Liu, Gulin und Reda (Liu et al., 2018) konnte ähnliche Ergebnisse liefern. Rekonstruktion auf 3D-Daten wurde von Yi, Li und Shao (Yi et al., 2017) durchgeführt. Diese Arbeiteten beinhaltet jedoch die Rekonstruktion von 2D Bildern auf 3D Modellen. Dabei wurde mit Hilfe von Autoencoder gearbeitet was auch für die folgenden Arbeit von Bedeutung sind(Yi et al., 2017). Da diese helfen durch die Dimensionreduktion eine einfachere Daten Weiterverarbeitung zu liefern.





Abb. 21: Bild Rekonstruktion eines Hundes welches durch ein Artefakt zerstört wurde

### 3 Methoden

In diesem Kapitel werden die Methoden für die Versuchsaufbauten 1 und 2 aufgezeigt, welche die Fragestellungen 1 und 2 überprüfen sollen, die unten festgehalten wurden.

**Fragestellung 1** Können durch GANs 3D-Punktwolken von Tabakblättern erlernt werden um neue Datensätze zu generieren?

**Fragestellung 2** Können durch GANs 3D-Punktwolken von Tabakblättern, die von ihrem Urzustand abgebracht wurden, rekonstruiert werden?

Die Fragestellung beinhaltet ob mit Hilfe von GANs der latente Objektraum eines Tabakblattes gelernt werden kann und anschließend durch den Generator des GAN Tabakblätter erzeugt werden können. Dieser Versuchsaufbau wird in Kapitel 3.3 behandelt. Die dazugehörigen Trainingsdaten werden in Kapitel 3.1 vorgestellt. Bei Fragestellung 2 soll überprüft werden, ob es mit Hilfe von GAN möglich ist Artefakte, von Tabakblätter zu entfernen und den Urzustand wieder her zustellen. Dieses Verfahren wurde in Kapitel 2.8 beschrieben. Die Trainingsdaten für diesen Versuchsaufbau können Kapitel 3.2 entnommen werden.

Versuchsaufbau 1.2 und 2.1 wurden auf einem Computer mit Ubuntu 14.05 Betriebssystem mit einen Intel® Core™ i7-7700k mit 4.50GHz einer Geforce GTX 1080 mit 8G Grafikspeicher durchgeführt. Training und Test wurden mit CUDA 9.0 und cudNN 7.1.1. Als Programmiersprache wurde Python 2.7 bzw. 3.5 für Datengenerierung verwendet. Als ANN Library wurde Tensorflow 1.5 genutzt und TFLearn 0.3.2. Für die EMD und Chamfer Loss des Autoencoder wurde die Implementierung von dfanhqme (<https://github.com/fanhqme/PointSetGeneration>) benutzt.

Versuchsaufbau 1.1 und 2.2 wurden auf einen Computer mit Windows Betriebssystem mit einen Intel® Core™ i7-7700k mit 4.50GHz einer Geforce GTX 1080 mit 8G Grafikspeicher durchgeführt. Training und Test wurden mit CUDA 9.0 und cudNN 7.1.1 durchgeführt. Also Programmiersprache wurde Python 3.5 verwendet. Als ANN Library wurde Tensorflow 1.10 genutzt.

### 3.1 Datensatz 1. Versuchsaufbau

Der erste Datensatz "Stühle" besteht aus 6778 Punktwolken, mit jeweils 2048 Punkten. Dieser wurde dem Shapenet Datensatz([www.shapenet.org](http://www.shapenet.org)) entnommen. Die Daten sind im PLY Datenformat abgespeichert. Ein Beispieldunktwolke kann aus Abb. 22 entnommen werden. Der Datensatz wurde auch in der Arbeit von Achlioptas, Diamanti, Mitliagkas und Guibas(Achlioptas et al., 2018) verwendet, was Kapitel 2.7 bereits vorgestellt wurde ist. In der hier liegenden Arbeit dient er als Validierungsdatensatz, um die Qualität der generierten Daten des GAN, welches mit Tabakblättern trainiert wurde, zu vergleichen.

Die Daten für den zweiten Datensatz "Blätter" stammen vom Fraunhofer-Institut und deren Gewinnung wurde in Kapitel 2.2 beschrieben. Der Grunddatensatz bestand aus mehreren 3D-Scans von Tabakpflanzen, bei den die Blätter der Pflanze zu einem Datensatz zusammengefügt wurden. Der "Blätter" Datensatz besteht aus 422 Punktwolken welche, dann durch ein Punktreduktionsverfahren auf jeweils 2048 Punkten je Punktwolke reduziert worden ist. Aus Komplexitätsgründen wurden der Farbchannel in dieser Arbeit außen vor gelassen, um den Informationsgehalt der Daten zu reduzieren und ein Training zu vereinfachen. Außerdem spielen Farben beim derzeitigen Ziel der Arbeit, Rekonstruktion von Tabakblättern keine Rolle und nehmen keinen Einfluss auf die Verwendbarkeit des Ergebnisses. Abgespeichert werden die Daten im PLY-Datenformat.



Abb. 22: 3D Punktwolke einer Tabakpflanze

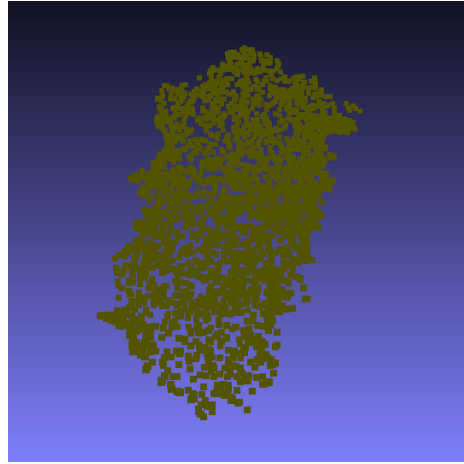


Abb. 23: 3D Punktwolke eines Tabakblattes mit 2048 Punkten

### 3.2 Datensatz 2. Versuchsaufbau

Für den zweiten Versuchsaufbau wird auf den in Datensatz Versuchsaufbau 1. Blatt Datensatz “Blätter“ zurück gegriffen. Der Grunddatensatz besteht aus 422 unterschiedlichen Blättern. Dieser Datensatz wird nun dahingehen verändert, das Löcher in die Blattoberfläche eingefügt werden. Diese Löcher simulieren das andere Blätter bei dem Scanverfahren die Sicht des Scanners auf das Blatt verdecken und dieses somit unvollständig eingescannt wird. Dieses Verdecken wird durch 3D-Sphären simuliert wie in Abb. 24 rot dargestellt. Die Sphären bestehen aus 10 000 Punkten, welche alle einen maximalen Radius von 15 mm haben und alle aus einer Normalverteilung entnommen werden, um eine gleichmäßige Verteilung der Punkte zu gewährleisten. Die Sphären werden nun vom Koordinatenursprung im euklidischen Raum bewegt, sodass sich 46 unterschiedliche Sphären im Raum ergeben. Diese werden dahin gehend im Raum bewegt um möglichst eine hohe Differenzmenge mit den 422 Blättern im Datensatz zu haben. In Abb. 25 sind alle Sphären symbolisch eingefügt, um dieses Vorgehen zu Veranschaulichen. Um nun das Verdecken zu simulieren, wird die Differenzmenge je Blatt mit einer der Sphären errechnet. Wobei die Differenzmenge eines Punktes der nächste Nachbar in einen Radius von 3mm befindet. Durch die dieses Verfahren entstehen kreisförmige Löcher in den Blättern welches in Abb. 26 dargestellt ist. Anschließend werden alle erzeugten “zerstörten“ Blätter nach der Anzahl ihrer übrig geblieben Punkte gefiltert um zu gewährleisten das genügen Punkte aus dem Urblatt entfernt worden sind und eine visuelle Diskrepanz zwischen den Blätter vorhanden ist. Dieses Vorgehen hat letztendlich zu 2047 zerstörten Blättern mit jeweils 2048 Punkten geführt.

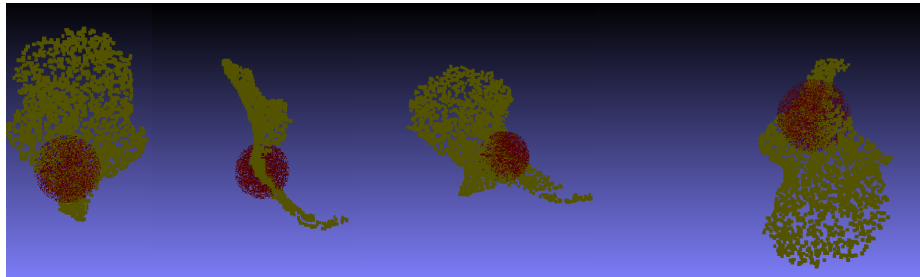


Abb. 24: Tabakblatt mit Sphäre welche die Differenzmenge bilden

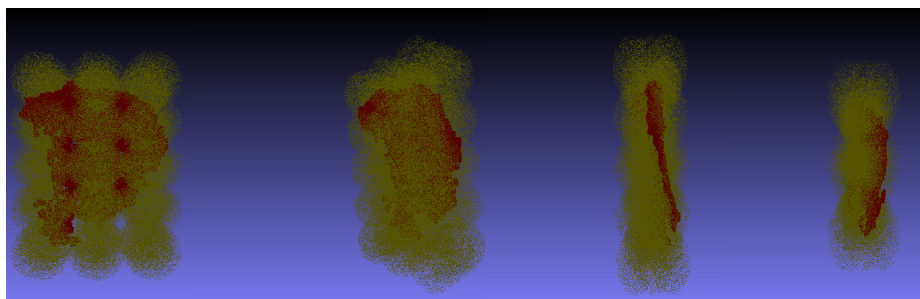


Abb. 25: Sphäre im Raum welche das verdecken der Blätter simulieren

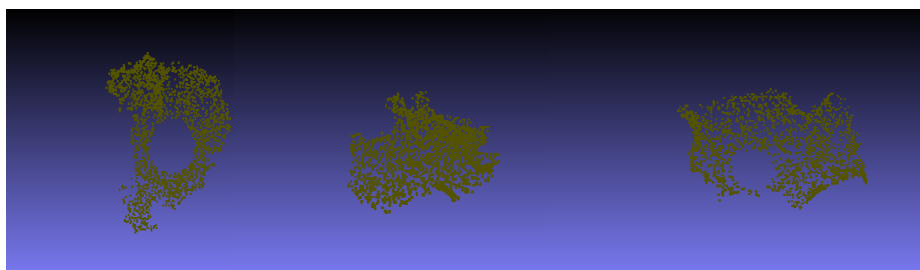


Abb. 26: 3D Punktwolke von zerstörten Tabakblättern

### 3.3 Versuchsaufbau 1 - GAN

Der Versuchsaufbau 1 besteht aus zwei unterschiedlichen Versuchen. Aufbau 1.1 ist das RAW-GAN((Achlioptas et al., 2018)), dabei wird auf den Vanilla-GAN Aufbau zurückgegriffen. Die Meta Trainingsvariablen für diesen Versuch sind Learningrate mit 0,0005 und einen Adam Optimizer, welcher zu den Mini-Batch Gradient Descent Verfahren aus Kapitel 2.1.2, und dabei auf das aus Kapitel 2.1.3, zurückgreift, mit einem Beta1 von 0,5 und einem Beta2 von 0,5. Die Batchgröße ist 64. Der Discriminator besteht aus 5-Layern, welche aus 1-Dimensionalen Convolutionen-Layer bestehen mit einer Filteranzahl [64, 128, 256, 256, 512] je Layer. Mit einer Kernel Größe von 1 und Stride von 1. Also Aktivierungsfunktion wird die ReLu-Funktion genutzt. Darauf folgen 3 Fully-Connected-Layer mit der Größe [128, 64, 1] alle mit einer ReLu-Funktion. Der Generator des RAW-GAN besteht auf 6 Fully-Connected-Layern mit [64, 128, 512, 1024 ,1536 ,6144] Neuronen, jeweils mit der Relu Aktivierungsfunktion. Der Noisevektor  $z$  wird aus einer 128-D Normal Verteilung entnommen mit  $\mu = 0$  und  $\sigma = 0,2$ . Als Zielfunktion wird die Wasserstein Metric gewählt da diese bessere Ergebnisse im Versuchsaufbau von Achlioptas, Panos und Diamanti(Achlioptas et al., 2018) erzielt hat. Das Training erfolgt jeweils mit den Datensätze SStühle und "Blätter" wie in Kapitel 3.1 beschrieben.

Der Versuchsaufbau 1.2 ist das Latent-GAN, welches in Kapitel 2.7 bereits vorgestellt wurde. Es wurde die Implementierung von (Achlioptas et al., 2018) verwendet ([www.github.com/optas/latent3d\\_points](https://www.github.com/optas/latent3d_points)). Zunächst wird dabei der Autoencoder mit den Trainingsdaten trainiert. Das Training erfolgt mit einer Learningrate von 0,0005 und einer Batchsize von 50. Der Encoder besteht dabei aus 4 1-D Convolutional-Layern mit [64, 128, 246, 1024] Filtern, eine Stride von 1 und Size von 1. Die Batchgröße ist 64. Als Aktivierungsfunktion wird die Relu verwendet. Der Encoder besteht aus 3 Fully-Connected-Layern mit [256, 256, 6144] Neuronen, alle mit der ReLu-Zielfunktion. Das Training erfolgt jeweils mit den Datensätze SStühle und "Blätter" wie in Kapitel 3.1 beschrieben und wurde auf 500 Epochen durchgeführt. Nach dem Training des Autoencoder kann der komprimierte Latent Code welcher von Encoder erstellt wird, dazu verwendet werden, das Latent-GAN zu trainieren. Alle Trainingsdaten werden nun durch den Encoder komprimiert, um ihre 128-D Latenten Code  $y$  zu erzeugen. Mit diesen wird nun das GAN trainiert, um eigene Latente Codes zu produzieren, welche vom Decoder wieder auf ihre ursprüngliche Dimension von 2048 Punkten projiziert werden. In Abb. 20 ist dieser Prozess dargestellt. Der Generator des Latent-GAN besteht aus 2 Fully-Connected-Layer mit [128, 128] Neuronen, welche als Aktivierungsfunktion eine ReLu-Funktion benutzen. Der Noisevektor  $z$  wird aus einer 128-D Normal Verteilung entnommen mit  $\mu = 0$  und  $\sigma = 0,2$ . Der Discriminator besteht aus [128, 64, 1] Fully-Connected-Layern welche ebenfalls ReLu-Funktion nutzen. Als Zielfunktion wird die Wasserstein Metrik gewählt, da diese bessere Ergebnisse im Versuchsaufbau von Achlioptas, Panos und Diamanti(Achlioptas et al., 2018) erzielt hat.

### 3.4 Versuchsaufbau 2 - CGAN für Punktwolkenrekonstruktion

Da im Versuchsaufbau 1 das Latent-CGAN bessere Ergebnisse beim Erlernen von Punktwolken Daten liefert, wird im Versuchsaufbau 2.1 Latent-CGAN das komprimieren von Trainingsdaten übernommen und dahin gehend verändert, dass Ziel von C-GAN zu übernehmen und bedingte Wahrscheinlichkeiten zu lernen. Zunächst werden wie bei Versuchsaufbau 1 die Trainingsdaten vgl.3.2 an einen Autoencoder trainiert. Dabei wird jeweils ein Autoencoder für zerstörte Blätter heran genommen und einer für unzerstörte im Urzustand.

Die Autoencoder folgen dabei den gleichen Aufbau wie in Versuchsaufbau 1.2. Der Encoder besteht dabei aus 4 1-D Convolutional-Layern mit [64, 128, 246, 1024] Filtern, eine Stride von 1 und Size von 1. Als Aktivierungsfunktion wird die ReLu-Funktion verwendet. Der Decoder besteht aus 3 Fully-Connected-Layern mit [256, 256, 6144] Neuronen, alle mit der ReLu-Zielfunktion. Als Batchsize wird 64 genommen. Als Zielfunktion wird jeweils einmal die CD getestet und die EMD. Für den Autoencoder wurde die Implementierung von (Achlioptas et al., 2018) verwendet([www.github.com/optas/latent3dpoints](http://www.github.com/optas/latent3dpoints)).

Der Generator des Latent-CGAN bekommt als Input  $z$  einen 128-D Vektor, welcher aus einer Normal Verteilung mit  $\mu = 0$  und  $\sigma = 0,2$  entnommen wurde.  $y$  welches zusätzlich zu  $z$  den Generator zum Training eingespeist wird sind die von Autoencoder codierten 128-D komprimierten zerstörte Tabakblätter. Der Generator des Latent-GAN besteht aus 2-Fully-Connected-Layer mit [128, 128] Neuronen, welche als Aktivierungsfunktion eine Leaky-ReLu-Funktion benutzen. Der Discriminator besteht aus 2 Fully-Connectected-Layern mit der Größe [128, 128]. Der Discriminator bekommt als Input entweder den vom Generator erzeugten latenten Code  $x'$  welcher vom Encoder des Autoencoder zu einen unzerstörten Blatt gemappt werden kann und zusätzlich  $y$  also  $(x',y)$ . Oder ein aus unseren Trainingsdaten stammendes  $(x,y)$  Paar, welche als reale-Datensätze gekennzeichnet sind. Als Ziel wird jeweils die Wasserstein Metrik und die Vanilla - GAN Metrik Loss-Function benutzt. Beim Trainingsaufbau 2.2 RAW-CGAN bekommt der Generator als Input Noisevektor  $z$  einen 128-D Vektor, aus einer Normal Verteilung entnommen mit  $\mu = 0$  und  $\sigma = 0,2$ . Des weiteren wird als  $y$  den Generator die zerstörte Blatt Punktwolke welche als ein 6144-D Vektor dargestellt werden als Input übergeben. Der Generator besteht aus Fully-Connected-Layern mit [6272, 3163, 1568, 3136, 4850, 5680, 6144] Neuronen je Layer. Als Aktivierungsfunktion wird die Leaky-ReLu genommen. Im letzten Layern wird keine Aktivierungsfunktion verwendet um den Output linear verarbeitbar zu machen und damit Euklidische Koordinaten generiert werden können. Des weiteren wird nach jedem Layer ein Batch-Normalisation-Layer eingesetzt mit Momentum 0,9 und einen Beta1 und Beta2 von 0,5 eingesetzt. Des Discriminator besteht aus [12288, 6144, 3072, 1536, 768, 384, 128, 32, 1] Fully-Connected-Layern mit Aktivierungsfunktion Leaky-ReLu bis auf den letzten Layer in den wird die Sigmoid-Funktion verwendet, wenn als Zielfunktion die Vanilla-GAN gewählt wird. Bei Verwendung der Wasserstein Metrik wird

auf eine Aktivierungsfunktion verzichtet. Des weiteren wird nach jedem Layer ein Batch-Normalisation-Layer eingesetzt mit Momentum 0.9 und einen Beta1 und Beta2 von 0,5. Der Input des Discriminator ist die vom Generator erstellten  $(x',y)$  oder die direkt aus dem Trainingsdaten stammenden Punktwolken Paare des zerstörten Blatt  $y$  und Urzustand Blatt  $x$  als Tupel  $(x,y)$ . Als Zielfunktion wird die Vanilla-GAN und W-GAN Zielfunktion getestet.

Im weiteren Versuchsaufbau für den RAW-CGAN wird der Discriminator umgebaut um ihn durch Convolutional-Layer bessere Lernfähigkeiten zu geben. Dies sorgt nicht nur für Vorteile für den Discriminator dadurch kann auch der Generator durch den Gradienten des Discriminators einen besseren latenten Raum lernen. Der Generator bleibt in diesen Aufbau gleich der Discriminator bekommt [64, 128, 256, 256, 512] 1-D-Convolutional-Layer, mit einer Kernel-Größe von 1 und Stride von 1 als Aktivierungsfunktion wird eine Leaky-ReLu verwendet. Die Batchsize beträgt 64. Anschließend folgen [128, 32, 1] Fully-Conncted-Layer mit Leaky-Relu. Zwischen jeden Layer befindet sich ein Batch-Normalisation-Layer mit Momentum von 0,9 und einen Beta1 und Beta2 von 0,5. Abb. 27 kann der Aufbau und Konnektivität der einzelnen Komponenten entnommen werden.

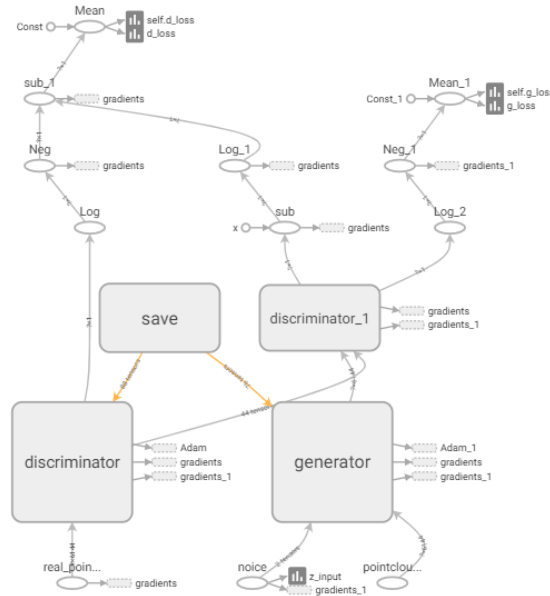


Abb. 27: Aufbau des RAW-CGAN

## 4 Evaluation und Ergebnisse

In diesem Kapitel wird auf die Ergebnisse von Versuchsaufbau 1 und 2 eingegangen. Zunächst wird Versuchsaufbau 1, vgl. 3.3 des Erlernen von latentem Raum von Punktwolken gezeigt. Dabei wird ein Vergleich zwischen den Modelldatensatz “Stühle” sowie einen aus der Praxis stammenden Datensatz “Blätter” aufgezeigt. Anschließend wird auf die Ergebnisse aus Versuchsaufbau 2 vgl. 3.4 eingegangen. Dabei wird evaluiert, ob es möglich ist die zerstörten Blattdaten in ihrem Urzustand wieder herzustellen und aussagekräftige Ergebnisse für die Anwendbarkeit in der Praxis vorhanden sind. Da es keine gängige Validierungsmethoden, beziehungsweise Validierungsmetriken gibt, wird sich bei der Prüfung der Testergebnisse auf das menschliche Auge verlassen, so dass die Prüfung in Form einer Sichtprüfung erfolgt.

### 4.1 Ergebnisse - Versuchsaufbau 1

Von RAW-GAN aus Versuchsaufbau 1.1 mit den Blattdaten lassen sich nach 500 Epochen Training Beispieldatensätze, welche vom Generator erzeugt wurden aus Abb.28 entnehmen. Wie zu erkennen ist, sind die Blätter nicht gut genug für eine Weiterverarbeitung. Der latente Raum, welcher von den GAN gelernt wurde und am Ende durch den Generator produziert wurde, zeigt keine Ähnlichkeit mit normalen Blättern, welche im Datensatz enthalten waren. Die Daten erinnern eher an einen hohlen Ball, als an Blättern. Es formen sich Punktansammlungen in der Nähe des Koordinatenursprungs und es lässt sich kein Anzeichen von Bildern einer Blattfläche erkennen. Die starke Punkthäufung

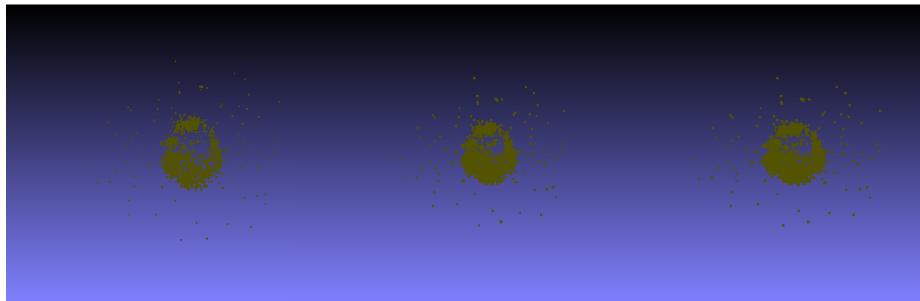


Abb. 28: Generierte Beispieldaten vom Generator des RAW-GAN mit den Blattdaten

am Koordinaten Ursprung ergeben sich dadurch, dass die realen Blattdaten an dieser Stelle häufig vertreten sind und sich an den Blattkanten eher unterscheiden. So lernt der Generator an diesen vermehrten Schnittpunkt mehr Punkte zu generieren als außerhalb. Dieser Effekt ist in Abb. 29 zu erkennen. Dargestellt



sind 24 Datensätze aus dem Blatt Datensatz, welche aus verschiedenen Blickwinkeln in einem Koordinatensystem geladen wurden sind. Die Punkte häufen sich in der Nähe des Koordinatenursprungs und sind im äußeren Bereich schwächer verteilt. Vergleicht man dies mit Abb. 28 sind dort dichte Ansammlungen eben-

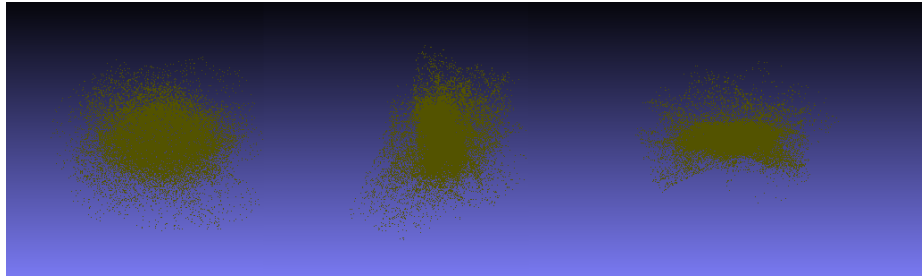


Abb. 29: 24 Punktwolken aus dem Blatt Datensatz in einem Koordinatensystem

falls zu erkennen. Der W-GAN Loss für den Generator und Discriminator kann in Abb. 30 entnommen werden. Der Loss des Discriminators konstant über mehrere Epochen bleibt, lässt sich daraus schließen, dass der Critic keine Schwierigkeiten hat die beiden Datenverteilungen zu unterscheiden. Es ist kein Anzeichen von Verbesserung zu erkennen wenn das Training länger als 500 Epochen läuft.

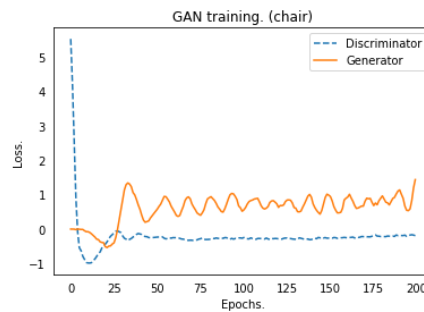


Abb. 30: Trainingsverlauf des RAW-GAN mit den Blatt Daten

Ähnliche Ergebnisse sind auch beim RAW-GAN von Versuchsaufbau 1.1 mit den Stuhl Daten zu beobachten. Wobei die Qualität der Stühle stärker an die im Trainingsdatensatz enthaltenen Daten erinnert vgl. Abb. 31. Es zeichnen sich Lehne und Stuhlbeine in den Daten ab, die vom GAN trainierten Generator stammen. Der Trainingsverlauf ist Abb. 32 zu entnehmen. Dabei ist zu sehen dass der Discriminator sich nicht mehr verbessert und der Generator keine bes-

seren Ergebnisse liefern kann. Diese Ergebnisse decken sich mit den von Achlioptas, Panos und Diamanti (Achlioptas et al., 2018) festgestellten Ergebnissen in den das RAW-GAN keine qualitativ gute Stuhldaten erzeugen konnte.

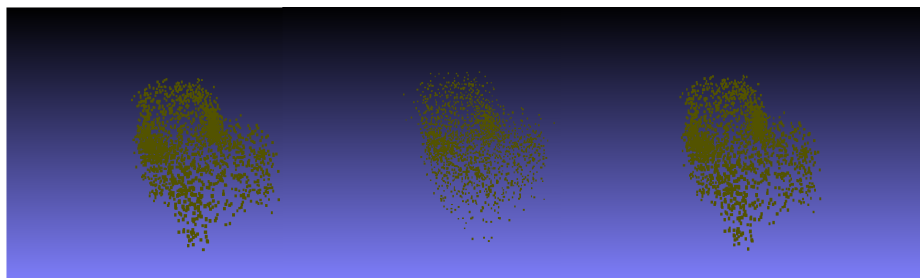


Abb. 31: Generierte Beispieldaten von Generator des RAW-GAN mit den Stuhldaten

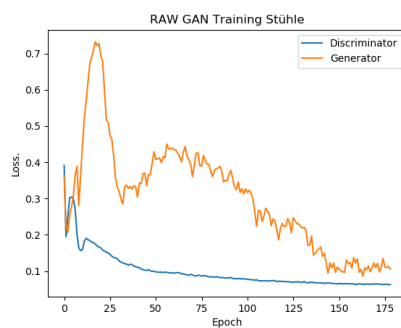


Abb. 32: Trainingsverlauf des RAW-GAN mit den Stuhldaten

Um nun die Ursache genauer zu beleuchten, was die Gründe für den Qualitätsunterschied zwischen den beiden generierten Datensätzen ausmacht und um festzustellen ob durch mehr Blattdaten bessere Ergebnisse erzeugt werden können, wurden die Stuhldaten auf 422 Trainingsdaten reduziert und ein erneutes Training des RAW-GAN durchgeführt. Die erzeugte Beispieldaten können aus Abb. 33 entnommen werden. Vergleicht man die Qualität des RAW-GAN mit 422 Stuhldaten und 6778 Stuhldaten (vgl. Abb.31) ist in Abb.33 eine Steigerung der Qualität klar festzustellen.

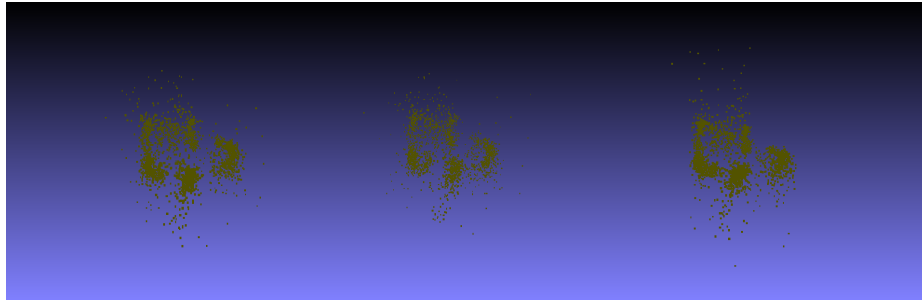


Abb. 33: Trainingsbeispiele des RAW-GAN mit den 422 Stuhldaten

Für Versuchsaufbau 1.2 Latent-GAN mit Stühlen kann der Trainingsverlauf aus Abb. 34 entnommen werden. Der Generator Loss verbessert sich zwar über mehrere Epochen gering, der Discriminator jedoch bleibt konstant auf einen seinen Ergebnisse und lässt keine starken Verbesserung des Generators mehr zu, da er die beiden Datenverteilungen perfekt unterscheiden kann. Dies führt dazu, dass der Generator kein Feedback über die Qualität seiner generierten Daten erhält und sich nicht verbessern kann. Beispieldatensätze, welche vom Generator nach

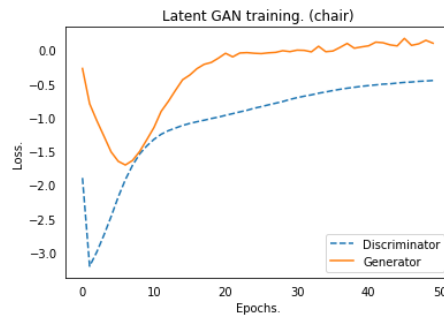


Abb. 34: Trainingsverlauf des RAW-GAN mit den 422 Stuhldaten

Das Latent-GAN welches mit den Blattdaten trainiert wurde, können exemplarisch generierte Daten aus der Abb. 36 entnommen werden. Diese zeigen eine qualitative Grundfläche, welche in der Trainingsdatengesamtheit enthalten ist. Jedoch ist festzustellen, dass es starke Punkthäufungen an gewissen Bereichen je Blatt gibt, diese sind beispielsweise im zweiten Blatt von Rechts, auf Abb. 36 zu erkennen und mit einem roten Kreis markiert. Der Trainingsverlauf in Abb. 37 zeigt, dass keine starken Veränderungen im weiteren Trainingsverlauf zu erwarten ist und das Training schon über mehrere Epochen stagniert.

44

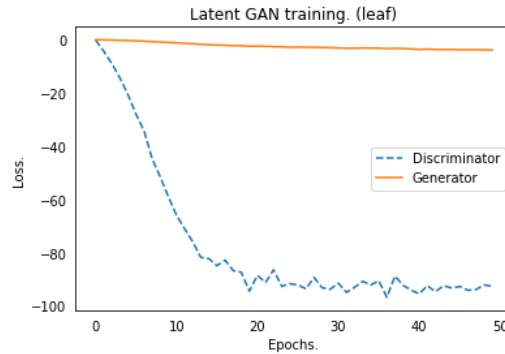


Abb. 37: Trainingsverlauf des Latent-GAN mit den Blattdaten

Um die Ergebnisse von Testaufbau 1.2 Latent-GAN mit Stuhl­daten besser vergleichbar zu machen und um ersichtlich zu machen, ob eine Erhöhung der Anzahl der Tabakblätter Daten zu einem besseren Ergebnis führen und die Punktanhäufung sich reproduzieren lässt, wurden 422 Stuhl­daten genommen um das Latent-GAN zu trainieren. In Abb. 38 sind die exemplarisch erzeugten Daten des Generator zu entnehmen. Zu erkennen ist ein ähnlich Phonemen wie bei den Blatt­daten. Hohe Punktanhäufungen finden sich an Kanten der Sitzfläche, zum Übergang der Stuhl­beine. Dadurch lassen sich Rückschlüsse darauf ziehen, dass eine Erhöhung der Daten zu einer besseren Qualität de kann. Der Trainingsverlauf kann aus Abb. 39 entnommen werden auch dieser zeigt eine Stagnation des Verlaufs schon über mehrere Epochen an und verspricht keine Erhöhung der Qualität beim weiteren Verlauf des Trainings.

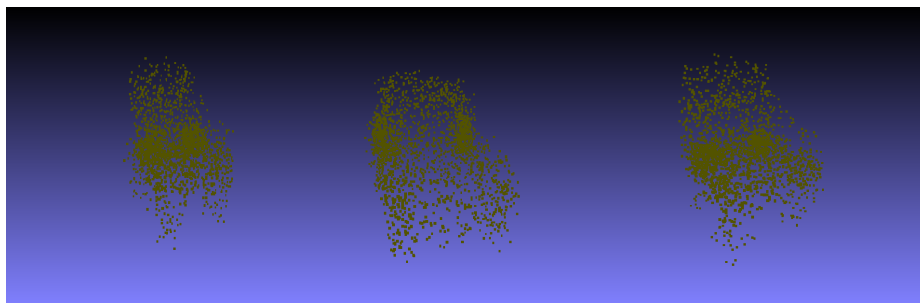


Abb. 38: Trainings­ergebnisse des Latent-GAN mit den Stuhl­daten und 400 Trainings­daten

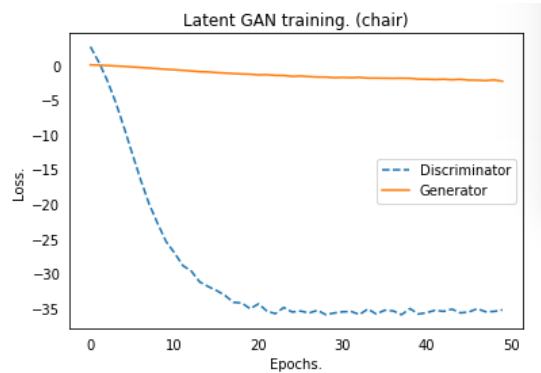


Abb. 39: Trainingsverlauf des Latent-GAN mit den Stuhl- und 400 Trainingsdaten

Zusammengefasst, primär in Hinblick auf den Blatt Datensatz, lässt sich feststellen, dass bei den Stuhl Datensatz die Ergebnisse von Achlioptas, Panos und Diamanti (Achlioptas et al., 2018) reproduziert werden konnte. Bei einem Datensatz, welcher durch Scanverfahren in diesem Beispiel Tabakblättern von realen Objekten durchgeführt wurde konnte gezeigt werden, dass durch das Latent-GAN eine erste Annäherung an qualitativ hochwertig generierten Daten zu sehen ist. Durch eine Erhöhung der Trainingsdaten lässt sich auch eine Steigerung der Qualität feststellen, wie bei den Stuhl- und 400 Trainingsdaten durch RAW-GAN und Latent-GAN gezeigt wurde. Des Weiteren können aber keine Aussagen darüber getroffen werden, in welchem Umfang eine Qualitätssteigerung statt finden kann.

Die Datensätze unterscheiden sich in ihrer Aufbereitung. Die Stuhl- und 400 Trainingsdaten sind besser auf einen Koordinatensprung geeicht wie in Abb. 40 festzustellen ist. Auf dieser Abbildung sind mehrere Stühle zum Vergleich in ein Koordinatensystem geladen. Zum Vergleich die das gleiche Konzept für die Blatt- und 400 Trainingsdaten in Abb. 53. Zu erkennen ist eine viel höhere Überschneidung der einzelnen Punktwolken aufeinander als bei den Blättern. Es sind dadurch gefestigtere Strukturen bei den Stühlen vorhanden und dadurch weniger Varianz bei den Punktwolken untereinander. Dies hilft den Suchraum für die GANs einzuschränken und erleichtert das Training bei den Stuhl- und 400 Trainingsdaten. Es müssten bessere Datenvorverarbeitungsschritte bei den Blatt- und 400 Trainingsdaten erhoben werden, wie beispielsweise ein besseres Festhalten des Massenzentrums aller Blätter auf den Koordinatensprung.

Außerdem sind die 6778 Stuhl- und 400 Trainingsdaten zum Vergleich zu den 422 Blatt- und 400 Trainingsdaten sehr gering und verschlechtern die Möglichkeit des Latent-GAN und RAW-GAN die Grundgesamtheit der Daten abzudecken. Aus technischer Sicht könnten andere Layer-Strukturen für bessere Ergebnisse sorgen. Die Bearbeitung mit Convolutional-Layern auf Bilddaten hat erst den Durchbruch bei der Rekonstruktion bei

Bilddaten möglich gemacht (Dong et al., 2016a). Da nun unser Generator nur aus Fully-Connected-Layern besteht, und 3D-Punktwolken komplexer sind als Bilder kann an dieser Stelle noch nachgearbeitet werden. Beispielsweise könnte ein 1D-Deconvolutional-Layer implementiert werden, welche bei Tensorflow 1.12 als Prototyp zur Verfügung steht. Eine andere Möglichkeit wäre auf die Arbeit von Cai, Zhongang und Yu (Cai, Yu, & Pham, 2018) zurückzugreifen. Diese postuliert eine Methode für 3D-Convolution Methode, welche auf 3D-Punktwolken arbeitet und eine Invarianz im Euklidischen Raum lernt, mit deren Hilfe das Erlernen von Rotationen der Punktwolken möglich gemacht werden soll.

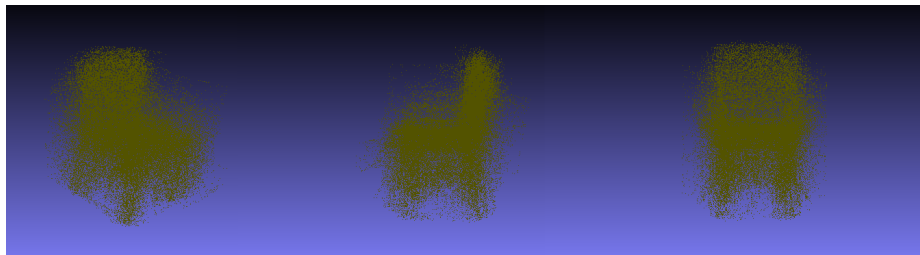


Abb. 40: 24 Stuhlzeiten in einem Koordinatensystem geladen

## 4.2 Ergebnisse - Versuchsaufbau 2

Für Versuchsaufbau 2.1, vgl. 3.4, wurde zunächst der Autoencoder mit den zerstörten Blattdaten auf 500 Epochen trainiert, um eine komprimierte Version auf 128-D zu erlernen. Dabei wurde zunächst die in Kapitel 2.4 vorgestellte Chamfer Distanz als Distanzmaß heran genommen. Beispiel Trainingsdatensätze können Abb. 42 entnommen werden, mit dem jeweils erzeugten Output des Decoders in Abb. 41. Wie zu erkennen ist, vervollständigt der Au-

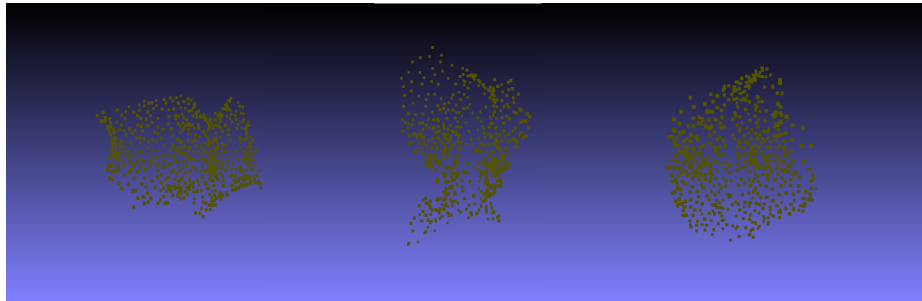


Abb. 41: Trainingsbeispiele des Autoencoder mit CD für zerstörte Blattdaten - Output

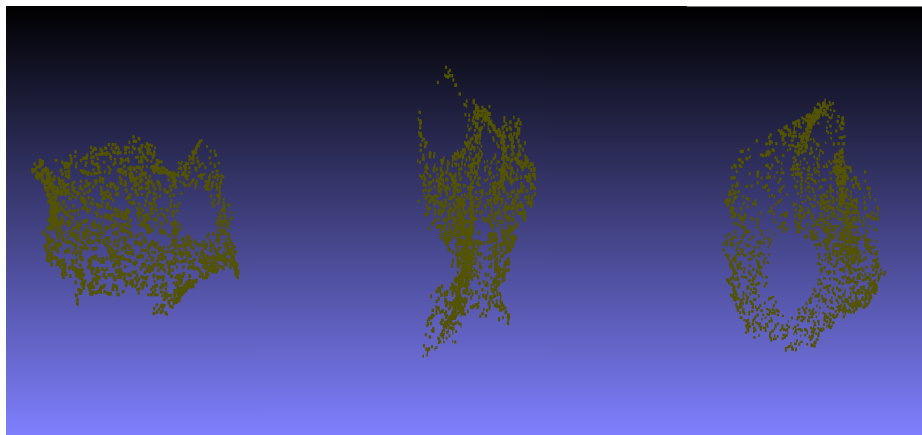


Abb. 42: Trainingsbeispiele des Autoencoder mit CD für zerstörte Blattdaten - Input

toencoder mit der CD die Blätter selbstständig, obwohl das nicht als Ziel in diesem Bearbeitungsschritt vorgeben war. Es ist zu erkennen, dass die Dichte,



in welcher die Punkte beim Output auf der Höhe der Löcher angeordnet sind stark von den üblichen Bereichen auf dem Blatt abweicht. Auch lässt der Trainingsverlauf in Abb. 43 darauf schließen, dass der Autoencoder selbst beim weiteren Training keine bessere Codierung erlernen wird, welche die Löcher in den Blättern erzeugt, da das Training seit Epoche 300 stagniert. Außerdem lässt der durch die CD berechnete Abstand zwischen den beiden Punktwolken keine starke Verbesserung mehr zu, da die durchschnittliche Diskrepanz nach dieser Metrik zwischen den Input  $x$  und den von Decoder erzeugten Output  $x'$  nur noch im 0.0005 Bereich liegt, wie in Abb.43 nach Epoche 500 abzulesen ist. Da keine verwendbaren Trainingsdaten für den Latenten C-GAN Versuchsaufbau 2.1 generiert werden konnten, soll an dieser Stelle der Versuchsaufbau 2.1 Latent C-GAN mit Chamfer Distanz nicht weiter geführt werden.

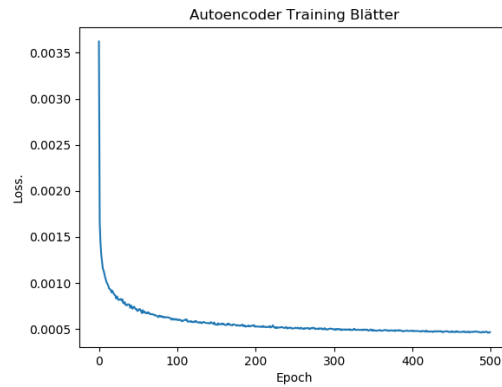


Abb. 43: Trainingsverlauf des Autoencoder mit CD für zerstörte Blattdaten

Um ein besseres Validierungsergebnisse zu erhalten wurde Versuchsaufbau 2.1 Latent-CGAN noch mit EMD Zielfunktion für Autoencoder getestet. Es wurden 800 Epochen trainiert. Die Ergebnisse sind ähnlich zu denen mit der CD vgl.2.4. Beispiel Trainingsinput Trainingsdatensätze können in Abb. 46 entnommen werden mit den jeweiligen erzeugten Output des Decoders in Abb. 44 entnommen. Wie zu erkennen ist, vervollständigt der Autoencoder mit der EMD die Blattflächen selbstständig, obwohl dies nicht als Ziel des Autoencoder ist, sondern lediglich das Erlernen einer Kodierung. Im Vergleich dazu sind die Blätter aber an der Stelle des Loches gleichmäßiger Verteilt und es gibt keine höheren Punktansammlungen auf der Blattfläche. Wenn man die Ergebnisse mit der Hinsicht der Zielfunktion berücksichtigt liefert die CD ein besseres Ergebnis. Die Begründung hierbei ist das auf der Blattoberfläche der Blätter die Punkte gleichmässiger verteilt sind. Möchte man nun die beiden Ergebnisse ebenfalls die Qualität des Rekonstruieren messen, obwohl dies nicht das eigentliche Ziel ist, sticht die EMD hervor, da im Bereich der Löcher die Punkte gleichmäßiger verteilt sind.

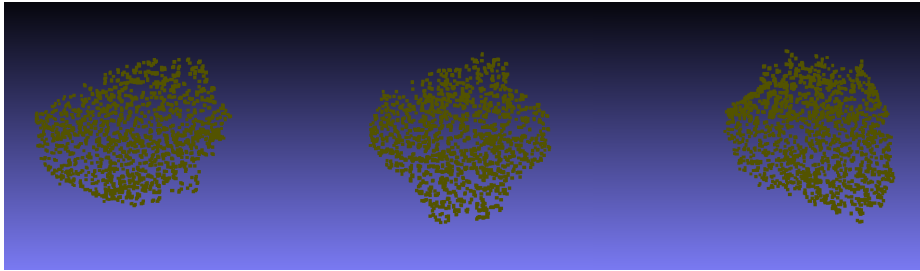


Abb. 44: Trainingsbeispiele des Autoencoder mit EMD für zerstörte Blattdaten  
- Output

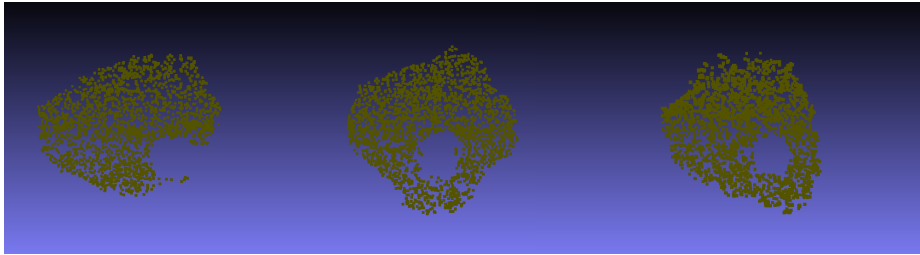


Abb. 45: Trainingsbeispiele des Autoencoder mit EMD für zerstörte Blattdaten  
- Input

Insgesamt sind die Distanzen niedriger und die Dichte an den Löchern nimmt ab. Jedoch liefert die EMD dem positiven Nebeneffekt, dass die Blätter realistischer rekonstruiert werden. Da seit Epoche 300 das Training stagniert ist darauf zu schließen, dass selbst beim weiteren Training nicht die Löcher, welche in den Blättern entstehen soll, herzustellen sind. Außerdem lässt das durch die Chamfer Distanz berechnete Abstand zwischen den beiden Punktwolken keine starke Verbesserung mehr zu, da die durchschnittliche Diskrepanz zwischen den Input  $x$  und den von Decoder erzeugten Output  $x'$  nur noch im 0.03 Bereich liegt wie auf Abb.43 in Epoche 500 abzulesen ist. Das Training entwickelt sich jedoch seit mehreren Epochen nicht mehr und stagniert. Auch mit dieser Anwendung kann das Latente C-GAN nicht weitergeführt werden, da keine erfolgreichen Trainingsdaten erzeugt werden konnten.

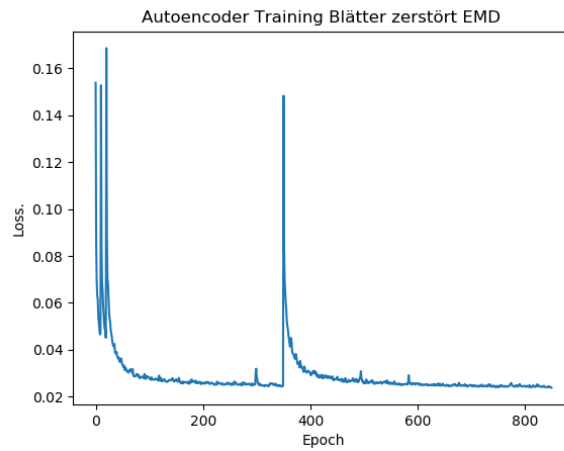


Abb. 46: Trainingsverlauf des Autoencoders mit EMD für zerstörte Blattdaten

Für den Versuchsaufbau 2.2 RAW-CGAN mit Fully-Connected-Layern mit EMD ist auf Abb. 49 rechts der Loss des Discriminators abgebildet und links der Loss des Generators. Der Loss des Generator sinkt zwar aber die generierten Trainingsbeispiele in Abb. 48 ähneln keine reparierten Blätter. Der dazugehörige Input ist in Abb.47 zu sehen. Da der Generator schon an sein Limit gerät und keine erhebliche Verbesserung mehr möglich ist, ist festzustellen, dass durch einen Generator mit Fully-Connected-Layern keine Blattdatenrekonstruktion stattfinden kann.

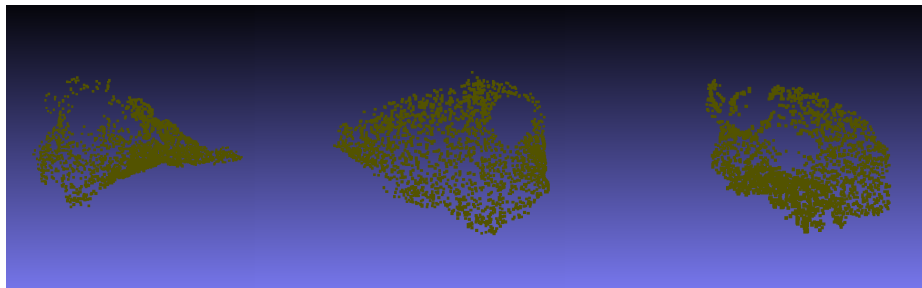


Abb. 47: Der Input aus welchen vom RAW-CGAN mit EMD die Daten rekonstruiert wurden

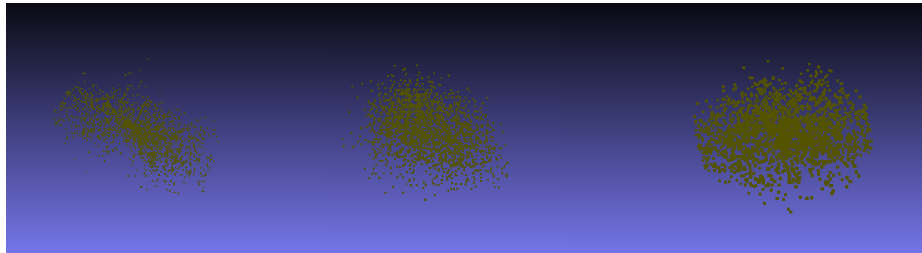


Abb. 48: Output des Generators vom RAW-CGAN mit EMD

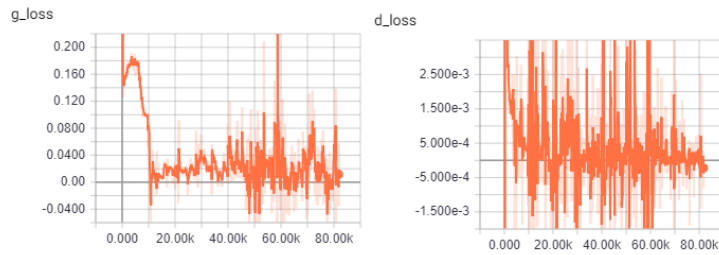


Abb. 49: Generator und Discriminator Loss des RAW-CGAN mit EMD

In Folgenden die Ergebnisse von Testaufbau 2.2 RAW-CGAN mit Convolutional Layer im Discriminator und der Vanilla-GAN-Loss vorgestellt. Der Trainingsverlauf vom Generator und Discriminator kann Abb. 52 entnommen werden. Die Sprünge sehen auf den ersten Blick sehr stark, aus beachtet man aber die Scalenwerte auf der Y-Achse des Generators, bewegt sich der Loss zwischen 0.691 und 0,696. Dies lässt auf keine starke Verbesserung schließen. Der Discriminator kann die Unterscheidung von Fake und Real Datensatz perfekt erkennen. Das Training läuft über mehrere Epoche konstant hinweg und lässt auch für einen längeren Durchlauf auf keine Verbesserung schließen. Einige Beispiel-tabakblätter welche vom Generator erstellt wurden sind können aus Abb. 50 zu entnehmen so wie der dazu gehörige Input aus Abb. 51. Diese sind jedoch keine hohe Ähnlichkeit zu Blätter zu erkennen, man erkennt eine Verbesserung zum RAW-CGAN mit Fully-Connected-Layern, da sich bessere Blattränder abbilden jedoch führen starke Punktansammlungen zu keinen dichten Blattflächen. Da es nach der Loss Funktion keine gravierenden Veränderungen mehr zu erwarten sind, kann festgehalten werden das keine Rekonstruktion von zerstörten Blattdaten mit den in Kapitel 3.4 beschriebenen RAW-CGAN mit Convolutional-Layer und Vanilla-GAN-Loss, möglich ist.

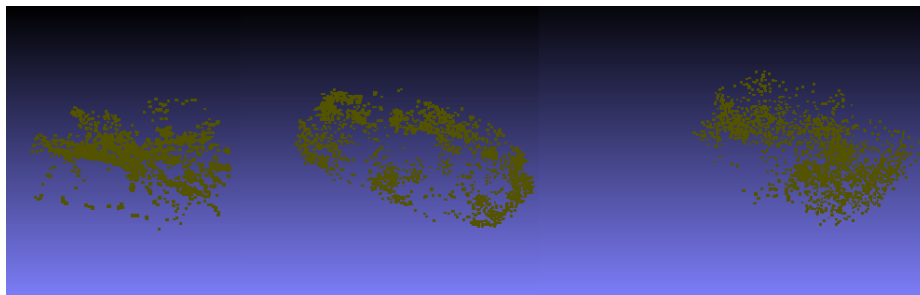


Abb. 50: Der Output aus welchen vom RAW-CGAN mit Convolutional Layern die Daten rekonstruiert wurde

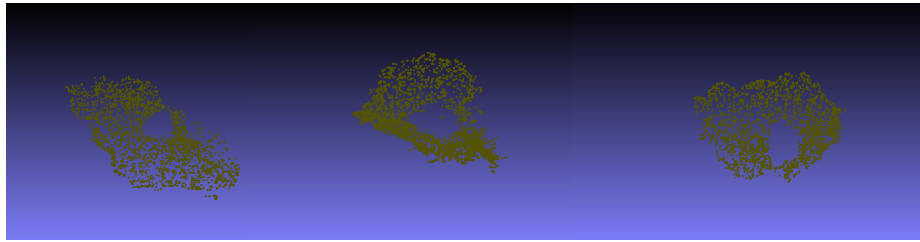


Abb. 51: Der Input aus welchen vom RAW-CGAN mit Convolutional Layern die Daten rekonstruiert wurden



Abb. 52: RAW-CGAN mit Convolutional-Layern und Vanilla-GAN-Loss

Zuletzt erfolgen die Ergebnisse von Testaufbau 2.2 RAW-CGAN mit Deconvolutional-Layer mit der Wasserstein-Metric, was mit den nicht komprimierten Daten trainiert wird. Der Trainingsverlauf vom Generator und Discriminator kann Abb. 53 entnommen werden. Der Verlauf des Discriminator Loss arbeitet sich konstant nach unten und hat zwei Ausbrecher welche aber kein Einfluss auf den weiteren Trainingsverlauf hat, da der Generator durch den Gradienten kein Aufschwung erlangt. Der Generator springt zwar noch am Ende des Trainings aber die Sprünge sind in einen so niedrigen Intervall das keine große Veränderung mehr erwartet werden kann. Die vom Generator generierten Outputs in Abb. 54, und deren dazugehöriger Input in Abb. 55 zeigen bereits die Umrisse der Blätter aber jedoch sind die Blattflächen nicht gebildet. Fast man Versuchsaufbau 2 zu-

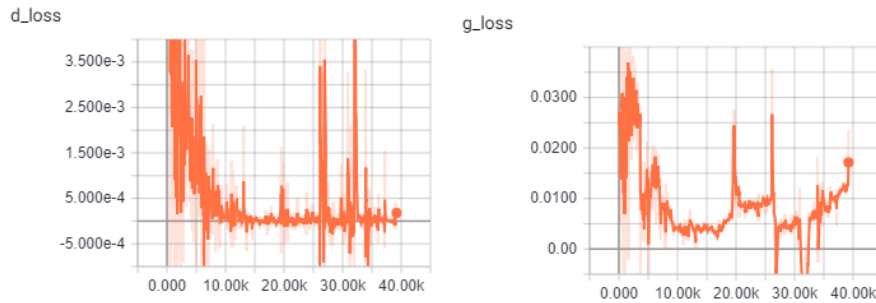


Abb. 53: Discriminator und Generator Loss RAW-DGAN mit Convolutional-Layer und Wasserstein Metrik

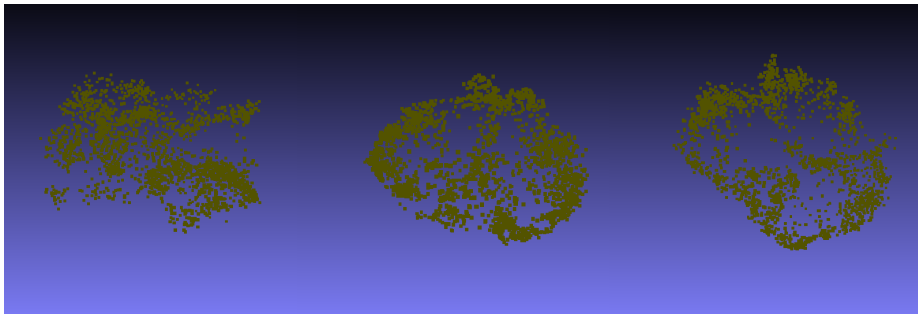


Abb. 54: Der Output aus welchen vom RAW-CGAN mit Convolutional Layern die Daten rekonstruiert wurde

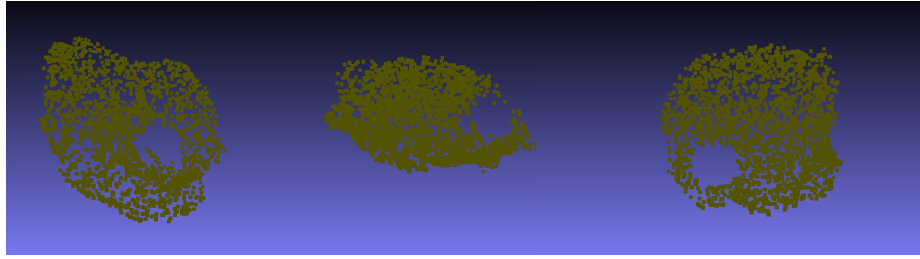


Abb. 55: Der Input aus welchen vom RAW-CGAN mit Convolutional Layern die Daten rekonstruiert wurden

sammen ist festzustellen das in Versuchsaufbau 2.1 abgebrochen werden musste da die Autoencoder nicht fähig waren, die zerstörten Trainingsdaten zu erlernen. Jedoch konnten sie die Daten rekonstruieren, obwohl dies nicht als Ziel für die Autoencoder definiert war. Ob eine praktische Anwendung hier durch ermöglicht ist und ob diese Rekonstruktion nur für die Form von Blätter möglich ist, oder Beispielsweise auch ähnliche Ergebnisse beim Stuhldaten Satz liefert, muss in folge Arbeiten geprüft werden.

Die beiden Zielfunktionen des Autoencoder, CD und EMD, hatten beide ein ähnliches Problem zu erkennen, welcher Punkt in Punktwolke  $x$  der dazugehörige Punkt in Punktwolke  $y$  ist, auf welchen die Optimierung statt finden soll. Beide lösten dies indem sie die nächsten Nachbarn eines Punktes aus der anderen Punktwolke suchen. Jedoch kann dieses Mapping zu Problemen führen, wie man in Abb.56 erkennen kann. Die rote Punktwolke ist der erste Versuch des Autoencoders die grüne Punktwolke zu kopieren. Eine Zuordnung in diesem Stadium, welcher Punkt in Rot zu Grün gehört ist sehr schwierig. Dadurch resultiert das Punkte in Zentrum eher als nächster Nachbar identifiziert werden. Im weiteren Trainingsverlauf dehnt die rote Punktwolke sich aus, lässt aber durch die Ansammlung im Zentrum und der bijektive Zuordnung der Punkte zueinander, nicht zu das sich die Löcher bilden. Für eine weitere Nutzung könnte an dieser Stelle angesetzt werden und andere Zielfunktionen für Autoencoder mit Punktwolken entwickelt werden. Oder den Encoder dahingehend umbauen das der Output frühere Lernepochen eher der Grundform eines Blattes ähnelt. Im Versuchsaufbau 2.2 zeigen sich mit der Verwendung von Convolutional-Layer im Discriminator bessere Ergebnisse. Ein qualitativer Unterschied zwischen dem Versuchsaufbau 2.2 RAW-CGAN mit Convolutional-Layer und Vanilla-GAN oder Wasserstein Metrik lässt sich nicht feststellen. Jedoch lässt die Convolutional-Layer das Erlernen der Latenten-Struktur besser gelingen. Eine Rekonstruktion ist aber auch mit diesen Model nicht möglich. Jedoch kann nicht ausgeschlossen sein das GAN durch andere Aufbauten wie Beispielsweise mit 1D-Deconvolutional-Layer, implementiert werden, welche bei Tensorflow 1.12 als Prototyp zur Verfügung steht. Eine weitere Möglichkeit wäre auf die Arbeit von Cai, Zhongang und Yu (Cai et al., 2018) anzusetzen dieser postuliert eine Heran-



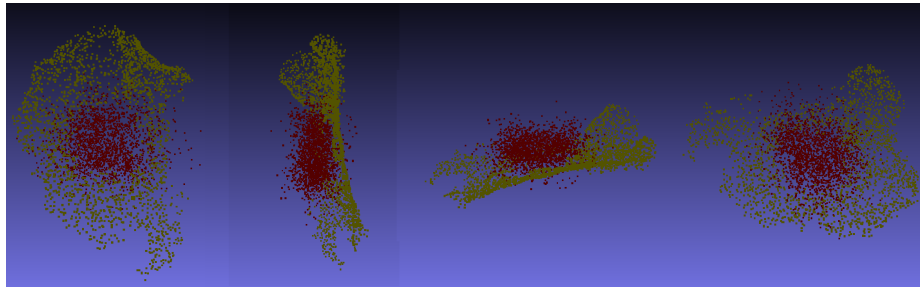


Abb. 56: Rot der Output des Autoencoders in Epoche 1, Grün der Input des Autoencoders

gehensweise für 3D-Convolution Methode, welche auf 3D-Punktwolken arbeitet und eine Invarianz im Euklidischen Raum lernt mit deren Hilfe es möglich sein soll Rotationen der Punktwolke zu erlernen. Durch diese neuen Module könnte das Erlernen des latenten Raumes von 3D-Punktwolken möglich gemacht werden und das letztendliche Ziel, 3D-Punktwolken zu rekonstruieren erreicht werden.

## 5 Zusammenfassung und Diskussion

3D-Daten bleiben eine Herausforderung zu den vergleichsweise anderen Daten wie Bilder oder Audio da ihr Suchraum erheblich höher ist, dieses Thema wurde auch schon in Kapitel 2.2.2 thematisiert. Durch den Informationsgehalt von Punktwolken erhöht sich der Suchraum für GANs und stellt Forscher vor neuen Herausforderungen. Die Fragestellungen welche in dieser Arbeit überprüft werden sollten

**Fragestellung 1** Können durch GANs 3D-Punktwolken von Tabakblättern erlernt werden um neue Datensätze zu generieren?

**Fragestellung 2** Können durch GANs 3D-Punktwolken von Tabakblättern, die von ihrem Urzustand abgebracht wurden, rekonstruiert werden?

ist festzuhalten das Fragestellung 1 (vgl. 3.3) durch GANs machbar ist jedoch noch Verbesserungen in der Qualität möglich sind. Der derzeitige Aufbau welcher die besten Ergebnisse hervorbringt ist das Latent-GAN. Aber auch beim RAW-GAN für Punktwolken ist eine Erhöhung der Qualität festzustellen sollten mehr Daten zur Verfügung stehen. Außerdem könnte eine systematischere Datenvorverarbeitung der Trainingsdaten bessere Ergebnisse hervor bringen, wie beispielsweise eine Zentrierung der Massepunkte jedes Blattes auf den Koordinaten Ursprung oder eine einheitlichere Ausrichtung der Blattflächen zu einem Fixpunkt im Euklidischen Raum. Bei den strukturellen Aufbau des GAN könnte ein 1D-Deconvolutional-Layer implementiert werden, welche bei Tensorflow 1.12 als Prototyp zur Verfügung steht. Eine andere Möglichkeit wäre auf die Arbeit von Cai, Zhongang und Yu (Cai et al., 2018) zurückzugreifen. Diese postuliert eine Methode für 3D-Convolution Methode, welche auf 3D-Punktwolken arbeitet und eine Invarianz im Euklidischen Raum lernt, mit deren Hilfe das Erlernen von Rotationen der Punktwolken möglich gemacht werden soll.

Für Fragestellung 2 welches in Kapitel 3.4 überprüft wurde ist festzuhalten das der Aufbau des Latent-CGAN abgebrochen werden musste, da der Autoencoder es nicht geschafft das die zerstörten Trainingsdaten zu erlernen. Und somit eine weitere Durchführung des Versuchsaufbau unmöglich macht. Jedoch war festzustellen das die Autoencoder die Löcher in den Blättern mit Punkte füllten und somit eine Rekonstruktion möglich machten. Dies ist auf die beiden Zielfunktionen des Autoencoder zurück zuführen welche eine durch das Distanzmaß zwischen 2 Punktwolken dafür sorgt das die Punktwolken gedehnt werden. Ob eine praktische Anwendung dadurch ermöglicht ist und ob diese Rekonstruktion nur für die Form für Blätter gilt oder auch Beispielsweise ähnliche Ergebnisse beim Stuhldaten Satz liefert muss in Zukünftigen Arbeiten geprüft werden. Dies kann durch die gleiche Datenverarbeitung mit den Spheren für Stuhldaten überprüft werden.

Das RAW-CGAN zeigte bessere Ergebnisse mit der Verwendung von Convolutional-Layern im Discriminator. Jedoch war auch mit diesen Versuchsaufbau die Re-

konstruktion von Daten nicht möglich. Es zeigte sich doch eine Verbesserung mit den Convolutional-Layern zum Vergleich der Fully-Connected Layern. Da bei der Rekonstruktion von Bildern erst durch den Einsatz verbesserter Convolution Methoden es erst Möglich gemacht wurde Bilder zu rekonstruieren und zu verändert (Dong et al., 2016a) und diese Convolution speziell auf 2D-Daten entwickelt wurde. Deshalb lässt sich nicht ausschließen das durch Verwendung von Layern welche speziell für 3D-Daten entwickelt wurden bessere Ergebnisse hervorbringen können. Auch hier können wie beim Fragestellung 2 beispielsweise mit 1D-Deconvolutional-Layer implementiert werden. Eine andere Möglichkeit wäre auf die Arbeit von Cai, Zhongang und Yu (Cai et al., 2018) anzusetzen dieser postuliert eine Methode für 3D-Convolution Methode welche auf 3D-Punktwolken arbeitet und eine Invarianz im Euklidischen Raum lernt mit deren Hilfe es leicht sein soll Rotationen der Punktwolke zu erlernen. Durch diese neuen Module könnte das Erlernen des latenten Raumes von 3d-Punktwolken möglich gemacht werden und das letztendliche Ziel, 3D-Punktwolken zu rekonstruieren möglich gemacht werden. Eine weitere Möglichkeit ist auch an der Zielfunktion von GAN zu ändern wie in der Arbeit von Zhu, Jun-Yan und Park (Zhu, Park, Isola, & Efros, 2017), mit ihren Pix2Pix Model. In dieser Verwendet sich die C-GAN Loss Funktion und ändern diese um Zusätzlich zu der üblichen Loss-Funktion die Differenz zwischen Generator Output und den dazugehörigen Urzustand Trainingsdatensatz zu berechnen und addieren dies zu der Loss-Funktion. Der Optimizer versucht nun dieses Differenz zusätzlich zu Minimieren. Man könnte dieses Verfahren nun für Punktwolken übernehmen und mit der CD oder EMD implementieren. Trotz das qualitativ Rekonstruktion ist potential die Forschungsfragen in Zukünftigen Arbeiten weiter zuverfolgen.

## Abbildungsverzeichnis

1	Künstliches Neuronales Netzwerk(Haque et al., 2018) . . . . .	8
2	künstliches Neuron . . . . .	9
3	Sigmoid Funktion . . . . .	9
4	Optimierung einer Zielfunktion . . . . .	11
5	Minimum,Maximum und Sattelpunkt einer Zielfunktion(Morabbi et al., 2018) . . . . .	11
6	Neuronales Netzwerk mit Batch-Normalization-Layer( <i>batchnorm</i> Chun, n.d.) . . . . .	14
7	Polygon File Format . . . . .	15
8	Visualisierte Punktwolke eines Tabakblattes . . . . .	15
9	Scankopf und Scanaufbau für die 3D-Punktwolken Gewinnung . . . . .	16
10	3D Punktwolke einer Tabakpflanze . . . . .	17
11	Convolutional Neural Network . . . . .	19
12	Convolution Beispiel(Dumoulin & Visin, 2016) . . . . .	20
13	Deconvolution Beispiel (Dumoulin & Visin, 2016) . . . . .	21
14	Autoencoder . . . . .	22
15	Earth Mover Distance . . . . .	23
16	Chamfer Distance . . . . .	24
17	Generativ Adversarial Network . . . . .	25
18	Deep Convolutional GAN( <i>dc-gan online book</i> , n.d.) . . . . .	26
19	Conditional Adversarial Network(Mirza & Osindero, 2014) . . . . .	30
20	Latent-GAN . . . . .	31
21	Bild Rekonstruktion eines Hundes welches durch ein Artefakt zerstört wurde . . . . .	33
22	3D Punktwolke einer Tabakpflanze . . . . .	34
23	3D Punktwolke eines Tabakblattes mit 2048 Punkten . . . . .	35
24	Tabakblatt mit Sphäre welche die Differenzmenge bilden . . . . .	36
25	Sphäre im Raum welche das verdecken der Blätter simulieren . . . . .	36
26	3D Punktwolke von zerstörten Tabakblättern . . . . .	36
27	Aufbau des RAW-CGAN . . . . .	39
28	Generierte Beispieldaten vom Generator des RAW-GAN mit den Blattdaten . . . . .	40
29	24 Punktwolken aus dem Blattdatensatz in einen Koordinatensystem . . . . .	41
30	Trainingsverlauf des RAW-GAN mit den Blattdaten . . . . .	41
31	Generierte Beispieldaten von Generator des RAW-GAN mit den Stuhldaten . . . . .	42
32	Trainingsverlauf des RAW-GAN mit den Stuhldaten . . . . .	42
33	Trainingsbeispiele des RAW-GAN mit den 422 Stuhldaten . . . . .	43
34	Trainingsverlauf des RAW-GAN mit den 422 Stuhldaten . . . . .	43
35	Trainingsbeispiele des RAW-GAN mit den 422 Stuhldaten . . . . .	44
36	Trainingsergebnisse des Latent-GAN mit den Blattdaten . . . . .	44
37	Trainingsverlauf des Latent-GAN mit den Blattdaten . . . . .	45
38	Trainingsergebnisse des Latent-GAN mit den Stuhldaten und 400 Trainingsdaten . . . . .	45

39	Trainingsverlauf des Latent-GAN mit den Stuhl- und 400	
	Trainingsdaten .....	46
40	24 Stuhl- und 400 in ein Koordinatensystem geladen .....	47
41	Trainingsbeispiele des Autoencoders mit CD für zerstörte Blatt- und 48	
	- Output .....	48
42	Trainingsbeispiele des Autoencoders mit CD für zerstörte Blatt- und 48	
	- Input .....	48
43	Trainingsverlauf des Autoencoders mit CD für zerstörte Blatt- und ...	49
44	Trainingsbeispiele des Autoencoders mit EMD für zerstörte	
	Blatt- und 48 - Output .....	50
45	Trainingsbeispiele des Autoencoders mit EMD für zerstörte	
	Blatt- und 48 - Input .....	50
46	Trainingsverlauf des Autoencoders mit EMD für zerstörte Blatt- und .	51
47	Der Input aus dem vom RAW-CGAN mit EMD die Daten	
	rekonstruiert wurden .....	52
48	Output des Generators vom RAW-CGAN mit EMD .....	52
49	Generator und Discriminator Loss des RAW-CGAN mit EMD .....	53
50	Der Output aus dem vom RAW-CGAN mit Convolutional	
	Layern die Daten rekonstruiert wurde .....	53
51	Der Input aus dem vom RAW-CGAN mit Convolutional Layern	
	die Daten rekonstruiert wurden .....	54
52	RAW-CGAN mit Convolutional-Layern und Vanilla-GAN-Loss .....	54
53	Discriminator und Generator Loss RAW-DGAN mit Convolutional-	
	Layer und Wasserstein Metrik .....	55
54	Der Output aus dem vom RAW-CGAN mit Convolutional	
	Layern die Daten rekonstruiert wurde .....	55
55	Der Input aus dem vom RAW-CGAN mit Convolutional Layern	
	die Daten rekonstruiert wurden .....	56
56	Der Output des Autoencoders in Epoche 1, Grün der Input des	
	Autoencoders .....	57

## Literatur

- Achlioptas, P., Diamanti, O., Mitliagkas, I., & Guibas, L. (2018). Learning representations and generative models for 3d point clouds.
- Ahmed, E., Saint, A., Shabayek, A. E. R., Cherenkova, K., Das, R., Gusev, G., ... Ottersten, B. (2018). Deep learning advances on different 3d data representations: A survey. *arXiv preprint arXiv:1808.01462*.
- Arjovsky, M., Chintala, S., & Bottou. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.
- Bang, D., & Shim, H. (2018). Improved training of generative adversarial networks using representative features. *arXiv preprint arXiv:1801.09195*.
- batchnorm chun*. (n.d.). <http://sanghyukchun.github.io/88/>. (Accessed: 2019-02-16)
- Cai, Z., Yu, C., & Pham, Q.-C. (2018). 3d convolution on rgb-d point clouds for accurate model-free object pose estimation. *arXiv preprint arXiv:1812.11284*.
- dc-gan online book*. (n.d.). [https://gluon.mxnet.io/chapter14\\_generative-adversarial-networks/dcgan.html](https://gluon.mxnet.io/chapter14_generative-adversarial-networks/dcgan.html). (Accessed: 2019-02-16)
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016a). Image super-resolution using deep convolutional networks. , *38*(2), 295–307.
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016b). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295–307.
- Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*.
- Fan, H., Su, H., & Guibas, L. (2017). A point set generation network for 3d object reconstruction from a single image. In *2017 IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 2463–2471).
- Golik, P., Doetsch, P., & Ney, H. (2013). Cross-entropy vs. squared error training: a theoretical and experimental comparison. In *Interspeech* (Vol. 13, pp. 1756–1760).
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
- Haque, M. R., Islam, M. M., Iqbal, H., Reza, M. S., & Hasan, M. K. (2018). Performance evaluation of random forests and artificial neural networks for the classification of liver disorder. In *2018 international conference on computer, communication, chemical, material and electronic engineering (ic4me2)* (pp. 1–5).
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504–507.
- Huszár, F. (2015). How (not) to train your generative model: Scheduled sampling, likelihood, adversary? *arXiv preprint arXiv:1511.05101*.

- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. *arXiv preprint arXiv:1804.07723*.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Morabbi, S., Keyvanpour, M., & Shojaedini, S. V. (2018). A new method for p300 detection in deep belief networks: Nesterov momentum and drop based learning rate. *Health and Technology*, 1–16.
- Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems* (pp. 841–848).
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Ravanbakhsh, S., Schneider, J., & Póczos, B. (2016). Deep learning with sets and point clouds. *arXiv preprint arXiv:1611.04500*.
- Ruder, S. (2016). *arXiv preprint arXiv:1609.04747*.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. In *Advances in neural information processing systems* (pp. 2234–2242).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International conference on machine learning* (pp. 1139–1147).
- Wu, J., Zhang, C., Xue, T., Freeman, B., & Tenenbaum, J. (2016). Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in neural information processing systems* (pp. 82–90).
- Yi, L., Shao, L., Savva, M., Huang, H., Zhou, Y., Wang, Q., ... others (2017). Large-scale 3d shape reconstruction and segmentation from shapenet core55. *arXiv preprint arXiv:1710.06104*.
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*.