

Deep Learning - Lernen von vielschichtigen neuronalen Netzen

KJARTAN FERSTL

SEMINARARBEIT

eingereicht am

Fachhochschul-Masterstudiengang

INFORMATION ENGINEERING UND -MANAGEMENT

in Hagenberg

im Juli 2014

Inhaltsverzeichnis

Kurzfassung	iii
Abstract	iv
1 Einleitung	1
1.1 Motivation	1
1.2 Ziele und Aufgaben	2
2 Grundlegendes	3
2.1 Definition	3
2.2 Entstehung	3
2.3 Hürden	4
2.4 Neuronale Netze	4
2.5 Überwachtes / unüberwachtes Lernen	6
3 Algorithmen und Strukturen	7
3.1 Backpropagation	7
3.2 Restricted Boltzmann Machines	9
3.2.1 Grundgedanke	10
3.2.2 Abkürzung	11
3.2.3 Begründung	11
3.2.4 Eignung	11
3.3 Deep Autoencoders	12
3.4 Faltungskodierte neurale Netze	13
3.5 Sparse coding	14

Inhaltsverzeichnis	ii
4 Anwendungen	15
4.0.1 Objekterkennung in Bildern	15
4.1 Zahlen erkennen	16
4.2 Spracherkennung	17
4.3 Hardware	18
4.4 Weitere Anwendungen	18
5 Zusammenfassung	19
Literaturverzeichnis	21

Kurzfassung

Deep Learning-Algorithmen bieten eine Möglichkeit vielschichtige neuronale Netze zu trainieren. Solche Netze kommen in aktuellen Anwendungen zu Erkennung von Objekten in Bildern, Text und Inhalt in Sprache und vielen anderen Gebieten zum Einsatz. Forschungen an den Algorithmen ermöglichen es, die sehr rechenintensiven Algorithmen zu vereinfachen und in großem Stil auf parallelisierten Computersystemen auszuführen. So ist es in einigen Gebieten bereits möglich, mit neuronalen Netzen bessere Ergebnisse als mit optimierten mathematischen Verfahren zu erzielen. Im Folgenden werden die wesentlichen Algorithmen und Strukturen von neuronalen Netzen aufgearbeitet, und die Schwierigkeiten und Lösungsansätze aufgezeigt. Außerdem werden einige aktuelle Anwendungen aus der Bild- und Sprachanalyse betrachtet.

Abstract

Deep learning algorithms make it possible to train multi layer neural networks. Such networks are used to recognize objects in images, to extract text and content in speech and for many other applications. Research in this area made it possible to reduce the complexity of such algorithms and to compute them in parallel on large scale run networks. Due to these optimizations, it is already possible, to produce better results with neural networks than with existing optimized mathematical algorithms for some applications. This paper focuses on the major algorithms and structures used for neural networks, the problems they come with and how they can be solved. Moreover, some applications from object recognition in images and voice analysis are examined.

Kapitel 1

Einleitung

1.1 Motivation

Komplexe neuronale Netze können bei einigen Problemstellungen, besonders in der Bild- und Spracherkennung ein sehr mächtiges Mittel zur Problemlösung sein. Das Lernen solcher Netze bringt Schwierigkeiten mit sich, die in den vergangenen Jahrzehnten schlecht oder gar nicht gelöst werden konnten. Deep Learning ist ein Ansatz, mit dem solche Netze trainiert werden können.

Aufgrund der Weiterentwicklung von PC-Hardware und der Verwendung der Grafik-Recheneinheit (GPU), sind seit den 00er-Jahren bereits bemerkenswerte Ergebnisse möglich. Momentan arbeiten sogar Chiphersteller an preiswerten, dedizierten Chips mit trainierbaren neuronalen Netzen. Qualcomm wird den ersten kommerziellen Chip mit einem integrierten neuronalen Netz noch dieses Jahr (2014) veröffentlichen [Talbot, 2013]. Neuronale Netze sind immer nur so gut, wie sie trainiert werden. Eines der wesentlichsten Themen im weiteren Fortschritt von neuronalen Netzen sind daher die Algorithmen zum Trainieren.

Aus der eingeleiteten Motivation ergeben sich daher folgende Kernaufgaben für die Seminararbeit:

- Einführung in die Problematik von Deep Learning mit kurzem Blick auf die bisherige Geschichte
- Analyse der Hürden im Deep Learning
- Analyse aktueller Deep Learning-Algorithmen

1.2 Ziele und Aufgaben

Folgende zentrale Fragestellungen sollen in der Seminararbeit beantwortet werden:

- Was ist Deep Learning?
- Welche Mustertypen können erkannt werden, und welche Anwendungen sind möglich
- Wie sind die gelernten Modelle konkret definiert?
- Welche Verfahren werden verwendet, um vielschichtige Modelle zu lernen?
- Was ist aktuell mit Deep Learning in Verbindung mit neuronalen Netzen bei der Mustererkennung in Bildern möglich?

Ziel dieser Arbeit ist es, dem interessierten Leser einen Einstieg in das Thema Deep Learning, mit besonderem Fokus auf neuronale Netze, zu vermitteln. Aufbauend auf diesen Erkenntnissen soll gezeigt werden, welche wesentlichen Algorithmen bis heute entwickelt wurden und welche technischen Problemstellungen mit diesen Algorithmen gelöst werden können oder sogar bereits heute damit gelöst werden. Ein wesentlicher Fokus soll dabei auf die Mustererkennung in Bildern gerichtet werden.

Kapitel 2

Grundlegendes

2.1 Definition

Deep Learning ist ein Ansatz aus dem maschinellen Lernen, der sich mit dem Trainieren von nichtlinearen Modellen, und hier meist mit vielschichtigen neuronalen Netzen, befasst.

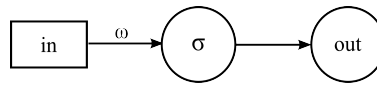
Deep Learning wird zum Beispiel zur Klassifizierung von Daten oder zur Extraktion von Merkmalen aus Daten eingesetzt.

2.2 Entstehung

Die Wurzeln der neuronalen Netze gehen zurück in die 1950er Jahre, als Frank Rosenblatt eine Maschine namens Perceptron [Rosenblatt, 1958] beschrieb. Diese Maschine war in der Lage, einige einfache Figuren wie Quadrate und Dreiecke zu erkennen. Für die damalige Zeit war das eine herausragende Maschine und regte den Gedanken, Maschinen zu bauen die Menschen imitieren können, weiter an.

Knapp 20 Jahre später schrieb Marvin Minsky [Minsky, 1969] ein Buch das die Limitierungen von Preceptrons aufzeigte und einige scheinbar fundamentale Probleme aufwarf. Diese Erscheinung ließ die neuronalen Netze wieder in den Hintergrund rücken. In den 1980er Jahren, hat der heute renommierte Wissenschaftler Geoff Hinton, neuronale Netze beschrieben [Rumelhart, 1985], die bereits eine versteckte Schicht enthielten und somit in der Lage waren, komplexere Aufgaben zu lösen. In dieser Zeit entstand auch der Backpropagation Algorithmus für das Trainieren von neuronalen Netzen. Das Trainieren dieser neuronalen Netze war sehr aufwendig und so verschwand auch sie bald wieder von der Bildfläche.

Erst mit einer Entdeckung 2006 [Teh, 2006], erneut durch Geoff Hinton, die

**Abbildung 2.1:** Neuron

den Backpropagation-Algorithmus erheblich vereinfachte, haben neuronale Netze mit mehreren versteckten Schichten wieder an Bedeutung gewonnen.

Heute stellen führende Unternehmen immer häufiger Teams rund um die Thematik der neuronalen Netze zusammen, um sich wirtschaftliche Vorteile zu sichern^{1,2}. Erste Erfolge sieht man an Microsofts Spracherkennung Cortana, Siri von Apple, Google Now oder der Bildersuche von Google, die anhand von Bildern ähnliche Bildern findet.

2.3 Hürden

Die aus der Vergangenheit bekannten Probleme des Lernens neuronaler Netze mit vielen versteckten Schichten strecken sich zum größten Teil bis heute durch. Zusätzliche versteckte Schichten erlauben es Modellen oft sich auf nicht repräsentative Abhängigkeiten der Trainingsdaten anzupassen und so tendieren vielschichtige neuronale Netze beim Training häufig zur Überanpassung. Problemstellen beim Training sind häufig auch die Unteranpassung durch zu wenig Trainingsdaten, lokalen Minima in denen die Trainingsalgorithmen hängen bleiben und die notwendige Rechenleistung.

Grundsätzlich kann man für heute verwendete Algorithmen sagen, dass die Leistung von neuronalen Netzen sehr stark von der Menge der Trainingsdaten und weniger von dem Trainingsalgorithmus selbst abhängt. Nicht zuletzt hat die Verfügbarkeit von immensen Datenmengen von Unternehmen und aus dem Internet einen sehr positiven Einfluss auf die Entwicklungen und damit das Interesse von Deep Learning.

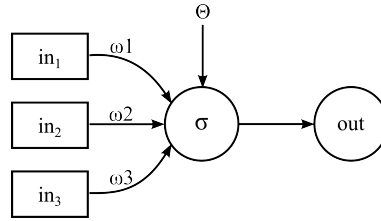
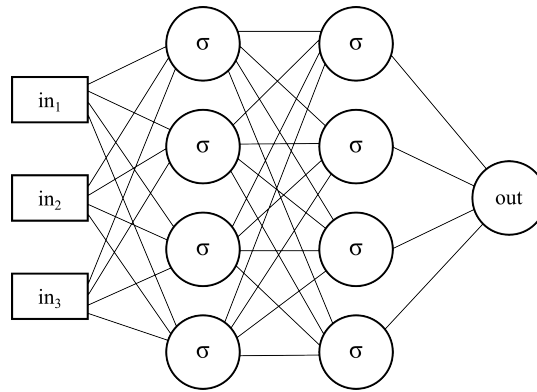
2.4 Neuronale Netze

Künstliche neuronale Netze sind Strukturen aus der Technik, die dem Nervensystem von Lebewesen ähneln. Es sind Netze die aus Eingangsdaten Ausgangsdaten über nicht lineare Zusammenhänge berechnen. Diese Netze müssen parametrisiert werden. Hier kommt Deep Learning zum Einsatz.

Neuronale Netze sind in der Technik zur Vereinfachung meist in mehreren

¹2013/3 Google stellt Geoffrey Hinton ein

²2013/12 Facebook stellt NYU Professor Yann LeCun ein

**Abbildung 2.2:** Neuron mit mehreren Eingängen**Abbildung 2.3:** Ein neuronales Netz

Schichten aufgebaut. Abbildung 2.1 zeigt ein einfaches Neuron mit nur einem Eingang. Der Ausgang wird aus der Multiplikation des Eingangswerts mit dem Gewicht und der Übertragungsfunktion σ errechnet. Als Übertragungsfunktion wird meist die Sigmoid-Funktion eingesetzt, sie lässt sich einfach differenzieren [sigmoid, 2014] und eignet sich daher besonders gut. Der Ausgang *out* dieses Neurons ergibt sich somit aus dem Eingang *in* multipliziert mit dem Gewicht ω in der Sigmoid-Funktion σ .

$$out = \sigma(\omega * in)$$

Ein Neuron hat in der Regel, wie in Abbildung 2.2 zu sehen, mehrere Eingänge. Außerdem wird ein Bias-Wert θ für die Sigmoid-Funktion hinzugefügt. Der Ausgang dieses Neurons kann somit wie folgt berechnet werden:

$$out = \sigma(\omega_1 * in_1 + \omega_2 * in_2 + \omega_3 * in_3 + \theta)$$

Um aus den einzelnen Neuronen ein Netzwerk zu bauen, werden Schichten gebildet. Eine Schicht ist eine Menge an Neuronen, die zur Vereinfachung untereinander meist nicht direkt verknüpft sind. Schichten werden hintereinander gehängt, in dem jedes Neuron einer Schicht als Eingang für jedes

Neuron der nächsten Schicht dient. Ein solches Netzwerk ist auch in Abbildung 2.3 zu sehen. Dieses Netzwerk hat eine Eingabeschicht, eine Ausgabeschicht und zwei Schichten dazwischen. Da die Ergebnisse dieser mittleren Schichten in der Regel nicht unmittelbar ein Ergebnis sind, werden sie als unsichtbar betrachtet und auch so bezeichnet. Es handelt sich hierbei also um die versteckten Schichten.

2.5 Überwachtes / unüberwachtes Lernen

Die Algorithmen des Deep Learnings werden im Groben in zwei Kategorien geteilt, in die überwachten und die unüberwachten Methoden. Überwacht bedeutet, dass Eingangsdaten an das System angelegt werden, für die die gewünschten Ausgangsdaten bekannt sind. Das könnte zum Beispiel heißen, dass das Netz mit Bildern trainiert wird, von denen einige Gesichter enthalten und einige nicht. Am Ausgang wird dann überprüft, ob das Netz die Gesichter richtig erkannt hat und wenn nicht, werden die Parameter des Netzes entsprechend angepasst.

Diese Art des Lernens scheint logisch, benötigt aber sehr viele Datensätze mit bekannten Ausprägungen. Deklarierte Datensätze sind selten in großen Mengen verfügbar und so tendiert diese Art der Algorithmen leicht zur Überanpassung [Tsoi, 1997], mehr dazu bei den jeweiligen Algorithmen.

Ein weiterer Ansatz ist das unüberwachte Lernen, es handelt sich dabei um Algorithmen, die Eingangsdaten aber keine Ausgabedaten zum Vergleich benötigen. Die grundlegende Idee dabei ist es, das Netz Merkmale aus den Eingangsdaten lernen zu lassen. Unüberwachtes Lernen ist besonders wegen der großen Verfügbarkeit unkategorisierter Daten interessant und wird häufig als Grundlage vor dem überwachten Lernen eingesetzt. Es hilft einige Probleme des überwachten Lernens zu verbessern, so können die Gewichte aus dem unüberwachten Lernen als Startwert für ein überwachtes Lernen eingesetzt werden [Bengio, 2010]. Diese Gewichte haben mehr Aussagekraft über Merkmale der Eingangsdaten als zufällige Werte und so können mit weniger kategorisierten Trainingsdaten oft wesentlich bessere Ergebnisse erzielt werden.

Kapitel 3

Algorithmen und Strukturen

3.1 Backpropagation

Backpropagation ist ein überwachter Algorithmus zum Trainieren von vereinfachten neuronalen Netzen. Ein solches vereinfachtes Netz ist in Abbildung 3.1 zu sehen und wird Feedforward-Perzeptron genannt. Überwacht bedeutet, dass das Netzwerk beim Lernen anhand von deklarierten Eingabedaten, Ausgabedaten errechnet. Der Fehler der Berechnung des Ausgangs wird beim Backpropagation-Algorithmus über lineare Algebra zur Adaptierung der Gewichte genutzt.

Der Trainingsprozess beginnt mit zufälligen Gewichten und arbeitet in den folgenden Schritten:

1. Ausgang des Netzes für bestimmte Eingabedaten berechnen

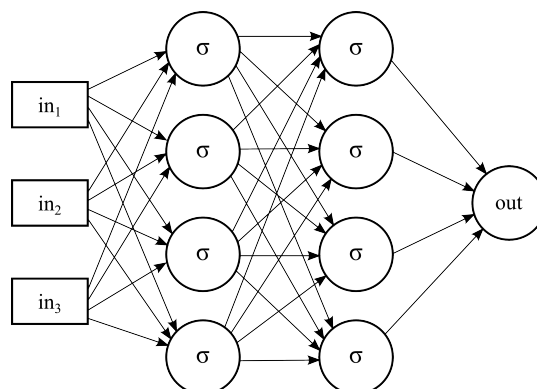


Abbildung 3.1: Ein Feedforward-Perzeptron

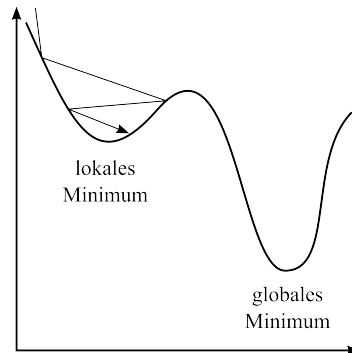


Abbildung 3.2: Lokales Minimum

2. Ausgang mit dem Sollwert vergleichen und daraus den Fehler errechnen

$$error = \frac{1}{K} \sum_{k \in K} (out_k - target_k)^2$$

<i>error</i>	...	Fehler, der später Einfluss auf die Gewichte nimmt
<i>out</i>	...	Wert, den der Ausgang erreicht hat
<i>target</i>	...	Wert, den der Ausgang erreichen sollte
<i>K</i>	...	Anzahl der Neuronen in der Ausgangsschicht

3. Gewichte abhängig von der Größe des Fehlers anpassen

Diese Schritte werden wiederholt, bis die Gewichte so angepasst wurden, dass die Ausgangsdaten möglichst gut den Vorgaben entsprechen oder eine maximale Anzahl von Iterationen erreicht ist. Die Anpassung der Gewichte errechnet sich im Wesentlichen anhand der Gradienten des Fehlers und der Gewichte. Zudem wird die Tiefe der Schicht des Gewichtes berücksichtigt.

Wie der Name des Algorithmus bereits verrät, wird der Fehler dabei im Wesentlichen über eine mathematische Funktion in die Gewichte zurückgeführt. Diese Rückführung bringt das Problem von lokalen Minima mit sich. Wird die Änderung des Fehlers geringer, so wird auch die Anpassung geringer. Befindet man sich auf dem Weg zu einem lokalen Minimum, so kann man durch den abnehmenden Fehler, wie in Abbildung 3.2 zu sehen, leicht darin hängen bleiben. Dieses lokale Minimum kann sehr weit von dem globalen Minimum entfernt sein. Die Wahrscheinlichkeit in dem lokalen Minimum hängen zu bleiben, steigt mit seiner Größe. Abhilfe kann hier der Neustart mit zufälligen Initialgewichten schaffen.

Der Algorithmus trainiert das Netz mit vordefinierte Ein- und Ausgabedaten mit dem Ziel für ähnliche Daten ähnliche Ergebnisse zu liefern. Trainiert man das Netz zu intensiv mit diesen Daten, so lernt es sie sehr genau. Dieser Effekt

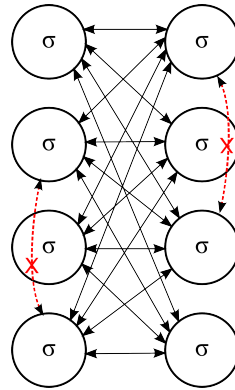


Abbildung 3.3: Eine Restricted Boltzmann Maschine

wird Überanpassung genannt und hat zur Folge, dass das Netz im späteren Betrieb schlechtere Ergebnisse liefert als hätte man es weniger lange trainiert. Abhilfe kann hier ein verfrühtes Ende des Trainings, mehr Trainingsdaten oder die Initialisierung der Gewichte durch unbeaufsichtigtes Lernen schaffen.

3.2 Restricted Boltzmann Machines

Neuronale Netze ohne Einschränkungen der Verbindungen sind schwierig und aufwendig zu trainieren. Restricted Boltzmann Maschinen, im weiteren als RBMs bezeichnet, sind eingeschränkte neuronale Netzwerke. Wie in Abbildung 3.3 dargestellt, existieren hierbei nur Verbindungen zwischen aneinander liegenden Schichten. Das Modell erlaubt keine Verbindungen von Neuronen zu sich selbst und alle Verbindungen müssen gleichermaßen in beide Richtungen verlaufen. So lässt sich unter Festlegung der Werte einer Schicht direkt auf die nächste versteckte Schicht weiter rechnen. RBMs arbeiten in der Regel mit Bool'schen Werten, lassen sich durch Erweiterung aber auch mit anderen Zahlenräumen verwenden.

RBMs wurden bereits 1986 von Paul Smolensky [Smolensky, 1986] unter dem Namen Harmonium beschrieben, fanden jedoch erst 1998, rund 10 Jahre später, durch die Entwicklung von effizienten Lernalgorithmen durch Geoffrey Hinton Anwendung.

Zum Trainieren von RBMs existieren verschiedene Algorithmen, die meist auf dem Prinzip beruhen, die Gewichte so anzupassen, dass das hin und her Rechnen zwischen zwei Schichten wieder die Ausgangsdaten ergibt. Im folgenden wird der Algorithmus von Geoffrey Hinton [Teh, 2006], der zugleich als der erste effizienten Lernalgorithmus für RBMs gesehen wird, erklärt.

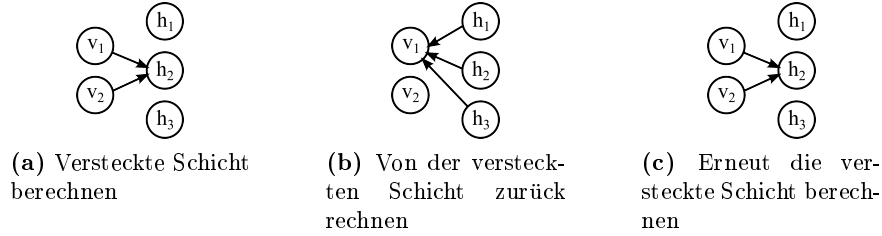


Abbildung 3.4: Berechnungsschritte zum Lernen von RBMs.

3.2.1 Grundgedanke

Sind die Werte einer Schicht fixiert, so kann einfach auf die nächste versteckte Schicht weitergerechnet werden. Zu Beginn wird ein Trainingsdatenset an die Eingänge angelegt und die folgenden Operationen, wie auch in Abbildung 3.4 zu sehen, wiederholt ausgeführt:

1. Wahrscheinlichkeit p für die versteckten Neuronen h_j anhand der sichtbaren Neuronen v_i und der Gewichte w_{ij} berechnen

$$p(h_j = 1) = \frac{1}{1 + e^{-(b_j + \sum_i (v_i * w_{ij}))}}$$

2. Gradientenmatrix $\langle v_i h_j \rangle$ über das dyadische Produkt errechnen

$$\langle v_i h_j \rangle = v * y^T$$

3. Ausgehend von den errechneten versteckten Neuronen, zurück auf die sichtbare Schicht rechnen

Wiederholt man die in der Liste angeführten Schritte sehr oft, so pendeln sich für die Neuronen Werte ein. Diese Werte hängen im Wesentlichen vom Modell und den festgelegten Wahrscheinlichkeiten ab, haben jedoch sehr wenig mit den Eingangsdaten zu tun. Anhand der Gradientenmatrix des letzten Durchlaufes lässt sich jedoch eine Differenz zu dem gesuchten Modell herausfinden, und so können die Gewichte wie folgt angepasst werden:

$$\Delta \omega_{ij} = \varepsilon (\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^\infty)$$

Diese Art der Ergebnisse lieferte sehr gute Modelle. Durch die vielen Iterationen benötigt der Algorithmus sehr viel Rechenzeit und ist daher nicht praxistauglich. Zudem ist es schwierig festzustellen, wie viele Iterationen notwendig sind, bis sich die Werte eingependelt haben.

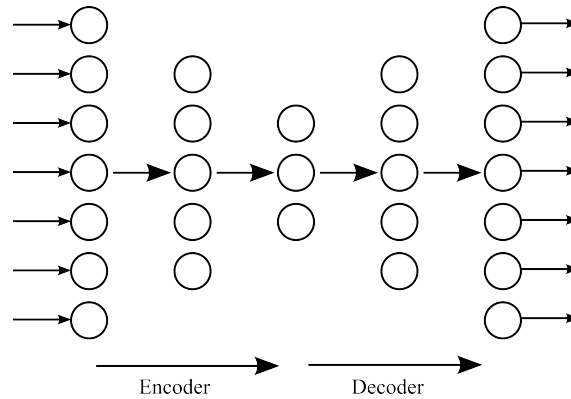


Abbildung 3.5: Deep Autoencoder

3.2.2 Abkürzung

Der oben genannte Algorithmus lässt sich in seiner Komplexität erheblich reduzieren, in dem die versteckte Schicht lediglich zweimal berechnet wird. Geoffrey Hinton hat mit der Vereinfachung gezeigt, dass es bereits mit dem Vergleich der ersten und zweiten Gradientenmatrix möglich ist, gute Ergebnisse zu erzielen [Hinton, 2002]. Diese Methode wird *Contrastive Divergence* genannt.

Die Regel zur Anpassung der Gewichte lautet dabei:

$$\Delta\omega_{ij} = \varepsilon(\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1)$$

3.2.3 Begründung

Der Grundgedanke von *Contrastive Divergence* ist es, dass das Modell mit Zufallsgewichten weg von den Eingabedaten, hin zu Daten die ihm besser gefallen, wandert. Wenn man erkennt wohin das Modell will, kann man die Gewichte so adaptieren, dass das Modell zu den Eingabedaten tendiert.

3.2.4 Eignung

Der Algorithmus benötigt, extrahiert aus den nicht deklarierten Eingabedaten, Merkmale die die Daten klassifizieren. Er eignet sich daher um Gemeinsamkeiten von Eingabedaten, wie zum Beispiel Bildern, zu finden.

3.3 Deep Autoencoders

Deep Autoencoder sind neuronale Netze aus mehreren Schichten, die sich in zwei wesentliche Teile trennen lassen. Im ersten Teil, dem Encoder, sinkt die Anzahl der Neuronen mit jeder Schicht. Im zweiten Teil, dem Decoder, steigt sie, bis am Ende wieder die Anzahl der Eingangsparameter erreicht ist. Abbildung 3.5 zeigt ein solches Netz. Im Inneren hat das Netz weniger Neuronen und somit geringere Speicherkapazität als Eingangsparameter. Es wird versucht das Netz so zu trainieren, dass es am Ausgang die gleichen Daten berechnet, wie am Eingang angelegt entstehen. Gelingt das Training, so muss das Netz Merkmale aus den Daten gelernt haben.

Richtig trainiert, komprimiert das Netz die Daten. So wäre es zum Beispiel möglich, Bilder durch das Netz zu schicken, die kleinste Schicht zu speichern, und jederzeit über den Decoder wiederherzustellen. Für Bilder, Ton und Videos gibt es einige Kompressionsalgorithmen, die sehr fein auf die Charakteristik menschlicher Organe abgestimmt sind, und Dinge die wir nicht hören oder sehen können vernachlässigen. Diese Algorithmen liefern für diesen Anwendungsfall daher meist bessere Ergebnisse. Dennoch demonstriert das Beispiel, dass damit unter Umständen erheblich Speicher und Verarbeitungsleistung eingespart werden kann. Oft entsprechen die Daten der Dekodierung sogar mehr dem gewünschten Ergebnis als zuvor. Ein Netz, das darauf trainiert ist Buchstaben zu lesen, könnte man zum Beispiel im Inneren so klein machen, dass es lediglich die einzelnen Buchstaben ohne Manipulationen speichert. Beim Wiederherstellen würde es dann den Buchstaben erzeugen, der seinem Muster am besten entspricht.

Deep Autoencoder liefern sehr vergleichbare Ergebnisse und werden daher gerne als Benchmark für Deep Learning-Algorithmen verwendet [Bengio, 2011]. Deep Autoencoder tendieren zu Unteranpassung (eng. underfitting), was auf die geringe Anzahl der Neuronen und damit der geringen Speichermöglichkeit zurückzuführen ist. Eine einfache und effektive Variante ein solches Netz zu trainieren ist es, jeweils zwei Layer separat als restricted Boltzmann Maschinen anzusehen.

Lernt man solche Netze mit Bildern und visualisiert die entstandenen Gewichte, so erkennt man, dass die Trainingsalgorithmen in erster Instanz meist Kanten und Farbintensitäten finden. In der zweiten Schicht findet man einfache Kombinationen dieser Elemente und ab der dritten Schicht werden meist schon sehr brauchbare Neuronen zur vollständigen Objekterkennung ausgebildet.

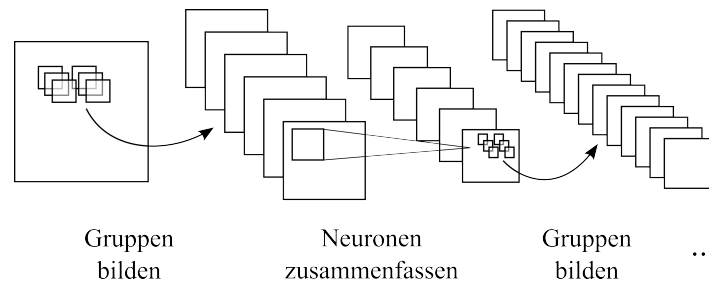


Abbildung 3.6: Faltungscodiertes neuronales Netz

3.4 Faltungskodierte neurale Netze

Faltungskodierte neuronale Netze ¹ sind neuronale Netze, die zur Reduktion der Gewichte in einzelne Teile zerteilt sind. Bei den bisher angeführten Netzen wurden jeweils alle Neuronen einer Schicht mit jedem Neuron der nächsten Schicht verbunden. Bei faltungskodierten Netzen geht man davon aus, dass weit voneinander entfernte Neuronen kaum etwas miteinander zu tun haben, während zwischen nahe aneinander liegenden Neuronen Zusammenhänge bestehen. Es werden daher Gruppen von aneinanderliegenden Neuronen gebildet. Diese Gruppen sind untereinander in die nächste Schicht verknüpft. Wie am Beispiel eines Bildes gut verständlich, wird es vorkommen, dass ein Merkmal nicht genau in eine dieser Gruppen passt. Es werden daher viele überlappende Gruppen gebildet.

Nach der Gruppenbildung gibt es eine Funktion, die die Ergebnisse aneinander liegender Werte für die nächste Schicht zusammenfasst. Dies geschieht meist über eine einfache mathematische oder statistische Funktion und hängt von der Implementierung ab. Diese Funktion könnte zum Beispiel die Kanten in mehrere Richtungen zusammenfassen, um die Rotation eines Bildes zu ignorieren.

Ein faltungskodiertes neuronales Netz wiederholt genau diese zwei Komponenten, die überlappende Gruppierung von Neuronen, sowie das Zusammenfassen von Neuronen auf jeder Schicht. Abbildung 3.6 zeigt ein solches Netz.

Dieser Typ von Netzwerk ist bereits seit 1980 [Fukushima, 1980] bekannt, wurde in den folgenden Jahrzehnten optimiert und erlebte einen Aufschwung durch die parallelisierte Berechnung auf einer GPU [Schmidhuber, 2010]. Nach weiteren Optimierungen in den vergangenen Jahren, wurde ein solches Netzwerk unter anderem für ein Projekt von Google [Le u. a., 2012] eingesetzt, das nicht zuletzt das so genannte face-Neuron ausprägte.

¹eng. Convolutional Neural Networks

3.5 Sparse coding

Sparse coding ist ein Algorithmus zum Trainieren eines neuronalen Netzes, mit dem Ansatz, möglichst aussagekräftige Merkmale zu finden. Anders als beim Deep Autoencoder wird hier nicht nur berücksichtigt wie gut das Ergebnis wiederhergestellt werden konnte, sondern auch wie eindeutig das Ergebnis war. Ziel ist es, an möglichst vielen Stellen im Ausgangsvektor den Wert Null zu berechnen und dennoch sehr gute Ergebnisse zu finden. Die Zielfunktion setzt sich aus zwei Faktoren zusammen und wird wie folgt berechnet:

$$\min \frac{1}{T} \sum_{t=1}^T \frac{1}{2} |x^{(t)} - D * h^{(t)}| + \lambda |h_1^{(t)}|$$

$x^{(t)}$...	Eingangswert
$h^{(t)}$...	Ausgangswert
D	...	Anzahl der Ausgänge
λ	...	Faktor zur Gewichtung

Die zwei zu minimierenden Bedingungen der Zielfunktion, Abweichung vom Ziel und die Zahl Nuller im Ergebnisvektor widersprechen einander teilweise und können über λ balanciert werden. Die Berechnung der Gewichte für die versteckten Schichten ist beim Sparse Coding wesentlich aufwendiger als bei Deep Autoencodern, da es sich hier um ein Optimierungsproblem, und nicht nur um eine lineare Funktion, handelt.

Kapitel 4

Anwendungen

In den vergangenen Jahren haben sich Unternehmen zunehmend an vielschichtigen neuronalen Netzen interessiert. Wie in Kapitel 2.2 beschrieben, sind darunter auch marktführende Firmen wie Google, Apple und Microsoft. Sie stellen renommierte Wissenschaftler ein und kaufen Unternehmen die in diesem Bereich tätig sind.

4.0.1 Objekterkennung in Bildern

Bei einem Projekt [Le u. a., 2012] 2012 trainierte Google das bisher größte Netz dieser Art mit 10 Millionen Bildern, aus dem Internet. Bekannt wurde das Projekt vor allem durch die Ausprägung des in Abbildung 4.1 zu sehenen Gesichts-, Katzen- und Körper-Neurons. Ein paar Eckdaten zum Netz und dem Training:

- 10 Millionen Trainingsbilder mit je 200x200 Pixel
- Autoencoder mit sparse coding
- 9 versteckte Schichten
- 1 Milliarde Verbindungen
- Training mit asynchronem stochastischem Gradientenverfahren
- Auf auf 1000 Maschinen mit jeweils 16 Rechenkernen drei Tage lang trainiert

Es wurde mit relativ kleinen Bildern, dafür aber mit sehr vielen, unbeaufsichtigt trainiert. Ein späterer Benchmark auf ImageNET, eine Datenbank mit Kategorisierten Bildern, ergab einer Trefferquote von 15,8%. Das entspricht einer Steigerung der bisherigen Bestleistung um 70%.



Abbildung 4.1: Markante Ausprägungen durch das Lernen von Bildern aus dem Internet [Le u. a., 2012]

Aus technischer Perspektive ist das Netz und damit die verwendete Technik weniger herausragend, als die Menge der Trainingsdaten und die damit verbundene Rechenleistung.

4.1 Zahlen erkennen

Google trainierte durch ihre Projekt Street View [Yuval Netzer, 2011] ein tiefes faltungskodiertes Netz, es lernte Hausnummern zu erkennen. Für die Adresssuche auf Google Maps und dem damit verbundenen Navigationssystem ist es von wesentlicher Bedeutung, Hausnummern richtig identifizieren zu können, besonders in den Gebieten, in denen Hausnummern nicht fortlaufend nummeriert sind.

Google trainierte mehrere Netze mit unterschiedlicher Anzahl an versteckten Schichten. Es stellte sich heraus, dass das tiefste Netz, mit elf versteckten Schichten, die Hausnummern am besten erkannte [Ian J. Goodfellow, 2013]. Das Netz wurde mit über 100 Millionen hochauflösenden Bildern trainiert und forderte entsprechend viel Rechenleistung. Zum Training wurde der DistBelief-Algorithmus [Jeffrey Dean und Ng, 2012] eingesetzt, ein Algorithmus der sich besonders gut auf verteilten System berechnen lässt.

Nummern, die das Netz nicht auflösen konnte, haben unbewusst Menschen gelöst. Hierzu hat Google ein Projekt ins Leben gerufen, das Besucher von Webseiten an sicherheitsrelevanten Stellen einen Code identifizieren lässt, um sicherzustellen, dass es sich dabei tatsächlich um einen Menschen und keinen Bot handelt.

Insgesamt hat das Netz bemerkenswert gut abgeschnitten [Ian J. Goodfellow, 2013] und erkennt mittlerweile 97,84% aller Hausnummern wie sie in Abbildung 4.2 zu sehen sind. Das Netz schaffte es sogar 99,8% aller Sicherheitsco-



Abbildung 4.2: Hausnummern die das neuronale Netz aus Google Street View erlernte.

des der schwierigsten Art¹, eingesetzt zum Schutz vor Bots auf Webseiten, zu lösen. Das Netz hat dabei besser als die meisten Menschen abgeschnitten, wodurch fraglich ist, wie gut diese Sicherheitscodes in Zukunft tatsächlich vor Bots² schützen werden.

4.2 Spracherkennung

Auch bei der Spracherkennung werden neuronale Netze in Kombination mit Deep Learning immer stärker und lösen nach und nach die Methode der Gaussian Mixture Models (GMM-Methoden) ab [J.P Willmore und Roberts, 2000]. Neuronale Netze für die Spracherkennung von Tonspuren funktionieren sehr ähnlich zu denen für die Bilderkennung. Wie bei der Bilderkennung prägen sich auf den ersten Schichten einfache Merkmale wie Konsonanten und Vokale aus. In höheren Schichten entstehen dann aussagekräftigere Merkmale bis hin zu vollständigen Wörtern, Phrasen und Stimmungen.

Ein heute bereits verbreitetes Feld der Anwendung ist die Sprachsteuerung von Geräten, die quasi auf allen großen Plattformen^{3,4,5,6} verfügbar ist. Durch die große Menge an verfügbaren Sprachdaten von modernen Mobiltelefonen, die zudem sehr gut auf Personentypen deklariert werden können, ist zu erwarten, dass diese Netze in Zukunft noch stark an Bedeutung und Anwendungen

¹<http://www.google.com/recaptcha>

²ein Computerprogramm das einen Menschen imitiert

³Google GOW Sprachsteuerung für Android-Geräte

⁴Apple Siri für iPhone-Geräte

⁵Microsoft Cortana für Windows Phone

⁶Diverse Systeme für Desktop-Betriebssysteme

gewinnen werden.

4.3 Hardware

2012 Hat Google mit einem neuronalen Netz das drei Tage auf 1.000 Servern mit insgesamt 16.000 Rechenkernen lief, alles bisher bekannten Netze in Komplexität und Rechenleistung übertroffen. Bereits 2013 Hat NVIDIA gemeinsam mit der Stanford Universität [Ng, 2013] ein 6,5-fach größeres neuronales Netz auf GPUs realisiert. Das Netz benötigt lediglich 16 Computer mit jeweils vier Hochleistungs-Grafikkarten, die aber aus dem Konsumentenbereich stammen. Es ist also gelungen, innerhalb von zwei Jahren Netze zu realisieren, die zuvor kaum möglich waren und dabei die Kosten für eine solche Implementierung weg von Supercomputern auf kleinere Servereinheiten bzw. Desktopcomputer zu holen.

Momentan entwickelt Qualcomm einen Controller, der eine relativ kleine neuronale Recheneinheit enthält [Talbot, 2013]. Der wesentliche Unterschied zu einer herkömmlichen CPU liegt dabei darin, dass alle Neuronen einer Schicht parallel rechnen. Das Netz benötigt daher nur wenige Zyklen bis zum Ergebnis, während herkömmliche CPUs jedes Neuron hintereinander berechnen müssen. In den frühen Entwicklungsstufen konnte das Netz bereits einige Bilderkennungsaufgaben ähnlich gut wie komplexe mathematische Algorithmen lösen. Der Chip soll 2014 in die Massenproduktion gehen und unter anderem in Mobiltelefonen zur Verfügung stehen.

4.4 Weitere Anwendungen

Weitere bemerkenswerte Anwendungsgebiete sind

- Erkennung von Verkehrsschildern, Benchmark bereits bei 99.91% Trefferquote [Igel, 2013]
- Komprimierung von Daten [Daugman, 1988]
- Alle Arten von Klassifizierungsaufgaben [Baxt, 1995] [Lippmann, 1989]
- Findung von Merkmalen in Daten [Sung, 2003]

Kapitel 5

Zusammenfassung

In der Arbeit wurden die wesentlichen Konzepte und Algorithmen rund um das Thema Deep Learning aufgearbeitet. Die verschiedenen Algorithmen des Deep Learnings basieren auf sehr ähnlichen Prinzipien und wurden mit der Entwicklung der Technik an diese angepasst. So sind moderne Algorithmen, wie der zum Beispiel der DistBelief-Algorithmus, im Wesentlichen darauf ausgelegt eine parallelisierte Berechnung zu ermöglichen und somit die verfügbare Hardware bestmöglich auszunutzen.

Aktuelle Netze sind meist sehr vielschichtig und werden mit unbeaufsichtigtem Lernen vorab trainiert. Häufig erübrigt sich dadurch aufwendiges Feature-Engineering durch Expertenwissen, und wird stattdessen automatisch durch das unbeaufsichtigte Lernen realisiert.

Bereits heute werden aktiv Netze zur Erkennung von Verkehrsschildern, Hausnummern und Gesichtern eingesetzt. Im Grundsatz zeigen momentan Benchmarks, dass die Ergebnisse von neuronalen Netzen im Wesentlichen von der Menge der Trainingsdaten und der Größe eines Netzes abhängen. Durch das Internet und die zunehmende Vernetzung existieren in vielen Bereichen Daten in sehr großen Mengen, die für unüberwachte Lernalgorithmen genutzt werden können. Mit dem Fortschritt der Technik und zunehmendem Interesse an diesem Forschungsgebiet sind bereits in naher Zukunft einige weitere interessante Anwendungen zu erwarten.

Persönlich finde ich die Entwicklungen in Bereich der Chipherstellung besonders interessant. Die Geschichte hat bereits mehrmals gezeigt, dass die Berechnung auf CPUs über Grafikkarten bis hin zu dedizierten Chips enorme Leistungssprünge bringt. So dürfte auch die Entwicklung von dedizierten Chips, wie sie zum Beispiel Qualcomm momentan entwickelt, einige Neuerungen bringen.

Die Leistungsfähigkeit neuronaler Netze ist oft schwierig zu verstehen und so stellt sich mitunter die etwas philosophische Frage, inwieweit ein solches Netz

an die Leistungsfähigkeit des Menschen heranreicht oder ihm gar eines Tages Konkurrenz macht. Gut trainiert, schlagen diese Netze den Menschen schon heute in der ein oder anderen Aufgabe. Doch sei an dieser Stelle erwähnt, dass bereits um 1940 die ersten Rechner dieser Welt, die *Zuse* Maschinen von *Konrad Zuse*, den Menschen in vielen Rechenaufgaben übertroffen haben. Bis neuronale Netze so groß werden, dass sie sich ein Bild von der Natur machen können, dürfte noch einige Zeit vergehen.

Literaturverzeichnis

- 1 **sigmoid 2014** : *Sigmoidfunktionen in neuronalen Netzwerken*. 05 2014.
– URL <http://de.wikipedia.org/wiki/Sigmoidfunktion>
- 2 **Baxt 1995** BAXT, W.G: Application of artificial neural networks to clinical medicine. In: *The Lancet* 346 (1995), S. 1135 – 1138
- 3 **Bengio 2010** BENGIO, Dumitru Erhan; Yoshua Bengio; Aaron Courville; Pierre-Antoine Manzagol; Pascal Vincent; S.: Why Does Unsupervised Pre-training Help Deep Learning? In: *Journal of Machine Learning Research* 11 (2010), S. 625–660
- 4 **Bengio 2011** BENGIO, Xavier Glorot; Antoine Bordes; Y.: Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach. In: *Proceedings of the 28 th International Conference on Machine Learning, Bellevue, WA, USA* (2011)
- 5 **Daugman 1988** DAUGMAN, J.G.: Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. In: *Acoustics, Speech and Signal Processing, IEEE Transactions* 36 (1988), S. 1169–1179
- 6 **Fukushima 1980** FUKUSHIMA, Kunihiko: *Biological Cybernetics*. Bd. 36. Kap. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, S. 193–202, Springer, 1980
- 7 **Hinton 2002** HINTON, Geoffrey: Training products of experts by minimizing contrastive divergence. In: *Neural Computation* 14 (2002), S. 1771–1800
- 8 **Ian J. Goodfellow 2013** IAN J. GOODFELLOW, Julian Ibarz-Sacha Arnaud Vinay S.: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. In: *arXiv:1312.6082* (2013)
- 9 **Igel 2013** IGEL, Sebastian Houben; Johannes Stallkamp; Jan Salmen; Marc Schlipsing; C.: Detection of Traffic Signs in Real-World Images: The

- German Traffic Sign Detection Benchmark. In: *Neural Networks (IJCNN), The 2013 International Joint Conference* (2013)
- 10 **Jeffrey Dean und Ng 2012** JEFFREY DEAN, Rajat Monga Kai Chen Matthieu Devin Quoc V. Le Mark Z. Mao Marc'Aurelio Ranzato Andrew Senior Paul Tucker Ke Y. ; NG, Andrew Y.: Large Scale Distributed Deep Networks. In: *NIPS 2012: Neural Information Processing Systems* (2012), S. 1232–1240
 - 11 **J.P Willmore und Roberts 2000** J.P WILLMORE, R.C. P. ; ROBERTS, W.J.J: Comparing Gaussian Mixture and Neural Network Modelling Approaches to Automatic Language Identification of Speech. In: *Aust. Int. Conf. Speech Sci. & Tech* (2000)
 - 12 **Le u. a. 2012** LE, Quoc ; RANZATO, Marc'Aurelio ; MONGA, Rajat ; DEVIN, Matthieu ; CHEN, Kai ; CORRADO, Greg ; DEAN, Jeff ; NG, Andrew: Building high-level features using large scale unsupervised learning. In: LANGFORD, John (Hrsg.) ; PINEAU, Joelle (Hrsg.): *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. New York, NY, USA : Omnipress, July 2012 (ICML '12), S. 81–88. – ISBN 978-1-4503-1285-1
 - 13 **Lippmann 1989** LIPPMANN, R.P.: Pattern classification using neural networks. In: *Communications Magazine, IEEE* 27 (1989), S. 47–50
 - 14 **Minsky 1969** MINSKY, Papert: *Perceptrons*. M.I.T. Press, 1969
 - 15 **Ng 2013** NG, Adam Coates; Brody Huval; Tao Wang; David J. Wu: Andrew Y.: Deep learning with COTS HPC systems. In: *International Conference on Machine Learning, Atlanta, Georgia, USA*. (2013)
 - 16 **Rosenblatt 1958** ROSENBLATT, Frank: The perceptron: A probabilistic model for information storage and organization in the brain. In: *Psychological Review* 65 (1958), S. 386–408
 - 17 **Rumelhart 1985** RUMELHART, Geoffrey E. ; Williams Ronald J.: *Parallel distributed processing: explorations in the microstructure of cognition*. Kap. Learning Internal Representations by Error Propagation, S. 318–362, MIT Press, 1985
 - 18 **Schmidhuber 2010** SCHMIDHUBER, Dan Claudiu Ciresan; Ueli Meier; Luca Maria Gambardella; J.: Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. In: *Neural Computation* 22 (2010)
 - 19 **Smolensky 1986** SMOLENSKY, Paul: *Parallel distributed processing: explorations in the microstructure of cognition*. Bd. 1. Kap. Information processing in dynamical systems: Foundations of harmony theory, S. 194–281, MIT Press, 1986

- 20 Sung 2003** SUNG, A.H.: Identifying important features for intrusion detection using support vector machines and neural networks. In: *Applications and the Internet, 2003. Proceedings. 2003 Symposium* (2003), S. 209–216
- 21 Talbot 2013** TALBOT, David: *Qualcomm to Build Neuro-Inspired Chips*. 2013. – URL <http://www.technologyreview.com/news/520211/qualcomm-to-build-neuro-inspired-chips/>
- 22 Teh 2006** TEH, Geoffrey E. Hinton; Simon Osindero; Yee-Whye: *Neural Computation*. Bd. 18. Kap. A Fast Learning Algorithm for Deep Belief Nets, S. 1527–1554, MIT Press, 2006
- 23 Tsoi 1997** TSOI, Steve Lawrence; C. Lee Giles; Ah C.: Lessons in Neural Network Training: Overfitting May be Harder than Expected. In: *In Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97* (1997)
- 24 Yuval Netzer 2011** YUVAL NETZER, Adam Coates Alessandro Bissacco Bo Wu Andrew Y. N.: Reading Digits in Natural Images with Unsupervised Feature Learning. In: *NIPS Workshop on Deep Learning and Unsupervised Feature Learning* (2011)