EBERHARD-KARLS-UNIVERSITÄT TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät
Institut für Neurobiologie
Lehrstuhl für Kognitive Neurowissenschaften

Bachelorarbeit

Hierarchische Trainingsinformation bei tiefen Neuronalen Netzen

Kornelius Raeth

Gutachter: **Prof. Dr. rer. nat. Hanspeter A. Mallot**
Institut für Neurobiologie
Kognitive Neurowissenschaften

Betreuer: **Gerrit Ecke**
Institut für Neurobiologie
Kognitive Neurowissenschaften

Raeth, Kornelius:
Hierarchische Trainingsinformation bei tiefen Neuronalen Netzen
Bachelorarbeit Kognitionswissenschaft
Eberhard-Karls-Universität Tübingen
Bearbeitungszeitraum: 6. Dezember 2017 – 6. April 2018

Erklärung

Hiermit erkläre ich, dass ich diese schriftliche Abschlussarbeit selbstständig verfasst habe, keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.

Tübingen, den 4. April 2018

Abstract

In the real world, objects have several designations. Humans are able to grasp and group the things surrounding them on different levels of abstraction and thus carry out classifications on an enormous range. Neural networks used for the classification of image data, on the other hand, are usually only trained with regard to a flat, hierarchy-free class structure. But isn't an idea of the underlying hierarchy necessary for the development of a good classification performance? The aim of this thesis was to investigate the extend to which the inclusion of hierarchical training information influences the classification behavior of deep neural networks. In addition to improved classification performance, patterns were expected to form in the network which reflect the additional hierarchical information and are therefore advantageous for classification. For this purpose, the classical neural network *AlexNet* was extended by two additional output layers and trained on a plant classification task. The three resulting output layers served as the output of the family, the genus and the species of the plant. The performance of the extended network was subsequently compared with the performance of the standard *AlexNet* version, which was trained only with the species information. In addition to a pure performance comparison in the classification task, further investigations were carried out to determine the formation of advantageous patterns within the network. A uniform pattern of findings was observed. The hierarchical training information led to a better classification performance in the extended network, but this was only shown by an increase of approx. 7%. For the extended network, better performance could also be observed in the following investigations, which aimed to prove the existence of advantageous patterns within the network. Here too, however, only a gradual improvement could be demonstrated. Thus, already with a flat class structure, neural networks seem to be able to capture higher hierarchical concepts and form corresponding patterns. However, additional hierarchical training information will allow these patterns to be further refined and thus improve the performance of the network.

Kurzfassung

In der echten Welt besitzen Dinge mehrere Bezeichnungen. Menschen sind in der Lage, die sie umgebenden Dinge auf verschieden Abstraktionsebenen zu erfassen und zu gruppieren und somit Klassifikationen in einer enormen Bandbreite durchzuführen. Neuronale Netze für die Klassifikation von Bilddaten hingegen werden meist nur bezüglich einer flachen, hierarchielosen Klassenstruktur trainiert. Ist für die Ausbildung einer guten Klassifikationsleistung aber nicht ein Konzept der zugrundliegenden Hierarchie von notwendig? Ziel dieser Arbeit war es zu untersuchen, inwiefern die Hinzunahme hierarchischer Trainingsinformation das Klassifikationsverhalten tiefer neuronaler Netze beeinflusst. Neben einer besseren Klassifikationsleistung wurde erwartet, dass sich Muster im Netzwerk ausbilden, welche die zusätzliche hierarchische Information widerspiegeln und somit vorteilhaft für die Klassifikation sind. Hierzu wurde das klassische neuronale Netzwerk *AlexNet* um zwei zusätzliche Ausgabeschichten erweitert und auf eine Pflanzenklassifikationsaufgabe trainiert. Die drei entstandenen Ausgabeschichten dienten der Ausgabe der Familie, der Gattung und der Art der Pflanze. Die Leistung dieses erweiterten Netzwerks wurde in der Folge mit der Leistung der klassischen *AlexNet* Variante verglichen, welche nur mit der Art-Information trainiert wurde. Neben einem reinen Leistungsvergleich in der Klassifikationsaufgabe wurden weitere Untersuchungen durchgeführt, um die Bildung vorteilhafter Muster innerhalb des Netzwerks zu eruieren. Es zeigte sich ein einheitliches Befundmuster. Die hierarchische Trainingsinformation führte zu einer besseren Klassifikationsleistung des erweiterten Netzwerks, jedoch handelte es sich nur um eine Zunahme um ca. 7%. Ebenso konnten beim erweiterten Netzwerk bessere Werte in den folgenden Untersuchungen beobachtet werden, welche darauf abzielten, die Existenz vorteilhafter Muster im Netzwerk nachzuweisen. Auch hier konnte jedoch lediglich eine graduelle Verbesserung für das erweiterte Netzwerk nachgewiesen werden. Neuronale Netze scheinen somit bereits mit einer flachen Klassenstruktur in der Lage zu sein, höhere hierarchische Konzepte zu erfassen und entsprechende Muster auszubilden. Durch zusätzliche hierarchische Trainingsinformation können diese Muster allerdings noch weiter verfeinert und somit die Leistung des Netzwerks gesteigert werden.

Inhaltsverzeichnis

1 Einleitung

Errungenschaften Neuronaler Netze

In den letzten Jahren konnten auf dem Forschungsgebiet der künstlichen Intelligenz immer wieder Durchbrüche verkündet werden. Hoch komplexe Probleme mit immensen Datenmengen können in der Zwischenzeit durch die Verwendung tiefer neuronaler Netze gelöst werden. Neuronale Netze sind informationsverarbeitende Systeme, welche aus einer Vielzahl einfacher vernetzter Strukturen bestehen, die untereinander Informationen in Form von Aktivierungen weiterleiten – für eine Einführung siehe Kapitel 2. Tiefe neuronale Netze konnten in den letzten Jahren in Bereichen wie der automatisierten Bildklassifizierung, der Objekterkennung und -verfolgung, der Posenschätzung, der Text- und Spracherkennung sowie vielen weiteren Disziplinen beachtliche Fortschritte erzielen (Gu et al., 2017). Im Oktober 2015 schlug AlphaGo von Google DeepMind den dreifachen Go-Europameister Fan Hui mit 5:0. Später sollte auch der als weltbester Go-Spieler geltende Leo Sedol von Alpha Go mit 4:1 besiegt werden. Die Architektur von AlphaGo basiert dabei zu großen Teilen auf tiefen neuronalen Netzen, welche vermutlich hauptverantwortlich für den großen Erfolg sein dürften. Go galt lange Zeit als eines der schwersten Spiele für künstliche Intelligenzen, vor allem wegen der sehr großen Anzahl möglicher Züge. Gerade deswegen begründete der Erfolg von AlphaGo einen großen Durchbruch für die einschlägige Forschungsgemeinde (Silver et al., 2016). Im Jahr 2011 erzielten Cireşan, Meier, Masci und Schmidhuber im Rahmen der IJCNN[1] mit ihrer Methode erstmals 'übermenschliche' Leistungen bei der Erkennung von Verkehrsschildern. Durch die Verwendung eines tiefen neuronalen Netzes wurde bei der Klassifizierung der Verkehrsschilder eine Fehlerrate von nur 0.56% erzielt. Zum Vergleich: Die durchschnittliche menschliche Fehlerrate lag bei 1.16%, weswegen die Leistung des Netzwerks als 'übermenschlich' bezeichnet wurde (Cireşan et al., 2012). Für das geläufige Training neuronaler Netze zur Klassifikation von Bilddaten liegt meist zu jedem Bild die korrespondierende Zielklasse vor, wodurch abhängig vom Fehler des Netzwerks entsprechende Anpassungen im Netzwerk durchgeführt werden, um die Genauigkeit im Laufe des Trainings immer weiter zu erhöhen.

Bei all ihren Erfolgen stellt sich jedoch nach wie vor die Frage, inwieweit neuronale Netze und allgemein künstliche Intelligenzen tatsächlich Intelligenz aufweisen und vielleicht sogar ein Verständnis von dem haben, was sie tun.[2] Oftmals wird versucht, solche Fragen zu beantworten, indem man neuronale Netze auf sogenannte interne Repräsentationen hin untersucht. Mit diesem Begriff aus der Kognitionswissenschaft werden Eigenschaften des Netzwerks bezeichnet, die in irgendeiner Weise (höhere) äußere Konzepte innerhalb des Netzwerks widerspiegeln. Beispielsweise könnte man von einer internen Repräsentation sprechen, wenn ein Teil des Netzwerks besonders stark auf Bilder einer der zu erkennenden Klassen reagiert. Oder wenn das Netzwerk in seinem Klassifikationsverhalten bestimmte Artefakte aufweist, die nahelegen, dass ihnen ein höheres Konzept in Form einer internen Repräsentation zugrunde

[1] International Joint Conference on Neural Networks

[2] Der Mangel an einer einheitlichen Definition des Intelligenzbegriffs macht die Beantwortung dieser Frage natürlich wesentlich schwerer.

liegen könnte. Meist wird davon ausgegangen, dass solche internen Repräsentationen in den Gewichten eines Netzwerks kodiert werden. Doch können neuronale Netzwerke solche internen Repräsentationen wirklich aufbauen? Und falls ja, wäre es möglich, dass neuronale Netze aufgrund solcher internen Repräsentationen tatsächlich ein Verständnis entwickeln? Fragen dieser Art beschäftigen Forscher im Bereich der künstlichen Intelligenz weltweit und die Antworten darauf fallen sehr unterschiedlich aus. Deshalb kann hier nur der Versuch unternommen werden, Teilaspekte der genannten Fragen näher zu beleuchten.[3] Hierzu soll im Folgenden zunächst betrachtet werden, wie Menschen und vor allem Kleinkinder lernen, Konzepte aufzubauen und Strukturen in der Welt zu erkennen.

Kategorien und Konzeptbildung

In der echten Welt besitzen Dinge mehrere Bezeichnungen Ein Küchenstuhl kann beispielsweise als Stuhl, als Möbel oder auch als Sitzgelegenheit bezeichnet werden. Alle Bezeichnungen umschreiben dasselbe Objekt. Erwachsene sind in der Lage Objekte parallel auf mehreren Ebenen zu klassifizieren, für jedes Objekt gibt es mehrere Abstraktionsstufen. Laut Rosch, Mervis, Gray, Johnson und Boyes-Braem (1976) gibt es eine sogenannte Basisebene der Abstraktion, auf welcher wir Objekte am häufigsten beschreiben. Bei dem genannten Beispiel wäre die Basisebene die Bezeichnung 'Stuhl'. Die Basisebene bildet dabei wie ihr Name nahelegt, den Ausgangspunkt für die Benennung desselben Objekts auf anderen Ebenen (Tafreschi, 2006). Es existieren übergeordnete Ebenen (Möbel, Sitzgelegenheit) und untergeordnete Ebenen (Küchenstuhl). Die Basiskategorie weist einige Merkmale auf, welche sie von anderen Abstraktionsstufen unterscheidet: Objekte, welche allesamt unter der Bezeichnung der Basisebene zusammengefasst werden können (Basisklasse), sind sich maximal ähnlich, gleichzeitig unterscheiden sie sich maximal von den Objekten anderer Basisklassen. Tische und Stühle bilden zwei Basisklassen unter der Oberklasse Möbel, Mitglieder der Basisklasse Stuhl (Schaukelstuhl, Küchenstuhl,..) sind sich untereinander sehr ähnlich, genau wie Mitglieder der Basisklasse Tisch (Schreibtisch, Esstisch, Couchtisch,..). Jedoch haben die Mitglieder der beiden Basisklassen über die Klassen hinweg wenig gemeinsam (Tafreschi, 2006). Weiter werden mit allen Mitglieder einer Basisklasse ähnliche Handlungen und Bewegungsabläufe assoziiert (Rosch et al., 1976). Sehen wir einen Schaukelstuhl oder einen Barhocker, so assoziieren wir mit beiden Gegenständen die Möglichkeit, sich darauf zu setzen. Des Weiteren lassen sich die Mitgliedern einer Basisklasse meist leicht mit einem prototypischen Beispiel zusammenfassen. Interessanterweise sind Basisklassennamen oft monomorphemisch und daher leicht auszusprechen und schnell zu lernen. Untergeordnete Klassennamen hingegen bestehen oft aus zusammengesetzten Wörtern und übergeordnete Klassennamen sind meist von deutlich abstrakterer Natur (Tafreschi, 2006). Am Beispiel *Küchenstuhl – Stuhl – Möbel* wird dies offensichtlich. Bei Kindern entwickeln sich die Basisklassennamen vor den übergeordneten und untergeordneten Klassennamen, also *Stuhl* vor *Möbel* (Tafreschi, 2006). Dies scheint angesichts der angeführten Punkte plausibel, betrachtet man jedoch die Bildung der zugrundeliegenden Konzepte und nicht die Fähigkeit, diese mit Namen zu benennen, so zeigt sich ein anderes Bild. Während der Entwicklung zeigt sich bei Kleinkindern ein sogenannter *global-to-basic-level shift* bezüglich der Kategorisierung von Objekten und Dingen. Zunächst werden Dinge in die Kategorien *Lebewesen* und *unbelebte Objekte* eingeordnet, bevor im Laufe der Entwicklung feinere Unterscheidungen getroffen werden (Pauen, 2002). Dies entspricht nicht wirklich den

[3] Die hier angeführte Argumentation entspricht dabei nur einer der möglichen Positionen hinsichtlich dieses Themas, ein Anspruch auf Vollständigkeit besteht selbstverständlich nicht.

Erwartungen, könnte man doch annehmen, dass zunächst eine Kategorienbildung bezüglich besonders ähnlicher Objekte – und damit auf der Ebene der Basisklassen – leichter fallen sollte. Dennoch liefert der betrachtete Ansatz somit eine Theorie der Kategorienbildung beim Menschen. Mithilfe der Gruppierung von Objekten und Lebewesen auf verschiedenen Abstraktionsstufen sind wir in der Lage, eine immense Bandbreite an verschiedenen Kategorien zu überblicken. In dieser Hinsicht ist der Mensch neuronalen Netzwerken weit voraus, denn diese klassifizieren Bilder höchstens bezüglich mehrerer tausend Kategorien.

Es bleibt jedoch die Frage offen, wie Konzepte entstehen, wie der Mensch also in der Lage ist, die Welt um sich zu begreifen und strukturiert wahrzunehmen. Die Theorie der verkörperten Kognition (engl. *embodied cognition*) ist ein Paradigma in der Kognitionswissenschaft und besagt, dass sämtliche kognitiven Fähigkeiten durch die sensomotorischen Erfahrungen konstituiert sind, welche ein verkörpertes Wesen in der kontinuierlichen Interaktion mit seiner Umwelt herbeiführt. Dass beispielsweise Wahrnehmung zwangsweise verkörpert ist, legt nach Fuchs (2017) bereits die verwendete Sprache nahe: *Wahr-nehmen* kann nur ein leibliches Wesen, welches sich in seiner Umwelt frei bewegen und Dinge ergreifen kann. Nur so ist es in der Lage, die es umgebende Welt und die Bedeutung der in ihr enthaltenen Dinge zu *begreifen* (Fuchs, 2017). Nach Lakoff und Johnson (1999) sind selbst Abstraktionen höherer Ordnung in ihrem Kern begründet durch Konzepte, welche aus der unmittelbaren Interaktion eines Körpers mit der ihn umgebenden Umwelt hervorgehen. Hierbei stellen abstrakte Konzepte eine metaphorische Ausdehnung bereits bekannter Konzepte dar (Jaekel & Meyer, 2013). Man könnte schlussfolgern, dass es uns als verkörperte Wesen auf diese Weise gelingt, Struktur in die Welt zu bringen und Objekte und Dinge in Konzepten zu erfassen. Durch die metaphorische Erweiterung bekannter Konzepte sind wir in der Lage, abstraktere Konzepte zu erzeugen und könnten diese dazu nutzen, Objekte auf verschiedenen hierarchischen Abstraktionsstufen zu gruppieren und zu benennen.

Implikationen für Neuronale Netze

In Anbetracht dieser Thesen ist die Frage nach Intelligenz und Verstand neuronaler Netze erneut zu bewerten. Neuronale Netze sind informationsverarbeitende, körperlose Systeme, die in der Lage sind, zu einer bestimmten Eingabe eine gewünschte Ausgabe zu liefern. Da neuronale Netze keine körperliche Ausdehnung aufweisen und lediglich in Form von Software auf einem Computer existieren, würde ihnen ein Verfechter der verkörperten Kognition wohl jegliche Form von Intelligenz absprechen. Ohne Körper und ohne Einbettung in eine Umwelt fehlt die Möglichkeit der Exploration und Interaktion und die daraus resultierende sensomotorische Erfahrung, auf welcher, nach der Verkörperungs-Theorie, sämtliche Formen von (höherer) Kognition beruhen. Hierauf aufbauend soll als nächstes der Begriff der *internen Repräsentation* genauer betrachtet und auf seine Tauglichkeit in Bezug auf neuronale Netze hin untersucht werden. Unter einer internen Repräsentation versteht man etwa die Fähigkeit neuronaler Netze, einen externen Sachverhalt – beispielsweise Merkmale der dargebotenen Daten – intern abzubilden und für die Verarbeitung nutzen zu können. Man könnte auch sagen, die interne Repräsentation steht stellvertretend für einen externen Sachverhalt – sie repräsentiert ihn.

Der folgende Abschnitt beruht auf den Ausführungen von Fuchs (2017, Kapitel 2). Hiernach stellt eine Repräsentation immer eine dreistellige Relation zwischen einem Repräsentat (Zeichen), dem Repräsentandum (Bezeichnetem) und einem Subjekt dar. *Etwas* stellt *für jemanden* ein Zeichen *für etwas* dar (entspr. Peirce, 1998). Repräsentationen bestehen somit an sich überhaupt nicht, erst Subjekte können

den Repräsentationszusammenhang von Sachverhalten erzeugen. Aber kann eine Repräsentation nicht auch *für* ein neuronales Netz bestehen? Die Verwendung der Präposition 'für' bezeichnet hier den Bezug auf ein Subjekt mit einem subjektiven Standpunkt, '*für* jemanden ist etwas wichtig'. Diese Verwendung ist jedoch im betrachteten Fall zurückzuweisen, denn es ist nicht davon auszugehen, dass ein neuronales Netz einen subjektiven Standpunkt besitzt. Als praktisches Beispiel zum besseren Verständnis der hier betrachteten Argumentation werden oft die Jahresringe eines Baumes angeführt. Die Ringe im Querschnitt des Baumes repräsentieren *für uns* sein Lebensalter, niemand würde wohl behaupten, dass die Jahresringe *für den Baum* selbst sein Lebensalter repräsentieren. Die Jahresringe stellen ein Repräsentat dar, welches aber erst durch uns als Subjekte als Repräsentation der Lebensjahre des Baumes interpretiert werden kann. Genauso wie der Baum seine Jahresringe erzeugt, kann ein neuronales Netz seine Art von Repräsentaten hervorbringen, doch für das neuronale Netz besitzen diese keinerlei Bedeutung, sie stehen auch nicht stellvertretend für etwas anderes. Erst durch uns als Subjekte kann ein Repräsentationszusammenhang erzeugt werden, welcher den Repräsentaten Bedeutung verleiht. Statt also von internen Repräsentationen zu sprechen, sollte der Begriff, wie von Fuchs vorgeschlagen, durch den des *Musters* ersetzt werden. Aus dieser Bezeichnung geht klar hervor, dass für die Deutung eines solchen Musters stets ein subjektiver Beobachterstandpunkt nötig ist. Aus diesem Grund wird in dieser Arbeit von Mustern und nicht von internen Repräsentationen die Rede sein.

Bedeutet das, dass neuronale Netze als körperlose, willen- und verstandlose informationsverarbeitende Systeme keinerlei Intelligenz aufweisen? Das hängt wohl von der Definition des Intelligenzbegriffs ab.[4] Wenn Intelligenz auf einer evolutionsbiologischen Ebene betrachtet wird, so steht der Begriff vor allem für eines: *Flexibilität*. Flexibles Verhalten und schnelle Anpassungen an variierende Umweltbedingungen, bzw. die Fähigkeit, die Umgebung für das eigene Überleben zu nutzen und entsprechend anzupassen, erhöht die Überlebenschancen und somit die Fitness eines Lebewesens (Fogel, 2006; Pfeifer & Scheier, 2001). Die Fähigkeit, auf bereits Gelerntem aufzubauen und das Gelernte auf neue Probleme anzuwenden, stellen dann höhere intellektuelle Fähigkeiten dar. In dieser Hinsicht weisen neuronale Netze tatsächlich kaum Intelligenz auf. Die Flexibilität neuronaler Netze scheint doch sehr gering auszufallen. Transferlernen liegt kaum vor, die meisten neuronalen Netze scheitern kläglich, wenn sie auf eine andere als die gelernte Aufgabe angesetzt werden Fogel (2006). Aber genau diese Fähigkeit würde Intelligenz nach obiger Ansicht in ihrem Kern begründen. Nichtsdestotrotz sind neuronale Netze in ihren Einsatzgebieten oft dem Menschen überlegen. Eventuell ist ihre Intelligenz grundlegend verschieden von der Intelligenz biologischer Lebewesen und muss deswegen von vornherein anders definiert werden. Betrachtet man Intelligenz als ein Zeichen schneller Informationsverarbeitung bei großen Datenmengen mit guten Ergebnissen in einer spezifischen Aufgabe und setzt dabei nicht voraus, dass die gezeigte Leistung auf andere Aufgaben übertragen werden kann, so weisen auch neuronale Netze Intelligenz auf. Es stellt sich die Frage, wie man diese Intelligenz und damit die Leistung neuronaler Netze weiter fördern kann.

Wie bereits erwähnt, werden neuronale Netze meist darauf trainiert, Objekte bzw. Bilder nur auf einer Ebene zu klassifizieren. Zu jedem Bild existiert genau eine Zielklasse, welche es vorherzusagen gilt. Es handelt sich somit um flache Klassenstrukturen. Vergleicht man dies mit dem Kategorienerwerb bei Menschen, so fällt auf, dass dieser deutlich vielschichtiger ist. Wir lernen schon früh Gegenstände auf verschiedenen Abstraktionsstufen zu benennen und sind vermutlich mitunter dadurch in der La-

[4] Natürlich wird der Intelligenzbegriff nach wie vor heftig diskutiert und die hier dargelegte Position ist nur eine von vielen.

ge, ein klareres Bild der zugrundeliegenden Konzepte auszubilden. Die Entwicklung von Kleinkindern genauer zu betrachten und die erhaltenen Erkenntnisse über die zugrundeliegenden Lernprozesse auf die Entwicklung künstlicher Intelligenzen zu übertragen, stellt nach Gopnik (2017) einen fruchtbaren Ansatz dar. Es wäre möglich, dass die Informationen, die wir neuronalen Netzen beim Training zur Verfügung stellen, auch bei einem Menschen nicht ausreichen würden, um eine gute Klassifikationsleistung zu erzielen. Ergo, vielleicht muss die Information, welche wir neuronalen Netzen bereitstellen, auch vielschichtiger sein, damit sich in den Netzwerken Muster bilden können, die vorteilhaft für die Kategorisierung sind. Deng, Berg, Li und Fei-Fei (2010) verfolgen hierzu einen neuartigen Ansatz. Indem die semantische Nähe bei der Bewertung einer fehlerhaften Klassifikation eines neuronalen Netzes miteinbezogen wird, lassen sich Falschklassifikationen in ihrer Häufigkeit und ihrem Ausmaß deutlich eingrenzen. Normalerweise wird eine Klassifikation lediglich als richtig oder falsch betrachtet, unabhängig von semantischen Verwandtschaften. Bei Deng et al. hingegen werden alle Kategorien bezüglich ihrer semantischen Ähnlichkeit in einem semantischen Baum angeordnet. Die semantische Nähe zweier Kategorien ergibt sich nun durch die Höhe des nächsten gemeinsamen Vorgängerknotens. Je höher dieser Knoten liegt, umso entfernter und damit semantisch verschiedener sind die beiden Kategorien. Bei einem Stimulus aus der Kategorie *Segelboot* ist eine Klassifizierung als *Boot* aufgrund der semantischen Nähe weniger schwerwiegend als eine Klassifizierung als *Mikrowelle*. Die Fehlerfunktion des neuronalen Netzwerks wurde dahingehend angepasst, dass sie die semantische Nähe der Kategorien berücksichtigt. Deng et al. konnten zeigen, dass ein Netzwerk, welches mit einer solchen Fehlerfunktion trainiert wird, deutlich seltener schwerwiegende Falschklassifikationen erzeugt.

Doch es existieren noch weitere Möglichkeiten, vielschichtige Information in das Training neuronaler Netze einfließen zu lassen. Ein neuer Ansatz könnte versuchen, Netzwerke darauf zu trainieren, Stimuli gleichzeitig auf mehreren Abstraktionsstufen zu kategorisieren. Hierzu müsste man beim Training eines solchen Netzes stets die Zielklassen bezüglich der jeweiligen Abstraktionsstufe bereitstellen. Hieraus könnten sich im Netzwerk Muster bilden, die als Repräsentationen der zugrundeliegenden hierarchischen Abstraktionsstufen interpretiert werden könnten. Diese Muster könnten sich als durchaus nützlich für die Klassifizierung erweisen.

Die Biologie liefert wohldefinierte hierarchische Strukturen – sogenannte Taxonomien – mit denen Lebewesen in einer hierarchischen Art und Weise systematisch erfasst werden. Taxonomien sind derart aufgebaut, dass eine Spezies auf jeder taxonomischen Stufe genau einer der möglichen Klassen zugeordnet wird. Die letzten drei taxonomischen Stufen werden als *Familie, Gattung* und *Art* bezeichnet. Diese drei Stufen könnten zu einem gewissen Grad mit den anfangs betrachteten Abstraktionsstufen bei Objekten (Bsp.: *Möbel - Stuhl - Küchenstuhl*) verglichen werden. Ein Netzwerk zur Pflanzenerkennung erhält üblicherweise als einzige Trainingsinformation die Arten der Pflanzen und führt somit Klassifikationen auf einer einzigen Taxonomie-Stufe durch. Bilal, Jourabloo, Ye, Liu und Ren (2018) trainierten ein *Convolutional Neural Network* darauf, Pflanzen- und Tierarten zu klassifizieren. Sie konnten zeigen, dass bereits das alleinige Training mit der Art-Information ausreicht, um Artefakte in der Klassifikation herbeizuführen, die nahelegen, dass im Netzwerk Muster entstanden sind, die von uns als Familienrepräsentationen interpretiert werden können. Das Netzwerk hat somit implizit Muster erzeugt, die als Konzepte höherer Ordnung interpretierbar sind, ohne dass die nötige Information für die Bildung dieser Muster explizit in den Trainingsdaten vorhanden war.

Fragestellung

Im Rahmen dieser Arbeit soll untersucht werden, ob die Verwendung vielschichtiger, hierarchischer Trainingsinformation zu einer besseren Klassifikationsleistung bei neuronalen Netzwerken führen kann. Als Grundlage wird hierfür ein neuronales Netz zur Pflanzenklassifikation verwendet. Statt das Netzwerk aber nur mit der Art-Information zu trainieren, wird zusätzlich noch die Information zur Gattung und zur Familie bereitgestellt. Die Hypothese ist, dass sich durch die vielschichtige Trainingsinformation Muster im Netzwerk ausbilden, welche zu einer besseren Leistung in der Klassifikationsaufgabe führen. Diese Muster könnten als Repräsentationen der dargebotenen hierarchischen Information gedeutet werden. Es werden somit Muster erwartet, die als Familienrepräsentationen oder als Gattungsrepräsentationen interpretierbar sind. Bilal et al. (2018) konnten zeigen, dass Ansätze solcher Muster bereits bei konventionell trainierten neuronalen Netzen vorzufinden sind. Durch die zusätzliche hierarchische Trainingsinformation könnten sich diese Muster jedoch noch weiter differenzieren und deutlicher herausbilden. Zunächst soll untersucht werden, inwieweit eine Leistungssteigerung aufgrund der hierarchischen Trainingsinformation vorliegt. Die Untersuchung des Netzwerks hinsichtlich der erwarteten Musterbildung soll im Rahmen weiterer Analysen realisiert werden.

Die Arbeit unterteilt sich folgendermaßen. Zunächst wird eine kurze Einführung in neuronale Netze dargeboten. Daraufhin werden im Methodenteil unter anderem die genauen Implementierungen und die durchgeführten Untersuchungen genannt. Im darauffolgenden Kapitel werden die Ergebnisse präsentiert, um diese schließlich im letzten Kapitel diskutieren und interpretieren zu können.

2 Theoretischer Hintergrund

2.1 Neuronale Netze

Neuronale Netze[1] können als informationsverarbeitende Systeme betrachtet werden, bestehend aus einer großen Menge einfacher Einheiten (Neuronen), welche sich Informationen in Form von Aktivierungen über gerichtete Verbindungen weiterleiten. Die Idee zu neuronalen Netzen entstammt dem Konnektionismus, einer wissenschaftlichen Strömung mit Beginn in den 1940er Jahren, welche besagt, dass viele einfache berechnende Einheiten als vernetzte Struktur in der Lage sind, intelligentes Verhalten hervorzubringen (Besold & Kühnberger, 2013). Sowohl die Bezeichnung als auch die Funktionsweise künstlicher neuronaler Netze legen eine Verwandtschaft zu neuronalen Netzen biologischer Systeme nahe. Jedoch handelt es sich bei künstlichen neuronalen Netzen vielmehr um eine stark abstrahierte, funktionale Modellierung von Informationsverarbeitung, die zwar in ihrem Kern durchaus inspiriert wurde von ihrem biologischen Pendant, aber dennoch nicht den Anspruch hat, die genauen neuronalen Dynamiken biologischer Organismen nachzubilden (Zell, 1994). Ein neuronales Netz kann dazu verwendet werden, eine bestimmte Aufgabe, wie zum Beispiel ein Klassifikationsproblem zu lösen. Der Vorteil neuronaler Netze gegenüber konventionellen Algorithmen besteht darin, dass das Netzwerk für die zu lernende Aufgabe nicht explizit programmiert werden muss, sondern allein aufgrund von Trainingsbeispielen selbstständig lernt, eine Aufgabe zu lösen (Zell, 1994). Neuronale Netze besitzen eine Eingabe- und eine Ausgabeschicht aus Neuronen. Die Aufgabe des Netzwerks ist es, zu einer bestimmten Eingabe, welche an die Eingabeschicht angelegt wird, eine gewünschte Ausgabe zu liefern. Möglich wird dies durch die Anpassung der Verbindungsstärken zwischen Einheiten innerhalb des Netzwerks. Trotz der recht simplen Struktur einzelner Einheiten sind große neuronale Netze in der Lage, komplexe Aufgaben zu lösen, welche auf konventionelle Weise schwer lösbar wären.

Ein grundlegendes Modell eines neuronalen Netzes stellt das sogenannte Perzeptron nach Rosenblatt (1958) dar. Hierbei handelt es sich im Falle eines einschichtigen Perzeptrons um ein einzelnes Ausgabeneuron, welches über gerichtete Verbindungen direkt mit der Eingabeschicht verbunden ist. Zunächst wird die gewichtete Summe der Aktivierungen e_i der Eingabeneuronen mit den Gewichten w_i gebildet. Eine nachgeschaltete Aktivierungsfunktion f erzeugt daraus die Ausgabe y (siehe Gleichung 2.1 und Abbildung 2.1).

$$y = f\big(\sum_i e_i \cdot w_i\big) \tag{2.1}$$

Einschichtige Perzeptren sind jedoch in ihren Fähigkeiten recht beschränkt (siehe z.B. Mallot, 2013). Durch die Verknüpfung mehrerer einschichtiger Perzeptren kann ein mehrschichtiges Perzeptron gebildet werden. Diese legen den Grundstein für komplexere Netztopologien (siehe Abbildung2.2), welche eine deutlich größere Bandbreite an Aufgaben lösen können. Durch weitere Verknüpfungen können be-

[1] Oft auch als *künstliche neuronale Netze* bezeichnet. In dieser Arbeit wird die Bezeichnung *neuronales Netz* gleichbedeutend verwendet.

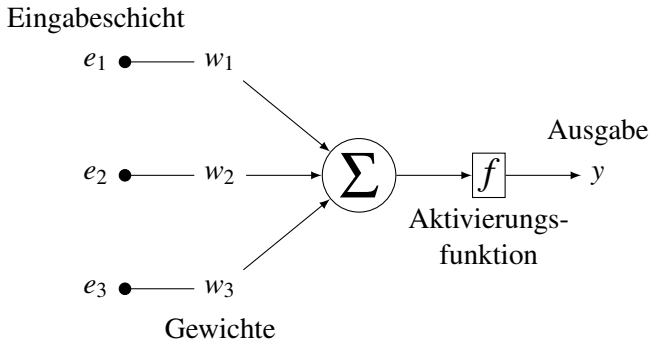


Abbildung 2.1: Darstellung der Funktionsweise eines einschichtigen Perzeptrons.[2]

liebig tiefe neuronale Netze erzeugt werden. Hierbei werden alle Schichten innerhalb des Netzwerks als verdeckte Schichten bezeichnet. Jedes Neuron in den verdeckten Schichten besitzt ein sogenanntes rezeptives Feld. Als solches wird jene Neuronenmenge der vorherigen Schicht(en) bezeichnet, die zur Aktivierung des betrachteten Neurons beiträgt. Der Begriff stammt von dem biologischen Pendant, dessen Existenz unter anderem durch die Forschung von Hubel und Wiesel (1962) an Katzen belegt werden konnte.

Die Lernverfahren neuronaler Netze lassen sich in die Bereiche *überwachtes Lernen* und *unüberwachtes Lernen* aufteilen. Bei überwachten Lernverfahren wird dem Netzwerk während des Trainings zusätzlich zur Eingabe eine gewünschte Ausgabe präsentiert. Beim unüberwachten Lernen hingegen bekommt das Netzwerk keinen Zielwert präsentiert, sondern muss selbstständig Muster in den präsentierten Stimuli erkennen. Da sich diese Arbeit mit überwachten Lernverfahren beschäftigt, werden im Folgenden nur diese genauer erläutert.

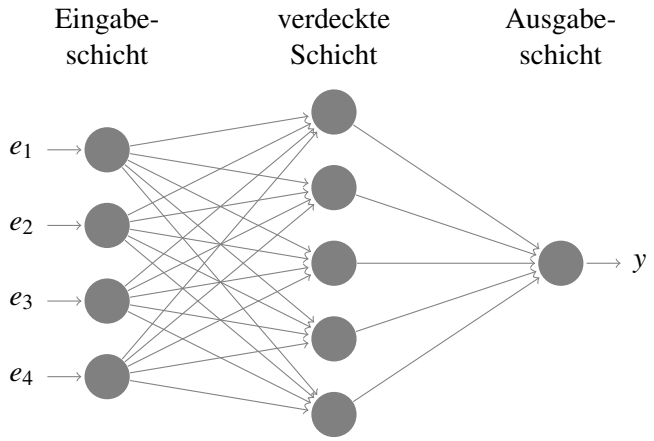


Abbildung 2.2: Darstellung eines zweischichtigen Perzeptrons.[3]

Beim überwachten Lernen wird die gewünschte Ausgabe (*teaching signal*) genutzt, um zusammen mit dem Ausgabevektor des Netzwerks einen Fehlervektor zu bilden. Eine Fehlerfunktion E berechnet mithilfe des Fehlervektors einen globalen Fehler. Diesen Fehler gilt es im Verlauf des Trainings zu minimieren. Nach Zell (1994) kann man die Fehlerfunktion auch als Funktion der Gewichte des Netz-

[2]Source Code der Grafik wurde adaptiert von: `https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network`, 05.03.2018

[3]Source Code der Grafik adaptiert von: `http://www.texample.net/tikz/examples/neural-network/`, 05.03.2018

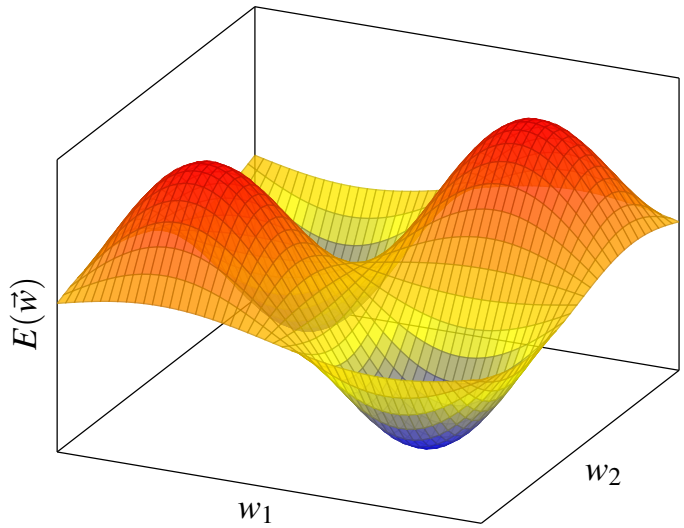


Abbildung 2.3: Mögliche Fehleroberfläche eines einfachen neuronalen Netzes in Abhängigkeit der Gewichte w_1 und w_2.

werks $E(\vec{w})$ betrachten[4]. Für jede Gewichtskonfiguration ergibt sich, über alle Stimuli gemittelt, ein Fehlerwert. Im einfachen Fall eines einschichtigen Perzeptrons mit zwei Eingabeneuronen und somit zwei Gewichten lässt sich solch eine Fehlerfunktion graphisch anschaulich darstellen (siehe Abbildung 2.3). Auf der entstandenen Fehleroberfläche kann nun ein Gradientenabstieg durchgeführt werden, um ein (globales) Minimum in der Fehlerlandschaft zu erreichen. Hierzu werden die partiellen Ableitungen $\frac{\partial E}{\partial w_i}$ der Fehlerfunktion bezüglich jedes Gewichts des Netzwerks gebildet und anschließend die Richtung des steilsten Abstiegs festgestellt. Der so berechnete Gradient multipliziert mit der Lernrate gibt an, wie stark die einzelnen Gewichte anzupassen sind und ermöglicht hierdurch einen stufenweisen Abstieg auf der Fehleroberfläche. Dieses Verfahren gestattet es, mit den an der Ausgabeschicht auftretenden Fehlern, Anpassungen der Gewichte innerhalb des Netzwerks vorzunehmen. Da die Gewichtsanpassungen von der Ausgabeschicht sukzessive in Richtung der Eingabeschicht durchgeführt werden und diese Richtung invers zur Propagierung eines Signals innerhalb des Netzwerks ist, wird hierbei auch von Fehlerrückführung oder *Backpropagation* gesprochen. Für eine genauere Ausführung siehe z.B. Mallot (2013) oder Zell (1994). Letztere Quelle liefert auch eine ausführliche Herleitung der konkreten *Backpropagation*-Regel.

2.2 Convolutional Neural Networks

2.2.1 Konvolution

Konvolution ist mathematisch als Faltungsoperation zweier Funktionen f, g definiert. Die nach Anwendung des Faltungsoperators resultierende Funktion $h = f * g$ entspricht gewissermaßen dem (punktweisen) Produkt der beiden Funktionen. Für eine ausführliche Darlegung siehe Mallot (2013). Um das Konzept von *Convolutional Neural Networks* (CNNs) zu verstehen, ist lediglich ein Verständnis der diskreten Konvolution relevant. Am besten lässt sich diskrete Konvolution anhand der Faltung eines Bildes[5]

[4] Hierbei steht $\vec{w}$ für den Gewichtsvektor des Netzwerks.
[5] Ein digitales Bild kann als diskrete, zweidimensionale Funktion betrachtet werden.

veranschaulichen. Hierbei entspricht die Konvolution der Anwendung eines Filters auf jedes Pixel des Bildes. Das resultierende Bild ist das Faltungsprodukt aus Originalbild und Filter. Ein solcher Filter, auch als Kernel bezeichnet, kann als Maske verstanden werden, welche pixelweise über das Originalbild bewegt wird. Für jedes Pixel wird die gewichtete Summe aus Originalbild und Filter gebildet. Der Filter liegt meist in Form einer $n \times n$ Matrix vor. Seine Einträge geben an, mit welcher Gewichtung die umliegenden Pixel in die gewichtete Summe einfließen. Seien im Folgenden ein Bild I mit seinen Pixelwerten und ein Filter f gegeben:

$$I = \begin{bmatrix} 1 & 3 & 1 & 2 & 4 \\ 2 & 4 & 2 & 1 & 3 \\ 5 & 2 & 2 & 4 & 1 \\ 3 & 1 & 5 & 2 & 1 \\ 2 & 3 & 1 & 5 & 4 \end{bmatrix} \qquad f = \begin{array}{|c|c|c|} \hline 0 & -1 & 0 \\ \hline -1 & 3 & -1 \\ \hline 0 & -1 & 0 \\ \hline \end{array}$$

Durch eine Faltung von I mit f entsteht ein neues Bild $I^* = I * f$. Es gilt $I^*[2,2] = 3$. Diesen Wert erhält man, indem man den Filter f so über I platziert, dass das Zentrum von f über dem Punkt $I[2,2]$ liegt. Alle Pixel unterhalb des Filters fließen nun entsprechend ihres korrespondierenden Filtereintrags in die gewichtete Summe ein. Somit ergibt sich $I^*[2,2] = 0*1+(-1)*3+0*1+(-1)*2+3*4+(-1)*2+0*5+(-1)*2+0*2 = 3$. Diese Berechnung wird für jedes Pixel durchgeführt, wodurch das Faltungsprodukt I^* entsteht.[6] Es existiert eine Vielzahl geläufiger Filter, manche eignen sich für die Hervorhebung von Kanten in Bildern, andere werden beispielsweise zum Weichzeichnen oder Schärfen von Bildern verwendet. Der Filter f stellt beispielsweise einen Schärfungsfilter dar.

2.2.2 CNN-Architekturen

Die Architektur eines *Convolutional Neural Network* kann mehrere verschiedenartige Schichten aufweisen. Diese lassen sich in drei Arten unterteilen: die *Convolutional Layer*, die *Pooling Layer* und die *Fully-connected Layer*. Im Folgenden werden alle drei Arten bezüglich ihres Aufbaus und ihrer Funktion genauer betrachtet.

Convolutional Layer

CNNs machen sich, wie ihr Name unschwer erkennen lässt, die Eigenschaften der Konvolution zu Nutze. Eine Schlüsselrolle nehmen hierbei die *Convolutional Layer*, zu Deutsch etwa 'Faltungsschichten', des Netzwerks ein. Diese realisieren das oben illustrierte Prinzip der diskreten Faltung innerhalb neuronaler Netze. Dabei sind die Filter implizit in Form von Gewichten im Netzwerk vorhanden, wobei jedes Neuron einer Schicht dieselben Eingangsgewichte aufweist (siehe Abbildung 2.4).
Die Größe des Filters, auch Kernelgröße genannt, entspricht dabei der Größe der rezeptiven Felder. Alle drei Neurone der zweiten Schicht in Abbildung 2.4 besitzen ein rezeptives Feld der Größe 3 und weisen die selben Eingangsgewichte auf. Jedes Neuron berechnet die gewichtete Summe aus seinen drei Eingaben, was vergleichbar ist mit der Faltung eines 5×1 Pixel Bildes unter Verwendung eines 3×1 Pixel Filters. Für die Faltung eines zweidimensionalen Bildes muss das dargestellte Schema in Abbildung 2.4

[6] Damit das Faltungsprodukt dieselbe Größe aufweist wie das Originalbild, wird in der Praxis meist ein Rahmen aus Nullen um das Bild erzeugt, wodurch die Anwendung des Filters auch im Randbereich des Bildes möglich wird.

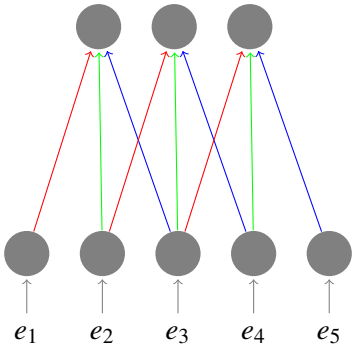


Abbildung 2.4: Schematische Darstellung eines *Convolutional Layer*. Gleichfarbige Verbindungen weisen dieselbe Gewichtung auf.

um eine Dimension erweitert werden. In CNNs wird in einem *Convolutional Layer* jedoch nicht nur ein Filteroperator angewendet, sondern meist mehrere solcher Operatoren gleichzeitig. Jeder Filter erzeugt durch Faltung eine sogenannte *Feature Map*, zu Deutsch etwa 'Merkmalskarte'. Der Name rührt daher, dass jeder Filter das Bild gewissermaßen auf besondere Merkmale, wie etwa Kanten, Kontrast oder gar auf bestimmte Formen hin untersucht. Das Faltungsprodukt hebt diese Merkmale hervor und wird deswegen als *Feature Map* bezeichnet. Dabei ist es egal, wo im Bild ein besonderes Merkmal zu finden ist. Unter anderem diese Translationsinvarianz macht *Convolutional Layer* so brauchbar. Mit mehreren *Feature Maps* besteht jeder *Convolutional Layer* aus einer weiteren dritten Dimension, die als Tiefe des *Layers* bezeichnet werden kann. Bei 5 Filtern und damit 5 verschiedenen *Feature Maps* hätte ein solcher *Convolutional Layer* folglich die Tiefe 5. Die Funktionsweise der *Convolutional Layer* weist eine gewisse Verwandtschaft zu bekannten neuronalen Mechanismen der visuellen Verarbeitung im menschlichen Gehirn auf. Im primären visuellen Kortex (V1) existieren ähnliche Filter wie jene, die sich bei CNNs in den ersten *Convolutional Layern* entwickeln (Kheradpisheh, Ghodrati, Ganjtabesh & Masquelier, 2016). Mitunter wegen dieser gewissen Verwandtschaft zu biologischen Vorbildern werden CNNs oft als prädestiniert für die Bilderkennung betrachtet.

Pooling Layer

Zusätzlich zu den *Convolutional Layern* besteht ein CNN auch noch aus sogenannten *Pooling Layern*, welche dazu beitragen, überflüssige Informationen zu reduzieren. Wie der englische Begriff *pooling* nahelegt, sorgen diese Schichten dafür, dass die verfügbare Informationsmenge gebündelt wird. Meist werden solche *Pooling Layer* einem *Convolutional Layer* nachgeschaltet, um die Information vor dem nächsten Verarbeitungsschritt zu reduzieren. Es existieren mehrere Formen der *Pooling*-Operation. Eine weitverbreitete Variante ist das sogenannte *Max-Pooling*, welches auch in dieser Arbeit angewandt wurde. Hierbei wird immer ein bestimmter Bereich der Neuronen eines *Convolutional Layers* betrachtet und nur die maximale Aktivierung innerhalb dieses Bereichs weitergeleitet, der Rest wird verworfen. Ähnlich wie bei der Konvolution gibt es bei dieser Operation auch eine rezeptive Feldgröße, die vorgibt, wie groß der Bereich ist, aus dem das Maximum gewählt wird. Ebenso kann der Grad der Überlappung zwischen den rezeptiven Feldern verändert werden, eine Überdeckung derselben kann somit auch unterbunden werden. Der Vorteil von *Pooling Layern* besteht darin, dass sie durch die Reduktion des Informationsgehalts die Verarbeitungsgeschwindigkeit des Netzwerks erhöhen, zum anderen führen sie dazu, dass die rezeptiven Felder der *Convolutional Layer* mit fortschreitender Verarbeitung absolut be-

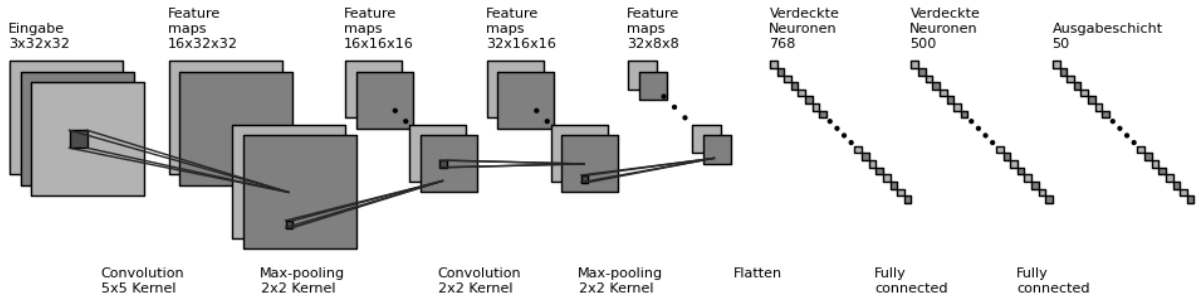


Abbildung 2.5: Gängige Darstellungsweise einer CNN-Architektur. Die Anmerkungen im oberen Teil spezifizieren den Inhalt der Schichten ($32 \times 16 \times 16$ impliziert 32 *Feature Maps* der Größe 16×16), die Anmerkungen im unteren Teil beschreiben die auf den Schichten durchgeführte Operation.[7]

trachtet größer werden. Nach Anwendung des *Max-Pooling* auf eine *Feature Map* weist diese nur noch ein Viertel ihrer ursprünglichen Größe auf. Wenn nun auf selbige ein *Convolutional Layer* mit Kernelgröße 2×2 folgt, so entspricht dies effektiv einer Konvolution mit Kernelgröße 4×4 auf der *Feature Map* vor dem *Max-Pooling*. Das rezeptive Feld ist somit um den Faktor 2 gewachsen, ohne dass sich die eigentliche Kernelgröße verändert hat.

Fully-connected Layer

Hierbei handelt es sich um Schichten im Netzwerk, die eine vollständige Verknüpfung zu ihrer Vorgängerschicht besitzen. Das bedeutet jedes Neuron aus der Vorgängerschicht ist über eine gewichtete Verbindung mit jedem Neuron der Nachfolgerschicht verbunden. Im Gegensatz zu den bisherigen Schichtarten sind die *Fully-connected Layer* eindimensional. Schichten dieser Art sind bereits aus anderen neuronalen Netzen bekannt. Das Perzeptron aus Abbildung 2.2 beispielsweise weist einen *Fully-connected Layer* auf.

Verknüpfung der Schichten

Durch die Verknüpfung mehrerer *Convolutional Layer* in einem CNN können bestimmte Merkmale eines Eingabebilds hervorgehoben werden. Diese Merkmale werden in Form von *Feature Maps* an die nächsten Schichten weitergereicht und weiterverarbeitet. Dabei werden die extrahierten Merkmale stets komplexer und spezifischer. In höheren Verarbeitungsstufen können sich Filter bilden, welche maximal auf ein bestimmtes Merkmal einer bestimmten Klasse an Eingabebildern reagieren. Über die *Fully-connected Layer* kann nun wiederum eine gewichtete Summe der Aktivierungen all dieser Merkmalsextraktoren gebildet werden um so in der Ausgabeschicht schließlich die Aktivierung für jede einzelne Klasse aufzuzeigen. Die Ausgabeschicht weist in diesen Fällen ein sogenanntes *one-hot encoding* auf, was bedeutet, dass für jede Kategorie ein eigenes Neuron existiert, welches die Aktivierung für seine Klasse kodiert. Für die Anordnung der Schichten innerhalb eines CNN gibt es keine allgemeine Festlegung, dennoch existieren Anordnungskombinationen, die in den meisten Architekturen wiederzufinden sind. Die ersten Schichten eines CNN bestehen meist aus *Convolutional Layern* und wahlweise *Pooling Layern*, die letzten Schichten bestehen generell aus *Fully-connected Layern*, zu welchen auch die Ausga-

[7] Python-Skript zur Erzeugung der Grafik adaptiert von `https://github.com/gwding/draw_convnet`, 09.03.2018

beschicht zählt. Für die Anwendungshäufigkeit der *Pooling Layer* gibt es ebenso wenig Bestimmungen wie für die Parametrisierung der *Convolutional Layer*. Es ist nicht notwendig und eventuell sogar hinderlich, nach jeder Konvolution eine *Pooling*-Operation durchzuführen. Des Weiteren ist die optimale Anzahl an *Feature Maps* und die ideale Kernelgröße eines *Convolutional Layers* meistens experimentell zu ermitteln. Abbildung 2.5 zeigt eine mögliche CNN Architektur unter Angabe der einzelnen Schichtarten, sowie deren Parametrisierung.

2.3 Verbesserungsmethoden

Um die Leistung neuronaler Netze und speziell von CNNs zu erhöhen, gibt es einige Verbesserungsmethoden. Hierbei handelt es sich einerseits um Methoden, die dafür sorgen, dass das Lernverfahren beschleunigt werden kann. Andere Methoden verhindern eine zu starke Anpassung des Netzwerks an die präsentierten Trainingsstimuli, ein Defekt, der auch als *Overfitting* bezeichnet wird. Im Folgenden sollen drei solcher Methoden kurz eingeführt werden.

Weight Decay

Unter dem Begriff *weight decay* versteht man Maßnahmen, welche die Entstehung sehr großer Gewichte im Netzwerk unterbinden. Der Term geht auf Werbos (1988) zurück, der dieses Konzept als erster einführte. Große Gewichte können dazu führen, dass sich das Netzwerk zu sehr an die Trainingsstimuli anpasst, was in einer schlechteren Generalisierungsfähigkeit auf neue Stimuli mündet. Des Weiteren führen große Gewichte dazu, dass die Neuronen aufgrund ihrer Aktivierungsfunktionen im Sättigungsbereich arbeiten, wodurch kaum Aktivierungsunterschiede zwischen den Neuronen feststellbar sind. Um große Gewichte zu verhindern, kann die Fehlerfunktion des Netzwerks so angepasst werden, dass große Gewichte bestraft werden. Beispielsweise kann dies durch einen quadratischen Einfluss der Gewichte auf die Fehlerfunktion realisiert werden (Zell, 1994). Häufig wird hierzu auf die ursprüngliche Fehlerfunktion noch die Summe aus den quadrierten Gewichten des Netzwerks aufaddiert. Hierdurch wirken sich große Gewichte automatisch negativ auf den Fehlerwert aus und werden in der Folge erst gar nicht gebildet.

Momentum

Wie bereits in Abschnitt 2.1 erläutert, wird ein Gradientenabstieg auf der Fehlerfunktion durchgeführt, um eine optimale Konfiguration der Gewichte zu erhalten. Hierbei kann es vorkommen, dass das Verfahren in lokalen Minima stagniert, oder auf flacheren Ebenen nur sehr langsam fortschreitet. Um dies zu verhindern, wird ein sogenannter Momentum Term (nach Rumelhart, Hinton & Williams, 1986) zu jeder Gewichtsanpassung addiert. Dieser Term beinhaltet immer einen Anteil der letzten Gewichtsanpassung. Betrachtet man den Gradientenabstieg bildlich als Abstieg in einer Berglandschaft, so kann man die durch den Momentum Term gewonnene Dynamik mit der einer Kugel vergleichen, die einen Berg hinunterrollt. Aufgrund ihrer Masse kann die Kugel kleinere Täler überwinden und auch auf flacheren Ebenen an Geschwindigkeit gewinnen. Dasselbe gilt im übertragenen Sinne für den Gradientenabstieg mit Momentum-Term: lokale Minima können überwunden werden und einer Verzögerung auf flachen Plateaus wird entgegengewirkt. Außerdem führt der Momentum Term zu einer Verlangsamung in stark zerklüfteten Fehleroberflächen, was das Überspringen globaler Minima verhindern kann (Zell, 1994).

Dropout

Dropout, zu Deutsch etwa 'Aussetzer', beschreibt nach Srivastava, Hinton, Krizhevsky, Sutskever und Salakhutdinov (2014) eine Methode, welche verhindern soll, dass durch die übermäßige Aktivierung einzelner Neurone andere Neurone davon abgehalten werden, einen Beitrag zu den Berechnungen beizusteuern. Falls ein Neuron bereits stark auf die Präsentation eines bestimmten Stimulus reagiert, kann es passieren, dass dieses Neuron die alleinige Entscheidungskraft auf sich zieht und als Expertenneuron für diese Art von Stimulus hervorgeht. Hierdurch werden die Gewichte anderer Neurone kaum optimiert. Expertenneurone machen ein Netzwerk somit ineffizient und können die Leistung des Netzwerks reduzieren. Aus diesem Grund gilt es, die Bildung solcher Expertenneurone zu verhindern. Die Verwendung von *dropout* führt dazu, dass ein bestimmter, zuvor definierter Anteil der Neurone einer Schicht für einen Berechnungsschritt inhibiert wird (Srivastava et al., 2014). Dabei fällt die Entscheidung, ob ein Neuron inhibiert wird, nach dem Zufallsprinzip, lediglich die Wahrscheinlichkeit einer Inhibierung ist festgelegt. Somit sind für jeden Berechnungsschritt andere Neurone inhibiert und folglich wird über den Trainingsverlauf jedes Neuron in die Berechnung mit einbezogen. Diese Maßnahme hat sich als äußerst effektiv erwiesen, um das Auftreten von *overfitting* zu reduzieren und in der Folge die Generalisierungsfähigkeit neuronaler Netze zu erhöhen.

3 Methoden

3.1 Rechnerarchitektur

Sämtliche Programmierarbeit wurde an einem Rechner am Lehrstuhl für kognitive Neurowissenschaften durchgeführt.

Hardware

Im Folgenden werden die wichtigsten Hardware-Spezifikationen der verwendeten Rechnerarchitektur genannt:

- Prozessoren: 4× Intel® Core™ i5-6500 CPU @ 3.20 GHz
- Speicher: 16 GB RAM
- Grafikkarte: NVIDIA® GeForce® GTX 1050 Ti graphics processing unit

Software

Folgende Software wurde eingesetzt:

- Betriebssystem: Kubuntu® Version 16.04
- Application Processing Interface (API): Tensorflow™ Version 1.4.1

3.2 Tensorflow

TensorFlow[1] ist eine im Jahre 2015 vom *Google Brain Team*[2] entwickelte Open Source Programmierbibliothek. Primär ist die Bibliothek ausgelegt für die Verwendung in den verschiedenen Teilbereichen des maschinellen Lernens. Durch seine Plattformunabhängigkeit und die Kompatibilität zu verschiedenen Programmiersprachen wie Python®, C++ oder Java® ist TensorFlow weit verbreitet und hat sich als Standard im Bereich des maschinellen Lernens etabliert. Die Berechnungen bei TensorFlow werden intern auf einem *computational graph* durchgeführt. Dieser Graph beschreibt den Datenfluss während der Berechnungen, wobei die Daten stets in Form sogenannter Tensoren vorliegen und weitergereicht werden. Ein Tensor ist vergleichbar mit einem multidimensionalen Array. Die Operationen in Tensorflow basieren darauf, Tensoren einzulesen, zu verarbeiten und weiterzureichen. Aus dem so entstehenden Datenfluss aus Tensoren leitet sich auch der englische Name TensorFlow ab (Abadi et al., 2015). Neuronale

[1] TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.
[2] `https://research.google.com/teams/brain/` , 26.02.2018

Netze, welche im Zentrum dieser Arbeit stehen, bilden einen großen Eckpfeiler des maschinellen Lernens und werden häufig mittels TensorFlow realisiert. Es schien somit sinnvoll, für die Implementierung der Netzwerke im Rahmen dieser Arbeit ebenfalls TensorFlow zu verwenden. Als Programmiersprache wurde für diese Arbeit Python 3.0 verwendet. Python wurde als erste Programmiersprache von TensorFlow unterstützt wurde und liefert nach wie vor die größte Bandbreite an Funktionen (Abadi et al., 2015).

3.3 Datensatz

Für das Training und die Evaluation der neuronalen Netze wurde der *PlantCLEF-2015* Datensatz verwendet, welcher im Rahmen der *LifeCLEF* Datensätze[3] entstanden ist. Diese Datensätze bestehen aus Fotografien von Fischen, Vögeln oder Pflanzen. Der hier verwendete *LifeCLEF* Datensatz enthält nur Pflanzenbilder, weswegen er auch als *PlantCLEF* Datensatz bezeichnet wird. Er enthält zusätzlich zu jeder Fotografie einer Pflanzenart auch eine Meta-Datei, welche unter anderem die Artinformation beinhaltet. Die *LifeCLEF* Datensätze werden jedes Jahr aktualisiert und für internationale Wettbewerbe im Bereich der automatisierten Bilderkennung verwendet. Der Datensatz für diese Arbeit stammt aus dem Jahr 2015.

Der Fokus bei der Erzeugung der Datensätze liegt laut den Betreibern darin, ein Szenario zu schaffen, welches dem einer tatsächlichen Pflanzenklassifizierung in freier Wildbahn nahe kommt. Hierbei spielen folgende Faktoren eine wichtige Rolle: Die Fotografien für den *PlantCLEF-2015* Datensatz wurden von insgesamt 8.960 Mitwirkenden angefertigt, wobei verschiedenste Kameras verwendet wurden. Zusätzlich wurden die Fotografien in unterschiedlichen Teilen Westeuropas aufgenommen und jedes Bild gehört zu genau einer der sieben möglichen Ansichtsarten. Diese beinhalten: *ganze Pflanze*, *Frucht*, *Blatt*, *Blüte*, *Stamm*, *Zweig* oder *Blatt-Scan*. Die Ansichtsart wird in der Meta-Datei eines Bildes spezifiziert. Somit enthält der Datensatz zu jeder Pflanzenart Bilder von mehreren Vertretern dieser Art aus unterschiedlichen Regionen ihres Verbreitungsgebiets. Des Weiteren können die Bilder zu unterschiedlichen Jahreszeiten entstanden sein und decken somit ein breites Spektrum der Stadien einer Pflanze ab. Insgesamt enthält der *PlantCLEF-2015* Datensatz 91758 Fotografien[4] von 1000 Pflanzenarten aus 516 Pflanzengattungen und 124 Pflanzenfamilien (Goëau, Joly & Pierre, 2015). Eine genaue Aufschlüsselung nach Ansichtsart ist in Tabelle 3.1 zu finden. Aufgrund der genannten Faktoren soll der Datensatz

Summe	Ganze Pflanze	Frucht	Blatt	Blüte	Stamm	Zweig	Blatt-Scan
91758	16235	7720	13367	28225	12605	8130	5476

Tabelle 3.1: Auflistung der Anzahl an Bildern pro Ansichtsart (nach Goëau et al., 2015)

nach Goëau et al. ein Spektrum an Variabilität aufweisen, welches der Vielfalt in der Natur am nächsten kommt. Die Aussagekraft der Klassifikationsleistung eines Programms zur Pflanzenerkennung, welches mittels des *PlantCLEF* Datensatzes evaluiert wurde, ist somit höher als bei der Verwendung eines konservativeren Datensatzes und entspricht eher den Ergebnissen, welche in der Natur zu erwarten wären.

[3] `http://www.imageclef.org/lifeclef/2015` , 01.03.2018
[4] In der zitierten Quelle werden fälschlicherweise 91759 Fotografien genannt, dies wurde hier korrigiert.

Die Meta-Datei, welche zu jedem Bild im Datensatz existiert, enthält die folgenden Attribute:

- **ObservationID**: Identifikationsnummer einer bestimmten Pflanze
- **MediaID**: Identifikationsnummer des JPG Bildes
- **Vote**: Die durchschnittliche Qualitätsbewertung des Bildes durch die Nutzer
- **Content**: Eine der sieben oben genannten Ansichtsarten
- **ClassID**: Identifikationsnummer der Spezies der Pflanze, welche nach der TelaBotanica vergeben wird[5]
- **Species**: Die Spezies der abgebildeten Pflanze
- **Genus**: Die Gattung der abgebildeten Pflanze
- **Family**: Die Familie der abgebildeten Pflanze
- **Author**: Name des Autors des Bildes
- **Location**: Ort der Observation
- **Longitude/Latitude**: Genaue Koordinaten der Observation
- **YearInCLEF**: Gibt an, in welchem Jahr das Bild in den Datensatz eingefügt wurde

3.4 Netzwerkarchitektur

Das für diese Arbeit implementierte neuronale Netzwerk orientiert sich stark an dem im Jahr 2012 vorgestellten Netzwerk *AlexNet*. Bei *AlexNet* handelt es sich um ein *Convolutional Neural Network* (CNN), das 2012 bei der *ImageNet Large Scale Visual Recognition Challenge*, kurz ILSVRC[6], eine Klassifikationsleistung erzielte, welche die Leistung aller anderen teilnehmenden Netzwerke bei weitem übertraf (Krizhevsky, Sutskever & Hinton, 2012). Die Klassifikationsleistung von *AlexNet* war derart gut, dass mit ihm eine Wiederkehr der etwas in Vergessenheit geratenen CNNs eingeläutet wurde. In einschlägigen Wissenschaftskreisen wurde wieder vermehrt auf die Verwendung von *Convolutional Neural Networks* gesetzt. Unter anderem in der automatisierten Bilderkennung sah man die Nutzung von CNNs als vielversprechend und zukunftsweisend an. Mittlerweile wurden weitere Netzwerke vorgestellt, welche auf den Erkenntnissen von *AlexNet* aufbauen und wiederum dessen Leistung übertreffen (siehe Gu et al., 2017). Allerdings gingen diese Leistungsverbesserungen stets mit deutlich komplexeren und vor allem tieferen Netzwerkarchitekturen einher, welche ein wesentlich längeres Training benötigen, um akzeptable Ergebnisse zu liefern. Mit einer tieferen Architektur nimmt auch die Anzahl der zu trainierenden Gewichte zu, weswegen für gute Leistungen auch ein größerer Trainingsdatensatz benötigt wird.

[5] Ein Netzwerk bestehend aus mehreren tausend französischen Botanikern, die es sich unter anderem zur Aufgabe gemacht haben, Daten zu sämtlichen heimischen Pflanzenarten zu sammeln und in entsprechenden Datenbanken einzupflegen. Siehe `http://www.tela-botanica.org/site:accueil` , 01.03.2018

[6] Hierbei handelt es sich um eine jährliche Konferenz auf der Algorithmen zur Bildklassifizierung und Objekterkennung aus aller Welt vorgestellt werden. Meist handelt es sich bei den Algorithmen um neuronale Netzwerke und speziell um *Convolutional Neural Networks*. Um die Leistung der vorgestellten Netzwerke besser vergleichen zu können, werden sie alle auf dem selben Testdatensatz evaluiert und analysiert. Für weitere Informationen siehe: `http://www.image-net.org/challenges/LSVRC/` , 01.03.2018

Für die angesetzte Untersuchung war es nicht primär notwendig, die bestmögliche Klassifikationsleistung zu erzielen. Vielmehr war es das Ziel, Leistungsunterschiede zu untersuchen. Hierfür sollte die absolute Leistung des Netzwerks zunächst zweitrangig sein. Der Vorteil von *AlexNet* liegt in seiner einfachen Architektur und der daraus resultierenden relativ geringen Anzahl an benötigten Trainingsdurchläufen. Für die betrachteten Untersuchungen eignet sich *AlexNet* somit besonders gut.

Im Folgenden wird die implementierte Netzwerkarchitektur genauer erläutert. Diese basiert grundlegend auf *AlexNet* und wurde nur soweit modifiziert, wie es für die Durchführung der angesetzten Untersuchungen nötig war. Wie für ein *Convolutional Neural Network* üblich, bestehen die ersten Schichten aus *Convolutional Layern*. Auf diese folgen eine Reihe *fully-connected Layer*, zu welchen auch die Ausgabeschicht gehört. Das Netzwerk besitzt acht zentrale Schichten, fünf *Convolutional Layer* und drei *Fully-connected Layer*. Alle *Fully-connected Layer* wurden stets mit einer *Dropout* Regularisierungsmaßnahme versehen, um *Overfitting* zu verhindern. Zwischen den *Convolutional Layern* weist das Netzwerk mehrere *Max-Pooling Layer* auf.[7] Als Aktivierungsfunktion der Neurone wurde die *rectifier function* verwendet, welche alle negativen Werte auf Null und alle positiven Werte linear abbildet, somit also mathematisch definiert ist als $f(x) = max(0,x)$. Die *rectifier function* stellt eine herkömmliche Aktivierungsfunktion im Bereich neuronaler Netze dar. Da diese Funktion jedoch für $x = 0$ nicht differenzierbar ist, wird stattdessen meist eine differenzierbare Approximation als Aktivierungsfunktion verwendet. Wie bei CNNs für Klassifikationsaufgaben üblich, weist auch das implementierte Netzwerk ein *one-hot encoding* in der Ausgabeschicht auf. Die Anzahl der Neurone in der Ausgabeschicht stimmt somit mit der Anzahl an verfügbaren Klassen überein. Jedes Ausgabeneuron kodiert eine bestimmte Klasse und das Neuron mit der höchsten Aktivierung entspricht der Vorhersage des Netzwerks.

Zur Durchführung der angesetzten Untersuchungen war es notwendig *AlexNet* zu erweitern. Es wurden zwei zusätzliche Ausgabeschichten zum ursprünglichen Modell hinzugefügt und diese genauso wie die bisherige Ausgabeschicht mit dem letzten *Fully-connected Layer* verbunden. Somit entstanden neue Ausgabeschichten für die Vorhersage der Pflanzenfamilie und der Pflanzengattung. Das erweiterte Modell enthält folglich Ausgabeschichten für Familie, Gattung und Spezies. Das Netzwerk mit den zusätzlichen Ausgabeschichten wird im weiteren Verlauf als *erweitertes Netzwerk* bezeichnet, wohingegen das unveränderte Netzwerk als *klassisches Netzwerk* bezeichnet wird. Alle Ausgabeschichten des erweiterten Netzwerks weisen ebenso ein *one-hot encoding* auf. Somit besitzt die Speziesausgabeschicht 1000 Neurone, die Gattungsausgabeschicht 516 Neurone und die Familienausgabeschicht 124 Neurone. Dies entspricht jeweils der Anzahl an verfügbaren Klassen. Für eine schematische Darstellung des implementierten neuronalen Netzwerks siehe Abbildung 3.1.

[7] Die *Max-Pooling Layer* werden meist als Folgeoperationen der *Convolutional Layer* betrachtet und deswegen nicht als eigenständige Schichten gezählt.

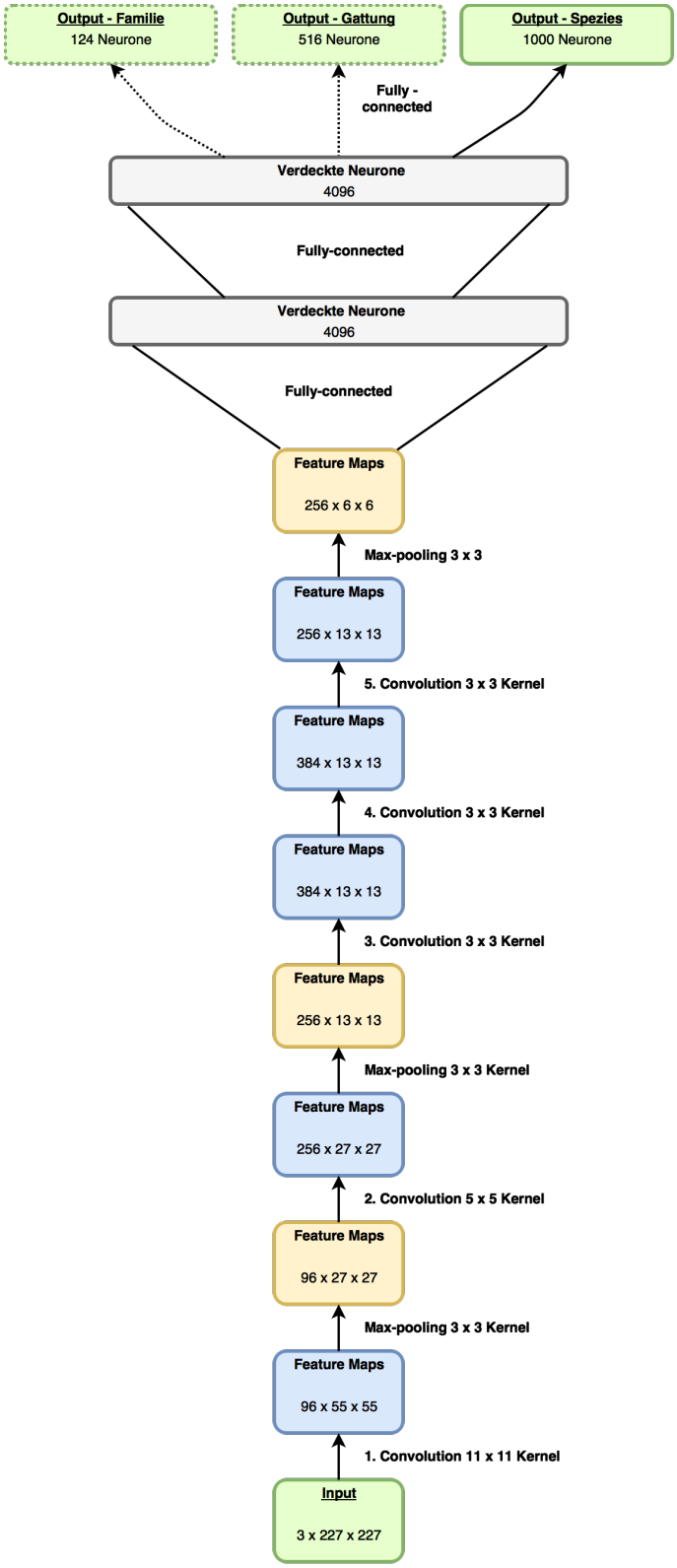


Abbildung 3.1: Graph der implementierten Netzwerkarchitektur. Das Eingabebild am unteren Ende wird über die verschiedenen Schichten durch das Netzwerk propagiert und dabei immer weiter verarbeitet. Die Beschriftungen neben den Pfeilen geben an, welche Operation durch die nächste Schicht realisiert wird. $256 \times 13 \times 13$ bezeichnet eine Schicht, bestehend aus 256 *Feature Maps* mit jeweils 13×13 Neuronen. In der Grafik sind bereits alle drei Ausgabeschichten des erweiterten Netzwerks eingezeichnet. Die für diesen Zweck zur klassischen Architektur hinzugefügten Ausgabeschichten wurden punktiert umrandet.

3.5 Training

Für ein optimales Ergebnis wurden über mehrere Trainingsdurchläufe Hyperparameter, wie die Anzahl an Iterationen, Lernrate, *dropout* und *weight decay* angepasst. Zuletzt wurden für die Hyperparameter folgende Werte festgesetzt:

- *dropout*: 0.3
- *weight decay*: 0.0005
- Iterationen: 150000
- Die Lernrate $\lambda(n)$ wurde als Stufenfunktion in Abhängigkeit der Iterationen n definiert. Für eine Visualisierung siehe Abbildung 3.2

$$\lambda(n) = \begin{cases} 0.001 & \text{für } 0 < n \leq 15000 \\ 0.0005 & \text{für } 15000 < n \leq 45000 \\ 0.0001 & \text{für } 45000 < n \leq 80000 \\ 0.000001 & \text{für } n > 80000 \end{cases}$$

Für Training und Evaluation des Netzwerks wurde der Datensatz aufgeteilt. Somit dienten 75650 Bilder als Trainingsdatensatz und 16109 Bilder als Testdatensatz. Als Metrik für die Fehlerfunktion wurde die Kreuzentropie verwendet.

Die Kreuzentropie stammt aus der Informationstheorie und kann vereinfacht ausgedrückt als Maß für die Ähnlichkeit zweier Wahrscheinlichkeitsverteilungen betrachtet werden (Goodfellow, Bengio, Courville & Bengio, 2016). Die Ausgabeschicht der hier verwendeten Netzwerke liefert für jede Klasse eine Aktivierung. Wendet man nun die sogenannte *softmax*-Funktion auf alle Aktivierungen an, so werden die Werte derart skaliert, dass ihre Summe Eins ergibt. Die resultierenden Aktivierungen können demgemäß als (diskrete) Wahrscheinlichkeitsverteilung verstanden werden. Das *teaching signal* besteht aus einer Eins und sonst aus Nullen, stellt also auch eine diskrete Wahrscheinlichkeitsverteilung dar. Die Kreuzentropie kann folglich verwendet werden, um die Ähnlichkeit zwischen der gewünschten Ausgabe und der tatsächlichen Ausgabe des Netzwerks zu beziffern (Goodfellow et al., 2016). Für diesen

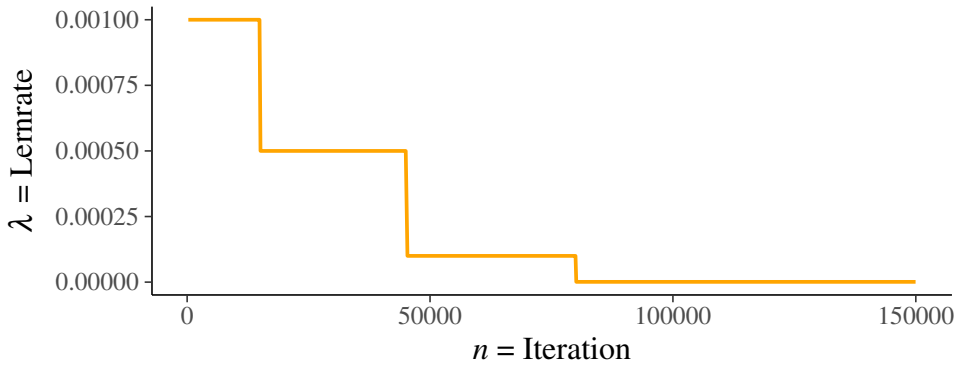


Abbildung 3.2: Verlauf der Lernrate über die Iterationen.

Zweck liefert Tensorflow eine Methode, die zunächst die *softmax*-Funktion anwendet und anschließend die Kreuzentropie berechnet.
Die Optimierung des Netzwerkes geschieht, indem die Kreuzentropie bezüglich der Aktivierung der Ausgabeschichten und des *teaching signals* reduziert wird. Dies wird klassisch über einen Gradientenabstieg auf der Fehleroberfläche realisiert. Gegenüber der geläufigen L^2-Norm[8] bietet die Kreuzentropie unter anderem den Vorteil, dass die in Folge von *Backpropagation* durchgeführten Gewichtsanpassungen im Netzwerk tatsächlich proportional zum vorgefunden Fehlervektor sind. Diese Eigenschaft ist bei der Verwendung der L^2-Norm nicht gegeben, da hier nicht die Aktivierungen der Neurone selbst, sondern lediglich deren differenzierte Aktivierungen in die Berechnung der Gewichtsanpassung einfließen. Bei starken Aktivierungen geht die differenzierte Aktivierung der Neurone aufgrund ihrer Aktivierungsfunktion allerdings gegen Null, was den Lernprozess stark verzögern kann (Nielsen, 2015, Kapitel 3)[9]. Da sich das Training zwischen dem klassischen und dem erweiterten Netzwerk leicht unterscheidet, wird es für beide Varianten getrennt erläutert.

Klassisches Netzwerk

In der Standardvariante von *AlexNet* wurde das Netzwerk nur auf die Erkennung der Pflanzenarten trainiert. Hierzu wurde zu jedem Trainingsbild ein *teaching signal* mit der zugehörigen Pflanzenart an die Ausgabeschicht der Spezies angelegt. Der Fehler des Netzwerks entsprach dann der Kreuzentropie aus der Aktivierung der Ausgabeschicht und dem angelegten *teaching signal*.

Erweitertes Netzwerk

Durch die Erweiterung der Architektur des Netzes war es möglich, das Netzwerk nicht nur auf die Erkennung einzelner Pflanzenarten, sondern zusätzlich auf die Erkennung der zugehörigen Pflanzenfamilie und der Pflanzengattung zu trainieren. Analog zum Training der Ausgabeschicht der Spezies wurde beim erweiterten Netzwerk zusätzlich ein *teaching signal* an die Ausgabeschicht der Familie und der Gattung angelegt, um das Netzwerk zu trainieren. Die Information über die Familie und Gattung eines Pflanzenbilds ist im *PlantCLEF* Datensatz zusätzlich zur Spezies enthalten und konnte mit den Bilddaten verknüpft werden. So war beim Anlegen eines bestimmten Bildes stets die Information über Familie, Gattung und Spezies vorhanden. Hierdurch wurde ein dezidiertes Training der Ausgabeschichten ermöglicht. Der Fehler jeder Ausgabeschicht ergab sich analog zur klassischen Netzwerkarchitektur. Zur Berechnung des gesamten Fehlers wurden die Fehlerwerte der einzelnen Ausgabeschichten summiert.

3.6 Untersuchungen

Zur Beantwortung der betrachteten Fragestellung wurden verschiedene Ansätze verfolgt. Im Folgenden werden diese Untersuchungsansätze einzeln angeführt und abgeleitete Hypothesen benannt. Für alle Untersuchungen wurde das erweiterte Netzwerk mit dem klassischen Netzwerk verglichen. Letzteres

[8] Auch als euklidische Norm bekannt. In 2D oder 3D euklidischen Vektorräumen entspricht dies der Länge oder dem Betrag eines Vektors. Im Falle einer Fehlerfunktion also der Länge des Fehlervektors.

[9] Diese Quelle liefert eine detailliertere Ausführung des Sachverhalts und kann auch als Online-Medium über folgenden Link besucht werden: `http://neuralnetworksanddeeplearning.com/chap3.html`

fungierte bei den Untersuchungen gewissermaßen als Kontrollbedingung. Man muss sich vor Augen führen, dass lediglich das erweiterte Netzwerk über zusätzliche Ausgabeschichten für Familie und Gattung verfügt. Um die Leistung beider Netzwerke zu vergleichen, ist man folglich auf die Speziesausgabeschicht beschränkt, weswegen sämtliche Informationen zum Vergleich beider Netzwerke den dort vorgefundenen Aktivierungen entnommen werden müssen.

3.6.1 Speziesgenauigkeit

Bei der ersten Untersuchung handelt es sich um den naheliegendsten Untersuchungsansatz. Diese Untersuchung beschäftigt sich mit der Frage, inwieweit die zusätzliche Bereitstellung von hierarchischer Trainingsinformation unmittelbar Auswirkungen auf die Pflanzenklassifikationsleistung des Netzwerks haben kann. Hierzu wurden das erweiterte Netzwerk und das klassische Netzwerk mit identischen Hyperparametern auf demselben Trainingsdatensatz trainiert, um gleiche Ausgangsbedingungen für einen späteren Vergleich zu schaffen. Nach dem abgeschlossenen Training wurden beide Netzwerke auf demselben Testdatensatz evaluiert. Hierbei wurde bei beiden Netzwerken die Erkennungsgenauigkeit der Spezies gemessen und als Vergleichsmaß der Klassifikationsleistung herangezogen. Zusätzlich wurde die sogenannte *top-5* Genauigkeit der Spezies gemessen, welche die Rate angibt, mit der sich die richtige Spezies unter den fünf Speziesneuronen mit der höchsten Aktivierung befindet.

Sollte sich die hierarchische Trainingsinformation unmittelbar positiv auf die Erkennungsleistung des Netzwerks auswirken, so sind für das erweiterte Netzwerk bessere Werte bezüglich der Speziesgenauigkeit und der *top-5* Speziesgenauigkeit zu erwarten.

3.6.2 Familienaktivierungen

Folgende Überlegung führte zu der Entwicklung dieses Ansatzes. Ein Training mit hierarchischer Information bezüglich Familie, Gattung und Spezies könnte dazu führen, dass im Netzwerk Verknüpfungen bzw. Muster entstehen, die für einen Forscher als interne Familienrepräsentationen deutbar wären. In diesem Fall könnte man annehmen, dass eine hohe Aktivierung einer bestimmten Familie dazu führt, dass die Speziesneurone, die dieser Familie angehören, ebenfalls stark aktiviert werden. Das Training mit Familie, Gattung und Spezies hätte somit einen Effekt auf die Aktivierungsverteilung in der Speziesausgabeschicht.

Um dies zu untersuchen, wurden wieder beide Netzwerke identisch trainiert. Nach dem Training wurden beide Netzwerke auf demselben Testdatensatz evaluiert. Hierbei wurden zunächst die fünf Speziesneurone mit der höchsten Aktivierung in der Speziesausgabeschicht selektiert. Da jedes Speziesneuron genau eine Spezies kodiert, konnte zu jeder der 5 Spezies die zugehörige Familie ermittelt werden. Der so entstandene fünfstellige Familienvektor konnte mit der tatsächlichen Familie verglichen werden. Als Genauigkeitsmaß wurde der Quotient berechnet, welcher beschreibt, wie viele der Einträge im Familienvektor die richtige Familie aufweisen. Sollten alle Einträge richtig sein, so entspräche dies einem Wert von $\frac{5}{5} = 1.0$, wären zwei der fünf Familien korrekt, so ergäbe sich ein Wert von $\frac{2}{5} = 0.4$. Man erhält somit ein Maß für den Einfluss der Familie auf die Aktivierungsverteilung in der Speziesausgabeschicht.

Es ist festzuhalten, dass der höchste erreichbare Wert in diesem Fall nicht 1.0 beträgt, sondern deutlich darunter liegt. Das rührt daher, dass manche Familien weniger als 5 Spezies beinhalten. Wäre das neuronale Netz im betrachteten Fall fehlerfrei, würde es also immer alle Speziesneurone der korrekten Familie

maximal aktivieren, läge der Wert bei 0.61. Dieser Wert sollte bei der Interpretation der Ergebnisse als Referenzwert herangezogen werden.

3.6.3 Familiengenauigkeit

Dieser Ansatz ähnelt in seinen Überlegungen dem vorherigen Ansatz, dennoch liegt ihm eine andere Hypothese zugrunde: Sollte das Netzwerk Muster erzeugen, die als Familienrepräsentationen interpretierbar sind, so wäre es möglich, dass zwar die richtige Familie erkannt wird, jedoch die falsche Art innerhalb dieser Familie durch das Netzwerk vorhergesagt wird. In diesem Fall ist die Erkennungsgenauigkeit der Spezies möglicherweise unverändert, die Fehler aber, welche das Netzwerk produziert, treten gehäuft innerhalb der richtigen Familie auf, wodurch die Familiengenauigkeit zunimmt. Da lediglich das angepasste Netzwerk eine dezidierte Familienausgabeschicht besitzt, war für den Vergleich beider Netzwerke bezüglich der Familiengenauigkeit ein Umweg notwendig. Zunächst wurde das Speziesneuron mit der höchsten Aktivierung in der Spezies-Ausgabeschicht – demnach die Speziesvorhersage – selektiert. Daraufhin wurde die zur Speziesvorhersage zugehörige Familie ermittelt und in der Folge mit der tatsächlichen Familie verglichen. Es ergab sich somit eine Familiengenauigkeit basierend auf den Speziesvorhersagen des Netzwerks. Sollte dieser Genauigkeitswert für das erweiterte Netzwerk größer sein als für das klassische Netzwerk, so wäre dies ein Indiz für einen positiven Einfluss der hierarchischen Trainingsinformation. Dieser wirkt sich zwar nicht notwendigerweise direkt auf die Erkennungsgenauigkeit der Spezies aus, führt aber im Netzwerk zu Mustern, welche die Schwere der begangenen Fehler zu reduzieren vermögen. Diese entstandenen Muster könnten für einen Betrachter ein gewisses Familienverständnis repräsentieren.

3.6.4 Reduzierte Trainingsinformation

Bei allen Untersuchungen wurde bisher nur die Spezies- und die Familiengenauigkeit in Betracht gezogen. Es stellt sich die Frage, inwieweit die Trainingsinformation bezüglich der Gattung einer Pflanzenart überhaupt relevant für die durchgeführten Untersuchungen ist. Wäre es nicht denkbar, dass die zusätzliche Information der Gattung das Netzwerk in seinen Anpassungen einschränkt statt fördert und somit eher hinderlich für das Entstehen von vorteilhaften Mustern für die Familienerkennung ist?

Um dies zu untersuchen, wurde das erweiterte Netzwerk nur mit Spezies und Familie trainiert und die Gattung außen vor gelassen. Bei der Evaluation des Netzwerks wurden alle bisherigen Untersuchungen erneut durchgeführt und ausgewertet. Sollte die Information bezüglich der Gattung einer Pflanzenart überflüssig sein und sogar einen negativen Einfluss auf die Leistung des Netzwerks haben, so wären hier positive Änderungen gegenüber den zuvor erhaltenen Werten zu erwarten.

3.6.5 Unbekannte Arten

Folgende Überlegung führte zur Entwicklung dieses Ansatzes. Sollten durch die hierarchische Trainingsinformation aus Familie, Gattung und Spezies im Netzwerk Muster entstehen, die für einen Betrachter als interne Familienrepräsentationen deutbar wären, so könnte dies dazu führen, dass bei der Präsentation bisher unbekannter Pflanzenspezies zumindest die Pflanzenfamilie richtig vorhergesagt werden kann. Es ist wichtig, hierbei den grundsätzlich neuen Charakter dieses Ansatzes zu allen vorherigen Ansätzen zu betonen. Zuvor wurden bei der Evaluation der Netzwerke zwar auch neue Stimuli

präsentiert, diese zeigten jedoch Bilder bereits bekannter Arten.

Bei dem hier verfolgten Ansatz sind nicht nur die präsentierten Bilder der Pflanzenarten neu für das Netzwerk, sondern auch die darauf abgebildeten Arten. Aus diesem Grund ist nicht davon auszugehen, dass das Netzwerk diese unbekannten Arten korrekt klassifizieren kann. Allerdings gehört jede unbekannte Art zu einer Familie mit mehreren anderen Vertretern. Die anderen Vertreter wurden dem Netzwerk während des Trainings präsentiert und haben, falls die Hypothese zutrifft, dazu beigetragen, dass sich entsprechende Muster im Netzwerk ausgebildet haben, die als Familienrepräsentationen gedeutet werden können. Auf der Ebene der Familie stellen die unbekannten Arten damit keinen komplett neuen Stimulus für das Netzwerk dar. Somit könnten die Familien der unbekannten Arten korrekt klassifiziert werden. Dies würde für eine gesteigerte Familien-Generalisierungsfähigkeit sprechen.

Um dies zu untersuchen, wurde der Datensatz bearbeitet. Alle Familien mit mindestens drei enthaltenen Spezies wurden selektiert und eine der beinhalteten Spezies aus dem Datensatz aussortiert. Alle Vorkommen dieser Spezies im ursprünglichen Datensatz wurden in einen neuen Datensatz verschoben. Die im ursprünglichen Datensatz verbliebenen Spezies dienten im Folgenden als Trainingsdatensatz, während die aussortierten Spezies als Testdatensatz fungierten. Somit entstand ein Trainingsdatensatz aus 82.984 Bildern und ein Testdatensatz aus 8.774 Bilder. Beide Netzwerke wurden auf dem Trainingsdatensatz mit identischen Parametern trainiert und daraufhin auf dem Testdatensatz evaluiert. Hierbei wurde, ähnlich zu Ansatz 3.6.3, die Familiengenauigkeit gemessen. Die Speziesvorhersage wurde festgehalten und die entsprechende Familie ermittelt. Somit ergab sich die Familiengenauigkeit basierend auf der Speziesvorhersage.

3.7 Analysemethoden

3.7.1 Teststatistik

Bei allen genannten Untersuchungen wurde die Leistung des erweiterten Netzwerks mit der des klassischen Netzwerks verglichen. Ein Unterschied zwischen den erhaltenen Werten kann als Indiz für einen tatsächlichen Leistungsunterschied dienen. Hierfür wird eine Teststatistik benötigt, mit welcher beide Werte auf Ungleichheit getestet werden können. Bei den gemessenen Werten handelt es sich immer um reellwertige Genauigkeitswerte im Intervall $[0,1]$. Diese können auch als Anteilswerte in zwei Grundgesamtheiten verstanden werden. Ein Hypothesentest, welcher sich für diese Art von Vergleich zweier Anteilswerte eignet, ist der *Two Proportion Z-Test*. Die Teststatistik ist gegeben durch:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2})}} \sim N(0,1)$$

Hierbei stehen p_1 und p_2 für die erhaltenen Genauigkeitswerte, während n_1 sowie n_2 die Größe der beiden Grundgesamtheiten angeben. Im betrachteten Fall entsprechen n_1 und n_2 der Anzahl an Testbildern, auf denen das Netzwerk evaluiert wird.

Es wird angenommen, dass die Genauigkeitswerte binomialverteilt sind. Die Binomialverteilung kann für große n duch die Normalverteilung gut approximiert werden. Aus diesem Grund wird hier die klassische Z-Statistik verwendet.[10]

[10] Streng genommen ist die Bedingung für eine Binomialverteilung der Genauigkeitswerte nicht gegeben. Die Binomialverteilung ist definiert als die Summe von unabhängigen, identisch verteilten Zufallsvariablen, welche einer Bernoulliverteilung

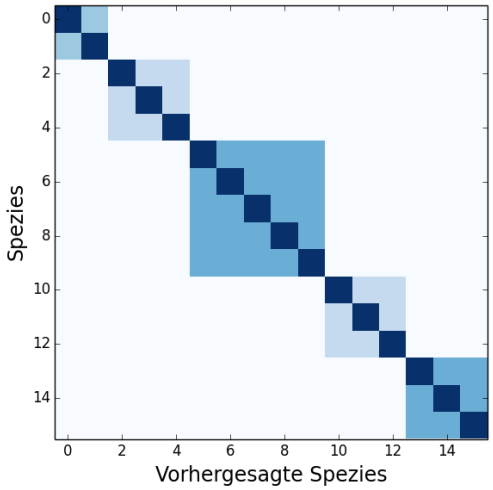


Abbildung 3.3: Schematische Darstellung möglicher Familiencluster innerhalb einer Konfusionsmatrix mit 16 familienweise angeordneten Spezies. Je dunkler die Färbung, desto höher die relative Häufigkeit einer bestimmten Klassifikation. Häufige Falschklassifikationen innerhalb einer Familie führen zur Bildung quadratischer Cluster entlang der Diagonalen der Matrix.

3.7.2 Konfusionsmatrix

Die Erzeugung einer Konfusionsmatrix ist eine weitverbreitete Methode, um die Leistung neuronaler Netze genauer zu analysieren. Sie eignet sich besonders gut, um das Klassifikationsverhalten und die dabei begangenen Fehler des Netzwerks besser zu verstehen. Insbesondere können mit ihr systematische Falschklassifizierungen aufgedeckt werden. Daher rührt auch die Bezeichnung derselben als Konfusionsmatrix

Wie auch in dieser Arbeit werden neuronale Netze meist darauf trainiert, Eingaben bestimmten Klassen zuzuordnen. In einer Konfusionsmatrix entspricht jede Zeile einer dieser Klassen. Die Spalten der Matrix entsprechen dabei den vom Netzwerk vorhergesagten Klassen. Ein Eintrag a_{ij} in der Konfusionsmatrix A gibt dann die relative Häufigkeit an, mit welcher das Netzwerk eine Eingabe aus der Klasse i der Klasse j zuordnet. Ist ein Netzwerk fehlerfrei, sagt es also für jede Eingabe der Klasse i stets die korrekte Klasse i vorher. So wären in diesem Fall in der Konfusionsmatrix alle Einträge $a_{ij} = 1$, für $i = j$ und sonst 0. Bei einem perfekten Klassifizierer würde demnach die Konfusionsmatrix in der Diagonalen nur Einsen aufweisen und sonst aus Nullen bestehen.

Zur besseren Visualisierung werden Konfusionsmatrizen mit n Klassen meist in Form einer *Heatmap* dargestellt. Das bedeutet, den relativen Häufigkeitswerten im Intervall $[0,1]$ wird ein bestimmter

unterliegen (Fisz, 1976). Die Bedingung identisch verteilter Zufallsvariablen bedeutet, dass die zugrundeliegende Erfolgswahrscheinlichkeit p bei allen Bernoulli-Versuchen identisch sein muss. Diese Bedingung ist im vorliegenden Fall nicht erfüllt, denn die Erfolgswahrscheinlichkeit – hier die Wahrscheinlichkeit einer korrekten Klassifizierung – ist für jede Spezies und damit für verschiedene Bernoulli-Versuche unterschiedlich. Eine Verteilung, welche mit dieser Einschränkung kompatibel ist, ist die sogenannte Verallgemeinerte Binomialverteilung. Hier kann jedem Versuch eine andere Erfolgswahrscheinlichkeit zugeordnet werden. Die Varianz der Verallgemeinerten Binomialverteilung ist dabei nach oben durch die der Binomialverteilung beschränkt. Bei Hypothesentests führt eine geringere Varianz zu signifikanteren Ergebnissen (Fisz, 1976). Somit können die aus obiger Teststatistik erhaltenen Z-Werte als untere Schranke der tatsächlich vorliegenden Z-Werte dienen. Wird also eine Differenz zweier Genauigkeitswerte unter Verwendung der vorliegenden Teststatistik signifikant, so ist dies auch unter der Annahme der verallgemeinerten Binomialverteilung der Fall. Die Teststatistik kann folglich trotz der Einschränkungen als untere Abschätzung verwendet werden.

Farbverlauf zugeordnet, wodurch aus der Konfusionsmatrix ein farbiges Pixelbild mit $n \times n$ Pixeln entsteht. Bei der Betrachtung einer Konfusionsmatrix ist zunächst zu beurteilen, zu welchem Grad eine Diagonale erkennbar ist. Je stärker sich die Diagonale vom Hintergrund abhebt, umso besser ist die Klassifikationsleistung des Netzwerks. Interessanter ist jedoch die Frage, wo in der Matrix außerhalb der Diagonalen Einträge existieren mit $a_{ij} > 0$. Diese stellen, abhängig von ihrem Ausmaß, mehr oder weniger schwerwiegende, systematische Falschklassifikationen des Netzwerks dar.

Für die genauere Analyse der Netzwerke in dieser Arbeit wurden mehrere Konfusionsmatrizen, sowohl für die Spezies- als auch für die Familienklassifikation angefertigt. Um die Familienvorhersage beider Netzwerke zu eruieren, wurde wie in Abschnitt 3.6.3 zu jeder Speziesvorhersage die zugehörige Familie ermittelt. Für die Konfusionsmatrix der Speziesklassifikation sollte die Anordnung der Spezies in der Matrix derart sein, dass alle Mitglieder einer Familie gruppiert werden. Wenn also eine Familie aus fünf Spezies besteht, so sollten diese fünf Spezies in der Matrix aufeinanderfolgend angeordnet sein. Bei mehr Falschklassifikationen innerhalb einer Familie dürfte sich dies somit durch quadratische Cluster entlang der Diagonalen der Matrix zeigen, wobei die Kantenlänge dieser Cluster durch die Familiengröße gegeben ist. Für eine schematische Darstellung siehe Abbildung 3.3.

Bilal et al. (2018) konnten für konventionell trainierte CNNs bereits zeigen, dass solch eine Clusterbildung großteils vorliegt. Die Frage ist, ob die Cluster noch stärker hervortreten, wenn vielschichtigere, hierarchische Trainingsinformationen vorliegen. Die Anordnung der Spezies innerhalb der Matrix entspricht den Labeln der einzelnen Spezies. Bei 1000 Spezies gibt es Label $0 - 999$. Die Nummerierung der einzelnen Spezies war für das Training zufällig festgelegt worden und folgte keiner Logik. Um die gewünschte Anordnung in der Matrix zu erhalten, wurden alle Spezies nun nach ihren Familien gruppiert und dann familienweise durchnummeriert.

4 Ergebnisse

Im Folgenden werden die Ergebnisse der durchgeführten Untersuchungen der Reihe nach dargelegt. Zusätzlich werden die erhaltenen Konfusionsmatrizen abgebildet.

4.1 Speziesgenauigkeit

Die Speziesgenauigkeit des klassischen Netzwerks auf dem Testdatensatz lag bei 37.06%. Das erweiterte Netzwerk erreichte hierbei eine Speziesgenauigkeit von 39.63%. Nach dem *Two Proportion Z-Test* ist das Ergebnis des erweiterten Netzwerks signifikant größer als das Ergebnis des klassischen Netzwerks ($Z = 4.73$, $p < .001$). Für eine Visualisierung der zeitlichen Entwicklung der Speziesgenauigkeit während des Trainings siehe Abbildung 4.1. Bezüglich der *top-5* Genauigkeit der Spezies wurden für das klassische Netzwerk 53.18% und für das erweiterte Netzwerk 57.25% gemessen. Auch hier ist nach dem *Two Proportion Z-Test* der Wert des erweiterten Netzwerks signifikant größer als der des klassischen Netzwerks ($Z = 7.33$, $p < .001$). Für das erweiterte Netzwerk wurden außerdem die Genauigkeitswerte für Familie und Gattung gemessen. Hierbei betrug die Familiengenauigkeit auf dem Testdatensatz 48.79%. Die Gattungsgenauigkeit lag bei 43.13%.

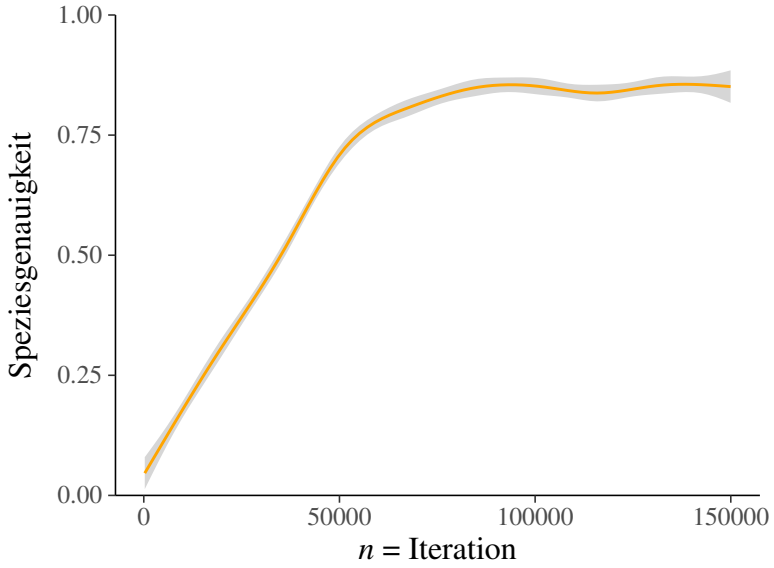


Abbildung 4.1: Entwicklung der Speziesgenauigkeit des erweiterten Netzwerks währen des Trainings. Der graue Bereich um die Kurve gibt die Streuung an.

	Training		
Messwert	Spezies	Familie & Spezies	Familie, Gattung & Spezies
Speziesgenauigkeit	37.05%	38.76%	39.63%
Top-5 Speziesgenauigkeit	53.18%	55.95%	57.25%
Familienaktivierung	22.39%	27.59%	28.82%
Familiengenauigkeit	45.42%	48.84%	50.04%

Tabelle 4.1: Überblick über die Messwerte, aufgeteilt nach Trainingsart. In der ersten Spalte befinden sich die Werte des klassischen Netzwerks, welches mit der Speziesinformation trainiert wurde. In der zweiten Spalte sind die Werte des erweiterten Netzwerks mit reduzierter Trainingsinformation abgebildet. Dieses Netzwerk wurde mit der Familien- und Speziesinformation trainiert. In der dritten Spalte sind die Werte des erweiterten Netzwerks aufgeführt, welches die vollständige Trainingsinformation bezüglich Familie, Gattung und Spezies erhalten hat.

4.2 Familienaktivierungen

Der Prozentsatz korrekter Familien unter den *top-5* Speziesneuronen belief sich beim klassischen Netzwerk auf 22.39% und beim erweiterten Netzwerk auf 28.82%. Der Wert des erweiterten Netzwerks ist nach dem *Two Proportion Z-Test* signifikant größer als der des klassischen Netzwerks ($Z = 13.21$, $p < .001$).

4.3 Familiengenauigkeit

Die Familiengenauigkeit auf Basis der Speziesvorhersage lag beim klassischen Netzwerk bei 45.42% und beim erweiterten Netzwerk bei 50.04%. Der Wert des erweiterten Netzwerks ist nach dem *Two Proportion Z-Test* mit $Z = 8.28$, $p < .001$ signifikant größer als der des klassischen Netzwerks.

4.4 Reduzierte Trainingsinformation

Für die Speziesgenauigkeit ergab sich ein Wert von 38.76%, die *top-5* Speziesgenauigkeit lag bei 55.95%. Bezüglich der Familienaktivierungen ergab sich ein Prozentsatz von 27.6% korrekter Familien unter den *top-5* Speziesneuronen. Die Familiengenauigkeit lag bei 48.84%. Alle erhaltenen Werte sind signifikant größer als die respektiven Werte des klassischen Netzwerks. Speziesgenauigkeit: $Z = 3.15$, $p < .001$; *top-5* Speziesgenauigkeit: $Z = 3.61$, $p < .001$; Familienaktivierung: $Z = 10.76$, $p < .001$; Familiengenauigkeit: $Z = 6.13$, $p < .001$.

Im Vergleich zum erweiterten Netzwerk ergab sich Folgendes: Die Speziesgenauigkeit ist nicht signifikant kleiner als die des erweiterten Netzwerks ($Z = 1.59$, $p = 0.056$). Die *top-5* Speziesgenauigkeit ist signifikant kleiner als die des erweiterten Netzwerks ($Z = 2.35$, $p = 0.009$). Auch die Familienaktivierung ist signifikant kleiner als die des erweiterten Netzwerks ($Z = 2.45$, $p = 0.007$). Die Familien-

genauigkeit ist auch signifikant kleiner als die des erweiterten Netzwerks ($Z = 2.15$, $p = 0.016$). Tabelle 4.1 gibt einen Überblick über die bisher genannten Messwerte.

4.5 Unbekannte Arten

Bei den unbekannten Arten ergaben sich folgende Werte für die Familiengenauigkeit: Das klassische Netzwerk erreichte eine Genauigkeit von 11.76% und das erweiterten Netzwerk eine Genauigkeit von 13.56%. Der Wert des erweiterten Netzwerks ist nach dem *Two Proportion Z-Test* signifikant größer als der des klassischen Netzwerks ($Z = 4.84$, $p < .001$).

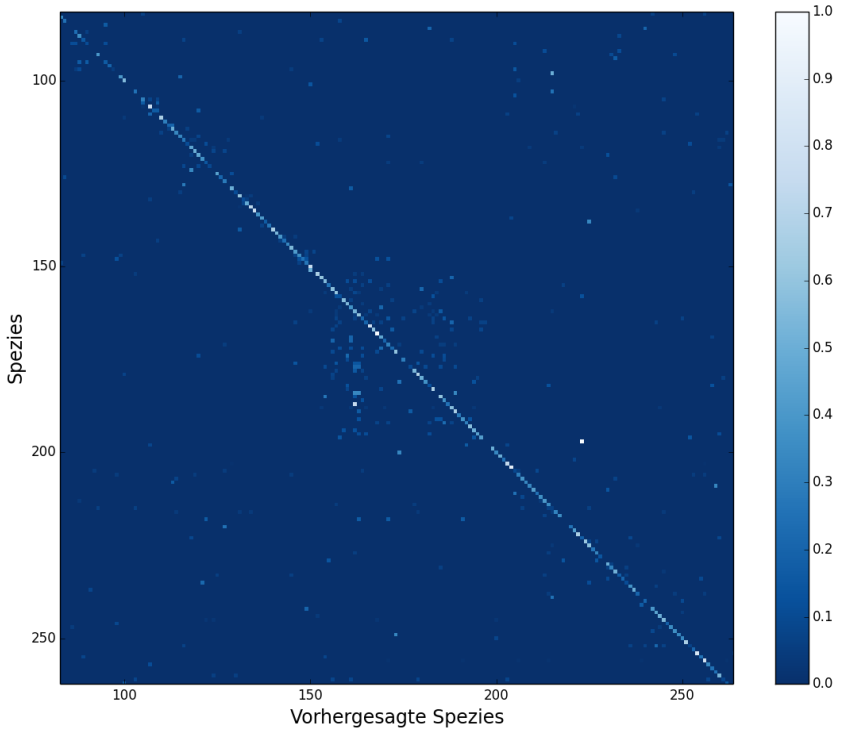
4.6 Konfusionsmatrix

Im Folgenden werden die Konfusionsmatrizen der Spezies- und Familienerkennung für das klassische und das erweiterte Netzwerk präsentiert.

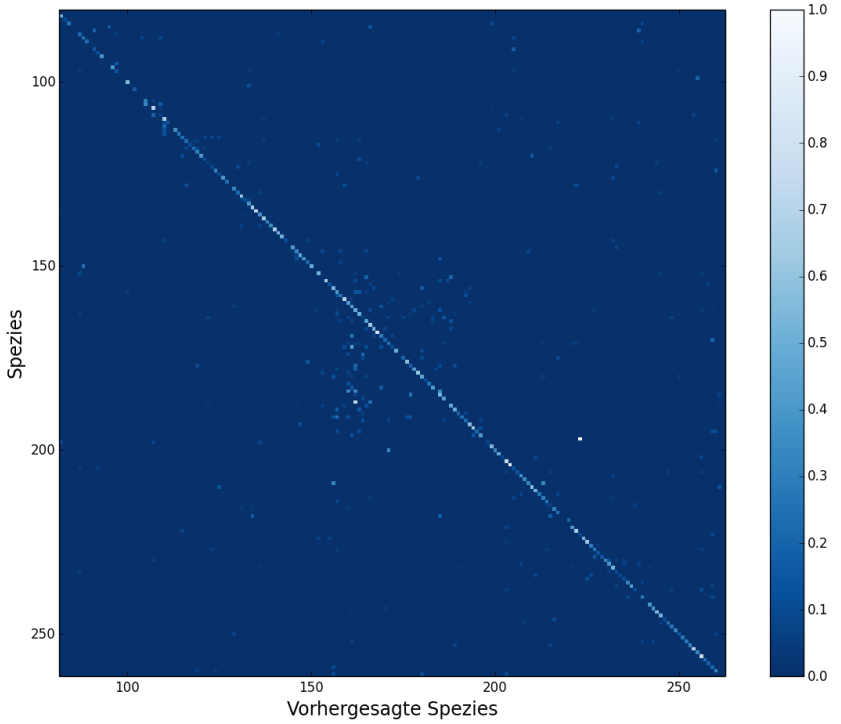
Die Darstellung der ganzen Konfusionsmatrix der Spezieserkennung gestaltet sich jedoch schwierig. Durch die große Anzahl von 1000 Spezies entsteht eine Matrix mit einer Million Einträgen. Aufgrund ihrer Größe ist es nicht möglich, die ganze Matrix abzubilden, ohne dass dabei die relevante Information über die Färbung einzelner Pixel verloren geht.[1] Stattdessen werden Ausschnitte der Konfusionsmatrix betrachtet. Von Interesse ist primär die Ausbildung von Familienclustern beim erweiterten Netzwerk. Die folgenden Ausschnitte aus der Konfusionsmatrix des erweiterten Netzwerks zeigen Kandidaten solcher Cluster. Zum Vergleich wurde zu jedem Ausschnitt ein entsprechender Ausschnitt aus der Konfusionsmatrix des klassischen Netzwerks angefertigt. Alle Abbildungen der Cluster sind auch in Form von .png Dateien auf der beigefügten CD enthalten.

Das erste Cluster befindet sich auf Höhe der Spezies 152 – 197, welche die Familie der Orchideen (Orchidaceae) bilden. Siehe hierzu Abbildung 4.2.

[1] In der beigefügten CD ist die gesamte Matrix als .png Datei enthalten.



(a) Erweitertes Netzwerk



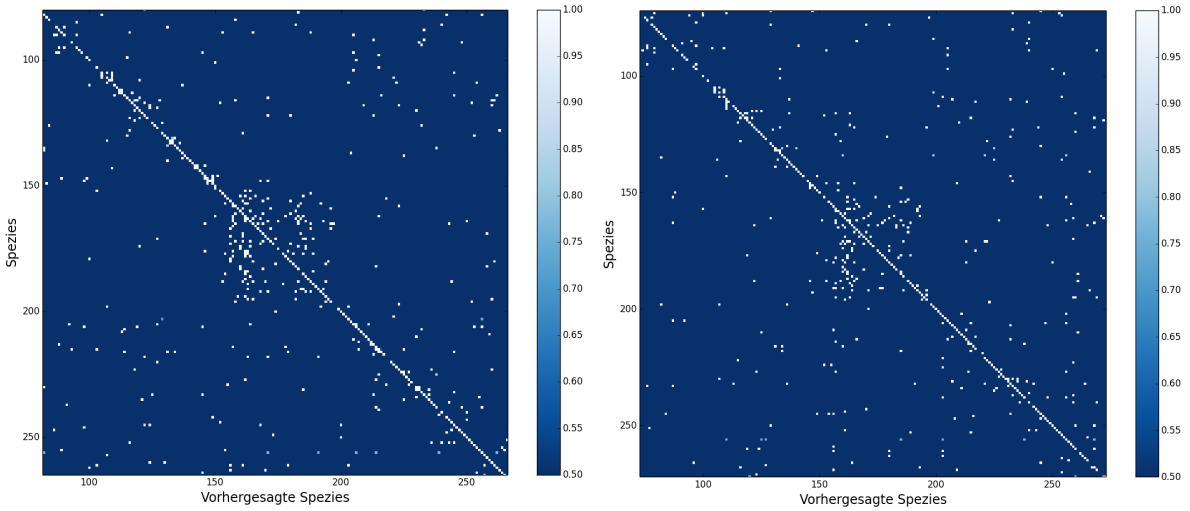
(b) Klassisches Netzwerk

Abbildung 4.2: Familiencluster der Orchideen bei beiden Netzwerken

Da sich die Cluster in dieser Ansicht nur sehr schwer erkennen lassen, wurden die Einträge in den Konfusionsmatrizen adaptiert. Durch Anwendung einer Schwellenfunktion[2] wurden alle Werte ab ca. 0.04 auf

[2]Es handelt sich um die Funktion $f(x) = \frac{1}{2} \cdot tanh(50x - 1) + \frac{1}{2}$. Diese approximiert eine Schwellenfunktion und lieferte die besten Ergebnisse. Gegenüber einer tatsächlichen Schwellenfunktion bietet diese Funktion den Vorteil, dass weiterhin ein kontinuierlicher Übergang zwischen den Werten 0 und 1 vorliegt.

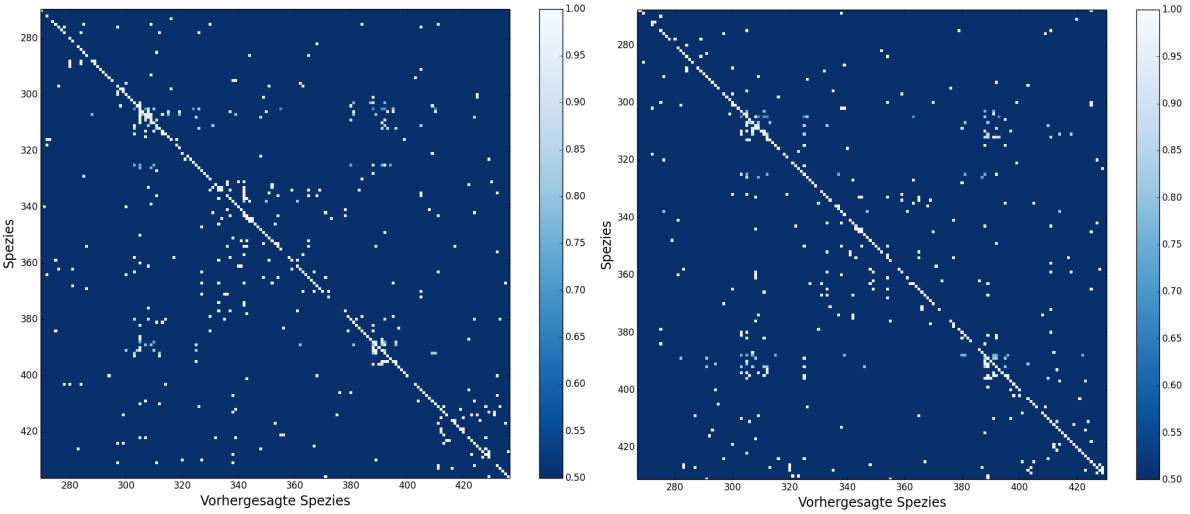
1.0 gesetzt. Hierdurch werden die Familiencluster besser sichtbar, allerdings nimmt auch das Rauschen innerhalb der Matrizen zu. Die Cluster aus Abbildung 4.2 sind nach Anwendung der Schwellenfunktion in Abbildung 4.3 abgebildet.



(a) Erweitertes Netzwerk (b) Klassisches Netzwerk

Abbildung 4.3: Ausschnitte von Abb. 4.2 nach Anwendung der Schwellenfunktion.

Aufgrund der besseren Erkennbarkeit werden im weiteren Verlauf bezüglich der Spezieserkennung nur noch Ausschnitte aus den adaptierten Konfusionsmatrizen gezeigt.[3]
Nachfolgend ist ein Cluster auf Höhe der Spezies 331 – 378 zu finden. Diese formen die Familie der Lippenblütler (Laminaceae). Für eine Visualisierung siehe Abbildung 4.4. Zusätzlich sind noch zwei kleine Cluster erkennbar. Das obere Cluster beinhaltet die Spezies 305 – 315, welche die Familie der Buchengewächse (Fagaceae) bilden. Das untere Cluster bildet mit den Spezies 388 – 396 die Familie der Birkengewächse (Betulaceae)

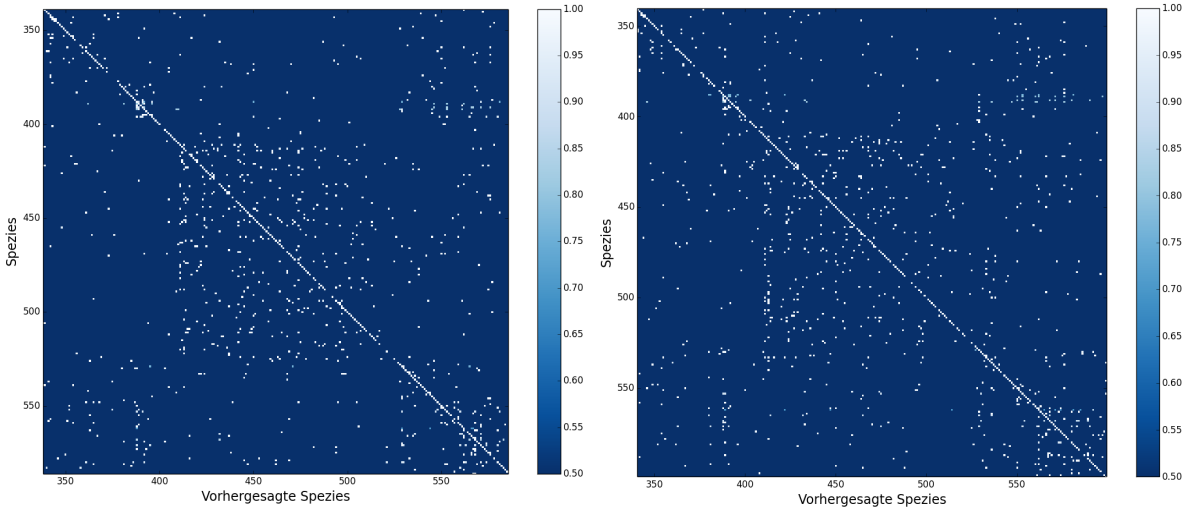


(a) Erweitertes Netzwerk (b) Klassisches Netzwerk

Abbildung 4.4: Cluster der Familie der Lippenblütler für beide Netzwerke.

[3]Dies hat zur Folge, dass die Information über Abstufungen innerhalb der Matrizen größtenteils verloren geht. Alle Ausschnitte aus den ursprünglichen Konfusionsmatrizen sind jedoch als .png Dateien auf der beigefügten CD enthalten.

Ein weiteres Cluster ist auf Höhe der Spezies 411 – 527 zu sehen. Diese Spezies bilden die Familie der Korbblütler (Asteraceae). Eine Visualisierung ist in Abbildung 4.5 zu finden.

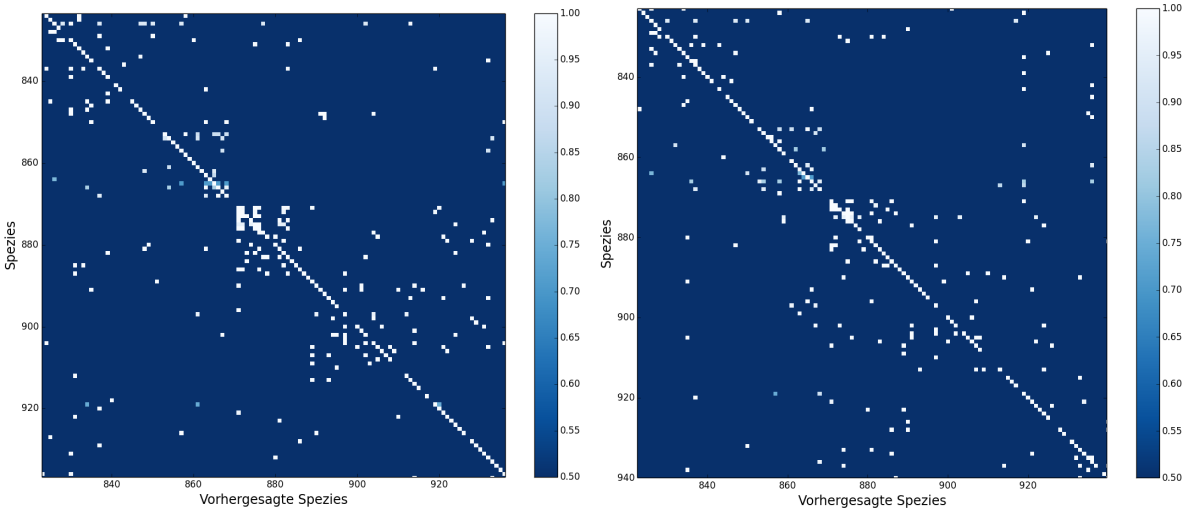


(a) Erweitertes Netzwerk

(b) Klassisches Netzwerk

Abbildung 4.5: Cluster der Familie der Korbblütler für beide Netzwerke.

Auf Höhe der Spezies 871 – 888 ist ein weiteres Cluster zu finden. Die genannten Spezies formen die Familie der Storchschnabelgewächse (Geraniaceae). Für eine Visualisierung siehe Abbildung 4.6.

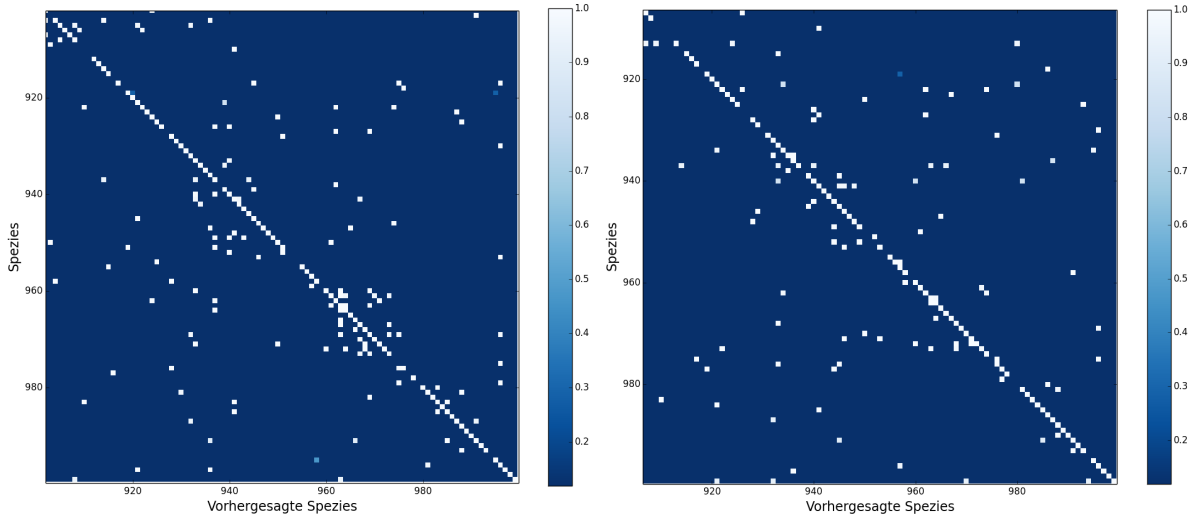


(a) Erweitertes Netzwerk

(b) Klassisches Netzwerk

Abbildung 4.6: Cluster der Familie der Storchschnabelgewächse für beide Netzwerke.

Schließlich befindet sich auf Höhe der Spezies 961 – 974 ein letztes Cluster. Diese Spezies bilden die Familie der Glockenblumengewächse (Campanulaceae). Für eine Visualisierung siehe Abbildung 4.7.

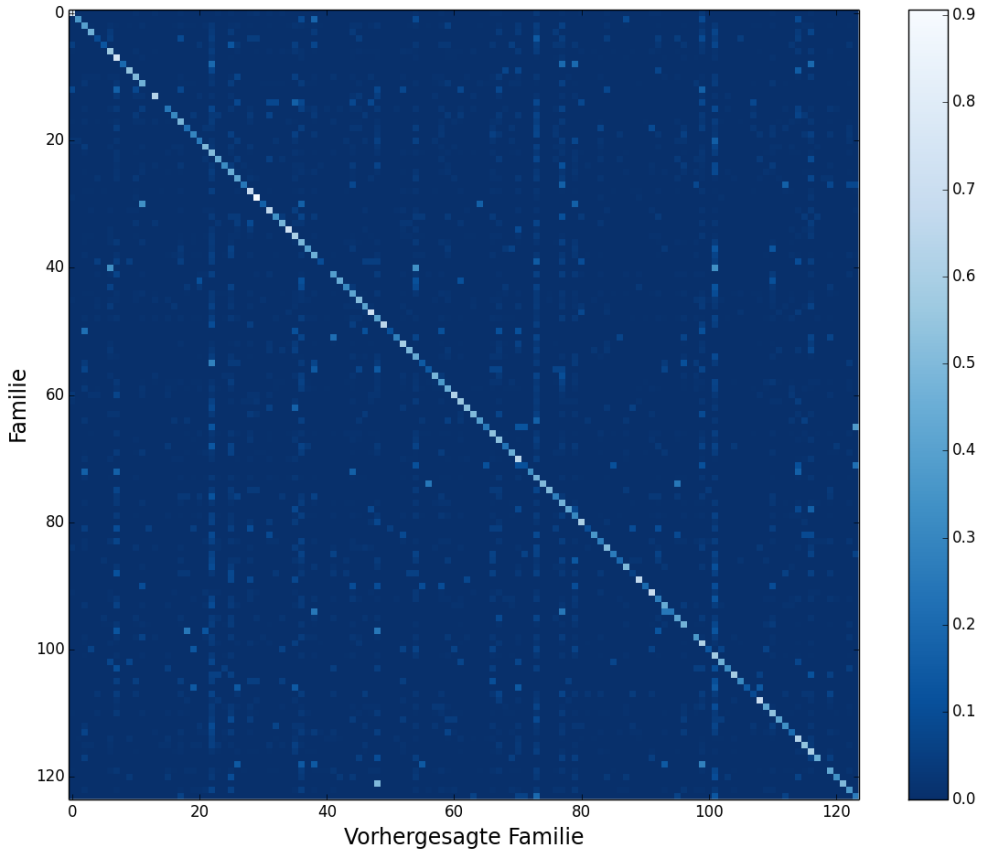


(a) Erweitertes Netzwerk

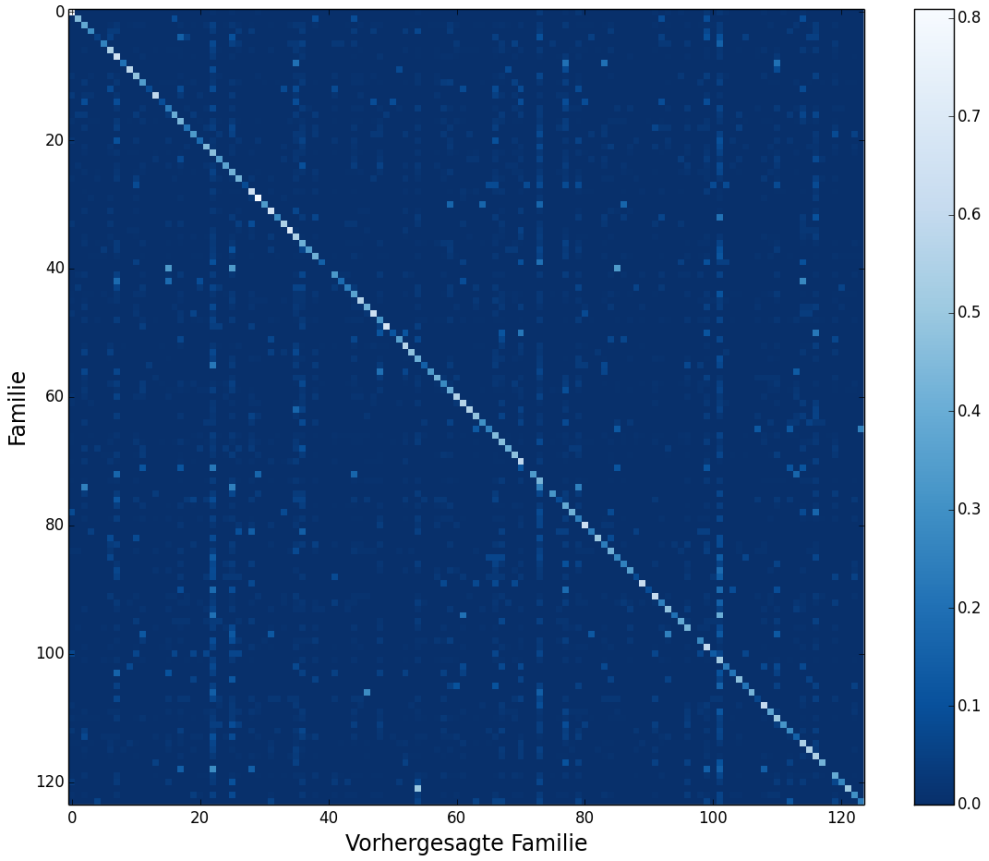
(b) Klassisches Netzwerk

Abbildung 4.7: Cluster der Familie der Glockenblumengewächse für beide Netzwerke.

Im Folgenden werden in Abbildung 4.8 die vollständigen Konfusionsmatrizen bezüglich der Familienerkennung abgebildet. Hierbei handelt es sich um die unveränderten Konfusionsmatrizen ohne Anwendung einer Schwellenfunktion auf die enthaltenen Einträge. Aufgrund der geringeren Anzahl von 124 Familien ist es möglich, die kompletten Matrizen abzubilden.



(a) Erweitertes Netzwerk



(b) Klassisches Netzwerk

Abbildung 4.8: Konfusionsmatrizen beider Netzwerke bezüglich der Familienerkennung.

5 Diskussion

Im Rahmen dieser Arbeit sollte untersucht werden, inwieweit die Verwendung hierarchischer Trainingsinformation bei neuronalen Netzen zu einer besseren Klassifikationsleistung führt. Für das Training eines neuronalen Netzes zur Pflanzenklassifikation wurde neben der üblichen Information der Pflanzenart auch die Information über Familie und Gattung bereitgestellt. Es war von Interesse zu untersuchen, ob sich durch die zusätzliche, hierarchische Trainingsinformation neue Muster im Netzwerk bilden, welche vorteilhaft für die Klassifikationsleistung sind. Im Folgenden werden die Ergebnisse jeder Untersuchung zunächst separat diskutiert, um sie daraufhin gesammelt zu betrachten.

5.1 Separate Betrachtung

Speziesgenauigkeit Der signifikante Unterschied in der Speziesgenauigkeit zwischen dem klassischen und dem erweiterten Netzwerk legt nahe, dass durch die hierarchische Trainingsinformation eine höhere Speziesgenauigkeit erreicht werden konnte. Mit 37.06% beim klassischen Netzwerk und 39.63% beim erweiterten Netzwerk fällt dieser Unterschied jedoch nicht besonders groß aus. Lediglich 2.57 Prozentpunkte trennen die beiden Genauigkeitswerte. Anders ausgedrückt liegt allerdings eine Steigerung von 6.9% bezüglich des Genauigkeitswerts des klassischen Netzwerks vor. Hinsichtlich der *top-5* Speziesgenauigkeit liegt mit 53.18% für das klassische Netzwerk und 57.25% für das erweiterte Netzwerk ein Unterschied von 4.07 Prozentpunkten und eine Zunahme von 7.6% vor. Vergleicht man die am Ende des Trainings erreichten Genauigkeitswerte von ca. 85% (siehe Abbildung 4.1) mit den hier genannten Genauigkeitswerten auf dem Testdatensatz, so fällt auf, dass das Netzwerk trotz der getroffenen Vorkehrungen, wie *dropout* und *weight decay*, weiterhin unter *overfitting* leidet. Es bleibt festzuhalten, dass ein Effekt vorzuliegen scheint, jedoch fällt die tatsächlich gemessene Verbesserung der Klassifikationsleistung eher marginal aus.

Familienaktivierungen Ziel dieser Untersuchungsmethode war es festzustellen, ob das Training mit Familie, Gattung und Art einen Einfluss auf die Aktivierungsverteilung in der Speziesausgabeschicht hat. Die Hypothese war, dass die vielschichtigere, hierarchische Information während des Trainings die Bildung von bestimmten Mustern im Netzwerk nach sich ziehen könnte. Manche dieser Muster könnten als Familienrepräsentationen interpretiert werden, die dafür sorgen, dass die Erkennung einer bestimmten Familie zur Folge hat, dass in der Speziesausgabeschicht vorzugsweise Spezies dieser Familie stark aktiviert werden. Gemessen wurde, zu welchem Anteil die fünf höchstaktivierten Neurone der Speziesausgabeschicht der korrekten Familie angehörten. Der vorgefundene signifikante Unterschied zwischen dem klassischen Netzwerk und dem erweiterten Netzwerk legt nahe, dass das Training mit hierarchischer Information einen Einfluss auf die Aktivierungsverteilung in der Speziesausgabeschicht hat. Mit 22.39% für das klassische Netzwerk und 28.82% für das erweiterte Netzwerk liegt ein Unterschied von 6.43 Prozentpunkten oder eine Zunahme von 28.7% bezüglich des Werts des klassischen Netzwerks

vor. Betrachtet man die erreichten Genauigkeitswerte nicht absolut, sondern bezüglich des berechneten maximal erreichbaren Werts von 61%, so ergibt sich für das klassische Netzwerk ein Wert von 36.7% und für das erweiterte Netzwerk ein Wert von 47.25%. Bezüglich dieses neuen absoluten Werts liegt eine Differenz von 10.55 Prozentpunkten vor. Die Ergebnisse zeigen, dass die Familienaktivierung beim erweiterten Netzwerk signifikant größer ausfällt als beim klassischen Netzwerk. Diese Befunde unterstützen die Hypothese, dass die hierarchische Trainingsinformation dazu führt, dass sich bestimmte interne Muster im Netzwerk ausbilden, welche einen Einfluss auf die Aktiverungsverteilung in der Speziesausgabeschicht haben. Diese Beeinflussung äußert sich darin, dass sich die Aktivierungen in der Speziesausgabeschicht gehäuft auf Spezies derselben Familie konzentrieren. In Übereinstimmung mit den Befunden von Bilal et al. (2018) liegt dieser Effekt jedoch in schwächerer Form bereits beim klassischen Netzwerk vor. Beim erweiterten Netzwerk scheint in diesem Aspekt somit keine Bildung gänzlich neuer Muster vorzuliegen, es findet vielmehr eine graduelle Erweiterung bereits bestehender Muster statt.

Familiengenauigkeit Bei dieser Untersuchung wurde betrachtet, inwieweit sich die Familiengenauigkeit beider Netzwerke unterscheidet. Hierzu wurde zu jeder Speziesvorhersage die Familie ermittelt und mit der tatsächlichen Familie verglichen. Entsprechend der Hypothese erreichte das erweiterte Netzwerk mit 50.04% einen signifkant höheren Genauigkeitswert als das klassische Netzwerk mit 45.42%. Der Unterschied beträgt 4.62 Prozentpunkte, was einer Zunahme von 10.2% entspricht. Der Befund legt nahe, dass sich hierarchische Trainingsinformation positiv auf die Familiengenauigkeit des Netzwerks auswirkt. Weiter sind die Ergebnisse ein Indiz dafür, dass sich durch die hierarchische Trainingsinformation im Netzwerk bestimmte Muster ausbildeten, welche als Familienrepräsentationen gedeutet werden könnten. Wiederum ist festzustellen, dass es sich hierbei lediglich um eine graduelle Verbesserung gegenüber dem klassischen Netzwerk handelt. Aus diesem Grund ist nicht davon auszugehen, dass die benannten Muster erst durch die hierarchische Trainingsinformation entstanden sind, vielmehr dürfte eine Differenzierung von Mustern vorliegen, die bereits im klassischen Netzwerk enthalten sind. Insofern unterstützen diese Befunde auch die Ergebnisse von Bilal et al. (2018).

Reduzierte Trainingsinformation Unter der Annahme, dass die Gattungsinformation für die angeführten Untersuchungen keinen unmittelbaren Informationsgehalt liefert, entstand die Hypothese, dass das alleinige Training mit Familie und Spezies zu besseren Leistungen in den betrachteten Untersuchungen führen könnte. Entgegen dieser Vermutung führte die Vorenthaltung der Gattungsinformation jedoch zu einer numerischen Verschlechterung in allen durchgeführten Untersuchungen verglichen mit dem erweiterten Netzwerk mit der vollen hierarchischen Trainingsinformation. Abgesehen von der Speziesgenauigkeit wurden alle Unterschiede statistisch signifikant. Verglichen mit dem klassischen Netzwerk sind alle erhaltenen Werte wiederum signifikant größer. Wie in Tabelle 4.1 gut ersichtlich, liegen die Werte des Netzwerks ohne Gattungsinformation für alle Untersuchungen zwischen denen des klassischen und des erweiterten Netzwerks. Es scheint somit ein gradueller Leistungsanstieg mit der Fülle an hierarchischer Trainingsinformation einherzugehen. Dies verleitet zur Hypothese, dass die Hinzunahme weiterer taxonomischer Informationen wie der Ordnung oder der Klasse zu einer erneuten Leistungssteigerung führen könnte. Auch wenn die Gattungsinformation für die betrachteten Metriken zuächst nicht unmittelbar relevant zu sein scheint, so hat sie offenbar dennoch positive Einflüsse auf die Bildung

vorteilhafter Muster innerhalb des Netzwerks.

Unbekannte Arten Bei dieser Untersuchung stand die Familien-Generalisierungsfähigkeit des Netzwerks im Vordergrund. Es wurde betrachtet, wie gut die Familien unbekannter Arten durch das Netzwerk erkannt werden. Mit einem signifikant größeren Genauigkeitswert scheint die Familien-Generalisierungsfähigkeit des erweiterten Netzwerks besser zu sein als die des klassischen Netzwerks. Der Wert des erweiterten Netzwerks liegt mit 13.56% nur 1.8 Prozentpunkte über dem des klassischen Netzwerks mit 11.76%. Dies entspricht allerdings einer Zunahme von 15.3% bezüglich des klassischen Netzwerks. Auch wenn der numerische Unterschied zwischen den Werten nicht sehr groß ist, so liegt doch eine nennenswerte prozentuale Zunahme vor. Die hierarchische Trainingsinformation scheint dazu zu führen, dass die Familien-Generalisierungsfähigkeit ansteigt. Wie bisher weist auch hier bereits das klassische Netzwerk diesen Effekt in abgeschwächter Form auf. Durch die hierarchische Trainingsinformation tritt somit nur eine graduelle Verbesserung der Familien-Generalisierungsfähigkeit ein.

Konfusionsmatrix Im Folgenden werden die erhaltenen Konfusionsmatrizen bezüglich der Spezies- und Familienerkennung genauer betrachtet und diskutiert. Zunächst werden die Konfusionsmatrizen bezüglich der Spezieserkennung diskutiert.

Die Konfusionsmatrizen hinsichtlich der Spezieserkennung wurden angefertigt, um die Falschklassifikationen der Netzwerke genauer zu untersuchen. Bei mehr Falschklassifikationen innerhalb der Familien wurde die Bildung quadratischer Familiencluster entlang der Diagonalen der Matrix erwartet. Sollte das erweiterte Netzwerk tatsächlich mehr familieninterne Falschklassifikation erzeugen als das klassische Netzwerk, so müsste sich dies in einer vermehrten Clusterbildung in der Konfusionsmatrix äußern. Sowohl für das klassische als auch für das erweiterte Netzwerk ist die Diagonale der Matrizen recht gut erkennbar. Sie weist allerdings auch einige Lücken auf. Es fällt auf, dass sich die abgebildeten Familiencluster für beide Netzwerke zeigen. Bereits die Konfusionsmatrix des klassischen Netzwerks weist quadratische Anordnungen entlang der Diagonalen auf. Dieser Befund deckt sich mit den Ergebnissen von Bilal et al. (2018). Bei einem Vergleich der Cluster beider Netzwerke kann man bei genauerer Betrachtung dennoch Unterschiede feststellen. Vergleicht man die Familiencluster der Orchideen (siehe Abb. 4.3), so fällt auf, dass das Cluster beim erweiterten Netzwerk breiter und somit quadratischer anmutet als beim klassischen Netzwerk. Das Cluster hebt sich durch klarere Grenzen besser vom Hintergrund ab und ist leichter auszumachen als beim klassischen Netzwerk. Bei einem Vergleich der Cluster der Lippenblütler (Abb. 4.4) ist das Cluster beim erweiterten Netzwerk klarer zu erkennen. Dies beruht auf klareren Kanten an der linken und oberen Kante des Clusters. Außerdem fällt auf, dass jenseits dieser Clustergrenzen weniger Rauschen aufgrund von Falschklassifikationen vorzufinden ist. Das trifft insbesondere auf die linke obere Ecke zu. Bei genauerer Betrachtung fallen vier weitere kleine Cluster auf, welche das Cluster der Lippenblütler einzurahmen scheinen. Hierbei handelt es sich um die Cluster der Buchengewächse und der Birkengewächse, welche untereinander offenbar häufiger falsch klassifiziert werden. Dies ist insofern interessant, als dass es sich bei beiden Arten um Laubbäume handelt. Bezüglich des Ausmaßes dieser vier Cluster scheint jedoch kein bedeutsamer Unterschied zwischen den beiden Netzwerken vorzuliegen. Betrachtet man das Familiencluster der Korbblütler (siehe Abb. 4.5), so ist das Cluster beim erweiterten Netzwerk klarer erkennbar, was vermutlich wieder auf einer prägnant linken und oberen Kante des Clusters beruht. Die Kanten heben sich aber hauptsächlich deshalb vom

Hintergrund ab, weil jenseits der genannten Clustergrenzen (in einem gewissen Bereich) kaum Rauschen vorzufinden ist. Besonders auffällig ist dies an der linken oberen Ecke des Clusters. Bei einem Vergleich der Familiencluster der Storchschnabelgewächse (siehe Abb. 4.6) zeigt sich das Cluster beim erweiterten Netzwerk geschlossener. Beim klassischen Netzwerk besteht das Cluster hauptsächlich aus einem linken und einem oberen Rand, wohingegen beim erweiterten Netzwerk zusätzlich ein unterer Rand vorzufinden ist. Außerdem wirkt das Cluster ausgefüllter, was auf einer erhöhte Anzahl familieninterner Falschklassifikationen hinweist. Bei dem Familiencluster der Glockenblumengewächse (Abbildung 4.7) fällt auf, dass dieses nur für das erweiterte Netzwerk sichtbar ist. Beim klassischen Netzwerk ist kein Ansatz eines entsprechenden Clusters zu erkennen. Scheinbar hat hier erst die hierarchische Trainingsinformation zur Bildung des Familienclusters und damit entsprechend zu einer erhöhten Anzahl familieninterner Falschklassifikationen geführt.

Zusammenfassend lässt sich folgendes zu den Konfusionsmatrizen bezüglich der Spezieserkennung sagen: Bereits beim klassischen Netzwerk liegt eine Clusterbildung bezüglich der Familien weitestgehend vor. Das Ausmaß dieser Clusterbildung ist beim erweiterten Netzwerk jedoch leicht erhöht. Manche Cluster, wie das der Glockenblumengewächse in Abbildung 4.7, sind zudem nur beim erweiterten Netzwerk vorzufinden, was in diesem einen Fall tatsächlich eine prinzipielle Veränderung darstellt. Es muss allerdings berücksichtigt werden, dass es sich hierbei um einen Einzelfall handelt und die graduellen Veränderungen hinsichtlich der anderen Familiencluster überwiegen. Interessant ist, dass die Mehrzahl der Cluster beim erweiterten Netzwerk mit einer klareren linken und oberen Kante und folglich einer besser erkennbaren linken oberen Ecke auffallen. Es scheint somit aufgrund der klarer abgegrenzten Cluster beim erweiterten Netzwerk eine höhere Anzahl an Falschklassifikationen innerhalb der Familien vorzuliegen als beim klassischen Netzwerk. Jedoch sind die Unterschiede in der Clusterbildung zwischen den beiden Netzwerken nur marginal und nicht unmittelbar ersichtlich. Hiermit werden die Befunde bezüglich der Familiengenauigkeit und der Familienaktivierung nochmals unterstrichen. Auch dort konnten lediglich graduelle Verbesserungen für das erweiterte Netzwerk festgestellt werden.

Im Folgenden werden die Konfusionsmatrizen beider Netzwerke bezüglich der Familiengenauigkeit verglichen (siehe Abb. 4.8). Zunächst fällt auf, dass die Diagonale für beide Netzwerke gut erkennbar ist, jedoch sind auch hier Lücken vorhanden. Größere Unterschiede fallen in der Diagonale kaum auf. Allerdings besteht auf Höhe der Familie 75 beim klassischen Netzwerk eine größere Lücke in der Diagonalen, welche beim erweiterten Netzwerk geschlossener erscheint. Dafür zeigt sich beim erweiterten Netzwerk auf Höhe der Familie 15 eine Lücke, welche beim klassischen Netzwerk in dieser Form nicht vorzufinden ist. Es fällt schwer, beide Netzwerke anhand der Ausgeprägtheit ihrer Diagonalen zu vergleichen. Dies deckt sich mit den Befunden zur Familiengenauigkeit beider Netzwerke, die in dieser Hinsicht lediglich eine Diskrepanz von knapp 5 Prozentpunkten aufwiesen – ein Unterschied, der aus dem Ausprägungsgrad der Diagonalen wohl kaum sichtbar hervorgehen dürfte. Bezüglich der Falschklassifikation fallen bei beiden Netzwerken markante vertikale Streifen in der Konfusionsmatrix auf. Diese deuten auf Familien hin, die stattdessen häufig fälschlicherweise klassifiziert werden. Die Streifen in der Matrix entstehen dadurch, dass diese Falschklassifikationen nicht nur eine bestimmte Familie betreffen, sondern bei einer ganzen Reihe von Familien auftreten. Unabhängig von der tatsächlichen Familie scheint es bestimmte Familien zu geben, die häufiger fälschlicherweise klassifiziert werden, man könnte sagen, ihre *false positive* Rate ist erhöht. Vergleicht man die Matrizen beider Netzwerke hinsichtlich der vertikalen Streifen so fällt auf, dass die meisten dieser Streifen bei beiden Netzwerken

gleichermaßen vorzufinden sind. Der eine oder andere Streifen scheint sich in seiner Ausprägung zu unterscheiden, jedoch lässt sich keine Systematik erkennen. Nachfolgend werden jene Familien genauer betrachtet, die eine erhöhte *false positive* Rate aufweisen und somit der Streifenbildung ursächlich zugrunde liegen. Der erste Streifen befindet sich auf Höhe der Familie 7, welche der Familie der Orchideen entspricht. Im Datensatz sind 46 Spezies enthalten, die den Orchideen zugeordnet werden. Der zweite Streifen befindet sich auf Höhe der Familie 22 und entspricht somit der Familie der Hülsenfrüchtler. Der Datensatz weist 72 Spezies auf, die dieser Familie angehören. Der nächste Streifen ist auf Höhe der Familie 25 zu finden, welche die Hahnenfußgewächse kodiert. Diese Familie weist 37 Spezies im Datensatz auf. Die Rosengewächse weisen als Familie 73 ebenfalls eine hohe *false positive* Rate auf und bilden somit einen gut sichtbaren Streifen in der Matrix. Der Datensatz enthält 52 Spezies, die den Rosengewächsen angehören. Ebenso eine hohe *false positive* Rate weist die Familie der Korbblütler mit der Nummer 101 auf. Im Datensatz werden 117 Spezies dieser Familie gelistet. Durchschnittlich enthält eine Familie 8 Spezies, alle genannten Familien sind somit mit überdurchschnittlich vielen Spezies im Datensatz repräsentiert. Die hohen *false positive* Werte mancher Familien scheinen somit nicht auf bestimmten Äußerlichkeiten ihrer Vertreter zu beruhen, sondern sind vermutlich vielmehr durch die Tatsache begründet, dass Vertreter dieser Familien während des Trainings überdurchschnittlich häufig präsentiert wurden.

5.2 Gesammelte Betrachtung

Betrachtet man die Ergebnisse der einzelnen Untersuchungen in ihrer Gesamtheit, so zeigt sich ein einheitliches Befundmuster. Das erweiterte Netzwerk erreicht durchweg bessere Werte als das klassische Netzwerk. Die Unterschiede fallen zwar für jede Untersuchung verschieden stark aus, aber sie sind bei jeder Untersuchung vorzufinden. Das klassische Netzwerk erreicht jedoch bei allen Untersuchungen stets vergleichbare Werte, die nur etwas schlechter sind als die des erweiterten Netzwerks. Es scheint sich somit um graduelle Verbesserungen zu handeln.

Es muss betont werden, dass die hohe statistische Signifikanz mancher Ergebnisse nicht zu voreiligen Schlüssen verleiten sollte. Die vorgefundenen Effekte sind stabil und es scheinen Unterschiede bezüglich der Leistung beider Netzwerke vorzuliegen. Man muss jedoch beachten, dass die Signifikanz der Ergebnisse im Allgemeinen abzugrenzen ist von ihrer Relevanz - somit also dem tatsächlichen Ausmaß der Verbesserung. Bei der verwendeten Teststatistik wirkt sich ein großes n und damit eine große Anzahl an Testbildern positiv auf die Signifikanz aus. Bei den meisten Untersuchungen wurden über 16000 Testbilder verwendet, wodurch selbst kleinere Unterschiede zwischen den Netzwerken statistisch signifikant werden. Hätte man die selbe Evaluation auf 32000 Testbildern durchgeführt, statt auf 16000 und hätten sich ähnliche Ergebnisse ergeben wie zuvor (was anzunehmen ist), so hätten dieselben numerischen Unterschiede zu deutlich signifikanteren Ergebnissen geführt. Die Relevanz der Leistungssteigerung würde man jedoch nach wie vor gleich bewerten.

Es bleibt somit, trotz der statistisch signifikanten Unterschiede, weiterhin die Frage bestehen, inwieweit die vorgefundenen Verbesserungen in der Praxis tatsächlich zu einem relevanten Leistungsunterschied führen. Die Ergebnisse legen zwar gemeinsam nahe, dass durch die Verwendung hierarchischer Trainingsinformation die Leistung neuronaler Netze gesteigert werden kann, diese Leistungssteigerung fällt jedoch mit einer Zunahme um knapp 7% in der Speziesgenauigkeit nicht besonders groß aus. Die

Ergebnisse legen allerdings auch nahe, dass die Genauigkeit des Netzwerks mit der Reichhaltigkeit der hierarchischen Trainingsinformation korreliert. Aus diesem Grund könnte eine Hinzunahme weiterer hierarchischer Informationen die Genauigkeit des Netzwerks möglicherweise erneut steigern. Die eingangs formulierte Hypothese, die Bildung vorteilhafter Muster aufgrund hierarchischer Trainingsinformation könnte mitunter dafür verantwortlich sein, dass die Leistung des Netzwerks zunimmt, scheint sich bestätigt zu haben. Die gebildeten Muster können aufgrund der angeführten Befunde als Familienrepräsentationen gedeutet werden. Das erweiterte Netzwerk scheint somit in der Lage zu sein, die zusätzlich dargebotene Information nutzen zu können, um Muster auszubilden, welche den hierarchischen Charakter der präsentierten Stimuli widerspiegeln. Es bleibt zu betonen, dass die Deutung der Muster nach Fuchs (2017) einem Beobachter mit einem subjektiven Standpunkt vorbehalten bleibt, für das Netzwerk selbst besitzen sie keinerlei Bedeutung. Die entstandenen Muster führen unter anderem dazu, dass das erweiterte Netzwerk eine höhere Familiengenauigkeit erreicht als das klassische Netzwerk, wodurch vermutlich auch die Gesamtleistung des Netzwerks gesteigert werden kann. Jedoch muss festgestellt werden, dass die Familiengenauigkeit, wie alle anderen erhobenen Messwerte, beim klassischen Netzwerk nicht wesentlich geringer ist als beim erweiterten Netzwerk. Die zugrunde liegenden Muster scheinen somit bereits beim klassischen Netzwerk in abgeschwächter Form vorzuliegen. Auch hier handelt es sich demnach um graduelle und nicht um prinzipielle Unterschiede. Die Konfusionsmatrizen konnten die Ergebnisse aus den vorangegangenen Untersuchungen bekräftigen und lieferten einen visuellen Vergleich beider Netzwerke. Die in den Konfusionsmatrizen vorgefundenen Familiencluster legen nahe, dass sich die Falschklassifikationen beim erweiterten Netzwerk in einem höheren Maß auf Spezies innerhalb der richtigen Familie beschränken – somit also schwerwiegendere Falschklassifikationen außerhalb der richtigen Familie seltener vorkommen als beim klassischen Netzwerk. Auch dieses Klassifikationsverhalten dürfte einer Verfeinerung entsprechender hierarchischer Muster im erweiterten Netzwerk geschuldet sein. Allerdings fällt der Unterschied hinsichtlich der Clusterbildung zwischen den Netzwerken gering aus. Die Ergebnisse der Konfusionsmatrizen deuten somit ebenso auf eine graduelle Weiterbildung bereits beim klassischen Netzwerk bestehender Muster hin, was mit den Befunden von Bilal et al. vereinbar ist.

5.3 Fazit und Ausblick

Bereits mit einer flachen Klassenstruktur, welche klassischerweise für das Training neuronaler Netze verwendet wird, ist das betrachtete Netzwerk in der Lage, hierarchische Konzepte zu erfassen und entsprechende Muster auszubilden, ohne dass die hierfür notwendige Information explizit vermittelt wurde. Die Hinzunahme von hierarchischer Trainingsinformation konnte jedoch zu einer graduellen Weiterbildung und Verfeinerung dieser Muster führen. Dies äußerte sich unter anderem in einer leichten Leistungssteigerung in der Spezieserkennung. Die Ausbildung prinzipiell neuer Muster aufgrund der hierarchischen Trainingsinformation scheint größtenteils nicht vorzuliegen.

Es bleibt die Frage bestehen, ob die vorgefundenen graduellen Verbesserungen wirklich als solche zu bezeichnen sind. Hier muss man allerdings abschwächend hinzufügen, dass die absolute Erkennungsleistung des verwendeten Netzwerks schon zu Beginn nicht sonderlich hoch ist. Bei einem Netzwerk mit einer höheren Erkennungsleistung würde eine Leistungszunahme um 7% im Vergleich zu einer größeren absoluten Verbesserung in der Spezieserkennung führen. Eine solche Verbesserung könnte dann mögli-

cherweise tatsächlich als relevante Steigerung in der Erkennungsleistung betrachtet werden.

Die Idee, den Kategorienerwerb beim Menschen genauer zu betrachten und die vorgefundenen Erkenntnisse auf das Training neuronaler Netze zu übertragen, scheint einen fruchtbaren Ansatz darzustellen, der weiter verfolgt werden sollte. Die durchgeführten Untersuchungen legen nahe, dass es – wie von Gopnik (2017) vorgeschlagen – trotz der offenkundigen Differenzen zwischen menschlicher und künstlicher Intelligenz von Vorteil sein kann, Faktoren, welche die Entstehung menschlicher Intelligenz begünstigen, auch bei der Konzeption künstlicher Intelligenzen zu beachten und anzuwenden.

Es wäre interessant in einer weiteren Studie den von Deng et al. (2010) verfolgten Ansatz der semantischen Nähe auf das betrachtete Problem der Pflanzenklassifikation anzuwenden. Falschklassifikationen der Spezies innerhalb der richtigen Gattung oder Familie wären dann weniger schwerwiegend als Falschklassifikationen der Spezies außerhalb der richtigen Familie. Die Hauptaufgabe bestünde darin, eine Fehlerfunktion zu implementieren, die das Prinzip der semantischen Nähe berücksichtigt. Im betrachteten Fall der Pflanzenklassifikation wäre die semantische Nähe gewissermaßen durch die taxonomische Struktur bereits vorgegeben. Ein Training mit einer solchen Fehlerfunktion könnte zu deutlich robusteren Spezieskassifikationen führen. Eventuell könnte es auch sinnvoll sein, diese Idee mit dem in dieser Arbeit verfolgten Ansatz zu verknüpfen. Somit hätte man eine Klassifikation bezüglich aller drei taxonomischen Stufen und simultan für die Speziesausgabeschicht eine Fehlerfunktion, welche die semantische bzw. taxonomische Nähe der Spezies berücksichtigt. Möglicherweise führt eine solche Kombination beider Ansätze zu einer deutlichen Steigerung der Klassifikationsleistung.

Literaturverzeichnis

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Zugriff auf `https://www.tensorflow.org/` (Software available from tensorflow.org)

Besold, T. R. & Kühnberger, K.-U. (2013). Konnektionismus, Neuronale Netze und *parallel distributed processing*. In A. Stephan & S. Walter (Hrsg.), *Handbuch Kognitionswissenschaft.* J.B.Metzler.

Bilal, A., Jourabloo, A., Ye, M., Liu, X. & Ren, L. (2018). Do convolutional neural networks learn class hierarchy? *IEEE transactions on visualization and computer graphics*, *24* (1), 152–162.

Cireşan, D., Meier, U., Masci, J. & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural networks*, *32*, 333–338.

Deng, J., Berg, A. C., Li, K. & Fei-Fei, L. (2010). What does classifying more than 10,000 image categories tell us? In *European conference on computer vision* (S. 71–84).

Fisz, M. (1976). *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Deutscher Verl. der Wissenschaften.

Fogel, D. B. (2006). *Evolutionary computation: toward a new philosophy of machine intelligence*. John Wiley & Sons.

Fuchs, T. (2017). *Das Gehirn - ein Beziehungsorgan: eine phänomenologisch-ökologische Konzeption* (5. Aufl.). W. Kohlhammer Verlag.

Goëau, H., Joly, A. & Pierre, B. (2015). Lifeclef plant identification task 2015. *CLEF working notes*.

Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. (2016). *Deep learning*. MIT press Cambridge.

Gopnik, A. (2017). Making AI more human. *Scientific American*, *316* (6), 60–65.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... Cai, J. (2017). Recent advances in convolutional neural networks. *Pattern Recognition*.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160* (1), 106–154.

Jaekel, F. & Meyer, U. (2013). Kategorisierung und Begriffe. In A. Stephan & S. Walter (Hrsg.), *Handbuch Kognitionswissenschaft.* J.B.Metzler.

Kheradpisheh, S. R., Ghodrati, M., Ganjtabesh, M. & Masquelier, T. (2016). Deep networks can resemble human feed-forward vision in invariant object recognition. *Scientific reports*, *6*, 32672.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (S. 1097–1105).

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.

Mallot, H. A. (2013). *Computational neuroscience: a first course*. Springer.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.

Pauen, S. (2002). The global-to-basic level shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioral Development*, *26* (6), 492–499.

Peirce, C. S. (1998). *The Essential Peirce: Selected Philosophical Writings (1893-1913)* (Bd. 2). Indiana University Press.

Pfeifer, R. & Scheier, C. (2001). *Understanding Intelligence*. MIT press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8* (3), 382–439.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65* (6), 386.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Bd. 1, S. 318–362).

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... Lanctot, M. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529* (7587), 484–489.

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.

Tafreschi, A. (2006). *Zur Benennung und Kategorisierung alltäglicher Gegenstände: Onomasiologie, Semasiologie und kognitive Semantik*. Kassel university press GmbH.

Werbos, P. J. (1988). Backpropagation: Past and future. *Proceedings of the International Conference on Neural Networks*, *1*, 343–353.

Zell, A. (1994). *Simulation Neuronaler Netze*. Addison-Wesley Bonn.