

ShroomNET

Neuronale Netze spezialisiert auf die Bestimmung von Pilzarten

Maturarbeit von Jan Obermeier
G15E | NKSA | 2018

Betreuende Lehrperson: Dr. Nicolas Ruh

Abstrakt

Diese Maturarbeit beschäftigt sich mit der Entwicklung und Optimierung eines Neuronalen Netzes für die Bestimmung von Pilzarten. Ziel der Arbeit ist es, die Umsetzbarkeit eines solchen Neuronalen Netzes zu prüfen. Zur Vereinfachung beschränkt sich die Erkennung vorerst auf die 20 häufigsten Pilzarten der Nordwestschweiz.

Im Theorienteil der Dokumentation soll ein grundlegendes Verständnis für *Neuronale Netze* und *Deep Learning* vermittelt werden, um die darauf folgende Dokumentation der Umsetzung nachvollziehbar gestalten zu können.

Bei der Umsetzung des Algorithmus werden verschiedene Techniken und Vorgehensweisen für den Aufbau und das Training des Neuronalen Netzes in Betracht gezogen, um eine möglichst akkurate Bestimmung der Pilzart ermöglichen zu können. Dabei werden verschiedene Netzwerkarchitekturen wie auch Datenaufbereitungsmethoden gegeneinander abgewägt und auf deren spezifische Vor- und Nachteile untersucht.

In einer Zusammenfassung soll schliesslich über die Anwendbarkeit, Zuverlässigkeit und Erweiterbarkeit des Algorithmus diskutiert werden.

Vorwort

Für *künstliche neuronale Netze* hatte ich schon immer eine gewisse Faszination. Ein im Grunde genommen statischer Code modelliert organisches Verhalten: Er kann lernen und sich dadurch verbessern. Und wie wir auch nicht genau wissen, wie unser Gehirn funktioniert, so wissen wir es auch nicht von den *neuronalen Netzen* — sie *sind* gewissermassen ein "elektronisches Hirn". Weiss man aber dieses "elektronische Hirn" zu belehren und einzusetzen, so eröffnen sich grenzenlose Möglichkeiten.

Codezeilen haben mich schon lange begleitet; die ersten Schritte machte ich mit meinem Vater: Variablen definieren, Werte daraus berechnen, erste Schleifen. Mit der Mittelschule und dem *infcom*-Kurs kamen Objekte, die Steuerung von grafischen Elementen und Game-Loops hinzu. Nebenbei hielt ich mich mit der aktuellen Technik auf dem Laufenden, welche aber zu viel mehr imstande war: Bilder und Sprache erkennen, später lernten sie auch, den Menschen in den komplexesten Brettspielen[1] und Computerspielen[2] zu schlagen. Weiter erschienen Videoaufnahmen von absurd Reden bekannter Staatsoberhäuptern, bei denen das Auge nicht mehr von echt oder gefälscht unterscheiden konnte [3]. Diese und weitere Möglichkeiten, aber auch die Genialität dieser Algorithmen haben mein Interesse an den *künstlichen neuronalen Netzen* geweckt, jedoch fehlten mir damals die Programmiergrundlagen, um ein solches Projekt umzusetzen. Mit der ersten grösseren Arbeit "EvoSim"[4] habe ich meine bisherigen Programmierkenntnisse festigen und erweitern können. Zudem habe ich mich in das Themengebiet der sogenannten *genetischen Algorithmen* begeben; eine rein durch Zufall und Selektion vorangetriebene Methode des *maschinellen Lernens*. Mit der Maturarbeit will ich somit einen Schritt weiter gehen und den Einstieg in die komplexere Welt der *künstlichen neuronalen Netze*, den "elektronischen Hirnen", machen.

Die Anwendung von *künstlichen neuronalen Netzen* kann sehr vielfältig sein; enorm stark vertreten sind sie im Bereich der Bilderkennung. Da viele bekannte Problemstellungen wie Gesichts- und Handschrifterkennung schon zu genüge behandelt worden sind, habe ich mich auf den Impuls von Dr. Nicolas Ruh auf ein für mich eher exotischeres Anwendungsgebiet eingelassen: die Erkennung von Pilzarten. Es ist bekannt, dass die zuverlässige Bestimmung von Pilzen langjährige Expertise voraussetzt, da sich Arten zum Teil nur anhand von wenigen Details auseinanderhalten lassen. Liesse sich dieser Prozess zu einem gewissen Teil von *neuronalen Netzen* übernehmen, so könnte man diese "Expertise" für jedermann mit einem Mobiltelefon in der Tasche zugänglich und nutzbar machen.

Daraus entstand der konkrete Plan, einen auf *künstlichen neuronalen Netzen* basierenden Algorithmus zu entwickeln, welcher für die Pilzartenbestimmung ausgelegt ist — daher auch der Name *ShroomNET*. Der Erkennungsalgorithmus soll der zentrale Punkt dieser Arbeit sein, wobei der Fokus auf die Entwicklung und Optimierung gelegt wird.

Während den Recherchen für diese Arbeit fanden wir ein Team von Studenten der Universität Helsinki, welches für ihr Projekt namens "Deep Shrooms"[5] ein ähnliches Konzept hatten. Jedoch war das Projekt leider erfolglos, weswegen ich in dieser Arbeit die selbe Hürde umso mehr zu nehmen versuchen will.

Inhaltsverzeichnis

1 Einführung	6
1.1 Zielsetzung	7
1.1.1 Datenbeschaffung & Datenaufbereitung	7
1.1.2 Entwicklung & Optimierung	7
1.1.3 Theorie	7
2 Grundlagen <i>KNNs & DNNs</i>	8
2.1 <i>KI, KNN, DL, usw.</i>	8
2.2 Einleitung	10
2.3 Maschinelles Lernen	11
2.3.1 Training	11
2.3.2 Verzerrung-Varianz-Dilemma	11
2.4 KNNs	12
2.4.1 Die Natur als Vorbild	12
2.4.2 Neuronen & Verbindungen	13
2.4.3 Layers	15
2.4.4 Leistung von KNNs messen	16
2.4.5 Training von KNNs	17
2.5 DNNs	18
2.6 Convolutional Neural Networks	18
2.6.1 Convolution Layer	19
2.6.2 Shared Weights & Feature Maps	20
2.6.3 Pooling	20
2.6.4 Aufbau eines CNNs	21
2.7 Zusätzliche Meta-Parameter und Modifikation	21
2.7.1 Bild-Normalisierung	21
2.7.2 ReLU-Layer & Batch-Normalization-Layer	21
2.7.3 Data Augmentation	22
2.7.4 Transfer Learning	22
3 Methode	23
3.1 Datenbeschaffung	23
3.1.1 Datenbank SwissFungi	23
3.1.2 Webseite ShroomNET	23
3.1.3 Internet Crawler	24
3.1.4 Videos	24
3.2 Datenaufbereitung	25
3.2.1 Bildbearbeitung	25
3.2.2 Aufteilung in Trainings- und Validierungsdaten	26
3.2.3 Bildnormalisierung	26
3.2.4 Farbfilter	27
3.2.5 Data Augmentation	27
3.2.6 Zusatzinformationen sammeln & aufbereiten	28
3.3 Entwicklungsprozess	29
3.3.1 Bezugswert	29

3.3.2	Basis-SNN	29
3.3.3	Basis-CNN	30
3.3.4	Data-Augmentation	30
3.3.5	Farbfilter	31
3.3.6	Justierung des CNNs	31
3.3.7	Transfer-Learning	32
3.3.8	Einspeisung von Zusatzinformationen	33
4	Darstellung der Ergebnisse	35
4.1	CNN ohne <i>Transfer-Learning</i>	35
4.2	CNN mit <i>Transfer-Learning</i>	37
4.3	CNN mit <i>Transfer-Learning</i> und Zusatzinformationen	39
5	Diskussion der Ergebnisse	43
5.1	Qualität der Trainingsdaten	43
5.2	Verbesserungs- und Erweiterungsvorschläge	43
6	Zusammenfassung	44
7	Danksagung	45
	Literatur	46
8	Anhang	49

1 Einführung

Wer behaupten will, er habe noch nie von *Deep Learning* gehört, muss gewissermassen unter einem Stein gelebt haben. Durch *Deep Learning* haben die *künstlichen Intelligenzen*¹ in den letzten Jahren ein Comeback erlebt und haben es auch schon einige Male in die Schlagzeilen geschafft:

”Googles AlphaGo KI¹ schlägt
Go-Weltmeister Lee Sedol 4-1” [6]

The Guardian, 2016

Das chinesische Brettspiel *Go* basiert auf ganz simplen Regeln: Schwarze bzw. weisse Spielsteine werden abwechslungsweise auf ein 19x19 Felder grosses Spielfeld gesetzt. Das Ziel des Spiels ist es, den Gegner einzukreisen und zu erobern [7]. Trotz den simplen Regeln gibt es fast unzählige Möglichkeiten: Insgesamt lassen sich die Steine in rund 10^{170} verschiedenen Arten anordnen. Ein kurzer Magnitudenvergleich: Schach hat etwa 10^{43} verschiedene Anordnungsmöglichkeiten [8], das beobachtbare Universum enthält schätzungsweise 10^{80} Atome [9].

Diese schiere Anzahl an Möglichkeiten machten *Go* bei KI-Programmierern sehr hoch angesehen, denn im Gegensatz zu Schach können nicht alle relevanten Züge in absehbarer Zeit vorausberechnet werden². Trainierte Go-Spieler können das auch nicht, dafür verlassen sie sich auf ihre Erfahrung und auf ihre Intuition, etwas, was Computern nur schwierig beizubringen ist. Und doch passierte es im März 2016: Der Go-Weltmeister Lee Sedol wurde geschlagen — von der *KI AlphaGo* [1].

AlphaGo unterscheidet sich grundlegend von konventionellen Spiele-KIs, weswegen es auch in *Go* gegen Lee Sedol antreten konnte. *AlphaGo* basiert auf dem Algorithmus namens *Deep Learning* (kurz *DL*). Diese Algorithmen sind in der Lage, aus Beispielen Muster zu erkennen, zu lernen und sich selber dadurch zu verbessern. Durch die Analyse von zahlreichen von Menschen gespielten Partien lernte *AlphaGo* die Grundlagen und Grundstrategien des Spiels. Daraufhin liess man den Algorithmus einige Tausend Male gegen sich selber spielen, wodurch sich *AlphaGo* sukzessive verbesserte und fortgeschrittenere Strategien entwickelte, sodass der Algorithmus schliesslich in der Lage war, gegen die besten Go-Spieler der Welt anzutreten.

Deep Learning (*DL*) gehört dem Überbegriff ”*künstliche neuronale Netze*“ (kurz *KNNs*) an. *KNNs* sind im Grunde genommen der Versuch der Informatik, den Aufbau und die Funktionsweise eines biologischen Gehirns zu modellieren und nachzuahmen, wobei *DL* eine Weiterentwicklung dieses Grundprinzips ist. Wie ein biologisches Gehirn muss auch ein *KNN* lernen, man muss es *trainieren*. Mit dieser Lernfähigkeit werden die Anwendungsmöglichkeiten enorm vielseitig und komplex: In Spiele-KIs, Spam-Filtern und der Bildbearbeitung, aber auch in der Gesichtserkennung und der Krebsdiagnose finden sie heutzutage Verwendung.

¹Künstliche Intelligenz: Modellierung/Nachahmung von ”intelligentem“ Verhalten durch Rechner, z.B. Schachcomputer, Übersetzungs-Tools oder Text-/Spracherkennung

²vgl. *Deep Blue*: Erster Schachcomputer (entwickelt von IBM), welcher 1997 den damaligen Weltmeister G. Kasparov geschlagen hat. *Deep Blue* verwendete eine *Brute-Force-Methode* (Ausprobieren aller Möglichkeiten), weswegen ein Entwickler sogar die ”Intelligenz“ hinter der KI *Deep Blue* bestritt[10].

1.1 Zielsetzung

Diese Arbeit behandelt die Entwicklung eines auf *KNNs* basierenden Algorithmus, welcher auf die Bestimmung von Pilzarten optimiert ist. Das Ziel der Arbeit ist es, durch Abwägung verschiedener Techniken und Vorgehensweisen einen möglichst zuverlässigen Bestimmungsalgorithmus für Pilzarten zu entwickeln. Als Grundlage für die Bestimmung dienen Fotografien von den zu bestimmenden Pilzen, sekundär können zusätzliche Eigenschaften wie Geruch und Standort des Pilzes angegeben werden, welche mit in die Bestimmung einfließen.

Da es sich im Rahmen dieser Arbeit hauptsächlich um die Evaluierung der Umsetzbarkeit geht, wird die Anzahl der bestimmbaren Pilzarten auf die 20 häufigsten Sorten der Nordwestschweiz beschränkt (siehe Tabelle 7). Neben den 20 ausgewählten Arten soll eine weitere Kategorie erstellt werden, welche nicht ausgewählte (*unbekannte*) Pilzarten beinhaltet. Dadurch soll gewährleistet werden, dass der Algorithmus nicht durch das Ausschlussverfahren Pilzarten bestimmen kann. Zwar wird sich durch diese Massnahme das Finden eines leistungsfähigen Algorithmus massgeblich erschweren, jedoch ist sie notwendig, um Schlüsse auf die praktische Anwendbarkeit ziehen zu können.

1.1.1 Datenbeschaffung & Datenaufbereitung

Ein wichtiger Teil für *KNNs* ist die Beschaffung von vielen Trainingsdaten. Aus den Kapiteln *Datenbeschaffung & Datenaufbereitung* geht hervor, wie die Trainingsdaten beschaffen worden sind und bearbeitet worden sind, um sich als Eingangsdaten für ein *KNN* zu eignen. Die für ergriffenen Massnahmen werden nachvollziehbar begründet und veranschaulicht.

1.1.2 Entwicklung & Optimierung

Bei der Umsetzung sollen verschiedene Methoden sowie Vorgehensweisen für den Aufbau eines *KNNs* genauer untersucht werden, um eine möglichst akkurate Bestimmung der Pilzart zu ermöglichen. Es werden verschiedene Netzwerkarchitekturen, aber auch andere Parameter wie Netzwerkgröße und verschiedene Trainingsdaten auf Stärken und Schwächen untersucht, um eine möglichst optimale Kombination für die Erkennung von Pilzarten evaluieren zu können. Im Kapitel 3 *Methode* wird nachvollziehbar auf die Verfahren und Massnahmen eingegangen, welche zu einer wesentlichen Verbesserung des Ergebnisses beigetragen haben.

1.1.3 Theorie

Für das erleichterte Verständnis der Dokumentation soll das Kapitel *Grundlagen KNNs & DNNs* einen Einblick in die Entwicklung der verwendeten Algorithmen geben sowie die grundlegende Funktionsweise derselben beschreiben. Für das Verständnis des Theorieteils wird ein solides technisches und mathematisches Grundwissen vorausgesetzt.

2 Grundlagen *KNNs & DNNs*

In diesem Kapitel werden die theoretischen Grundlagen für die in dieser Arbeit verwendeten Algorithmen behandelt. In einer Einleitung werden die Begrifflichkeiten erläutert sowie die Motivation wie auch Schwierigkeiten für die Entwicklung von *maschinellen Lernalgorithmen (ML)* geschildert. Der Hauptteil behandelt den technischen Aspekt von *KNNs* und *DNNs*, worin das Funktionsprinzip, verschiedene Architekturen als auch Umsetzungstechniken vorgestellt werden.

Es ist anzumerken, dass es sich bei *maschinellem Lernen* und *künstlichen neuronalen Netzen* um ein sehr komplexes Themengebiet handelt. Für das Verständnis des Theorieteils wird somit ein solides Grundwissen für technische als auch mathematische Konzepte vorausgesetzt, um die relevanten Informationen in einem angemessenen Umfang behandeln zu können.

2.1 *KI, KNN, DL, usw.*

In den Medien wie auch im Internet hört und liest man die Ausdrücke und Abkürzungen immer wieder: *KIs*, *Deep Learning* und *neuronale Netze*, *AIs* und *DNNs*. Man könnte meinen, sie seien ähnlich, jedoch bedeuten sie nicht alle dasselbe. Folgender Überblick soll helfen, die verwirrenden Bezeichnungen und Fachausdrücke korrekt zu interpretieren. Der Inhalt greift dabei etwas vor, um späteres Nachschlagen zu ermöglichen.

KI: Künstliche Intelligenz (*engl. artificial intelligence, AI*):

Überbegriff in der Informatik für automatisierte Prozesse, welche "intelligentes" Verhalten modellieren/nachahmen. Mangels einer klaren Definition von "Intelligenz" ist der Begriff sehr weit anwendbar[11], d.h. einen Roboter auf zwei Beinen gehen zu lassen gehört genau so zur *KI* wie die Erkennung von Gesichtern.

ML: Maschinelles Lernen (*engl. machine learning*):

Teilbereich der *künstlichen Intelligenzen*, welcher sich mit künstlichen Lernprozessen befasst. Konventionellen Programmen gibt man einen Satz an Regeln mit, um z.B. Muster erkennen zu können. Beim *maschinellem Lernen* ist es hingegen das Ziel, den Algorithmus durch Training die Regeln und Gesetzmäßigkeiten selber erkennen und bestimmen zu lassen. Für den Trainingsprozess werden meistens bezeichnete Trainingsdaten verwendet[12].

KNN/NN: Künstliches neuronales Netz (*engl. artificial neural network, ANN*):

Programmtechnik aus dem Gebiet des *maschinellen Lernens*, welche inspiriert biologischen neuronalen Netzen (Gehirnen) Neuronen und deren Interaktionen zu modellieren versucht. Die *Neuronen* sind in Schichten angeordnet (sog. *Layers*) und untereinander verbunden. *KNNs* sind in der Lage aus Trainingsdaten zu lernen[13]. Im Kontext der Informatik wird zwischen *künstlichen neuronalen Netzen* und *neuronalen Netzen (NNs)* nicht differenziert.

SNN: Shallow Neural Network:

Eher selten gebrauchter Begriff, *KNN* impliziert z.T. ein *SNN*³. "Shallow" weist dabei auf die geringe Anzahl der *Layers* hin (i.d.R. 1, höchstens 2 *Hidden Layers*). Wegen den wenigen *Layers* können *SNNs* nur einfachere Probleme lösen[13].

³Um Missverständnisse zu meiden, werden in dieser Arbeit die Begriffe *KNN* und *NN* als **Überbegriff** von *SNNs* und *DNNs* verwendet.

DL/DNN: Deep Learning/Deep Neural Network: Im Gegensatz zu *SNNs* bezeichnet man mit *Deep Learning* Netzarchitekturen, welche 2 oder mehr *Hidden Layers* besitzen. Mit speziellen *Layers* können dem *KNN* andere Eigenschaften oder Charakteristiken geben werden, so z.B. verbesserte Verallgemeinerung oder ein "Kurzzeitgedächtnis". *Deep Neural Network* weist spezifisch auf die Umsetzung eines *DL*-Algorithmus mit *NNs* hin, jedoch werden die beiden Begriffe meist synonym verwendet[14].

CNN: Convolutional Neural Network:

Spezielle Art eines *Deep Neural Networks*, welche sich v.a. in der Bilderkennung bewährt. Mit verschiedenen *Layers* werden die relevanten Informationen gefiltert und auf immer kleinere *Layers* konzentriert. Diese Massnahme ermöglicht sehr komplexe und leistungsfähige Netze in absehbarer Zeit zu trainieren.

2.2 Einleitung



Abbildung 1: Finger

Sie haben wahrscheinlich ohne zu zögern das Bild gesehen und sofort erkannt, dass darauf eine Hand mit 3 ausgestreckten Fingern gezeigt wird. Es scheint auf den ersten Blick ein simpler Prozess zu sein. So simpel, dass sogar Kleinkinder dazu fähig sind. Mit dieser Vorstellung waren ersten KI-Forscher der 1960er der Meinung, sie könnten Rechnern die visuellen Interpretation von Bildern wie auch andere menschliche Fähigkeiten beibringen — in der Zeitspanne von nur zwanzig Jahren. Wie es sich herausstellte, haben sie die Hürde massiv unterschätzt. Erst heute mit der steigenden Rechenleistung und neuen Entwicklungen beginnen die Computer langsam die Versprechen der 60er zu erfüllen. In anderen Worten: Ihr visueller Kortex hat für die Erkennung dieser Finger soeben das geleistet, wofür Rechner Jahrzehnte an Forschung und Entwicklung benötigt haben.

”Maschinen werden in 20 Jahren zu jeder Arbeit
fähig sein, die ein Mensch verrichten kann.” [15]

*Herbert A. Simon
KI-Pionier, 1965*

Um die Schwierigkeiten der Bilderkennung zu verstehen, muss das Problem aus der sich von Rechnern betrachtet werden. Was wir als ein Bild von einer Hand erkennen sind für einen Computer Zahlen. Jeder Bildpunkt (*Pixel*) besteht aus drei Intensitätswerten zwischen 0 und 255 für die Grundfarben rot, grün und blau; im Raster angeordnet ergeben sie das Bild. Die Frage besteht in diesem Beispiel darin, ob eine mathematische Funktion⁴ für diese Pixelwerte gefunden werden kann, welche etwas über die Anzahl der gezeigten Finger aussagen kann. In Worten formuliert sähe die Funktion wie folgt aus: ”Ein ausgestreckter Finger ist leicht rötlich, länglich und von der Hand gespreizt. Diese gilt es zu zählen”. Will man nun aber einen Algorithmus designen, welche diese Eigenschaften erkennen kann, so verliert man sich schnell in unzähligen Regeln und komplexen Ausnahmen. Es sei denn, man lässt den Algorithmus anhand von Beispieldaten die Muster und Regeln selber finden:

⁴Laut Alan Turing kann jede mathematische Funktion als Computerprogramm umgesetzt werden[16], weswegen das Finden dieser Funktion das eigentliche Ziel ist.

2.3 Maschinelles Lernen

Maschinelle Lernalgorithmen sind auf den künstlichen Wissenserwerb spezialisiert. Dabei ist es wichtig, nicht durch "Auswendiglernen", sondern durch Erkennen und Verallgemeinern von Gesetzmäßigkeiten Daten zu verarbeiten. Es sind viele verschiedene Umsetzungsansätze für lernfähige Algorithmen bekannt; so z.B. *genetische Algorithmen*, *Entscheidungsbäume* und *künstliche neuronale Netze*.

2.3.1 Training

Um "Wissen zu erwerben" werden maschinelle Lernalgorithmen *trainiert*. Abstrakt kann man das Ziel vom Training folgendermassen formulieren: Eine unbekannte Funktion $f(x)$ kann fehlerlos jedem x das korrekte y zuordnen, seien x und y Zahlenwerte oder Bilder (x) und die dazugehörigen Kategorien (y). Für diese unbekannte Funktion $f(x)$ gilt es nun im Training, mithilfe von wenigen bekannten Datenpunkten (x, y) (den *Trainingsdaten*) eine möglichst genaue Modellierung von $f(x)$ zu finden.

In der Praxis sieht der *Trainingsprozess* generell folgendermassen aus: Ein Datenpunkt x aus den *Trainingsdaten* wird in den zu trainierenden Algorithmus $F(x)$ eingespeist, worauf das erhaltene Ergebnis mit dem erwünschten Ergebnis y verglichen wird. Durch Anpassung der *Parameter* des Algorithmus wird versucht, das Resultat zu verbessern. Dieser Prozess wird mit allen Datenpunkten der Trainingsdaten wiederholt, wodurch der Algorithmus eine möglichst genaue Repräsentation für $f(x)$ finden soll: $F(x) \approx f(x)$.

2.3.2 Verzerrung-Varianz-Dilemma

Das Verzerrung-Varianz-Dilemma[17] ist eine grundlegende Schwierigkeit von Lernalgorithmen. Im Grunde genommen will man, dass der Algorithmus mit den *Trainingsdaten* als auch mit komplett neuen Daten akkurate Ergebnisse liefern kann. Dazu muss man beim Designen eines maschinellen Lernalgorithmus die richtige Balance zwischen Genauigkeit (*Varianz*) und Toleranz (*Verzerrung*) finden. Ist das Gleichgewicht jedoch nicht bewerkstelligt, so äussert sich das mit den Symptomen der *Überanpassung* oder *Unteranpassung*.

Überanpassung Erhält man ein verdächtig akkurates Ergebnis mit den Trainingsdaten, läuft man Gefahr der *Überanpassung*: Der Algorithmus erstellt zu detaillierte Regeln für die Trainingsdaten spezifisch und verallgemeinert schlecht (er "lernt auswendig"). *Überanpassung* führt dazu, dass der Algorithmus zwar ausserordentliche Leistungen innerhalb der Trainingsdaten erzielt, jedoch ungelernte Daten nur mit einer sehr grossen Fehlerquote verarbeiten kann.

Um eine Überanpassung des Algorithmus zu vermeiden, verwendet man bei maschinellen Lernalgorithmen zwei sehr menschliche Eigenschaften:

- Nicht perfekt sein
- Annahme, dass manche Details ignoriert werden können

Diese Eigenschaften erlauben es dem Algorithmus durch Weglassen von Details das Rauschen in den Trainingsdaten zu glätten. Es werden mehr Informationen ignoriert, wodurch die Funktion weniger komplex wird. Die Reduktion der Komplexität führt aus rein probabilistischen Gründen dazu, dass der Algorithmus toleranter wird und somit verallgemeinert wird[18].

Unteranpassung Wenn man nun aber zu viele Annahmen und Fehler zulässt, unterliegt man der Gefahr der *Unteranpassung*, bei der die Leistung des Algorithmus über alle gelernten sowie ungelernten Daten konsistent schlecht ist: Der Algorithmus verallgemeinert zwar sehr gut, aber macht zu viele/falsche Annahmen und ignoriert für die Bestimmung relevante Details. Unteranpassung tritt dann auf, wenn der Algorithmus zu simpel ist oder unpassend für die Problemstellung designet worden ist.

Validierung Um die Balance zwischen Genauigkeit und Toleranz zu gewährleisten, werden während dem Training regelmässig Validierungsdaten eingespeist, auf welche der Algorithmus **nicht** trainiert ist. Mit den durch die Validierungsdaten gelieferten Ergebnissen kann auf die Leistung des Algorithmus ausserhalb der Trainingsumgebung geschlossen werden. Sobald die Leistung mit den Validierungsdaten sinkt während die der Trainingsdaten weiterhin steigt, kann man davon ausgehen, dass der Algorithmus ein Optimum erreicht hat und dass weiteres Trainieren zu einer Überanpassung führen wird. Wenn der trainierte Algorithmus in seinem Optimum trotzdem keine zufriedenstellenden Ergebnisse liefert, ist von einer *Unteranpassung* auszugehen. Durch Ändern der Trainingsparameter⁵ oder durch Umgestaltung des Algorithmus kann der *Unteranpassung* wie auch der *Überanpassung* entgegengewirkt werden.

2.4 Künstliche neuronale Netze

Obwohl die *KNNs* erst in den letzten Jahren ihren grossen Durchbruch erlebt hatten, lässt sich die ursprüngliche Idee von *künstlichen neuronalen Netzen* bis in die 40er Jahre zurückdatieren. In den 50er Jahren wurden die ersten Versuche mit Rechnern der damaligen Zeit durchgeführt, wobei mit dem Vorgänger des *Neurons*, dem *Perceptron*, nur müsiger Fortschritt erzielt worden ist. Erst mit der Entwicklung der sehr effizienten *Gradient Descent-Trainingsmethode* und der Einführung der heute gängigen *Neuronen* kam die Forschung an *KNNs* in den 70er Jahren wieder in Fahrt. Der richtige Durchbruch der *KNNs* kam aber erst in den letzten beiden Jahrzehnten mit der rapiden Steigerung der Prozessorleistung, aber auch dem Aufkommen des Internets und der schieren Menge an gespeicherten und zugänglichen Daten. Das ermöglichte mit neuen Entwicklungen im Bereich des *Deep Learnings* höchst komplexe Systeme. So waren solche *DNNs* in der Lage, Aufgaben zu meistern, welche für Algorithmen lange als sehr schwierig oder unmöglich erachtet worden sind. Mit neusten Erfolgen in der Bilderkennung, aber auch der Entscheidungsfindung (siehe Kapitel 1 *AlphaGo*) und vielen praktischen Anwendungen hat *Deep Learning* öffentliche Aufmerksamkeit erhalten[13].

Die technischen Informationen zu *KNNs* und *DNNs* in den folgenden Abschnitten sowie im Kapitel 2.5 basieren auf dem öffentlich verfügbaren Lernmittel *Neural Networks and Deep Learning* von Michael Nielsen[19].

2.4.1 Die Natur als Vorbild

Um die in Kapitel 2.3 erwähnten Lernverhalten simulieren zu können, bedienen sich die *KNNs* dem Vorbild der Natur: das Gehirn. Der Aufbau des Gehirns lässt sich in wenigen Worten folgenderma-

⁵Nicht zu verwechseln mit den *Parametern* des Algorithmus! Trainingsparameter beeinflussen den Trainingsprozess, d.h. Dauer des Trainings, Anzahl Trainings-/Validierungsdaten, etc. Die *Parameter* des Algorithmus werden trainiert und berechnen aus den Eingangsdaten Ausgangsdaten. Um die Verwechslung dieser beiden relevanten Begriffe zu vermeiden, wird für die Trainingsparameter wie auch für die Architektur des Algorithmus der Begriff **Meta-Parameter** verwendet.

sen formulieren: Das Gehirn besteht aus einer Vielzahl an Neuronen, welche durch Verbindungen miteinander vernetzt sind. Diese Neuronen sind in der Lage, Signale aus den eingehenden Verbindungen mittels Ionenkanäle zu gewichten und mithilfe der in der Zelle herrschenden Ladung miteinander zu verrechnen. Sobald die Ladung einen gewissen Schwellenwert übersteigt, feuert das Neuron und gibt das Signal an die folgenden Neuronen weiter, welche den selben Prozess wiederholen. Der Lernprozess wird durch Belohnung von nützlichen und häufig verwendeten Verbindungen bewerkstelligt, wobei diese Verbindungen verstärkt und stärker gewichtet werden und ungebrauchte oder ungünstige Verbindungen verkümmern[20].

Auf diesen Grundlagen basiert die Idee der *KNNS*, weswegen sie auch noch eine weitere wichtige Eigenschaft gemeinsam haben: Als Aussenstehender ist einem genaue das genaue Funktionsprinzip unbekannt. So sind wir wie auch *KNNS* dazu fähig, handschriftliche Ziffern voneinander zu unterscheiden, wie es aber im Detail bewerkstelligt wird, ist und bleibt unklar⁶. Für die algorithmische Umsetzung wurden jedoch einige Anpassungen vorgenommen, daher sind heutige *KNNS* nur noch im entferntesten Sinne mit biologischen neuronalen Netzen verwandt.

2.4.2 Neuronen & Verbindungen

Das *Neuron* ist der Grundbaustein von *KNNS*. Wie das biologische Vorbild enthält es mehrere Signaleingänge sowie einen Signalausgang, alle Eingangssignale werden mit einem Faktor Gewichtet und zudem hat jedes *Neuron* seinen Schwellenwert. Die Gewichtungen wie auch der Schwellenwert sind die zu trainierenden *Parameter* des *KNNS*.

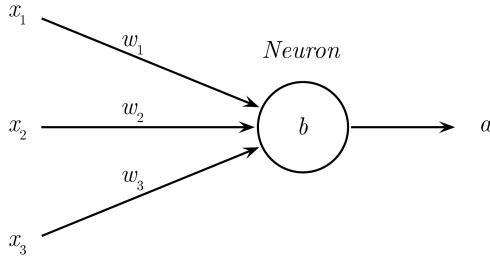


Abbildung 2: Schema von einem künstlichen Neuron

Den Eingangssignalen werden die Variablen x_1, x_2, \dots, x_n gegeben, wobei sie jeden Wert zwischen 1 und 0 annehmen können. Zur Verallgemeinerung der Formeln wird n für die Anzahl der Eingangssignale verwendet. Jeder Verbindung wird ein Gewichtungs-*Parameter* w_1, w_2, \dots, w_n ⁷ gegeben, dessen Zahlenwert nicht begrenzt ist und daher auch negativ sein kann. Die Eingangssignale werden mit dem jeweiligen Gewichtungs-*Parameter* multipliziert ($x_i \cdot w_i$), wodurch man das gewichtete Signal vom i -ten Eingang erhält. Im nächsten Schritt gilt es, alle gewichteten Signale zu summieren, welches in folgender Formel bewerkstelligt wird:

$$\sum_{i=1}^n x_i \cdot w_i \quad (1)$$

Nun soll der Schwellenwert-*Parameter*⁷ als $b \equiv -\text{Schwellenwert}$ implementiert werden. b kann somit ganz simpel mit dem Ergebnis der Formel 1 addiert werden, wodurch man $\sum_{i=1}^n x_i \cdot w_i + b$ erhält.

⁶Ironischerweise war eine Motivation für die Entwicklung von *KNNS*, uns ein besseres Verständnis für die Funktionsweise von intelligentem Verhalten zu geben. Jedoch stellte es sich heraus, dass weder *KNNS* noch das biologische Vorbild im Detail verstanden werden können.

⁷In den folgenden Kapiteln werden die Gewichtungen w_1, w_2, \dots, w_n und der Schwellenwert b eines Neurons allgemein als **Parameter** bezeichnet. Nicht zu verwechseln mit den **Meta-Parametern**(siehe Fussnote 5)

An dieser Stelle soll zur Übersicht dieser gesamte Term gleich zur Variable z zusammengefasst werden:

$$\sum_{i=1}^n x_i \cdot w_i + b \equiv z \quad (2)$$

Wenn z kleiner als 0 ist, bedeutet dies eine Unterschreitung des Schwellenwertes, ist z grösser als 0, eine Überschreitung. Um für das Ergebnis a vom *Neuron* wieder einen Wert im Bereich $0 \leq a \leq 1$ zu erhalten, wird die stufenlose *Sigmoid*-Funktion⁸ zur Standardisierung von z angewandt. Diese Funktion lässt die Werte für a im Unendlichen gegen 1 resp. 0 gehen, jedoch ermöglicht sie im Bereich um $z = 0$ eine Feinjustierung des Ausgangswertes. Zudem eignet sich die *Sigmoid*-Funktion auch, weil sie sich differenzieren lässt.

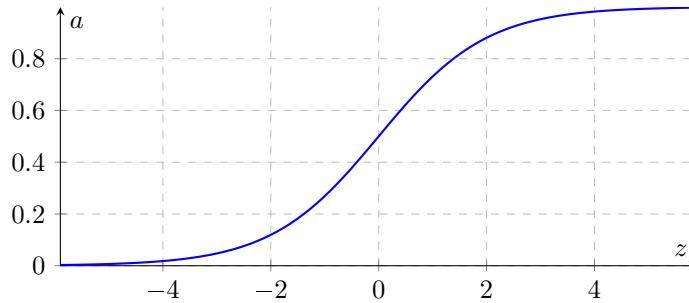


Abbildung 3: Sigmoid-Funktion $\sigma(z) = a$

Mit der *Sigmoid*-Funktion als ideale Kurve zur Standardisierung von z gilt es nur noch, alle Elemente miteinander zu verknüpfen:

$$a = \sigma \left(\sum_{i=1}^n x_i \cdot w_i + b \right) \quad \text{oder} \quad a = \sigma(z) \quad (3)$$

wobei x_1, \dots, x_n den Eingangswerten und a dem Ausgangswert a entspricht. Die Variablen w_1, \dots, w_n und b sind *Parameter* und verändern das Verhalten des Neurons.

Dies ist die Formel eines jeden *Neurons* und bildet (z.T. mit geringfügigen Modifikationen, siehe Kapitel 2.7.2) den Grundbaustein aller *KNNs*.

Anzumerken ist, dass die Funktion 3 aufgrund der Differenzierbarkeit der *Sigmoid*-Funktion und der Linearität der restlichen Operationen ebenfalls differenzierbar ist. Diese Eigenschaft wird später essenziell, um einen effizienten Trainingsalgorithmus für die Justierung der *Parameter* zu entwickeln.

⁸Definition der *Sigmoid*-Funktion: $\sigma(x) = \frac{1}{1+e^{-x}}$

2.4.3 Layers

Da ein einzelnes *Neuron* alleine nicht viel leisten kann, werden sie miteinander vernetzt — daher auch der Name *künstliches neuronales Netz*. Die einfachste Art der Vernetzung schlägt folgende Struktur vor:

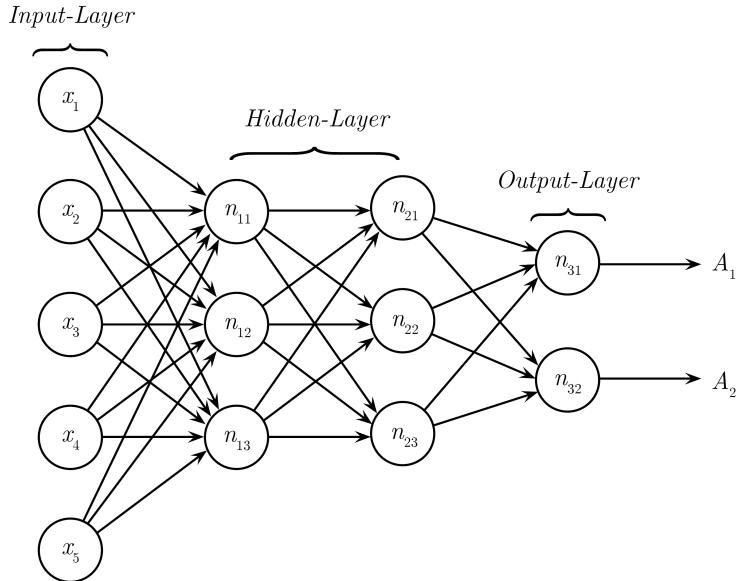


Abbildung 4: Schema von einem kleinen *KNN*

Die *Neuronen* im Netz werden in Schichten strukturiert, in sog. *Layers*⁹. Die *Neuronen* eines *Layers* werden mit allen Ausgangssignalen der *Neuronen* im vorherigen *Layer* verbunden (ein sog. *Fully-Connected-Layer*), wobei jede Verbindung separat gewichtet wird. Anzumerken ist, dass die vielen Ausgänge eines *Neurons* alle dem selben Ausgangswert a (siehe Kapitel 2.4.2) entsprechen.

Generell gibt es drei Typen von *Layers*:

- Eingangs-*Layers*, sog. *Input-Layers*
- Zwischen-*Layers*, sog. *Hidden-Layers*
- Ausgangs-*Layers*, sog. *Output-Layers*

Jedes *KNN* hat genau einen *Input-Layer* sowie einen *Output-Layer*. Dazwischen befinden sich ein oder mehrere *Hidden-Layers*.

Die Eingangsdaten x_1, \dots, x_n (in Abbildung 4 $n = 5$) werden in die *Neuronen* des *Input-Layers* geschrieben¹⁰. Dies sind z.B. Testwerte von einem Versuch oder Pixel eines Bildes (wobei n der Pixelanzahl entsprechen muss).

Darauf werden die Werte durch die *Neuronen* der *Hidden-Layers* verarbeitet. Zwar bestehen die Rechenoperationen nur aus denen im letzten Kapitel 2.4.2 beschriebenen Funktion 3, jedoch können durch richtige Justierung der *Parameter* w und b im Trainingsprozess (siehe Kapitel 2.4.5) auch sehr komplexe Funktionen modelliert werden.

Die *Neuronen* im *Output-Layer* geben das berechnete Resultat A_1, \dots, A_m (in Abbildung 4 $m = 2$) aus. Wird das *KNN* zur Kategorisierung von Daten, z.B. Bildern, verwendet, wird jedem

⁹Da die Literatur der *KNNs* praktisch ausschließlich in der englischen Sprache niedergeschrieben wird, wird auf weitere deutsche Übersetzungen der Fachausdrücke verzichtet.

¹⁰Die *Neuronen* des *Input-Layers* sind daher an sich keine *Neuronen*, da sie nur den Eingangswert annehmen und direkt an die folgenden *Neuronen* weitergeben ($x = a$).

A_i eine Kategorie zugeordnet, z.B. $A_1 \equiv$ Hund und $A_2 \equiv$ Katze. Die erhaltenen Werte für A_1, \dots, A_m entsprechen dabei der Wahrscheinlichkeit, dass die Eingangsdaten zur jeweiligen Kategorie gehören.

Die Grösse der *Input-* und *Output-Layers* müssen den Eingangsdaten respektive der Anzahl Kategorien angepasst werden. Die Grösse und Anzahl der *Hidden-Layers* ist hingegen frei wählbar und gehört zu den *Meta-Parametern*¹¹ des *KNNs*. Kleinere Probleme lassen sich noch mit einem einzelnen *Hidden-Layer* bewerkstelligen, wobei von einem *Shallow Neural Network* die Rede ist. Bei komplexeren Problemen empfiehlt sich jedoch die Verwendung eines *DNNs*, in welchem mehrere (oft spezialisierte) *Hidden-Layers* in Serie geschaltet werden.

Mathematisch gesehen werden bei der Vernetzung der *Neuronen* die Formeln ineinander verschachtelt, wodurch sich das ganze *KNN* als mathematische Funktion schreiben lässt. Die gesamte Formel ist zwar sehr komplex und unübersichtlich zu notieren, jedoch werden die ursprünglichen Eigenschaften der Funktion 3 beibehalten, d.h. es bleibt eine differenzierbare mathematische Funktion. Das *KNN* soll als Funktion zusammengefasst geschrieben werden als:

$$N(x_1, \dots, x_n) = A_1, \dots, A_m \quad (4)$$

wobei $N(\dots)$ das *KNN* als Funktion darstellt, n die Anzahl der Eingangswerte x_1, \dots, x_n (sowie der *Input-Neuronen*) ist und m die Anzahl der Ausgangswerte A_1, \dots, A_m (sowie der *Output-Neuronen*) ist. Selbstverständlich sind die *w-* und *b-Parameter* nach wie vor Teil des *KNNs*, einfacheitshalber werden sie in der allgemeinen Darstellung weggelassen, das sie in der Regel konstant sind.

2.4.4 Leistung von KNNs messen

Das *künstliche neuronale Netz* ist erstellt, doch bevor es mit trainiert werden kann, wird eine Methode zur Evaluation der Leistung des *KNNs* benötigt. Eine Art dies zu bewerkstelligen, ist die *Quadratic-Cost*-Funktion:

$$C(A_1, \dots, A_m, y_1, \dots, y_m) = \sum_{i=1}^m (A_i - y_i)^2 = Err \quad (5)$$

wobei m die Anzahl der *Output-Neuronen* ist. A_1, \dots, A_m sind die durch das *KNN* erhaltene Ergebnisse und y_1, \dots, y_m die korrekten Ergebnisse der Trainingsdaten. Die neue Variable *Err* steht für den erhaltenen Fehler.

Sind die Differenzen zwischen A_i und y_i gross, so erhält man einen grossen Wert für *Err* und vice versa. Das bedeutet, wenn *Err* ≈ 0 ist, hat das *KNN* das Ergebnis y_1, \dots, y_m mithilfe der Eingangsdaten sehr genau vorausbestimmen können. In anderen Worten: Beim Training gilt es diesen Wert *Err* zu minimieren um die Leistung zu maximieren.

Wie in den vorherigen Kapiteln 2.4.2 und 2.4.3 betont wurde, kann das gesamte *KNN* auch als Funktion dargestellt werden. Dadurch lässt sich der Fehler *Err* auch als Funktion der Eingangsdaten x_1, \dots, x_n (sowie aller Gewichtungen *w* und Schwellenwerten *b*) eines *KNNs* $N(\dots)$ und dem erwarteten Ergebnis y_1, \dots, y_m ausdrücken:

$$Err = C(N(x_1, \dots, x_n), y_1, \dots, y_m) \quad (6)$$

¹¹Parameter, die die Architektur des *KNNs* (Grösse, Tiefe, etc.) aber auch den Trainingsprozess (Dauer, Daten, etc.) beeinflussen

2.4.5 Training von KNNS

Bis jetzt wurden die *Parameter* b und w der *Neuronen* als nicht-veränderliche Werte angenommen. Im Trainingsprozess geht es nun um die Justierung dieser Werte, um den Fehlerwert Err des Netzes zu minimieren. Alle *Parameter* werden am Anfang des Trainingsprozesses zufällig initialisiert, somit wird der Fehler Err entsprechend gross sein.

Eine Möglichkeit, den Fehler zu minimieren besteht darin, einen einzelnen *Parameter* p aus dem Netz ein wenig zu ändern um dann dessen Auswirkung auf Err zu erkennen zu können. Wird Err grösser, versucht man die andere Richtung, wird Err kleiner, ist man dem Ziel einen Schritt näher und wiederholt dasselbe mit allen anderen Parameter. Zwar klingt diese Methode verständlich und machbar, jedoch ist sie sehr ineffizient, da viel Ausprobieren involviert ist. Wie also kann man die Auswirkung einer Parameteränderung auf den Fehler Err am effizientesten berechnen?

Um sich zuerst ein intuitives Bild von dieser Problemstellung machen zu können, muss man es sich als Parameterlandschaft vorstellen: In drei Dimensionen entsprächen die Ortskoordinaten zwei *Parametern*, die Höhe stünde für den zu den *Parametern* zugehörige Fehlerwert Err , die "Landschaft" mit ihren Bergen und Tälern wird durch die Trainingsdaten bestimmt (entspricht der Funktion $f(x)$ aus Kapitel 2.3.1). Da *KNNS* weithin mehr als zwei *Parameter* besitzen, enthält deren Parameterlandschaft so viele "Ortsdimensionen" wie das *KNN Parameter* hat. Das Ziel ist es in dieser Landschaft, ein tiefes Tal zu finden — mit einer Schwierigkeit: Man ist blind ($f(x)$ ist unbekannt).

In diesem Gedankenmodell sähe die vorher erwähnte Brute-Force-Strategie zum Finden des Minimums folgendermassen aus: Man macht einen Stritt in eine beliebige Richtung. Ist man höher als vorher, geht man zwei Schritte in die entgegengesetzte Richtung und wiederholt den Prozess für die anderen Dimension(en)/*Parameter*, bis man in einem Tal angekommen ist (d.h. eine Bewegung in jede Richtung bewirkt eine Verschlechterung des Ergebnisses).

Gradient Descent Eine elegantere Lösung lässt sich mithilfe der Analysis finden, weswegen bei der Herleitung der Formeln stets auf die Differenzierbarkeit geachtet wurde. Durch *partielles Ableiten* der Funktion 6 nach allen *Parametern* lässt sich die Steigung in der Parameterlandschaft im momentanen Punkt approximieren. Der erhaltene Richtungsvektor soll dabei als \vec{v} bezeichnet werden, bei dem jede Dimension einem *Parameter* entspricht. Auf das Gedankenmodell angewandt zeigt dieser Vektor \vec{v} in die Richtung, in der man am meisten Höhe verliert, obwohl die umliegende Landschaft nicht bekannt ist. Da man dem Gradienten entlang absteigt, nennt man dieses Optimierungsverfahren *Gradient Descent* (zu Deutsch *Gradientenverfahren*). Das Ausprobieren fällt weg, wodurch sich das Training massiv beschleunigt.

Bei der *Gradient Descent*-Trainingsmethode lassen sich einige solcher *Meta-Parameter* einstellen. Diese beeinflussen die Lerngeschwindigkeit, Genauigkeit und die Dauer des Trainings.

- **Lernrate η** bestimmt die Grösse der "Schritte" in der Parameterlandschaft ($\vec{v} \cdot \eta$). Bei zu kleinen Werten wird das Training nur langsam vorangehen; bei zu grossen Stritten hingegen riskiert man das Überspringen eines Minimums und z.T. sogar eine Verschlechterung der Leistung.

- **Mini-Batch** Der Richtungsvektor \vec{v} wird in der Regel über mehrere Trainingsexemplare gemittelt, wofür alle Trainingsdaten zufällig in sog. *Mini-Batches* unterteilt werden. Der Trainingsalgorithmus berechnet für alle Exemplare eines *Mini-Batches* die Steigung und mittelt diese, um die Richtung \vec{v} zu erhalten. Erst dann werden die *Parameter* angepasst (ein "Schritt"). Die-

ser Prozess wird mit dem nächsten *Mini-Batch* für die neue Position in der Parameterlandschaft wiederholt.

Je grösser die Mini-Batches sind, umso genauer repräsentiert \vec{v} die Parameterlandschaft aller Trainingsdaten, jedoch müssen für einen "Schritt" mehr Berechnungen angestellt werden, welche das Training verlangsamen.

- **Lernepochen** Sind alle *Mini-Batches* schon einmal für die Berechnung des Gradienten verwendet worden, ist eine Lernepoche vorüber. Die Mini-Batches werden neu gemischt und können wieder in den Trainingsalgorithmus eingespeist werden.

Je mehr Epochen das Training durcharbeitet, umso genauer können die *Parameter* eingestellt werden, jedoch läuft man bei zu vielen Epochen Gefahr der *Überanpassung*, welches man mithilfe von Validierungsdaten zu verhindern versucht (siehe Kapitel 2.3.2: Validierung).

2.5 Deep Neural Networks

Theoretisch liesse sich jegliche (kontinuierliche) Funktion $f(x)$ mit den oben beschriebenen *SNNs* beliebig genau approximieren ($SNN(x) \approx f(x)$), notwendig dafür ist mindestens ein einziger, entsprechend grosser *Hidden-Layer*[21]. Man nennt Algorithmen mit dieser Eigenschaft *universelle Approximatoren*. Daraus folgt, dass zusätzliche *Hidden-Layers* den Bereich der modellierbaren Funktionen nicht erweitern und daher nicht zur Komplexität eines neuronalen Netzes beitragen¹².

Wendet man aber nun ein breites *SNN* auf komplexere Probleme wie z.B. der Bilderkennung an, so wird man jedoch feststellen, dass es praktisch unmöglich ist, die gesuchte Funktion $f(x)$ mittels *Gradient-Descent* zu finden. Das Problem besteht darin, dass die Parameterlandschaft aufgrund der vielen *Parameter* enorm gross ist und viele lokale Minima aufweist, welche das Training frühzeitig beenden.

Um brauchbare Ergebnisse mittels *Neuronalen Netzen* zu erzielen, wurden in der Vergangenheit die wesentlichen Informationen manuell aus den Eingangsdaten gefiltert, um die Anzahl der Neuronen und Verbindungen und somit die Grösse der Parameterlandschaft zu reduzieren. Die grundlegende Idee hinter *DNNs* ist es nun, diesen ersten Schritt der Vorverarbeitung mittels zusätzlichen, spezialisierten *Layers* zu erlernen. Somit werden die rohen Informationen der Eingangsdaten von den ersten *Layers* zuerst gefiltert, bevor diese durch einen *Fully-Connected-Layer* verarbeitet werden.

In der Praxis gibt es einige verbreitete *DNN*-Typen: *RNNs* (*Recurrent Neural Networks*), *LSTM RNNs* (*Long short-Term Memory RNN*), *Deep Belief Networks*, etc., jedoch ist das *Convolutional Neural Network* (*CNN*) der geeignetste Typ für Bilderkennungsprobleme. Da es in der Arbeit Pilze anhand von Bildern zu kategorisieren gilt, wird im nächsten Kapitel der Fokus nur auf *CNNs* gelegt.

2.6 Convolutional Neural Networks

Will man ein *KNN* für die Erkennung von Bildern verwenden, so sind *SNNs* mit einem *Fully-Connected-Layer* aus mehreren Gründen ungeeignet, u.a. werden räumlichen Zusammenhänge der Eingangs-Pixel komplett ignoriert, wodurch das erkennen von Mustern und Formen nicht unmöglich wird, aber deutlich erschwert wird. *CNNs* verschaffen mit *Convolution-Layers* Abhilfe,

¹²Im Gegenteil entstehen durch weitere *Fully-Connected-Layers* zusätzliche Probleme beim Training mittels *Gradient-Descent*, da sog. *Unstable Gradients* den Trainingsprozess destabilisieren.

wobei mit sog. *Feature-Maps* Muster erkannt und gefiltert werden können. Zusätzliche *Pooling-Layers* konzentrieren zudem die von den *Convolution-Layers* erhaltenen Informationen weiter.

Durch Hinzufügen dieser Schichten (daher "Deep Neural Network") erhofft man sich eine Konzentration der für die Bilderkennung relevanten Informationen (sog. *Bias* der *Layers*), welche einfacher von einem letzten, kleineren *Fully-Connected-Layer* verarbeitet werden können. Die Leistung ist primär von diesen *Meta-Parametern* abhängig, weswegen das Finden der optimalen Kombination und Konfiguration dieser Layers die eigentliche Kunst ist.

2.6.1 Convolution Layer

Um räumliche Zusammenhänge erkennen zu können, werden einem Neuron eines *Convolution-Layers* die Eingangssignale auf einen lokalisierten Bereich des vorhergehenden Layers eingeengt. Diesen Bereich nennt man *Local Receptive Field*. Hat man z.B. den *Input-Layer* und einen darauf folgenden *Convolution-Layer*, auf die eine *Local-Receptive-Field*-Grösse von 5×5 *Neuronen* angewandt wird, so ist ein *Neuron* des *Convolution-Layers* nur mit einem 5×5 *Neuronen* grossen Bereich des *Input-Layers* verbunden. Für das erste *Neuron* des *Convolution-Layers* wird es wie in Abbildung 5a gezeigt aussehen.

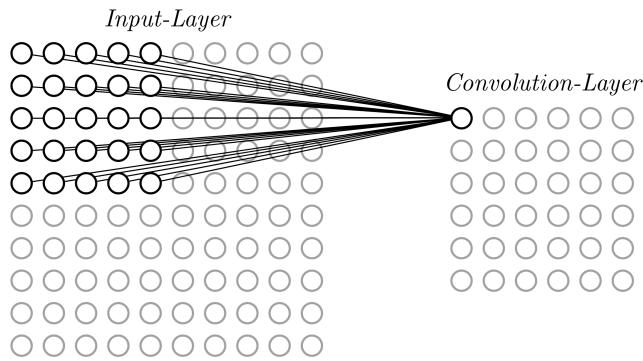


Abbildung 5a: Beispiel für das *Local-Receptive-Field* vom 1. *Neurons* des *Convolution-Layers*

Für das nächste *Neuron* des *Convolution-Layers* verschiebt sich die gesamte Maske einen Schritt in die jeweilige Richtung. Wie die Grösse des *Local-Receptive-Fields* ist auch die Schrittgrösse ein *Meta-Parameter* und kann angepasst werden, in diesem Beispiel wird eine Schrittgrösse von einem *Neuron*(/Pixel) verwendet:

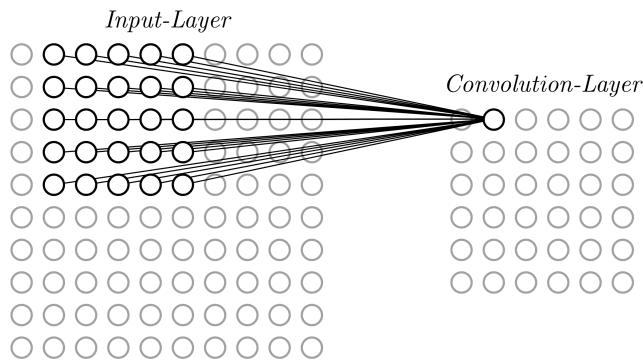


Abbildung 5b: Beispiel für das *Local-Receptive-Field* vom 2. *Neurons* des *Convolution-Layers*

2.6.2 Shared Weights & Feature Maps

Ein weiterer verwendeter Trick ist, dass alle Neuronen des *Convolution-Layers* dieselben *Parameter* verwenden. Dabei ist die Rede von sog. *Shared weights*, d.h. das *Neuron_{ij}* im *Local-Receptive-Field* wird für jedes *Convolution-Neuron* mit w_{ij} gewichtet, wobei i und j die Indexe für die horizontale resp. vertikale Position im *Local-Receptive-Field* bezeichnet. Im 5×5 -Beispiel von oben gibt es somit nur 25 verschiedene Gewichtungen für die Verbindungen, der Schwellenwert b ist überall der selbe.

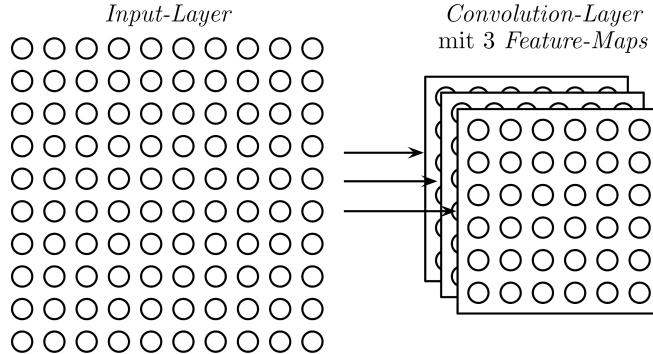


Abbildung 6: Veranschaulichung von *Feature-Maps*

Intuitiv kann man sich den *Convolution Layer* als Schablone für ein durch die *Parameter* bestimmte Muster vorstellen, welche Schrittweise über das Eingangsbild geschoben wird. Entspricht das Bild dem Muster, wird ein hoher Wert zurückgegeben und vice versa. Je nach *Parameter* kann man z.B. Kanten oder Ecken erkennen. Man hat somit eine sog. *Feature Map* erstellt, welches das Vorkommen eines bestimmtes Musters im Bild kartiert. Da ein Bild nicht nur aus einem einzigen Muster besteht, werden in einem *Convolution-Layer* sogleich mehrere *Feature-Maps* erstellt, welche sich auf verschiedene Muster spezialisieren können.

2.6.3 Pooling

Um die vom *Convolution-Layer* erhaltenen *Feature-Maps* weiter zu konzentrieren, folgt in der Regel ein sog. *Pooling-Layer*. Dabei wird wiederum aus einem Bereich des vorherigen *Layers* nur der höchste Wert (*max-Pooling*) oder der durchschnittliche Wert (*average-Pooling*) an den nächsten *Layer* weitergegeben. Da *Pooling-Layers* ebenfalls nach dem Prinzip der *Local-Receptive-Fields* funktionieren, haben auch sie eine Filter- wie auch Schrittgrösse.

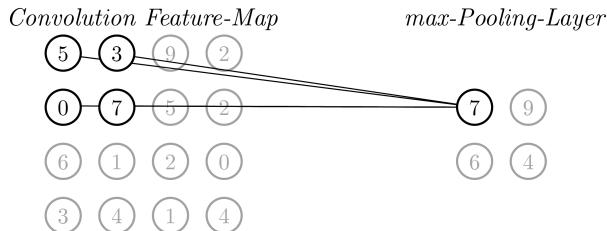


Abbildung 7: Funktionsweise eines *max-Pooling-Layers* mit Filtergrösse 2×2 und Schrittgrösse 2

2.6.4 Aufbau eines CNNs

Beim Aufbau eines *CNNs* werden einem viele Freiheiten gelassen, jedoch hat sich folgende Grundstruktur für viele Bildklassifizierungsprobleme bewährt: Nach dem *Input-Layer* werden ein *Convolution-Layer* und ein *Pooling Layer* hintereinandergeschaltet, z.T. werden noch weitere *Layers* (z.B. *ReLU-Layer*, siehe Kapitel 2.7.2) hinzugefügt. Zusammen bilden diese eine Einheit, welche einige Male wiederholt wird. Die Grösse der Layers nimmt mit der Tiefe ab, wodurch sich der charakteristische trichterförmige Aufbau von *CNNs* bildet. Auf die letzte Einheit folgt ein *Fully-Connected-Layer*, wodurch die erkannten Muster miteinander kombiniert und ausgewertet werden. Schliesslich werden mithilfe eines sog. *Softmax-Layers* die Ausgangssignale auf Prozentwerte normalisiert, um anschliessend die Klassifizierung durchzuführen.

2.7 Zusätzliche Meta-Parameter und Modifikation

Die in den vorherigen Kapiteln beschriebenen *KNNs* und *DNNs* als auch deren Training sind keineswegs starr und lassen sich in vielen Aspekten modifizieren und justieren. Auch diese *Meta-Parameter* sind für spätere Kapitel relevant, weswegen sie in diesem Abschnitt kurz beschrieben werden.

2.7.1 Bild-Normalisierung

Da Pixelwerte in der Regel zwischen 0 und 255 liegen, würden bei direkter Einspeisung der Bilder sehr grosse, wenn auch wenig aussagekräftige Werte in das *KNN* eingespeist werden. Aus diesem Grund ist eine erste einfache Massnahme, um redundante Informationen zu minimieren, das anfertigen eines Durchschnittsbildes. Von allen Trainingsdaten wird der Mittelwert berechnet, welcher dann vor der Einspeisung von den Eingangsdaten subtrahiert werden. Dadurch nehmen ähnliche bleibende Regionen (wie z.B. der Hintergrund im Randbereich) einen neutralen Wert an, während grosse Abweichungen vom Durchschnitt stärker gewichtet werden[22].

2.7.2 ReLU-Layer & Batch-Normalization-Layer

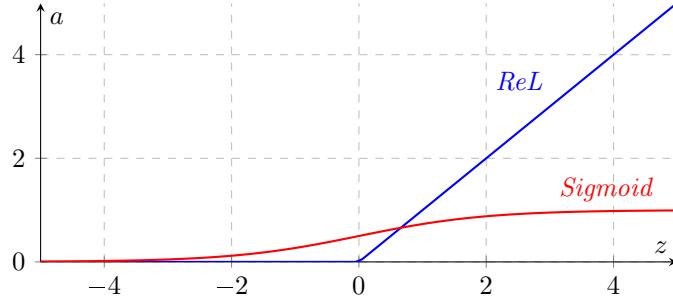
Die *Sigmoid*-Kurve aus Kapitel 2.4.2 ist nicht die einzige Funktion, die zur Standardisierung des gewichteten Ausgangssignals z eines Neurons (siehe Formel 2) verwendet wird. Neben der *Sigmoid*-Funktion haben sich auch die sog. *Rectified-Linear*-Funktion bewährt. Ein *Neuron*, welches diese Funktion verwendet, nennt man *Rectified-Linear-Unit* (kurz *ReLU*); ein *ReLU-Layer* ist somit eine Schicht zusammengesetzt aus diesen *ReLU-Neuronen*.

Im Gegensatz zur *Sigmoid*-Kurve liefert diese Funktion einen ganz anderen Ansatz: Bei $z > 0$ verhält sie sich linear und bei $z \leq 0$ gibt sie einen konstanten Wert von 0 zurück, d.h. der Ausgangswert ist nach oben nicht begrenzt und immer positiv.

$$\sigma(z) = \begin{cases} z, & \text{wenn } z > 0 \\ 0, & \text{wenn } z \leq 0 \end{cases} \quad (7)$$

Als Graph sieht die in der Formel 7 beschriebene *Rectified-Linear*-Funktion verglichen mit der *Sigmoid*-Funktion folgendermassen aus:

Aufgrund der Linearität kann auch für sehr grosse positive z -Werte eine signifikante Steigung für die Justierung der *Parameter* berechnet werden, wo hingegen die *Sigmoid*-Funktion in diesem Bereich nur eine verschwindend kleine Steigung liefert. Für negative Werte ist die Steigung hin-



gegen konstant 0, wodurch die *Parameter* jenes Neurons nicht mehr durch das Training justiert werden können und gewissermassen "einfrieren". Dies führt dazu, dass je nach Eingangsdaten ein Teil der Neuronen nicht trainiert wird, welches wiederum die Parameterlandschaft kleiner macht und dadurch der Überanpassung entgegenwirkt. Zudem sind *ReLU-Layers* aufgrund der einfachen Operationen nicht sehr rechenintensiv.

Da das Resultat nach oben nicht begrenzt ist, können die Ergebnisbereiche zwischen einzelnen *ReLU-Neuronen* stark variieren. Um zu grosse Werte zu vermeiden, werden die Ergebnisse eines *ReLU-Layers* mit einem sog. *Batch-Normalization-Layer* auf gemässigte Werte herunterskaliert.

2.7.3 Data Augmentation

Ein wichtiger Aspekt für die Entwicklung eines zuverlässigen maschinellen Lernalgorithmus ist der Umfang der Trainingsdaten. Wenn jedoch zu wenige vorhanden sind, ist das Training sehr unzuverlässig und man riskiert eine Überanpassung an die wenigen Trainingsexemplare.

Mithilfe von *Data Augmentation* lässt sich die Anzahl der Trainingsexemplare künstlich vergrössern. Dazu werden die Trainingsbilder vervielfacht und um einige Pixel verschoben, gespiegelt, ein wenig rotiert oder verzerrt. Man generiert dadurch viele weitere, voneinander verschiedene Trainingsexemplare, welche für das Training eines *KNNs* verwendet werden können.

2.7.4 Transfer Learning

Ein *KNN* kann man stets von Grund auf neu aufbauen und trainieren, jedoch kann auch das *Wissen* von bereits trainierten *DNNs* verwendet werden, um es auf ähnliche Probleme zu spezialisieren. Durch sogenanntes *Transfer Learning* wird ein bereits trainiertes *KNN* (z.B. *GoogLeNet* oder *AlexNET*) nochmals mit neuen Trainingsdaten und den dazugehörigen Kategorien trainiert. Man erhofft sich dabei, dass häufig vorkommende Muster wie Kanten und Formen nicht von Neuem gelernt werden müssen. Dadurch soll das Training einerseits beschleunigt und andererseits soll dessen Leistung gesteigert werden.

3 Methode

In diesem Kapitel wird der Entwicklungsprozess des auf *CNNs* basierenden Pilzerkennungsalgorithmus behandelt (siehe Kapitel 1.1). Vorerst wird auf den Datenbeschaffungsprozess sowie die Datenvorverarbeitung eingegangen. Die darauf folgenden Abschnitte behandeln den Optimierungsprozess der *Meta-Parameter* (siehe Kapitel 2.4.5, 2.6 & 2.7) und erwähnen kurz diejenigen Massnahmen, welche zu einer signifikanten Verbesserung der Leistung des Erkennungsalgorithmus führten.

In diesem Projekt wird für die Datenaufbereitung der Trainingsdaten wie auch für das Erstellen und Trainieren von *KNNs* die Entwicklungsumgebung MatLab mit der *Deep Learning Toolbox*[23] verwendet. Die Entscheidung fiel auf MatLab aufgrund des effizienten und einfach zu modifizierenden Grundgerüstes für *Deep Learning* und *KNNs*. Die Unterstützung von Mehrkernprozessoren sowie die Einbindung von Grafikkarten beschleunigt zudem den Trainingsprozess merklich. Auch sind in MatLab vortrainierte *KNNs* für das *Transfer-Learning* (siehe Kapitel 2.7.4) einfach zu implementieren und modifizieren.

3.1 Datenbeschaffung

Trainingsdaten bilden die Grundlage jedes auf *Machine Learning* basierenden Algorithmus, weshalb man ein solides Fundament aus vielen Trainingsdaten benötigt (siehe Kapitel 2.3.1). Da die Datenlage für Pilzbilder hingegen eher schlecht ist, ist die Datenbeschaffung ein wichtiger Bestandteil dieser Arbeit. Dabei gilt es, möglichst viele korrekt bestimmte Pilzbilder zu sammeln.

Für das Training sollen von jeder der 20 ausgewählten Pilzarten (siehe Tabelle 7) mindestens 230 verschiedene Bilder beschaffen werden, von den unbekannten 1840 (das Achtfache). Um die Leistung des Algorithmus mittels unabhängigen Daten zu testen (siehe *Validierung* in Kapitel 2.3.2), werden zusätzlich 20 Bilder pro bekannte Art resp. 160 für die unbekannte Kategorie gesammelt. Somit mussten 250 Bilder pro ausgewählte Art sowie 2000 Bilder von unbekannten Arten gesammelt werden, insgesamt also etwa 7000 Bilder. Die Anzahl der bekannten Arten begründet sich mit den schwer verfügbaren Bilddaten, die der unbekannten durch eine erfahrungsgestützte Abschätzung, um im Rahmen einer Maturarbeit zu bleiben und trotzdem noch aussagekräftige Ergebnisse erhalten zu können.

Folgende Massnahmen wurden ergriffen, um möglichst viele, reine Trainingsdaten zu beschaffen:

3.1.1 Datenbank SwissFungi

Um einen Grunddatensatz zu erhalten, wurde die Fotodatenbank vom offiziellen SwissFungi Pilzatlas der WSL Schweiz verwendet[24]. Von den 20 ausgewählten Arten konnten dadurch je einige Dutzend Bilder gesammelt werden; von den unbekannten Arten konnte die geforderte Anzahl von 2000 Exemplaren gedeckt werden. Diese Bilder sind alle professionell bestimmt und bedürfen daher keiner weiteren Kontrolle um die Datenreinheit zu garantieren.

3.1.2 Webseite ShroomNET

Um Pilzsammler und Pilzfotografen aus der Region einfach zu erreichen, wurde die Webseite www.obermeier.ch mit einem Hochladeformular für Pilzbilder eingerichtet. Aus einer Liste der ausgewählten Arten lässt sich die Kategorie auswählen, wobei die hochzuladenden Bilder mittels Drag-and-Drop direkt hochgeladen werden. Neben der Hochladefunktion wurde um die Datenreinheit zu gewährleisten zusätzlich ein *Quiz* erstellt, bei dem Besucher die hochgeladenen Pilzbilder

klassifizieren können. Mithilfe dieser Webseite lassen sich somit einfach weitere Bilder zusammentragen, welche auch direkt von erfahrenen Sammlern verifiziert werden können (siehe Abbildung 26).

Verbreitet wurde die Webseite über Telefonate und Mail-Verkehr mit Pilzkontrolleuren aus der Region Brugg-Aarau sowie in einem Artikel der Schweizerischen Zeitschrift für Pilzkunde SZP[25], welche in der ganzen Schweiz versendet wurde. Mittels der Webseite konnten nochmals insgesamt etwa 300 Bilder gesammelt werden.

3.1.3 Internet Crawler

Mithilfe eines Python-Skriptes[26] konnten automatisiert mit der Google-Suchmaschine nochmals etwa 400 Bilder pro Art gefunden und heruntergeladen werden, gesucht wurde dabei nach dem wissenschaftlichen Namen der Pilze (siehe Tabelle 7 im Anhang). Aufgrund vieler Unreinheiten waren pro Kategorie nur etwa 150 weitere Bilder brauchbar. Da die Google-Suche nicht immer korrekte Ergebnisse liefert, wurden diese Bilder auf die Webseite hochgeladen, um sie im *Quiz* nochmals von Pilzsammlern verifizieren zu lassen.

3.1.4 Videos

Eine weitere Möglichkeit, um an viele Bilddaten zu kommen, ist das Extrahieren von Standbildern aus Videos. Bei sich ändernden Kameraperspektive lassen sich mehrere Standbilder von selben Pilz anfertigen, welche nicht identisch sind und sich daher als Trainingsdaten eignen. Dafür wurde das von Michael Bachmeier[27] zur Verfügung gestellte Videomaterial verwendet. Insgesamt liessen sich durch die Bearbeitung von Videos weitere 1500 Bilder zur Datenbank hinzufügen. Auch diese Daten bedürfen aufgrund der zuverlässigen Quelle keiner weiteren Kontrolle.



Abbildung 8: Beispielbilder von gesammelten Rohdaten der ersten 12 ausgewählten Pilzarten

3.2 Datenaufbereitung

Nach dem Sammeln der rohen Bilddaten gilt es in der Aufbereitung darum, diese durch einige Vorverarbeitungsmethoden auf das Training anzupassen.

3.2.1 Bildbearbeitung

Im ersten Schritt der Datenaufbereitung wurden die Eingangsdaten durch Zentrieren sowie Skalieren auf den Pilz zugeschnitten¹³. Dafür wurde eine Mindestauflösung von $200 \times 200\text{px}$ festgelegt, welches eine gute Balance zwischen Details im Bild und Komplexität des Netzes bildet. Ein zu diesem Zwecke mit MatLab programmiertes Hilfsprogramm (siehe Abbildung 10) beschleunigte den manuellen Bearbeitungsprozess von Bildern wie auch Videos.

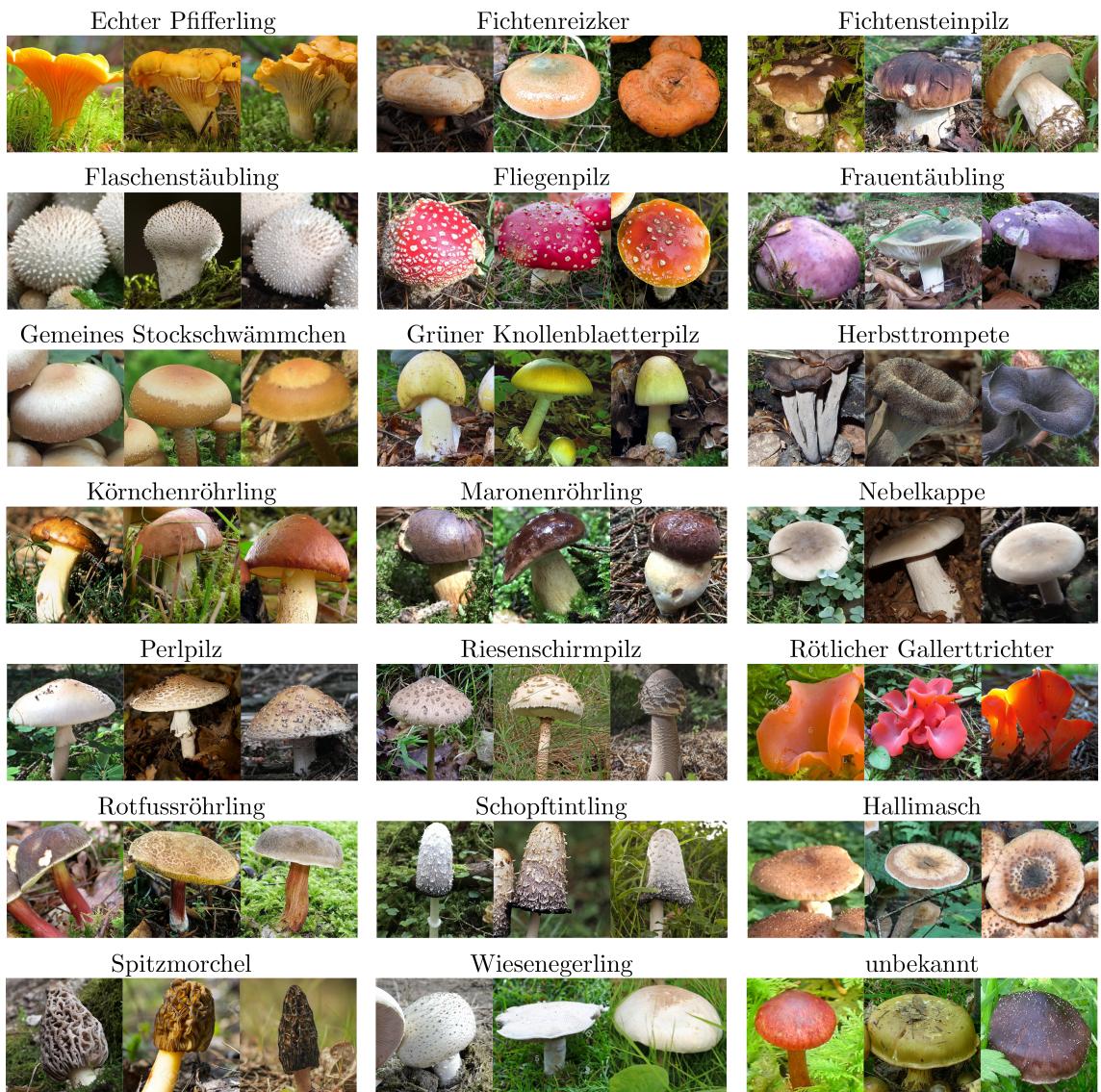


Abbildung 9: Je drei bearbeitete Beispielbilder aus den 21 Kategorien.

¹³Zwar könnten die Daten auch unbearbeitet in das Training eingespeist werden, jedoch erschwert die zusätzliche Lokalisierung des Pilzes den Lernprozess und führt unweigerlich zu einer Verschlechterung der Erkennungsleistung. Das Problem der Objektfindung in Bildern soll nicht Teil der Problemstellung sein, weswegen dieser Schritt hier manuell verrichtet wird.

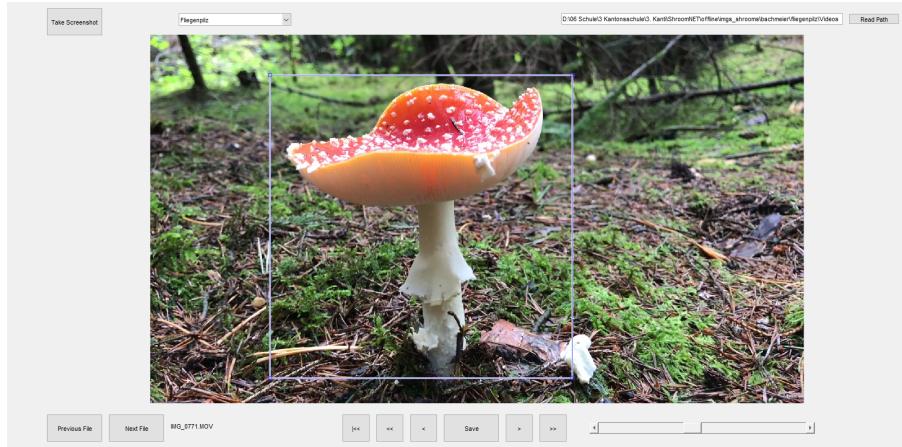


Abbildung 10: Screenshot vom Hilfsprogramm beim Bearbeiten eines Videos: Oben links kann die Kategorie ausgewählt werden, der blaue Rahmen legt den Bildausschnitt fest. Unten befinden sich Steuerelemente zur Navigation sowie Speicherung des zugeschnittenen Bildes.

3.2.2 Aufteilung in Trainings- und Validierungsdaten

Die zugeschnittenen Daten wurden in den vorher festgelegten Proportionen zufällig in Trainings- und Validierungsdaten unterteilt. Die aus Videos extrahierten Trainingsdaten eignen sich aufgrund der in Abbildung 11 ersichtlichen Ähnlichkeit nicht zur Validierung¹⁴, stattdessen werden diese ausschliesslich als Trainingsdaten verwendet.



Abbildung 11: Ähnlichkeit von 8 Standbilder aus 2 Videos

Von gewissen Pilzarten waren weitaus mehr als 230 Bilder vorhanden. Dieser Überschuss musste ausgeglichen werden, damit von jeder Art gleich viele Trainingsbilder vorhanden sind. Statt diese überzähligen Exemplare aus dem Training auszuschliessen, wurden die Trainingsbilder mit Kopien zufälliger Exemplare der selben Kategorie auf 400 Bilder pro Pilzart ergänzt¹⁵. Um die Proportionen der Kategorien beizubehalten, wurden die Anzahl der unbekannten Trainingsbilder entsprechend auf 3200 erweitert.

3.2.3 Bildnormalisierung

Bei MatLab wird die Bildnormalisierung standardmässig vor jedem Training durchgeführt, jedoch muss um spätere Versuche mit den Zusatzinformationen (siehe Kapitel 3.2.6) separat durchgeführt werden. Dafür wird, wie in Kapitel 2.7.1 erklärt, der Durchschnitt aller Trainingsbilder berechnet (siehe Abbildung 12). Dieser wird dann von den Eingangsdaten subtrahiert, bevor diese im Netz eingespeist werden.

¹⁴Die Ähnlichkeit führte in ersten Tests zu einer Verfälschung der Ergebnisse.

¹⁵Die Programmierumgebung von MatLab lässt keine benutzerdefinierte Verhältnisse der Kategorien zu, ohne tiefliegende Modifikationen an der *Neural Network Library* vornehmen zu müssen. Daher diese Problemumgehung mittels Duplizieren einiger Trainingsdaten.

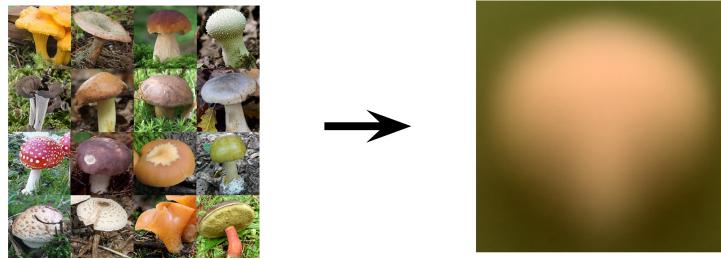


Abbildung 12: Das für die Bildnormalisierung verwendete Durchschnittsbild aller Trainingsdaten

3.2.4 Farbfilter

Um dem Algorithmus zusätzlich zu unterstützen, kann man durch Hinzufügen eines einfachen Farbfilters Farbbereiche löschen, welche keine relevanten Informationen tragen. Im Fall der Pilzerkennung ist Grün in grossen Mengen im Hintergrund vertreten, welches mittels eines Grünfilters auf den für das *CNN* neutralen Wert 0 (nach der Anwendung der Bildnormalisierung) gesetzt werden kann¹⁶.



Abbildung 13: Grünfilter angewandt auf ein Trainingsbild¹⁷

3.2.5 Data Augmentation

Wie in Kapitel 2.7.3 beschrieben, können durch leichtes Transformieren der Bilder künstlich "neue" Trainingsdaten generiert werden. Hierfür stellt MatLab ein entsprechendes *Data-Augmentation-Tool* (sog. *augmentedImageDatastore*) zur Verfügung. Wie in Abbildung 14 zu erkennen ist, liefert *Data-Augmentation* durch Verschieben, Rotieren und Spiegeln neue Variationen des Ursprungsbildes, wobei die Hoffnung ist, Überanpassung (siehe 2.3.2) an die Trainingsdaten zu vermeiden¹⁸.

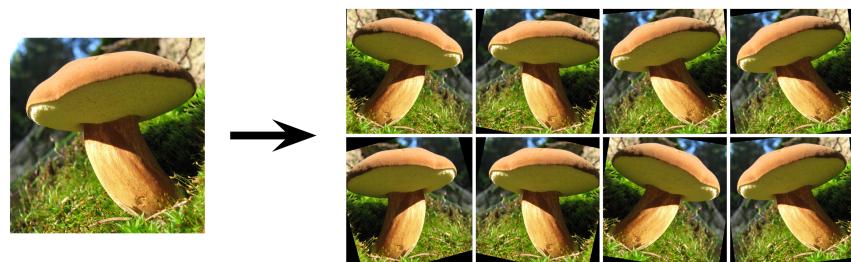


Abbildung 14: *Data-Augmentation* angewandt auf ein Bild

¹⁶Da durch den Farbfilter nun viel weniger Grüntöne in den Daten enthalten sind, muss für die Kombination von Grünfilter und Bildnormalisierung das Durchschnittsbild entsprechend neu kalkuliert werden.

¹⁷Das gezeigte Bild ist nur eine Veranschaulichung des Grünfilters und wiedergibt das Eingangsbild nicht genau. Hinzu kommt die Bildnormalisierung bevor die maskierten Bereiche auf 0 gesetzt werden.

¹⁸Wie auch beim Grünfilter muss wegen den schwarzen Rändern das Durchschnittsbild für die Bildnormalisierung neu kalkuliert werden.

3.2.6 Zusatzinformationen sammeln & aufbereiten

Da auch ein Experte für die akkurate Pilzartenbestimmung mehr als nur das Aussehen aus einer einzigen Perspektive benötigt, dürfte man dasselbe auch nicht von einem Bestimmungsalgorithmus erwarten. Deswegen sollen auch zusätzliche Informationen mit in die Bestimmung einfließen können. Folgende Eigenschaften wurden aufgrund der einfachen Bestimbarkeit und Implementierbarkeit in die Bestimmung aufgenommen:

Eigenschaft	Optionen
Standort	Laubwald, Mischwald, Nadelwald, Wiese
Durchmesser Hut/Fruchtkörper	klein (<3cm), mittel (3-10cm), gross (>10cm)
Oberfläche Hut/Fruchtkörper	trocken, feucht
Hutrand	glatt, wellig, gerieft, kein
Sporenanlage	Lamellen, Röhren, Fruchtschicht, Fruchtmasse
Lamellenhaltung	frei, herablaufend, angewachsen, kein
Ring	hängend, aufsteigend, doppelt, gerieft, fleischig, kein
Stiel	voll, wattig, hohl, kein
Geruch	angenehm, geruchlos, stinkend

Tabelle 1: Zusatzinformationen für genauere Pilzartenbestimmung

Von den ausgewählten Pilzarten wurden diese Informationen mithilfe des Pilznachschlagewerks *Welcher Pilz ist das?* von Markus Flück [28] zusammengetragen. Eine Tabelle mit den Eigenschaften aller Pilzarten findet sich im Anhang (Tabelle 8).

Da die *Deep Learning Toolbox* von Matlab nicht auf mehrere Eingangssignale ausgelegt ist, mussten diese Zusatzinformationen in den Bilddaten selber hinterlegt werden. Hierzu codiert jedes Pixel der ersten Bildzeile eine Option der Pilzeigenschaften. Wenn die Informationen einer Eigenschaft gegeben sind, nehmen die zu den Optionen entsprechenden Pixel Werte von 1 resp. -1 an; ist eine Information nicht gegeben, so werden diese Pixel auf den neutralen Wert 0 gesetzt. Mittels des benutzerdefinierten *Extraction-Layers* können diese Informationen abgegriffen werden, um tiefer im Netz direkt verarbeitet werden zu können. Ein vor das *KNN* geschaltete *Cover-Layer* überschreibt diese Pixel der ersten Zeile, damit diese keinen Einfluss auf den Bilderkennungsteil haben.

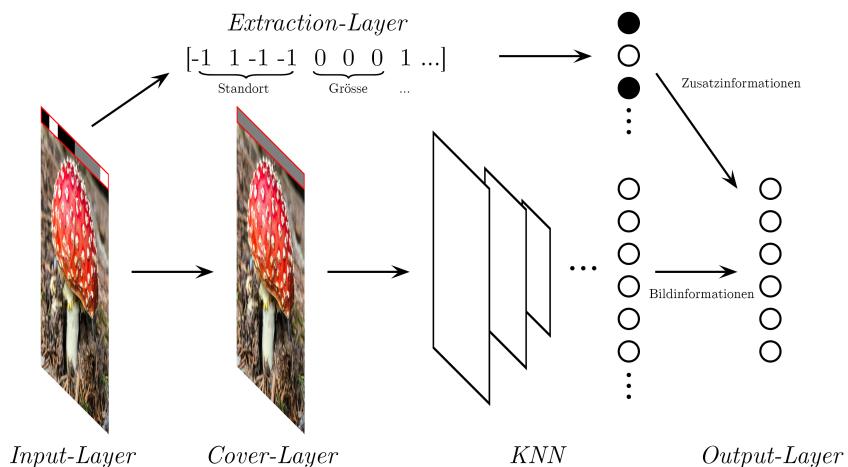


Abbildung 15: Extraktion und Einspeisung von Zusatzinformationen

3.3 Entwicklungsprozess

Mit den aufbereiteten Trainingsdaten gilt es im nächsten Schritt den Algorithmus für die Pilzarternerkennung zu erarbeiten. In den folgenden Abschnitten wird beschrieben, welche Modifikationen zu einer signifikanten Verbesserung der Bestimmungsleistung beigetragen haben. Dabei wird in diesem Kapitel der Prozentsatz der richtig erkannten Validierungsdaten als Metrik für den Vergleich der Algorithmen verwendet. Durch *Data-Augmentation* können z.T. auch unterschiedliche Resultate erhalten werden. Um trotzdem einen zuverlässigen Mittelwert für die Diskussion zu erhalten, werden die Validierungsdaten je zehn mal eingespeist und kategorisiert.

3.3.1 Bezugswert

Um bestimmen zu können, ob ein Algorithmus tatsächlich "intelligentes" Verhalten aufweist und nicht durch Zufall die korrekte Lösung errät, gilt es im ersten Schritt einen Richtwert für die Leistung zu finden. Diesen Wert gilt es von den Algorithmen zu übertreffen. Mit den in Kapitel 3.2 gegebenen Klassenverhältnissen ist es naheliegend, jedes Bild der unbekannten Kategorie zuzuteilen. Mit dieser Strategie erhält man eine Genauigkeit von $\frac{8}{20+8} \approx 28.6\%$, welcher als Bezugswert für den zu entwickelnden Algorithmus verwendet werden soll.

3.3.2 Basis-SNN

Der erste Algorithmus besteht aus einem einfachen *SNN* mit einem *Fully-Connected-Layer* als *Hidden-Layer* und einem weiteren *Fully-Connected-Layer* als *Output-Layer*. Für das Training werden die zugeschnittenen Bilddaten nur mittels Bildnormalisierung vorverarbeitet; *Data-Augmentation*, der Grünfilter sowie die Zusatzinformationen werden erst in den folgenden Abschnitten implementiert.

Um ein grobes Bild der Leistung zu erhalten, werden folgende *Meta-Parameter* des *SNNs* variiert: Die Eingangsdatengrößen ($200 \times 200\text{px}$, $50 \times 50\text{px}$ oder $10 \times 10\text{px}$ ¹⁹) sowie die *Hidden-Layer*-Grösse (20, 60 oder 100 Neuronen). Für die restlichen *Meta-Parameter* dienen vorerst Werte, welche sich erfahrungsgemäss für ähnliche Bilderkennungsprobleme bewährt haben. Alle *Meta-Parameter* sind in der Tabelle 9 im Anhang zu finden. Folgend die Ergebnisse:

Bildauflösung	Anzahl <i>Hidden-Neuronen</i>		
	20	60	100
$200 \times 200\text{px}$	3.6%	3.6%	3.6%
$50 \times 50\text{px}$	12.5%	13.6%	20.4%
$10 \times 10\text{px}$	41.3%	43.6%	40.2%

Tabelle 2: Bestimmungsleistung von *SNNs* im einem *Hidden-Layer*

Bemerkenswert ist das Ergebnis für das *SNN* mit 60 *Hidden-Neuronen*, welches die kleinsten Eingangsdaten erhalten hat: Mit nur 0.25% der vorhandenen Daten leistet es mehr als das Zehnfache wie die Netze, welche alle $200 \times 200 \times 3$ (Auflösung \times Farbkanäle) Datenpunkte erhalten haben, und fast das Doppelte des Bezugswertes. Der nächste Schritt ist es nun, die Komprimierung der Informationen nicht durch verlustbehaftetes Skalieren¹⁹ zu erreichen, sondern durch Hinzufügen von zusätzlichen *Convolution-Layers*.

¹⁹Skalierung der Bilder wird mittels *bikubischer Interpolation* durchgeführt.

3.3.3 Basis-CNN

Als Grundlage für das *CNNs* dient ein nur aus *Convolution-Layers*, *Pooling-Layers* und einem *Fully-Connected-Layer* (entspricht dem *Output-Layer*) zusammengesetztes Netz (komplette Liste der *Meta-Parameter* siehe Tabelle 10). Als grobe Vorlage für Einstellungen der drei *Convolution-Pooling-Einheiten* diente *AlexNet*[29]. Die Auflösung der Eingangsdaten wird für das Training des *CNNs* bei $200 \times 200\text{px}$ belassen, da die Filterung von den relevanten Informationen den *Convolution-Layers* überlassen werden dann.

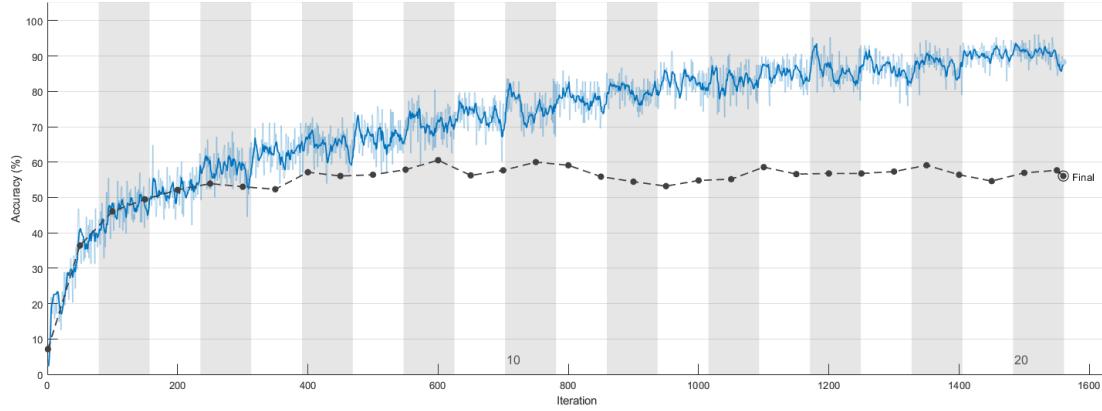


Abbildung 16: Trainingsverlauf vom Basis-CNN: Leistung mit Trainingsdaten in blau, Leistung mit Validierungsdaten in grau

Nach dem Training erreichte das *Basis-CNN* mit den Trainingsdaten eine Genauigkeit von 88.3%. Mit den Testdaten etwa ein Drittel weniger: **56.1%**. Wie in der Abbildung 16 zu erkennen ist, verbessert sich die Leistung der Trainingsdaten (blau) kontinuierlich, die der Validierungsdaten (grau) jedoch stagniert ab Epoche 8 (Iteration 600). Dies ist ein typisches Verhalten eines überangepassten Algorithmus(siehe *Überanpassung* in Kapitel 2.3.2).

3.3.4 Data-Augmentation

Um der Überanpassung von Basis-*SNN* aus dem letzten Abschnitt entgegenzuwirken, sollen die durch *Data-Augmentation* künstlich erweiterten Trainingsdaten verwendet werden. Alle Einstel-

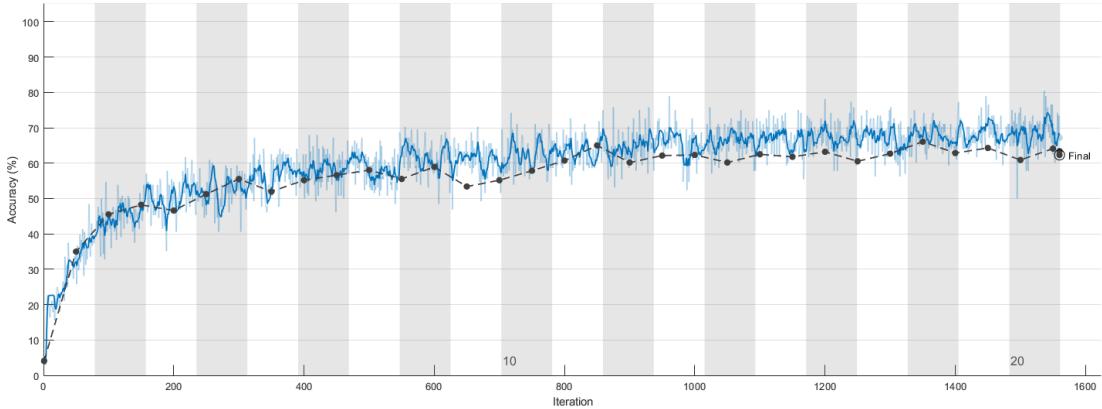


Abbildung 17: Trainingsverlauf vom Basis-CNN mit *Data-Augmentation*: Leistung mit Trainingsdaten in blau, Leistung mit Validierungsdaten in grau

lungen zur *Data-Augmentation* sowie *Meta-Parameter* sind in der Tabelle 11 im Anhang zu finden.

Nur mit künstlich erweiterten Trainingsdaten erreichte das ansonsten gleiche *CNN* eine Leistung von **63.4%**. Zudem ist in Abbildung 17 auch deutlich zu erkennen, dass die Differenz zwischen Validierungsdaten und Trainingsdaten (67.2%) massiv kleiner ist.

3.3.5 Farbfilter

Der folgende Testlauf wird mit den selben *Meta-Parameter* wie *CNN* aus 3.3.4 ausgeführt, jedoch wurde auf alle Eingangsdaten der im Kapitel 3.2.4 eingeführte Grünfilter angewandt.

Nach dem Training erreichte man ein ähnliches Resultat wie ohne Filter (60.71%), jedoch unterschieden sich die Methoden in der Laufzeit: Das Filtern der Grün-Töne für jedes Trainingsexemplar verdoppelte fast die Berechnungszeit (45 gegenüber 27 Minuten).

Aus dem ähnlichen Ergebnis lässt sich schliessen, dass das *CNN* die für die Bestimmung relevanten Farben ohne externer Unterstützung filtern kann, was wiederum auf eine robuste Netzarchitektur deutet. Aufgrund des grossen zusätzlichen Zeitaufwandes wird diese Massnahme nicht in den folgenden *CNNs* implementiert.

3.3.6 Justierung des CNNs

Da die *Meta-Parameter* des Basis-*CNNs* nur durch eine erfahrungsgestützte Schätzung eingestellt worden sind, gilt es in diesem Abschnitt die *Meta-Parameter* zu justieren. Dabei sollen weitere *Layers* und *Layer*-Typen verwendet werden, um das *CNN* weiter auf die Pilzartenerkennung zu spezialisieren. Da es für die Zusammenstellung sowie Einstellung von *Layern* unzählige Kombinationsmöglichkeiten gibt, soll nur ein kleiner Bereich möglichst systematisch untersucht werden.

Eine erste signifikante Verbesserung wurde durch die Erweiterung jeder *Convolution-Pooling-Einheit* mit je einem *ReLU*- sowie *Batch-Normalization-Layer* erreicht (68.2%). Durch weitere Exploration des Parameterraumes konnten durch Anpassung der *Feature-Map*-Anzahl noch 6 weitere Prozentpunkte gewonnen werden, was sich zu einem finalen Ergebnis von **74.6%** summiert. Zu beachten gilt es jedoch auch das Resultat der Trainingsdaten (90.6%), welche auf eine ausgeprägte Überanpassung deutet. Folgende *Meta-Parameter* wurden verwendet:

Trainings-Einstellungen	
Parameter	Wert
Lernrate η	Epochen 1-10: 0.01 Epochen 11-20: 0.001
Mini-Batch-Grösse	128
Anz. Lernepochen	20
L2 Regularisierung	0.002
Trainingsdaten mischen	nach Epoche
Bildnormalisierung	✓
<i>Data-Augmentation</i>	Drehung: $\pm 25^\circ$ X-Versetzung: $\pm 10px$ Y-Versetzung: $\pm 10px$ X-Spiegelung: ✓ Y-Spiegelung: ×
Zusatzinformationen	×

Tabelle 3: Trainings-Parameter des justierten *CNNs*

Netzaufbau	
Layer-Typ	Grösse
Bildeingabe-Layer	$200 \times 200 \times 3$
Convolution-Layer	Filter: 10×10 Schrittweite: 4 Anz. Feature Maps: 40
ReLU-Layer & Batch-Norm.-Layer	
max-Pooling-Layer	Filter: 3×3 Schrittweite: 2
Convolution-Layer	Filter: 5×5 Schrittweite: 1 Anz. Feature Maps: 80
ReLU-Layer & Batch-Norm.-Layer	
max-Pooling-Layer	Filter: 3×3 Schrittweite: 2
Convolution-Layer	Filter: 2×2 Schrittweite: 1 Anz. Feature Maps: 100
ReLU-Layer & Batch-Norm.-Layer	
max-Pooling-Layer	Filter: 2×2 Schrittweite: 2
Output-Layer	21

Tabelle 4: Aufbau des justierten CNNs

3.3.7 Transfer-Learning

In Kapitel 2.7.4 wurde erwähnt, dass bereits vortrainierte CNNs auch auf neue Datensätze umtrainiert werden können. Im Rahmen dieser Arbeit sollen daher zwei hochperformante CNNs der Bilderkennung auf das Pilzartenerkennungsproblem für *Transfer-Learning* verwendet werden: AlexNET[29] und GoogLeNet[30]. Beides sind Ergebnisse der *ImageNet Large Scale Visual Recognition Challenge*[31] und wurden mit etwa 1.2 Millionen verschiedenen Trainingsbildern auf 1000 verschiedene Kategorien trainiert.

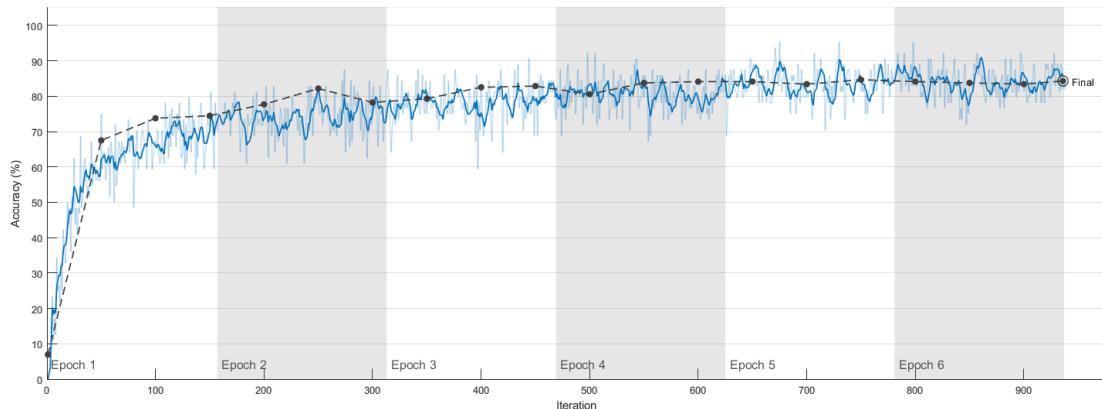
Um AlexNET resp. GoogLeNet auf das Pilzerkennungsproblem anzupassen, musste nur der letzte *Fully-Connected-Layer* (*Output-Layer*) ersetzt werden durch einen *Fully-Connected-Layer* mit 21 Neuronen. Für *Transfer-Learning* empfiehlt es sich zudem, etwas modifizierte Trainingseinstellungen zu verwenden: Der neue *Output-Layer* soll einiges stärker trainiert werden als die transferierten *Layer* aus dem trainierten Netz, weswegen die Lernraten η individuell auf 20 (*Output-Layer*) resp. 0.0001 (transferierte *Layer*) gesetzt werden. In Tabelle 5 sind nochmals alle *Meta-Parameter* aufgelistet.

Netzaufbau		Trainings-Einstellungen	
Layer-Typ	Grösse	Parameter	Wert
Bildeingabe-Layer	AlexNET: $227 \times 227 \times 3$	Lernrate η	<i>Output-Layer:</i> 20 transferiert: 0.0001
	GoogLeNet: $224 \times 224 \times 3$	Mini-Batch-Grösse	64
transferierte Layers	AlexNET: 21 Layers	Anz. Lernepochen	6
	GoogLeNet: 25 Layers	L2 Regularisierung	0.0001
<i>Output-Layer</i>	21	Trainingsdaten mischen	nach Epoche
		Bildnormalisierung	✓
		<i>Data-Augmentation</i>	Drehung: $\pm 25^\circ$ X-Versetzung: $\pm 10px$ Y-Versetzung: $\pm 10px$ X-Spiegelung: ✓ Y-Spiegelung: ×
		Zusatzinformationen	×

Tabelle 5: *Meta-Parameter für Transfer-Learning*

Durch die vortrainierten *CNNs* konnten wieder signifikant bessere Ergebnisse erzielt werden: Das mit AlexNET modifizierte *CNN* erreichte eine Genauigkeit von 83.8%, das mit GoogLeNet modifizierte sogar **84.3%**.

Wie im Trainingsverlauf des *GoogLeNet-Transfer-CNNs* (Abbildung 18) zu erkennen ist, leidet dieses *CNN* im Gegensatz zum dem aus vorherigem Abschnitt in keiner Weise an Überanpassung.

Abbildung 18: Trainingsverlauf von *Transfer-Learning* mit GoogLeNet: Leistung mit Trainingsdaten in blau, Leistung mit Validierungsdaten in grau

3.3.8 Einspeisung von Zusatzinformationen

Bis jetzt konnten schon eindrückliche Ergebnisse erzielt werden, wenn man betrachtet, dass für die Bestimmung nur mit den Informationen eines Bildes gearbeitet wurde. Um die Fehlerquote weiter einzugrenzen, sollen die in Kapitel 3.2.6 eingeführten Zusatzinformationen mit in die Pilzartenbestimmung mit einfließen. Hierfür wird das im vorherigen Abschnitt erarbeitete *GoogLeNet*-

Transfer-CNN mit dem *Information-Extraction-* und *Information-Cover-Layer* so modifiziert, damit die in den Eingangsbildern enthaltenen Zusatzinformationen direkt im *fully-connected Output-Layer* verarbeitet werden können (vgl. Abbildung 15). Da der *Output-Layer* standardmäßig Bildinformationen von über Eintausend Neuronen des letzten transferierten *Layers* erhält, wurden die vergleichsweise wenigen Neuronen der Zusatzinformationen (34) in ihrer Anzahl verzehnfacht, um in der Umenge an Bildinformationsneuronen schneller als wichtige Informationsträger detektiert zu werden.

Um das *CNN* nicht primär von den Zusatzinformationen abhängig zu machen, werden diese nicht immer zu Verfügung gestellt. Um dies zu bewerkstelligen, werden 50% der Trainingssexemplare ganz ohne Zusatzinformationen eingespeist, 10% erhalten die Informationen aller Eigenschaften und für die restlichen 40% der Bilder werden zwischen 1 und 10 der 11 Eigenschaften angegeben.

Alle *Meta-Parameter* finden sich in folgender Tabelle:

Netzaufbau		Trainings-Einstellungen	
Layer-Typ	Grösse	Parameter	Wert
Bildeingabe-Layer	$224 \times 224 \times 3$	Lernrate η	<i>Output-Layer</i> : 20 transferiert: 0.0001
<i>Extraction-Layer</i>	Abgreifen von 34 Zusatzinformationen	Mini-Batch-Grösse	48 (GPU-Limitation)
transferierte <i>Layers</i>	GoogLeNet: 25 <i>Layers</i>	Anz. Lernepochen	11
<i>Extraction-Layer</i>	Einspeisung und Verzehnfachung der Zusatzinformationen	L2 Regularisierung Trainingsdaten mischen Bildnormalisierung	0.0001 nach Epoche ✓
<i>Output-Layer</i>	21	<i>Data-Augmentation</i> Zusatzinformationen	Drehung: $\pm 25^\circ$ X-Versetzung: $\pm 10px$ Y-Versetzung: $\pm 10px$ X-Spiegelung: ✓ Y-Spiegelung: ✗ keine: 50% teilweise: 40% alle: 10%

Tabelle 6: *Meta-Parameter* für *Transfer-Learning* mit GoogLeNet und Zusatzinformationen

Durch die unzuverlässige Einspeisung der Zusatzinformationen konnte die Lücke zur 100%-Marke doch noch halbiert werden: Das finale Trainingsresultat für dieses *CNN* beläuft sich auf **92.3%** Genauigkeit. Da das *CNN* auch die Trainingsdaten mit 93.8% Genauigkeit erkennt, kann auf eine sehr geringen Überanpassung gedeutet werden. Diese finale Version wie auch die zwei Vorgänger werden als Ergebnis dieser Arbeit genauer analysiert.

4 Darstellung der Ergebnisse

In diesem Kapitel werden die in der Methode erarbeiteten Erkennungsalgorithmen veranschaulicht. Hierzu gehören das ohne *Transfer-Learning* entwickelte *CNN* aus Kapitel 3.3.6, das mittels GoogLeNet verbesserte *CNN* (*Transfer-Learning*) aus Kapitel 3.3.7 sowie die finale Version mit *Transfer-Learning* und Zusatzinformationen aus Kapitel 3.3.8.

4.1 CNN ohne *Transfer-Learning*

Dieses *CNN* wurde von Grund auf neu konfiguriert und mit den künstlich erweiterten Trainingsdaten trainiert. Es besteht insgesamt aus 16 *Layers* und mehr als 140'000 lernbaren *Parametern* (siehe Kapitel 2.4.2) und besteht hauptsächlich aus drei *Convolution-Pooling*-Einheiten mit je einem *ReLU*- und *Normalization-Layer*. Durch die Justierung einiger *Meta-Parameter* konnte dieses *CNNs* weiter auf die Pilzartenerkennung angepasst werden (*Meta-Parameter* zum Aufbau und Training in Tabellen 4 und 3 auf Seite 32 f.).

Folgende Bestimmungsgenauigkeiten konnten durch dieses justierte *CNN* erreicht werden:

Leistung mit Trainingsdaten	Leistung mit Validierungsdaten
89.4%	76.0% ²⁰

Um einen besseren Einblick in die Zusammensetzung der Fehlerquote zu erhalten, dient die in Abbildung 20 dargestellte Vertauschungsmatrix, bei der alle 560 Validierungsdaten je zehnmal vom *CNN* bestimmt und entsprechend eingetragen worden sind.

Zwar ist die genaue Funktionsweise eines trainierten *KNNs* wenn überhaupt nur schwierig nachzuvollziehen, jedoch lässt sich der erste Schritt der Informationsfilterung noch veranschaulichen. In Abbildung 19 sind die Gewichtungen der 40 *Feature-Maps* des ersten *Convolution-Layers* abgebildet. Diese zeigen, auf welche Strukturen und Muster die jeweilige *Feature-Map* anspricht.



Abbildung 19: 40 10×10px *Feature-Maps* vom erstem *Convolution-Layer* des justierten *CNNs*

²⁰Leistung gemittelt über 10 Testreihen, da durch grenzwertige Trainingsexemplare mit *Data-Augmentation* verschiedene Resultate erzielt werden können. Um trotzdem ein zuverlässiges Resultat zu erhalten, wurden alle Validierungsdaten je 10 Mal eingespeist.

	Echter Pfifferling	168	1			8		10			1			26	
	Fichteneiherz	6	153			1		2	1	4				78.5% 21.5%	
	Fichtensteinpilz		166	1				12		6		3	5	18 73.9% 26.1%	
	Flaschersäubling		6	133			1		6	8	1	2	6	1 83.0% 17.0%	
	Fliegenpilz			200			1			9			13 70.4% 29.6%	95.2% 4.8%	
	Frauentäubling	7	4		120		1	2	17			3		27 66.3% 33.7%	
	Gemeines Stockschwämmchen	3			156		9	1		2				8 87.2% 12.8%	
	Gruener Knollenblätterpilz			4	4	166	6		1	1				8 82.2% 17.8%	
	Herbsttrampete			2		146		2	1		1	3		24 80.7% 19.3%	
	Koerntchenrohring	3	2	1	9	7		111	8		1	6		2 50.2% 49.8%	
	Maronenohrling		2		9		9	2	141			18		9 73.4% 26.6%	
	Nebelkappe			1	19	9	15	1	120					5 58.5% 41.5%	
	Perlpilz		2	1		2			129	11		12		2 74.6% 25.4%	
	Riesenschimpfliz		1	7					22	165		1	9	5 68.8% 31.3%	
	Roetlicher Gallerttrichter			1		1	2			188				7 90.4% 9.6%	
	Rottusrohring	3	2		14	1		9	3	3		99		1 60.7% 39.3%	
	Schopftintling					3	1			1				2 94.1% 5.9%	
	Hallimasch		2	1		7		9		9	10	1	7		13 67.5% 32.5%
	Spitzmorchel										2	4			15 90.1% 9.9%
	Wieseneggerling			5	25		1			12	9	3			161 62.9% 37.1%
	unbekannt	18	30	15	7		36	15	13	25	52	24	37	11 76.8% 23.2%	
tatsächliche Pilzart															

Abbildung 20: Vertauschungsmatrix des justierten CNNs, Validierungsdaten je zehn mal bestimmt: Auf der Diagonalen befinden sich die Anzahl korrekt klassifizierten Validierungsdaten, die restlichen sind falsch bestimmt worden.

vorhergesagte Pilzart

4.2 CNN mit *Transfer-Learning*

Als Ausgangslage dient GoogLeNet[30], welches mittels *Transfer-Learning* auf die Pilzerkennung umtrainiert worden ist. Hierzu wurde vor dem Training der letzte *Fully-Connected-Layer* durch einen auf die Anzahl der Kategorien angepasster *Fully-Connected-Layer* ersetzt und mit zufälligen Werten initialisiert. GoogLeNet verwendet neben konventionellen *Convolution-Layers* auch sog. *Inception-Module* (detaillierter Aufbau siehe [30]), durch welche die ausserordentliche Leistung von GoogLeNet erreicht werden kann. Mit insgesamt 9 solcher *Inception-Module* weist das *Transfer-CNN* eine totale Tiefe von 27 *Layers* auf und enthält etwa 5.8 Millionen lernbare *Parameter*. Beim Training wurde mittels verschiedenen Lernraten bewerkstelligt, dass die *Parameter* des neuen *Fully-Connected-Layers* (ca. 21'000) viel schneller angepasst werden als die der transferierten *Layers*. Trainiert wurde das *Transfer-CNN* ebenfalls mit Bildnormalisierung und *Data-Augmentation* (alle *Meta-Parameter* zum Aufbau und Training siehe Tabelle 5 auf Seite 33).

folgende Ergebnisse erzielt werden:

Leistung mit Trainingsdaten

85.9%

Leistung mit Validierungsdaten

84.8%²¹

Auch für das *Transfer-CNN* ist in Abbildung 22 zur Analyse der Fehlerquote die entsprechende Vertauschungsmatrix abgebildet, wobei die Validierungsdaten wiederum je zehn Mal bestimmt worden sind.

Für den Vergleich des *Transfer-CNNs* mit dem justierten *CNN* aus vorherigem Abschnitt werden zudem wieder die *Feature-Maps* des ersten *Convolution-Layers* in folgender Abbildung aufgezeigt.

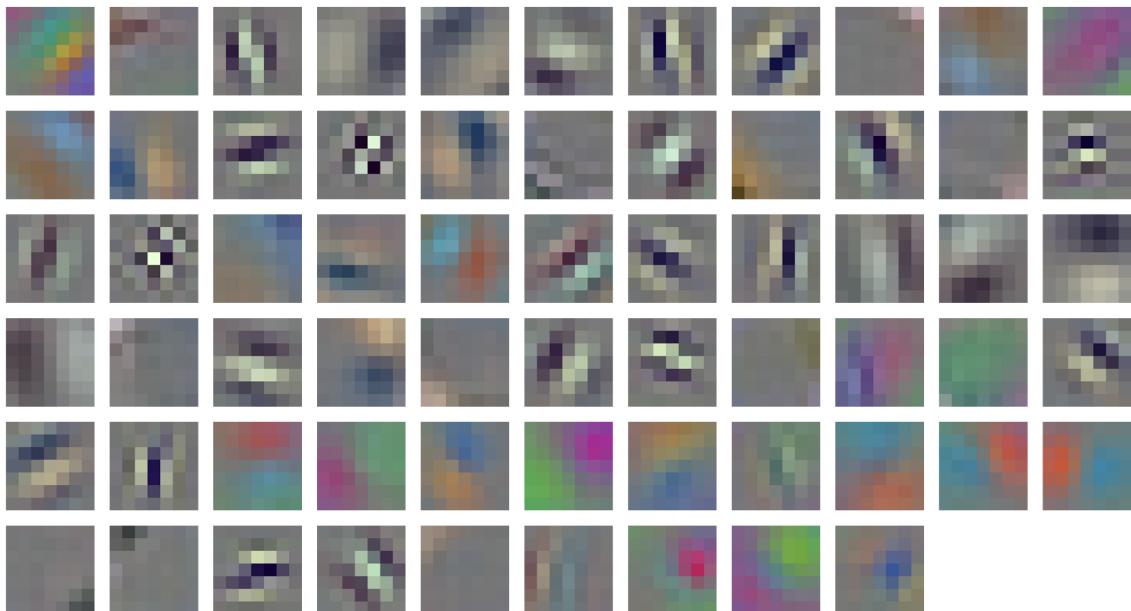
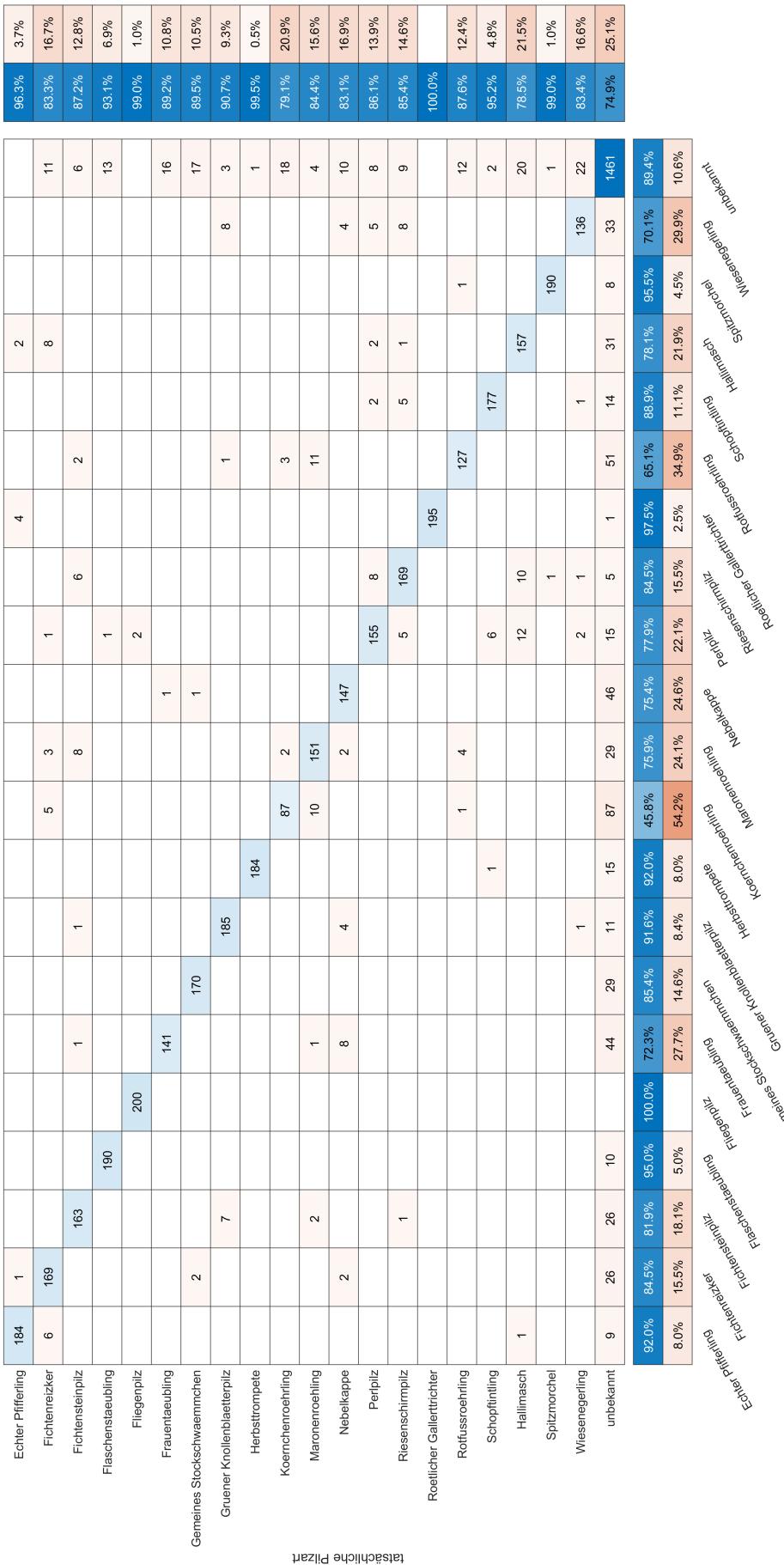


Abbildung 21: 64 7×7px *Feature-Maps* vom ersten *Convolution-Layers* des *Transfer-CNNs*

²¹Auch hier wurde das Ergebnis über 10 Testreihen gemittelt, siehe Fussnote 20



vorhergesagte Pilzart

Abbildung 22: Vertauschungsmatrix des Transfer-CNNs, Validierungsdaten je zehn mal bestimmt: Auf der Diagonalen befinden sich die Anzahl korrekt klassifizierten Validierungsdaten, die restlichen sind falsch bestimmt worden.

4.3 CNN mit *Transfer-Learning* und Zusatzinformationen

Leistung mit Trainingsdaten	Leistung mit Validierungsdaten		
50% keine, 40% teilweise, 10% alle Inf.	keine Inf.	50% der Inf.	alle Inf.
91.5%	88.0%	95.4%	98.5%

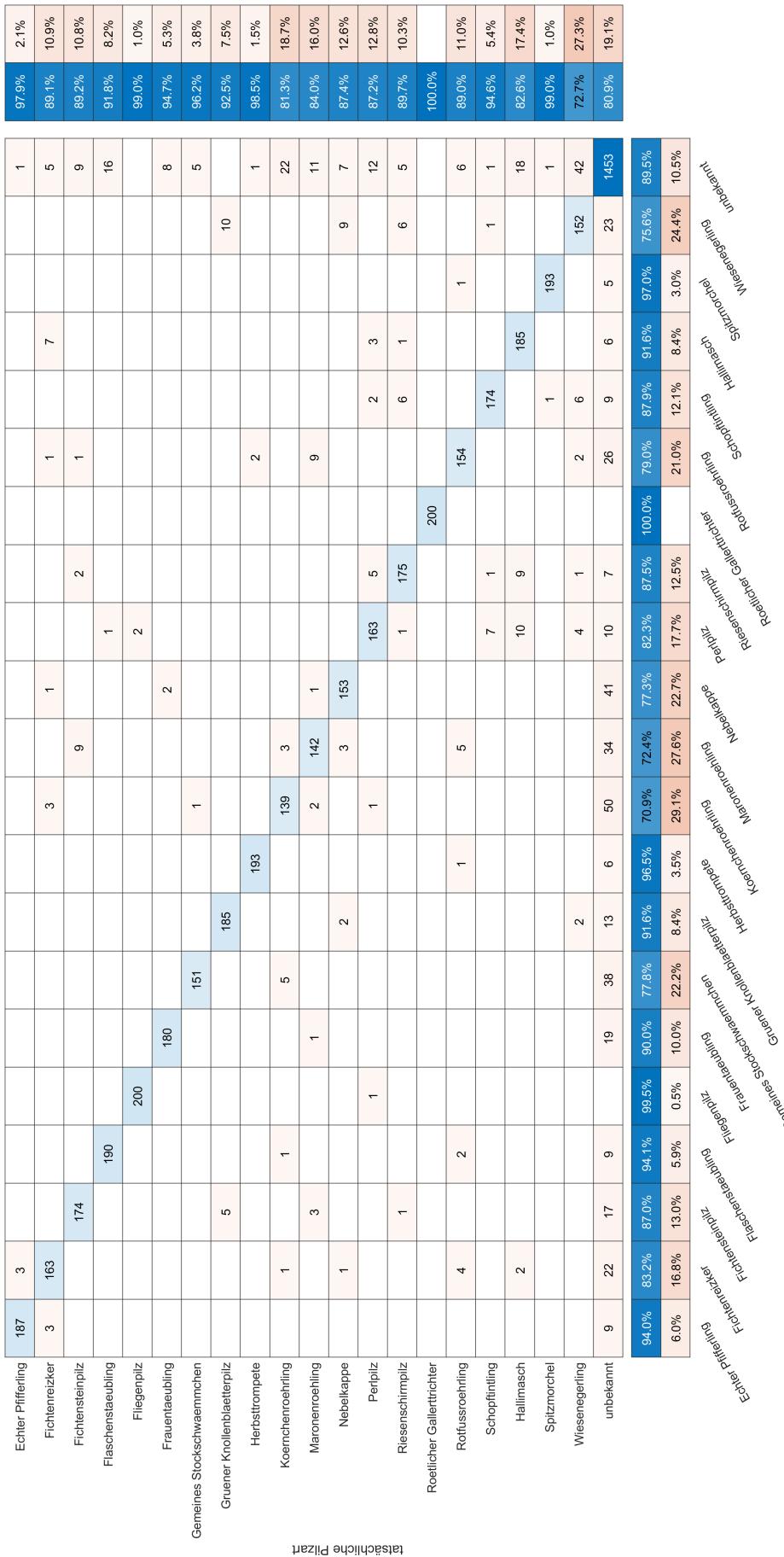


Abbildung 23: Vertauschungsmatrix des Zusatzinformationen-Transfer-CNNs, Bestimmung ohne Zusatzinformationen durchgeführt, Validierungsdaten je zehn mal bestimmt: Auf der Diagonalen befinden sich die Anzahl korrekt klassifizierten Validierungsdaten, die restlichen sind falsch bestimmt worden.

vorhergesagte Pilzart

Abbildung 24: Vertauschungsmatrix des Zusatzinformationen-Transfer-CNNs, für Bestimmung 50% der Zusatzinformationen geben, Validierungsdaten je zehn mal bestimmt: Auf der Diagonalen befinden sich die Anzahl korrekt klassifizierten Validierungsdaten, die restlichen sind falsch bestimmt worden.

	Echter Pfefferling	Fichtenreizker	Fichtensteinpilz	Flaschenstaubling	Fliegenpilz	Frauentäubling	Gemeines Stockschwämmbchen	Gruener Kropfbretterpilz	Herbsttrampete	Koerchencrährling	Maronenrohling	Nebelkappe	Perlpilz	Riesenschirmpilz	Roetlicher Gallerttrichter	Rottfussrohrling	Schopftintling	Hallimasch	Spitzmorchel	Wiesenegerling	unbekannt			
	199																						100.0%	
		198																					2	99.0% 1.0%
			189																				1	92.2% 7.8%
				193																			7	96.5% 3.5%
					200																			100.0%
						199																		1
							200																	99.5% 0.5%
								200																100.0%
									200															99.5% 0.5%
										193														97.0% 3.0%
											181													1
												200												94.3% 5.7%
													200											100.0%
														199										99.5% 0.5%
															200									99.5% 0.5%
																200								100.0%
																	193							99.5% 0.5%
																		200						100.0%
																			193					1
																				200				5
																								97.1% 2.9%
																								99.0% 1.0%
																								99.5% 0.5%
																								99.5% 0.5%
																								97.1% 2.9%
																								98.1% 1.9%

vorhergesagte Pilzart

Abbildung 25: Vertauschungsmatrix des Zusatzinformationen-Transfer-CNNs, alle Zusatzinformationen für Bestimmung gegeben, Validierungsdaten je zehnmal bestimmt: Auf der Diagonalen befinden sich die Anzahl korrekt klassifizierten Validierungsdaten, die restlichen sind falsch bestimmt worden.

5 Diskussion der Ergebnisse

5.1 Qualität der Trainingsdaten

Nicht repräsentativ

5.2 Verbesserungs- und Erweiterungsvorschläge

6 Zusammenfassung

7 Danksagung

Literatur

- [1] AlphaGo. <https://deepmind.com/research/alphago/>, 2017. Abgerufen am 30.07.2018.
- [2] Openai five. <https://blog.openai.com/openai-five/>, 2018. Abgerufen am 09.08.2018.
- [3] anonym. Deepfake. <https://www.deepfakes.club/>, 2017. Abgerufen am 09.08.2018.
- [4] Jan Obermeier Nils Schlatter. Evosim, 2018.
- [5] Jonas Harjunpää Teemu Koivisto, Tuomo Nieminen. Deep shrooms.
<https://tuomonieminen.github.io/deep-shrooms/>, 2017. Abgerufen am 25.07.2018.
- [6] Steven Borowiec. Alphago seals 4-1 victory over go grandmaster lee sedol.
<https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol>, 2016. Abgerufen am 09.08.2018.
- [7] British Go Association. Play go. <https://www.britgo.org/files/pubs/playgo.pdf>, 2014.
Abgerufen am 30.07.2018.
- [8] Claude E. Shannon. Programming a computer for playing chess. Kapitel General Considerations, 1949. Bell Telephone Laboratories, Inc., Murray Hill, N.J.
- [9] Wikipedia. Observabile universe. https://en.wikipedia.org/wiki/Observable_universe.
Abgerufen am 30.07.2018.
- [10] Wikipedia. Deep blue. [https://en.wikipedia.org/wiki/Deep_Blue_\(chess_computer\)](https://en.wikipedia.org/wiki/Deep_Blue_(chess_computer)), 2017.
Abgerufen am 09.08.2018.
- [11] Wikipedia. Artificial intelligence. https://en.wikipedia.org/wiki/Artificial_intelligence.
Abgerufen am 10.08.2018.
- [12] Wikipedia. Machine learning. https://en.wikipedia.org/wiki/Machine_learning. Abgerufen am 10.08.2018.
- [13] Wikipedia. Artificial neural network.
https://en.wikipedia.org/wiki/Artificial_neural_network. Abgerufen am 10.08.2018.
- [14] Wikipedia. Deep learning. https://en.wikipedia.org/wiki/Deep_learning. Abgerufen am 10.08.2018.
- [15] H. A. Simon. *The Shape of Automation for Men and Management*. 1965. New York: Harper & Row, S. 96.
- [16] Alan Turing. On computable numbers, with an application to the entscheidungsproblem, 1936. Proceedings of the London Mathematical Society.
- [17] Stuart Geman René Doursat, Elie Bienenstock. Neural networks and the dias/variance dilemma, 1992. Massachusetts Institute of Technology.
- [18] Welch Labs. Learning to see : Part 8. <https://www.youtube.com/user/Taylorn34/>, 2017.
Abgerufen am 11.08.2018.
- [19] Michael A. Nielsen. Neural networks and deep learning.
<http://neuralnetworksanddeeplearning.com/>. Abgerufen am 05.06.2018.

- [20] Roman Claus et al. *Natura, Biologie für Gymnasien*. 1991. Klett, ISBN 3-12-042200-2, S. 252 ff.
- [21] Halbert White Kurt Hornik, Maxwell Stinchcombe. Multilayer feedforward networks are universal approximators.
http://www.cs.cmu.edu/~epxing/Class/10715/reading/Kornick_et_al.pdf, 1989.
- [22] MatLab. imageinputlayer, normalization. <https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.imageinputlayer.html>. Abgerufen am 05.09.2018.
- [23] MathWorks. Matlab: Deep learning toolbox.
<https://www.mathworks.com/products/deep-learning.html>. Abgerufen am 05.08.2018.
- [24] Dr. Stefan Blaser Dr. Andrin Gross. Swissfungi, verbreitungsatlas der schweiz.
<https://www.wsl.ch/>, 2018.
- [25] Jan Obermeier. Pilzbestimmung in codezeilen gefasst. Ausgabe 3/2018.
- [26] hardikvasa. Google images download.
<https://github.com/hardikvasa/google-images-download>. Abgerufen am 12.09.2018.
- [27] Michael Bachmeier. Pilzwelten.
https://www.youtube.com/channel/UCezNep13jlv8Wrsaxv_saeA. Abgerufen am 20.07.2018.
- [28] Markus Flück. *Welcher Pilz ist das?* 2016. Kosmos, ISBN 978-3-440-14983-6.
- [29] Alex Krizhevsky. Imagenet classification with deep convolutional neural networks.
<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>, 2012. Abgerufen am 30.09.2018.
- [30] Christian Szegedy et al. Going deeper with convolutions.
<https://arxiv.org/pdf/1409.4842.pdf>, 2014. Abgerufen am 30.09.2018.
- [31] Prof. Li Fei-Fei et al. Imagenet. <http://image-net.org/index>. Abgerufen am 30.09.2018.

Abbildungsverzeichnis

1	Finger, https://img.livestrongcdn.com/ , 2017, Artikel 541001	10
2	Schema von einem künstlichen Neuron	13
3	Sigmoid-Funktion	14
4	Schema von kleinem <i>KNN</i>	15
5a	Beispiel für das <i>Local-Receptive-Field 1. Neuron</i>	19
5b	Beispiel für das <i>Local-Receptive-Field 2. Neuron</i>	19
6	<i>Feature-Maps</i>	20
7	<i>max-Pooling-Layer</i>	20
8	Rohdaten	24
9	<i>Beispiele Trainingsbilder</i>	25
10	Hilfsprogramm	26
11	Video-Frames	26
12	Durchschnittsbild für Bildnormalisierung	27
13	Beispiel Grünfilter	27

14	Beispiel zu <i>Data-Augmentation</i>	27
15	Extraktion und Einspeisung von Zusatzinformationen	28
16	Trainingsverlauf Basis- <i>CNN</i>	30
17	Trainingsverlauf Basis- <i>CNN</i> mit <i>Data-Augmentation</i>	30
18	Trainingsverlauf <i>Transfer-Learning</i>	33
19	<i>Feature-Maps</i> vom justierten <i>CNN</i>	35
20	Vertauschungsmatrix des justierten <i>CNNs</i>	36
21	<i>Feature-Maps</i> vom <i>Transfer-CNN</i>	37
22	Vertauschungsmatrix des transferierten <i>CNNs</i>	38
23	Vertauschungsmatrix des Zusatzinformationen- <i>Transfer-CNNs</i> ohne Zusatzinformationen	40
24	Vertauschungsmatrix des Zusatzinformationen- <i>Transfer-CNNs</i> mit 50% der Zusatzinformationen	41
25	Vertauschungsmatrix des Zusatzinformationen- <i>Transfer-CNNs</i> mit allen Zusatzinformationen	42
26	Hochladeformular & Quiz	53

Tabellenverzeichnis

1	Zusatzinformationen für genauere Pilzartenbestimmung	28
2	Bestimmungsleistung von <i>SNNs</i>	29
3	Trainings-Parameter des justierten <i>CNNs</i>	31
4	Aufbau des justierten <i>CNNs</i>	32
5	<i>Meta-Parameter Transfer-Learning</i>	33
7	ausgewählte Pilzarten	49
8	Eigenschaften der ausgewählten Pilzarten	52
9	<i>Meta-Parameter</i> des Basis- <i>SNNs</i>	54
10	<i>Meta-Parameter</i> des Basis- <i>CNNs</i>	54
11	<i>Meta-Parameter</i> des Basis- <i>CNNs</i> mit <i>Data-Augmentation</i>	55

8 Anhang

Pilzart	Wissenschaftlicher Name	Essbarkeit
Echter Pfifferling/Eierschwamm	<i>Cantharellus cibarius</i>	essbar
Fichtenreizker	<i>Lactarius deterrimus</i>	essbar
Fichtensteinpilz	<i>Boletus edulis</i>	essbar
Flaschenstäubling	<i>Lycoperdon perlatum</i>	essbar
Fliegenpilz	<i>Amanita muscaria</i>	giftig
Frauentäubling	<i>Russula cyanoxantha</i>	essbar
Gemeines Stockschwämmlchen	<i>Kuehneromyces mutabilis</i>	essbar
Grüner Knollenblätterpilz	<i>Amanita phalloides</i>	giftig
Herbsttrompete/Totentrompete	<i>Carterellus cornucopioides</i>	essbar
Körnchenröhrling	<i>Stiellus granulatus</i>	essbar
Maronenröhrling	<i>Xerocomus badius</i>	essbar
Nebelkarpe/Nebelgrauer Trichterling	<i>Clitocybe nebularis</i>	giftig
Perlpilz	<i>Amanita rubescens</i>	essbar
Riesenschirmmpilz/Parasolpilz	<i>Macrolepiota procera</i>	essbar
Rötlicher Gallertrichter	<i>Tremiscus helvelloides</i>	essbar
Rotfussröhrling	<i>Xerocomus chrysenteron</i>	essbar
Schopftintling	<i>Coprinus comatus</i>	essbar
Hallimasch	<i>Armillaria ostoyae</i>	essbar
Spitzmorchsel	<i>Morchella conica</i>	essbar
Wiesenegerling/Wiesen Champignon	<i>Agaricus campestris</i>	essbar

Tabelle 7: ausgewählte Pilzarten

MatLab Code für *Transfer-Learning* mit Zusatzinformationen

Transfer-learning (GoogLeNet) with extra information

Initialise training data

get training data

```
[imgsTrain, imgsTest, ~, ~, ~] = snippets.getShroomData;
```

get and resize average image for normalisation

```
normImg = imresize(snippets.getAugAvg, [224,224]);
```

get extra information

```
eInf = snippets.readExtraInfo;
```

set up data for training with size, normalization, data augmentation and extra information (with probabilities)

```
trainingData = customAugmentedImageDatastore([224 224 3],imgsTrain, ...
    normImg, snippets.augmentor, eInf, 0.5, 0.4, 0.1);
validationData = customAugmentedImageDatastore([224 224 3],imgsTest, ...
    normImg, snippets.augmentor, eInf, 0.5, 0.4, 0.1);
```

Set up net for transfer learning

get pretrained net (GoogLeNet)

```
tNet = googlenet;
transferLayers = layerGraph(tNet);
```

specify extra information mutiplier

```
eInfMultiplier = 10;
```

specify new input layer and cover layer

```
input = [imageInputLayer([224 224 3], 'Normalization', 'none', 'Name', 'data')
    coverLayer(snippets.numExtraInfoTrain, 0, 'cover')];
```

specify new output layer and set learning rates for new fully-connected Layer

```
fc = [depthConcatenationLayer(1 + eInfMultiplier, 'Name', 'dc1')
    fullyConnectedLayer(21,'WeightLearnRateFactor',20, ...
    'BiasLearnRateFactor',20, 'Name', 'loss3-classifier')
];
```

specify new classification layer

```
class = classificationLayer('Name', 'output');
```

specify and add information extraction layer

```
extract = extractionLayer(snippets.numExtraInfoTrain, 'extr');
transferLayers = addLayers(transferLayers, extract);
```

replace previously specified layers in pretrained net

```
transferLayers = replaceLayer(transferLayers, 'data', input);
transferLayers = replaceLayer(transferLayers, 'loss3-classifier', fc);
transferLayers = replaceLayer(transferLayers, 'output', class);
```

connect extraction layer to input and to last fully-connected layer (10x)

```
transferLayers = connectLayers(transferLayers, 'data/out', 'extr/in');
for i = 2:eInfMultiplier + 1
    transferLayers = connectLayers(transferLayers, ...
        'extr/out', strcat('dc1/in', num2str(i)));
end
```

Set training options

```
options = trainingOptions('sgdm', ...
    'MiniBatchSize', 48, ...
    'InitialLearnRate', 1e-4, ...
    'Shuffle', 'every-epoch', ...
    'ValidationData', validationData, ...
    'ValidationFrequency', 50, ...
    'ValidationPatience', 10, ...
    'Plots', 'training-progress');
```

Train network

```
[transferNet, trainingInfo] = trainNetwork(trainingData, transferLayers, options);
```

Pilzart	Standort	Hut-grösse	Hutober-fläche	Hutrand	Sporenanlage	Lamellen-haltung	Ring	Stiel	Geruch
Echter Pfifferling	Laub- & Nadelwald	2-10cm	trocken	wellig	Lamellen	herablaufend	kein	voll	angenehm
Fichtenreizker	Nadelwald	3-10cm	trocken	glatt	Lamellen	angewachsen	kein	hohl	angenehm
Fichtensteinpilz	Mischwald	6-20cm	feucht	glatt	Röhren	angewachsen	kein	voll	angenehm
Flaschenstäubling	Laub- & Nadelwald	2-6cm	trocken	kein	Fruchtmasse	kein	kein	hohl	geruchlos
Fliegenpilz	Laub- & Nadelwald	5-15cm	trocken	glatt	Lamellen	frei	hängend	voll	geruchlos
Frauentäubling	Laub- & Nadelwald	4-15cm	feucht	glatt	Lamellen	angewachsen	kein	voll	geruchlos
Stockschwämmchen	Laub- & Nadelwald	3-6cm	trocken	gerieft	herablaufend	aufsteigend	kein	voll	angenehm
Knollenblätterpilz	Laubwald	4-12cm	trocken	glatt	Lamellen	frei	hängend, gerieft	wattig	angenehm
Herbsttrompete	Laubwald	4-12cm	trocken	wellig	Fruchtschicht	kein	kein	hohl	angenehm
Körnchenröhrling	Nadelwald	2-9cm	feucht	glatt	Röhren	angewachsen	kein	voll	angenehm
Maronenröhrling	Nadel- & Mischwald	3-15cm	feucht	glatt	Röhren	angewachsen	kein	voll	angenehm
Nebelkappe	Laub- & Nadelwald	6-15cm	trocken	glatt	Lamellen	herablaufend	kein	wattig	angenehm
Perlpilz	Nadel- & Mischwald	5-15cm	trocken	glatt	Lamellen	frei	hängend, gerieft	wattig	geruchlos
Riesenschirmpilz	Laubwald	10-25cm	trocken	glatt	Lamellen	frei	doppelt	voll	angenehm
Röt. Gallerttrichter	Laub- & Nadelwald	2-5cm	trocken	wellig	Fruchtschicht	kein	kein	kein	geruchlos
Rotfussröhrling	Mischwald	3-10cm	trocken	glatt	Röhren	herablaufend	kein	voll	angenehm
Schopftintling	Wiese	3-6cm	trocken	glatt	Lamellen	frei	kein	voll	geruchlos
Hallimasch	Mischwald	5-12cm	trocken	glatt	Lamellen	angewachsen	fleischig	voll	angenehm
Spitzmorchel	Nadelwald	3-5cm	trocken	kein	Fruchtschicht	kein	kein	hohl	geruchlos
Wiesenergerling	Wiese	3-10cm	trocken	glatt	Lamellen	frei	kein	voll	angenehm

Tabelle 8: Eigenschaften der ausgewählten Pilzarten

Bilder hochladen

Schritt 1: Foto auswählen

Ziehen Sie ein oder mehrere Fotos in den untenstehenden Rahmen um diese für das Hochladen bereitzustellen. Bei mehreren Fotos empfiehlt es sich, Fotos der gleichen Art miteinander hochzuladen, da die Art nur einmal pro Hochladen spezifiziert werden kann.

Die Bilder sollen folgende Eigenschaften haben:

- Die Aufnahme bildet einen ungepfückten Fruchtkörper vollkommen ab, Perspektive wenn möglich von schräg oben.
- Mindestauflösung von 800 x 800px
- maximale Dateigröße von 5MB
- Dateiformat: JPG, JPEG, .BMP, .PNG



Schritt 2: Art auswählen

Klicken Sie die passende Pilzart an. Falls Sie eine hier nicht aufgeführte Art hochladen wollen, können Sie zuerst die "andere Pilzart" anwählen und diese mit dem deutschen Namen benennen. Bei mehreren Fotos werden ALLE Fotos mit der selben Art registriert!

Echter Pfifferling / Eierschwamm
Cantharellus cibarius

Fichten-Steinpilz
Boletus edulis

Fliegenpilz
Amanita muscaria

Gemeines Stockschwämmchen
Kuehneromyces mutabilis

Herbstrompete / Totentrompete
Carterellus cornucopioides

Maronen-Röhrling
Xerocomus badus

Perlpilz
Amanita rubescens

Rötlicher Gallertrichter
Tremicus helvelloides

Schopf-Tintling
Coprinus comatus

Spitz-Morchel
Morchella conica

nicht spezifiziert
(wenn möglich vermeiden)

Fichten-Reizker
Lactarius deterrimus

Flaschen-Stäubling
Lycoperdon perlatum

Frauen-Täubling
Russula cyanoxantha

Grüner Knollenblätterpilz
Amanita phalloides

Körnchen-Röhrling
Suillus granulatus

Nebelkappe / Nebelgrauer Trichterling
Citocybe nebularis

Riesenschrimpfz / Parasolpilz
Macrolepiota procera

Rotfuss-Röhrling
Xerocomus chrysenteron

Hallimasch
Armillaria ostoyae

Wiesen-Egerling / Wiesen-Champignon
Agaricus campestris

andere Pilzart



Bilder verifizieren

- | | |
|--|--|
| <input type="radio"/> Echter Pfifferling / Eierschwamm
<i>Cantharellus cibarius</i> | <input type="radio"/> Fichten-Reizker
<i>Lactarius deterrimus</i> |
| <input type="radio"/> Fichten-Steinpilz
<i>Boletus edulis</i> | <input type="radio"/> Flaschen-Stäubling
<i>Lycoperdon perlatum</i> |
| <input type="radio"/> Fliegenpilz
<i>Amanita muscaria</i> | <input type="radio"/> Frauen-Täubling
<i>Russula cyanoxantha</i> |
| <input type="radio"/> Gemeines Stockschwämmchen
<i>Kuehneromyces mutabilis</i> | <input type="radio"/> Grüner Knollenblätterpilz
<i>Amanita phalloides</i> |
| <input type="radio"/> Herbstrompete / Totentrompete
<i>Carterellus cornucopioides</i> | <input type="radio"/> Körnchen-Röhrling
<i>Suillus granulatus</i> |
| <input type="radio"/> Maronen-Röhrling
<i>Xerocomus badus</i> | <input type="radio"/> Nebelkappe / Nebelgrauer Trichterling
<i>Citocybe nebularis</i> |
| <input type="radio"/> Perlpilz
<i>Amanita rubescens</i> | <input type="radio"/> Riesenschrimpfz / Parasolpilz
<i>Macrolepiota procera</i> |
| <input type="radio"/> Rötlicher Gallertrichter
<i>Tremicus helvelloides</i> | <input type="radio"/> Rotfuss-Röhrling
<i>Xerocomus chrysenteron</i> |
| <input type="radio"/> Schopf-Tintling
<i>Coprinus comatus</i> | <input type="radio"/> Hallimasch
<i>Armillaria ostoyae</i> |
| <input type="radio"/> Spitz-Morchel
<i>Morchella conica</i> | <input type="radio"/> Wiesen-Egerling / Wiesen-Champignon
<i>Agaricus campestris</i> |
| <input type="radio"/> andere Pilzart | |

BESTÄTIGEN **ÜBERSPRINGEN**

deutscher Name:

Schritt 3: Hochladen

Wenn Sie oben alles ausgefüllt haben, können sie mit einem Klick unten das Hochladen starten. Sollte alles funktionieren, wird nochmals ein Bestätigungsfenster auftreten, worauf die Fotos hochgeladen werden. Status wird im Feld oben angezeigt. Falls Sie wollen, können Sie unten ihren Namen angeben, damit ich Sie in der Danksagung meiner Arbeit erwähnen kann.

Ihr Name (optional):

HOCHLADEN

Abbildung 26: Hochladeformular (links) und Quiz (rechts) von www.obermeier.ch

Netzaufbau		Trainings-Einstellungen	
Layer-Typ	Grösse	Parameter	Wert
Bildeingabe-Layer	$200 \times 200 \times 3$ oder	Lernrate η	0.001
	$50 \times 50 \times 3$ oder	Mini-Batch-Grösse	128
	$10 \times 10 \times 3$	Anz. Lernepochen	7
Fully-ConnectedLayer	100 oder	L2 Regularisierung	0.0002
	60 oder	Trainingsdaten mischen	nach Epoche
	20	Bildnormalisierung	✓
Output-Layer	21	Data-Augmentation	✗
		Zusatzinformationen	✗

Tabelle 9: *Meta-Parameter* des Basis-SNNs

Netzaufbau		Trainings-Einstellungen	
Layer-Typ	Grösse	Parameter	Wert
Bildeingabe-Layer	$200 \times 200 \times 3$	Lernrate η	0.001
Convolution-Layer	Filter: 10×10	Mini-Batch-Grösse	128
	Schrittweite: 4	Anz. Lernepochen	20
	Anz. Feature Maps: 20	L2 Regularisierung	0.0002
max-Pooling-Layer	Filter: 3×3	Trainingsdaten mischen	nach Epoche
	Schrittweite: 2	Bildnormalisierung	✓
Convolution-Layer	Filter: 5×5	Data-Augmentation	✗
	Schrittweite: 1	Zusatzinformationen	✗
	Anz. Feature Maps: 40		
max-Pooling-Layer	Filter: 3×3		
	Schrittweite: 2		
Convolution-Layer	Filter: 2×2		
	Schrittweite: 1		
	Anz. Feature Maps: 60		
max-Pooling-Layer	Filter: 2×2		
	Schrittweite: 2		
Output-Layer	21		

Tabelle 10: *Meta-Parameter* des Basis-CNNs

Netzaufbau		Trainings-Einstellungen	
Layer-Typ	Grösse	Parameter	Wert
	siehe Tabelle 10	Lernrate η	0.001
	:	Mini-Batch-Grösse	128
		Anz. Lernepochen	20
		L2 Regularisierung	0.0002
		Trainingsdaten mischen	nach Epoche
		Bildnormalisierung	✓
		Data-Augmentation	Drehung: $\pm 25^\circ$ X-Versetzung: $\pm 10px$ Y-Versetzung: $\pm 10px$ X-Spiegelung: ✓ Y-Spiegelung: ×
	Zusatzinformationen		×

Tabelle 11: *Meta-Parameter* des Basis-CNNs mit *Data-Augmentation*