

KI IN DER SCHMUTZWÄSCHE- SORTIERUNG

Entwicklung einer intelligenten
Mensch-Maschine-Interaktion zur
Identifikation von Waschkategorien

*Nico Rabethge
Kurt-Georg Ciesinger*

**WORKING
PAPER #4**



ÜBER DAS KOMPETENZZENTRUM ARBEITSWELT.PLUS

Wie wird Künstliche Intelligenz die Arbeitswelt verändern? Wie gelingt es, Veränderungen der Arbeitswelt gemeinsam zu gestalten? Und wie können Beschäftigte auf den Wandel eigentlich vorbereitet werden? Antworten auf diese Fragen liefern wir als Kompetenzzentrum Arbeitswelt.Plus.

Unserem gemeinsamen Leitmotiv **Mensch. Industrie. Morgen.** entsprechend entwickeln Hochschulen und Unternehmen aus OstWestfalenLippe im Kompetenzzentrum gemeinsam mit der IG Metall Ansätze für die Einführung von Künstlicher Intelligenz in der Arbeitswelt, beispielsweise im Hinblick auf die Arbeitsplatzgestaltung und die Qualifizierung von Mitarbeiter:innen.

ÜBER DIE WORKING-PAPER-REIHE

Damit die Ausprägung der künftigen Arbeitswelt nicht allein technologisch geprägt wird, braucht es eine **ganzheitliche Gestaltung**. Deshalb führt das Kompetenzzentrum Arbeitswelt.Plus Erkenntnisse der Arbeitsforschung im Kontext von KI-Anwendungen zusammen und entwickelt daraus passende Lösungen für mittelständische Unternehmen.

Mit dieser **Working-Paper-Reihe** geben wir Einblicke in die laufende Forschung der Wissenschaftler:innen des Kompetenzzentrums und möchten gleichzeitig einen Beitrag zur Diskussion rund um aktuelle Themen aus den Feldern Künstliche Intelligenz und Arbeitsforschung leisten.

ÜBER DIE AUTOR:INNEN



Nico Rabethge

M. Sc. Nico Rabethge hat an der WWU Informatik studiert und ist seit 2020 wissenschaftlicher Mitarbeiter im Forschungsinstitut für Kognition und Robotik (CoR-Lab) der Universität Bielefeld. Er forscht in den Bereichen Künstliche Intelligenz in der Bildverarbeitung, Erklärbare künstliche Intelligenz und Aktives Lernen.



Kurt-Georg Ciesinger

Dipl.-Psych. Dipl.-Arb.wiss. Kurt-Georg Ciesinger ist Leiter der Abteilung Forschung und Entwicklung der Deutschen Angestellten-Akademie DAA NRW. Seine Schwerpunkte sind Arbeitsgestaltungs- und Dienstleistungsforschung in den Branchen Medien und IT, Gesundheitswirtschaft, Einzelhandel, Öffentliche Verwaltung, Polizei.

ABSTRACT

Im Leuchtturmprojekt KI in der Schmutzwäschesortierung beschäftigen wir uns mit der Entwicklung eines humanzentrierten KI-gesteuerten Identifikationsmoduls zur Sortierung von Wäschestücken nach Waschkategorien. Diese Entwicklung soll in der Praxis dazu beitragen, dass die hochbelastende und potenziell gesundheitsgefährdende manuelle Sortierung von Schmutzwäsche nicht mehr wie im bisherigen Umfang von Menschen geleistet werden muss. Eine KI-gestützte Identifikation, verbunden mit einem Handhabungsgerät könnte die Sortierung in weiten Teilen autonom übernehmen.

Neben der optimalen Unterstützung von Mitarbeitenden sind die Erklärung und Transferierbarkeit der KI, sowie die Entwicklung eines Schulungsprogramms, um Akzeptanz sowie Verständnis bei den Nutzer:innen zu verstärken, Forschungsschwerpunkte des Projekts. Es soll eine Mensch-Maschine-Interaktion entwickelt werden, die es Mitarbeitenden ermöglicht die KI selbst nachzutrainieren und zu überwachen. Hierdurch ergeben sich neue, zukunftsfähige Arbeitsfelder für die Beschäftigten in Industriegeschereien.

Zuerst wurde anhand des Versuchsstandes ein initialer Datensatz aus 9405 Bildern erstellt. Zu jedem der Bilder existieren Informationen über die Farbe, den Typ und zumeist auch über die Kontamination, die Schäden und das Material der Schmutzwäsche. Anschließend wurden neuronale Netze sowohl zur Extraktion der Region des Wäschestücks als auch zu der Klassifikation der Waschkategorien trainiert, d. h. die Netze lernten zu erkennen, was zum Wäschestück gehört und was nicht sowie Wäschestücke, Farben und Verschmutzungen zu erkennen. Durch einen am aktiven Lernen orientierten Prozess konnten dabei 2091 Segmentierungen automatisch erstellt werden, d. h. ohne menschliche Eingabe der Annotationen. Für die Optimierung dieser neuronalen Netze wurden zusätzlich auch noch vortrainierte Netze nachtrainiert, was allgemein als Transfer Learning (transferiertes Lernen) bezeichnet wird. Hinsichtlich des Aktiven Lernens wurde der CEAL-Algorithmus erläutert und damit die Sicherheit der Netze bei der Klassifikation bestimmt. Für visuelle Erklärungen wurde Layerwise Relevance Propagation umgesetzt.

Wir haben erste Erkenntnisse bezüglich der Sortierung erlangen können und für einige Kategorien bereits sehr zufriedenstellende Ergebnisse erlangt. Durch die beiden neuronalen Netze konnten gute Erkennungsgenauigkeiten für alle Kategorien, insbesondere der Farbe, erzielt werden. Hierbei muss jedoch beachtet werden, dass die Datensätze und dementsprechend vor allem die Validierungsdatensätze klein sind. Es ist unbedingt notwendig, den Datensatz zu vergrößern. Dabei sollte höchste Priorität sein, die Aufnahme und Annotation in einer Wäscherei zu verlagern, um möglichst automatisiert und möglichst schnell Daten aufnehmen und annotieren zu können.

1 Einleitung

Problemstellung

In der industriellen Wäschereitechnik wird die Sortierung nach Waschkategorien von Schmutzwäsche bisher entweder gar nicht, anhand von Barcodes und RFID-Chips oder unter Sicherheitsmaßnahmen, händisch von Mitarbeiter:innen vorgenommen. Letzteres erfordert den Kontakt des Menschen mit der verschmutzten und oftmals auch kontaminierten Wäsche, was eine große physische und psychische Belastung für die Mitarbeiter:innen darstellt und zudem ein erhebliches gesundheitliches Risiko in sich birgt. Diese Problematik ergibt sich bei fast allen industriellen Wäschereien (allein in Deutschland mehrere hundert Unternehmen), sodass eine Unterstützung der Wäschereien in diesem Problemfeld eine gewaltige Verbesserung der Arbeitsbedingungen für abertausende Mitarbeiter:innen in diesen Betrieben bedeutet.

Die Anwendung der KI in der Wäschereibranche ist aktuell zudem sehr problematisch. Wie in anderen Branchen auch ist der Begriff der KI extrem negativ behaftet, weil die generelle Befürchtung besteht, dass durch KI Arbeitsplätze vernichtet werden. Einer Einführung von KI wird also seitens der Beschäftigten und deren Vertretungen Skepsis entgegengebracht. Diese ist in weiten Teilen berechtigt und muss Gegenstand betrieblicher Einführungsstrategien sein.

In einigen Bereichen beruht die Skepsis aber auch auf Unkenntnis und falschen Vorstellungen von KI und deren praktischen Einsatzmöglichkeiten. Das tatsächliche Ziele der und die tatsächlichen Veränderungen durch die KI im Arbeitsprozess müssen verdeutlicht werden. Eine Aufklärung durch Weiterbildung der Beschäftigten ist damit eine zentrale Voraussetzung für gelingende KI-Einführung.

Zielsetzung und Lösungsansatz

Ziel ist, den Kontakt der arbeitenden Menschen mit verschmutzter und evtl. kontaminierter Wäsche durch die automatische Klassifikation der in der Wäscherei angelieferten verschmutzten Wäsche zu minimieren, sodass diese ohne direkten menschlichen Kontakt möglichst sortenrein behandelt werden kann.

Aufbauend auf einer vorhandenen automatischen Maschine zur Wäschevereinzelnung soll eine primär auf Kamerabilder und KI basierende Lösung zur Schmutzwäschesortierung gemäß Waschprogrammen entwickelt werden.

Hierbei muss berücksichtigt werden, dass in unterschiedlichen Wäschereien verschiedene Arten von Wäsche gewaschen werden, sodass zudem ein Mensch-Maschine-Interaktions-Modul zum intuitiven Eintrainieren der konkreten Anforderungen einer spezifischen Wäscherei mittels Verfahren des „Aktiven Lernens“ mitentwickelt wird.

Außerdem soll das gelernte Wissen des hierbei verwendeten künstlichen neuronalen Netzes nicht als „Black-Box“ fungieren, sondern es sollen aktuelle Ansätze der „Explainable Artificial Intelligence“ (Erklärbare künstliche Intelligenz) auf die Bedürfnisse der Mitarbeiter:innen in mittelständischen Firmen angepasst und weiterentwickelt werden.

Dadurch entsteht eine neue Arbeitsteilung zwischen Mensch und KI-gesteuerter Maschine. Es gilt, durch eine überzeugende ganzheitliche Gestaltung und einen fairen Einbezug von KI in die Verbesserung der Arbeitsgestaltung aktiv positive Aufklärungsarbeit zu leisten, sodass die Möglichkeiten der Anwendung von Methoden der KI in der Arbeitswelt von arbeitenden Menschen als positiver Nutzen erkannt werden. Dieser Einführungsprozess der KI-Steuerung soll durch einen Qualifizierungsprozess unterstützt werden, der die lernungewohnten Mitarbeiter:innen auf die Kooperation mit dem KI-System vorbereitet und schult. Hierbei stehen nicht nur die fachlichen Schulungen im Vordergrund, sondern es soll auch ein grundlegendes Verständnis der Funktionsweise dieses KI-Systems vermittelt werden, um dadurch die Vorbehalte gegen den Einsatz dieses KI-basierten Systems auf eine rationale Basis zu stellen.

Durch die Erweiterung von Waschstraßen um eine Komponente zur Schmutzwäsche-klassifikation verbessern sich die Arbeitsbedingungen im Bereich der Schmutzwäsche-sortierung deutlich. Dies ist für die Mitarbeiter:innen in industriellen Wäschereien vorteilhaft, erleichtert den meist mittelständischen Unternehmen die Gewinnung von Arbeitskräften für diesen nicht sonderlich beliebten Arbeitsbereich und festigt die Stellung der Firma Kannegiesser als technologisch äußerst innovatives Unternehmen.

2 Methodik und Konzeption

In diesem Kapitel wird zunächst die Konzeption des Projekts sowie der Anlage besprochen. Anschließend wird auf die KI-Methoden eingegangen, auf welche wir in diesem Projekt zurückgreifen. Zusätzlich werden weitere Methoden zum Umgang mit der KI und Erklärung der KI erläutert, die es auch Personen ohne Fachwissen ermöglichen sollen mit dieser zu interagieren.

2.1 Aufbau eines Versuchsstandes

Als erster Schritt wurde für einen Versuchsstand eine Anlage konzipiert, welche für die Erkennung und Klassifikation von einzelnen Wäschestücken verwendet werden soll.

Die erste Version dieses Moduls, welches zur Aufnahme eines initialen Datensatzes verwendet wurde, ist in Abbildung 1 zu sehen. Inmitten dieses Moduls befindet sich ein dunkelgraues Förderband, über welches die Wäschestücke vereinzelt durch das Modul fahren. Die Geschwindigkeit des Förderbandes beträgt dabei 0.6m/s. Über dem

Förderband ist eine FRAMOS D415e Tiefenkamera mit einer Auflösung von 1280x720 Pixeln montiert.

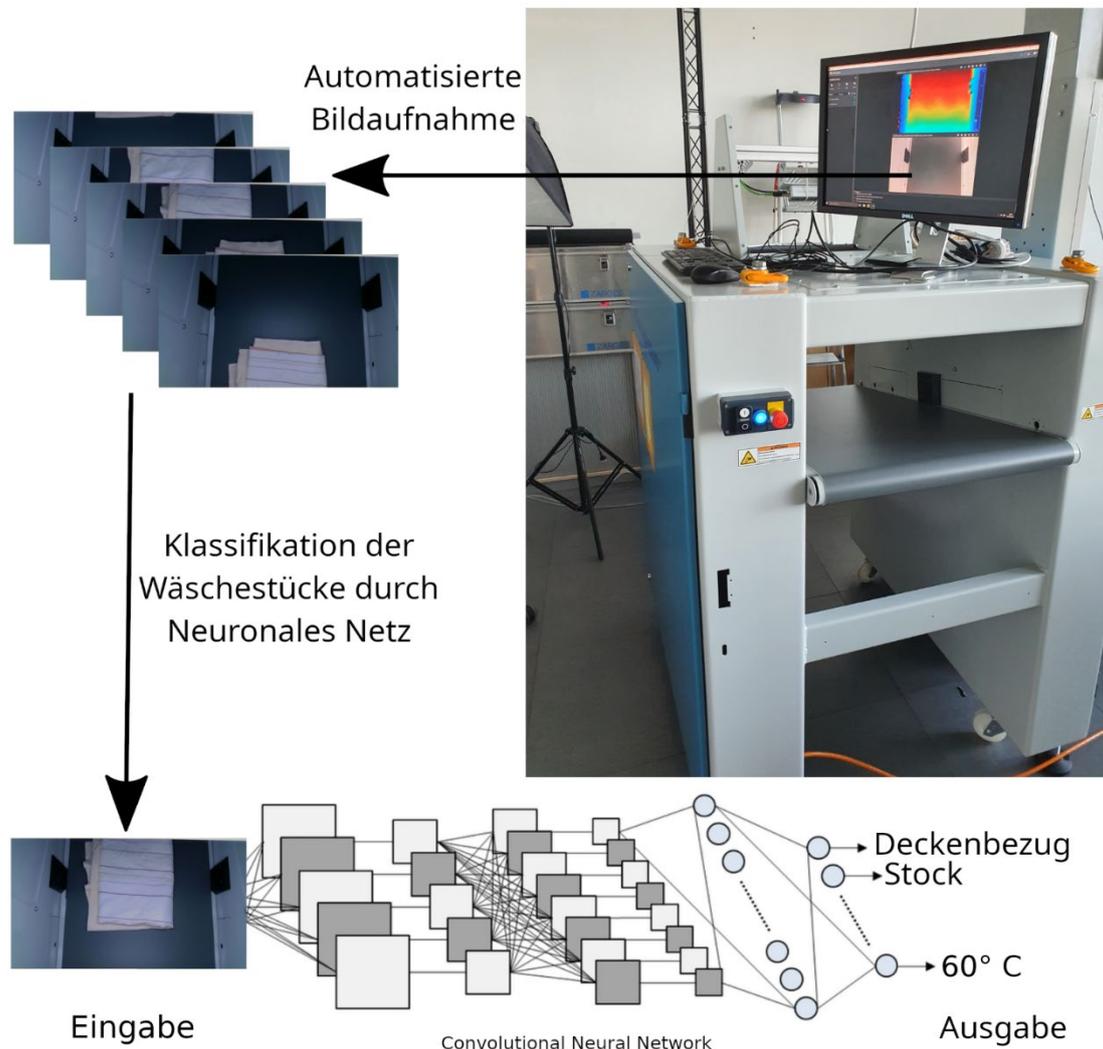


Abbildung 1: Aufbau des Versuchszustandes. Oben rechts ist das Erkennungsmodul abgebildet. Läuft ein Wäschestück über das Förderband, werden fünf verschiedene Bilder abgespeichert (oben links). Anschließend werden diese einzeln in ein neuronales Netz eingegeben, welches die Klassen des Wäschestückes vorhersagt (unten).

Um Umwelteinflüsse, wie die Tageszeit oder die Gegebenheiten des Standortes herauszufiltern ist die Förderbandoberfläche von oben, unten und den Seiten durch die Wände und das Förderband selbst isoliert. Für die Beleuchtung sorgen zwei LED-Streifen vor und hinter der Kamera.

Für die Aufnahme werden die Wäschestücke vereinzelt und an drei unterschiedlichen Positionen auf das Förderband fallen gelassen. Dies soll den Abwurf eines Greifarmes simulieren, welcher in Zukunft die Vereinzlung und den Abwurf der Wäschestücke übernimmt. Anschließend läuft das Wäschestück über das Förderband und erreicht den Aufnahmebereich der Kamera. Wenn die Kamera nun ein Wäschestück detektiert, werden

mehrere Bilder in gleichmäßigen Abständen von dem Wäschestück aufgenommen. Dies ist notwendig, da einige Wäschestücke zu groß für den Aufnahmebereich sind. Zudem ist es dadurch möglich, schneller an mehr Daten zu gelangen. Dies war von Vorteil, da für die ersten Versuche wenig Kapazität an Wäschestücken und Personal für den Aufnahmeprozess zur Verfügung standen. Ein weiterer Nachteil dieser Methodik ist, dass das Wäschestück nur von einer Seite fotografiert wird und zudem auch noch durch den Abwurf zusammengeknüllt sein kann. Eigenschaften wie die Farbe, der Typ und das Material sollten jedoch trotzdem erkennbar sein. Bei Verschmutzungen, die jedoch nur auf der Unterseite sichtbar sind, ist es nicht möglich, diese von oben zu fotografieren. Demnach wären die Verschmutzungen auf den Bildern nicht zu sehen. Diese Einschränkung musste jedoch getätigt werden, da in vielen Wäschereien nicht genügend Platz vorhanden ist, um die Wäsche aufzuhängen. Eine mögliche Lösung für diese Problematik könnte eine Vorrichtung sein, welche das Wäschestück auf die andere Seite dreht, woraufhin ein weiteres Bild aufgenommen wird.

Im nächsten Schritt müssen zu den Bildern die jeweiligen Klassen, nach denen sie später sortiert werden, gespeichert werden. Dieser Prozess wird auch Annotation genannt. Die Kategorien, nach denen die Wäsche sortiert wird, lauten zunächst: Farbe (weiß, schwarz, Rottöne, Blautöne, ...), Typ (T-Shirt, Hose, Kittel, ...), Verschmutzungsgrad (keine, stark, infektiös, ...), Verschmutzungsart (Tinte, Stock, Blut), Schäden (keine, chemisch, mechanisch) und Waschtemperatur (30°, 60° und 90 (75)°). Dabei sind die Klassen sehr fein gewählt. Eine Anpassung für gröbere Klassen zur Sortierung ist dadurch auch möglich.

Dazu wurde zusammen mit einem Textilmeister ein Leitfadens erarbeitet, in dem die möglichen Klassen aller Kategorien aufgelistet und beschrieben wurden. Anhand dieses Leitfadens werden die aufgenommenen Bilder annotiert.

Insgesamt wurden an dem Versuchsstand 9405 Bilder aufgenommen und annotiert. Davon sind 6855 Bilder (72.89%) von Wäschestücken, welche aus Pflegeheimen stammen und 2550 Bilder (27.11%) von Arbeitsbekleidung. Da bei jedem Durchlauf eines Wäschestücks fünf Bilder gespeichert werden und jedes Wäschestück drei Mal von unterschiedlichen Startpositionen aus über das Förderband fährt, sind insgesamt 1881 Durchläufe mit 627 verschiedenen Wäschestücken getätigt worden.

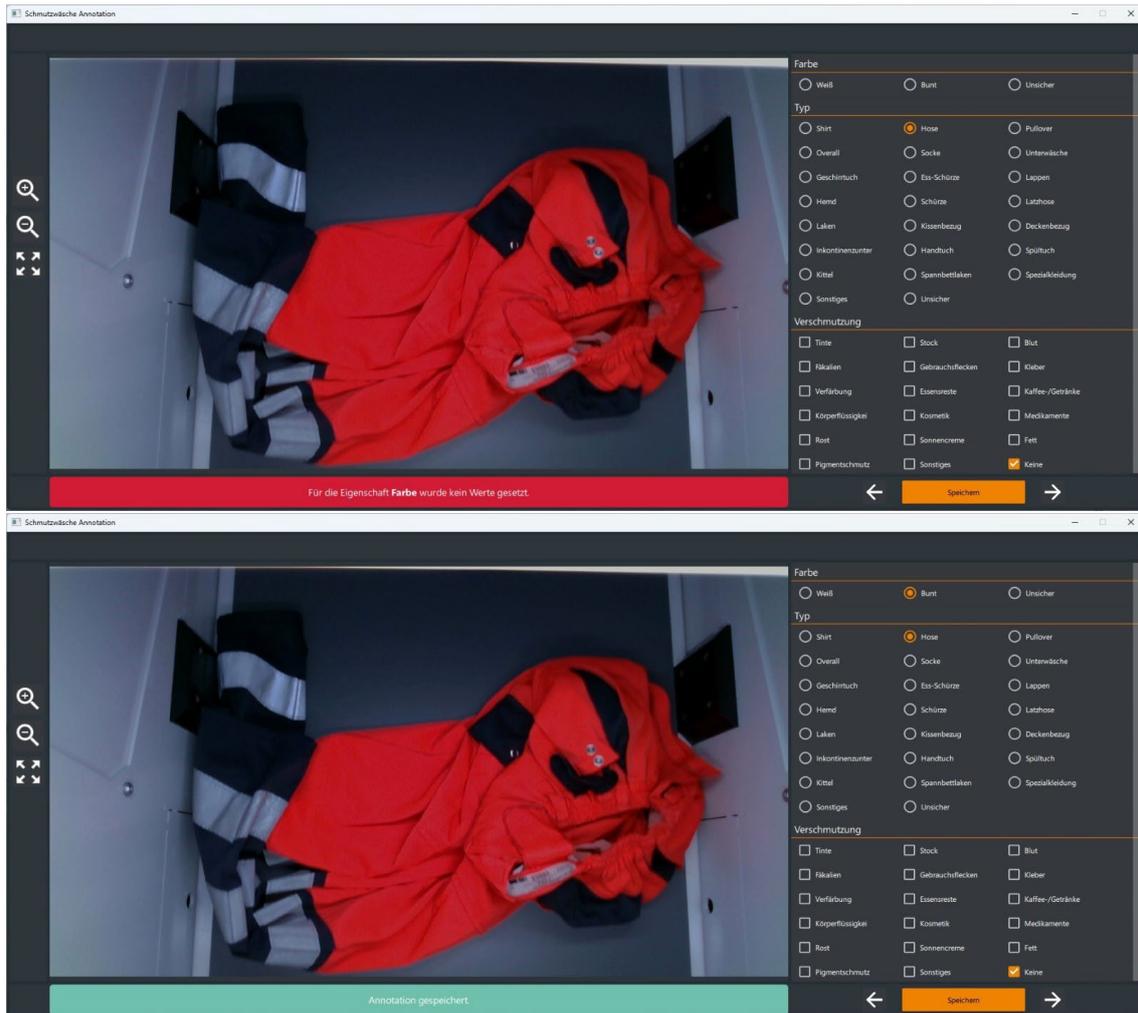


Abbildung 2: Software zur Annotation der Wäschebilder. Zentral ist das Bild des Wäschestückes zu sehen. Es ist möglich, das Bild zu verschieben oder zu zoomen, um beispielsweise Mängel besser erkennen zu können. Auf der rechten Seite ist die Einordnung in die verschiedenen Klassen vorzunehmen. Ist dies erfolgt, klickt man auf den Speichern-Button unten rechts. Wenn keine reguläre Auswahl getroffen wurde (oben), färbt sich die Informationsleiste unten rot und zeigt den Grund an. In diesem Fall korrigiert man die Annotation und bei erfolgreicher Korrektur färbt sich die Informationsleiste grün und die Annotation wird gespeichert (unten).

2.2 KI in der Bildverarbeitung

Da wir uns im Kompetenzzentrum Arbeitswelt.Plus mit modernen KI-Methoden beschäftigen, fokussieren wir uns auf die Verwendung von Convolutional Neural Networks (CNN) (vgl. LeCun, 1989). Dies sind künstliche neuronale Netze, welche durch ihre Konzeption besonders für Bilddaten geeignet sind.

Um den Umgang mit diesen Methoden den Mitarbeiter:innen zu vereinfachen, beschäftigen wir uns ebenso mit Methoden des Aktiven Lernens und der Erklärbarkeit von künstlicher Intelligenz.

2.2.1 Convolutional Neural Networks

Ein künstliches neuronales Netz besteht typischerweise aus einer Aneinanderreihung von mehreren Schichten. Kennzeichnend für ein CNN sind die sogenannten Convolutional- (Faltungen) und Pooling-Schichten. Unter einer Faltung versteht man Filter, welche wie ein Fenster über die jeweilige Eingabe gehen und optimalerweise besonders stark auf ein Muster reagieren. Beispielsweise werden bei einer Faltung auf Bilddaten in den ersten Schichten einfache Muster wie Linien oder Kanten erkannt. Aufbauend auf diesen Kanten werden in der folgenden Schicht Ecken detektiert – bis hin zu komplexen Mustern, wie beispielsweise Latzhosen, in den hinteren Schichten. Dies ist exemplarisch in Abbildung 3 dargestellt.

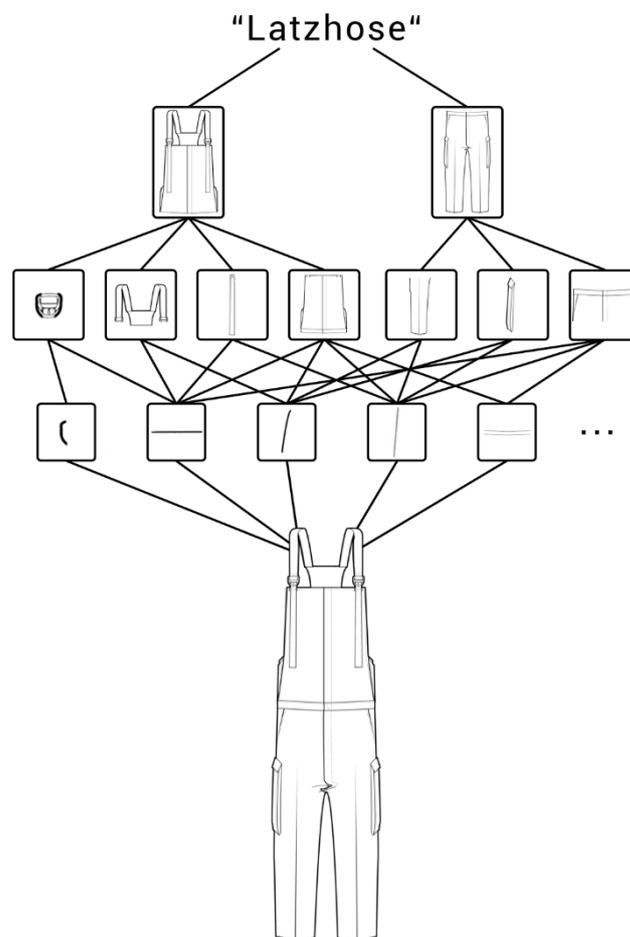


Abbildung 3: Die visuelle Welt bildet eine räumliche Hierarchie von visuellen Modulen: Elementare Linien oder Texturen verbinden sich zu einfachen Objekten wie Taschen oder Trägern, die sich zu übergeordneten Konzepten wie „Latzhose“ verbinden.

Die Pooling-Schichten sind hinter den Convolutions angeordnet und verkleinern die Ausgabe der vorherigen Schicht, indem sie diese in Bereiche aufteilen und zusammenfassen. Dazu wird üblicherweise das Maximum oder der Durchschnitt der Werte des jeweiligen Bereichs betrachtet. Nach diesen Kombinationen an Schichten werden abschließend vollverbundene Schichten aus künstlichen Neuronen zur Klassifikation verwendet.

Die letzte Schicht enthält dabei üblicherweise so viele Neuronen, wie es verschiedene Klassen der zu bestimmenden Kategorie gibt, welche das spezifische Netz unterscheiden soll. Ab einer höheren Anzahl dieser Schichten spricht man von einem tiefen neuronalen Netz.

Überwachtes Training neuronaler Netze

Neben der Art der Schichten und des Aufbaus ist auch die Spezifikation des Lernalgorithmus erforderlich. Zu Beginn werden die Parameter der Schichten zufällig initialisiert. Die Ausgabe des Netzes obliegt demnach ebenso dem Zufall. Während des sogenannten überwachten Trainings werden dem Netz Eingabedaten gezeigt, zu denen es bereits die korrekte Ausgabe gibt. Durch den Vergleich der Ausgabe des neuronalen Netzes und der gewünschten Ausgabe wird der Fehler für das aktuelle Beispiel bestimmt. Dieser Fehler wird durch das Netz „zurückpropagiert“, d.h. die Einflüsse der verschiedenen Parameter auf den Fehler werden bestimmt. Abhängig davon werden die Parameter der Schichten und Neuronen so angepasst, dass der Fehler (abhängig von dem aktuellen Trainingsbeispiel) beim nächsten Durchlauf geringer ist. Dies wird üblicherweise wiederholt, bis sich der Fehler nicht mehr signifikant verändert und das Netz eine gute Approximation (Näherung) der gewünschten Ausgabe produziert.

Das Training wird von mehreren Parametern beeinflusst, den sogenannten Hyperparametern, welche im Gegensatz zu den Parametern des Modells vor dem Training definiert werden und die Optimierung steuern. Beispiele sind die Lernrate, welche die Stärke der Anpassung pro Lernschritt steuert oder die Batch Size, welche die Anzahl der Trainingsbeispiele angibt, die pro Optimierungsschritt betrachtet werden. Häufig findet im Rahmen einer Hyperparameteroptimierung eine Rastersuche durch verschiedene Werte für die jeweiligen Hyperparameter statt, um eine optimale Kombination dieser zu erlangen.

Die Performanz eines neuronalen Netzes wird dann abschließend auf einem Validierungsdatensatz bestimmt. Dieser beinhaltet Daten, welche nicht während des Trainings verwendet wurden.

Das Netz lernt ohne die Verwendung von explizitem Wissen, wie es von der Eingabe eine Approximation der echten Lösung berechnet. Es bedarf keiner Analyse oder technischen Aufarbeitung der Eingabedaten, sondern nur der korrekten Ausgaben. Dies wird als End-to-End-Learning bezeichnet.

Mit der Verwendung von neuronalen Netzen entstehen jedoch auch Herausforderungen und Hindernisse. Nachteile sind unter anderem der Bedarf an einer enormen Menge von Trainingsdaten, das Finden einer Architektur und das oben beschriebene Suchen nach einer guten Kombination von Hyperparametern für ein optimales Training des Netzes sowie das Verstehen des Entscheidungsweges des trainierten Netzes. Auf diese Herausforderungen gehen wir im Folgenden ein.

Architekturen

Die Art und Weise, in welcher die Neuronen angeordnet und verbunden sind wird als Architektur des künstlichen neuronalen Netzes bezeichnet.

In diesem Projekt haben wir uns bisher mit Architekturen befasst, welche zur Zeit ihrer Publikation die beste Performanz (vgl. <https://paperswithcode.com/sota/image-classification-on-imagenet>) auf dem ImageNet-Datensatz¹ erzielt haben. Der ImageNet-Datensatz besteht aus über einer Million Bildern aus 1000 verschiedenen Klassen und diente jahrelang als Grundlage für den renommierten ILSVRC² Wettbewerb.

Für die Klassifikation wurden zunächst folgende CNN-Architekturen implementiert:

- VGGNet (vgl. Simonyan & Zisserman, 2015)
- InceptionNet (vgl. Szegedy, 2014)
- ResNet (vgl. He, 2015)
- DenseNet (vgl. Huang, 2016)
- EfficientNetB0-B7 (vgl. Tan & Le, 2019)

Jede Architektur bringt dabei bedeutsame Neuerungen für das Forschungsgebiet mit sich und steigt bis auf das EfficientNet in der Komplexität und Größe zu den Vorgängern.

Um erste Erkenntnisse zu sammeln wurde ein VGGNet (vgl. Simonyan & Zisserman, 2015) und ein DenseNet (vgl. Huang, 2016) trainiert. Grundlage für diese Entscheidung war die niedrige Komplexität des VGGNet und die Parametereffizienz des DenseNets. Für jede dieser Architekturen wurde jeweils ein Netzwerk pro Kategorie trainiert und optimiert, um die Komplexität anfangs möglichst gering zu halten.

Transfer-Learning

Für die Optimierung wurden zusätzlich auch noch vortrainierte Netze nachtrainiert, welches allgemein als Transfer Learning (transferiertes Lernen) bezeichnet wird. Man verwendet dabei ein Modell, welches bereits für eine andere, bestenfalls ähnliche, Problemstellung optimiert wurde. Dies basiert auf der Annahme, dass ein Teil des Lernfortschrittes dieses Modells auch für die neue Problemstellung erworben werden müsste. Man startet demnach nicht von Null, da grundlegende Konzepte bereits erlernt wurden. Dadurch ergeben sich Vorteile wie weniger Ressourceneinsatz und eine schnellere Erstellung des Modells.

¹ <https://www.image-net.org/index.php>

² <https://www.image-net.org/challenges/LSVRC/>

Für das Vortraining wurde der ImageNet-Datensatz verwendet. Für das Nachtrainieren war es notwendig, die letzte Schicht auszutauschen, da wie bereits erwähnt die Anzahl der Neurone der Anzahl der möglichen Klassen pro Kategorie entspricht. Je nach Kategorie liegen diese zwischen 2 (Farbe) und 22 (Typ). Anschließend wird diese Schicht und der restliche Teil der vollvernetzten Schichten am Ende des Netzwerks nochmal für einige Zeit nachtrainiert, damit die bereits erlernten Konzepte zunächst bestmöglich den neuen Klassen zugeordnet werden.

Anschließend werden einige weitere Schichten im hinteren Teil der Faltungsschichten ebenso nachtrainiert, um die erlernten Konzepte der hinteren Schichten ebenso auf die neue Problemstellung anzupassen.

2.2.2 Aktives Lernen

Bei der Anforderung an die Datenmenge befindet man sich abhängig von der Komplexität der Aufgabenstellung und Anzahl der Klassen idealerweise im Stellenbereich von Tausenden bis Millionen annotierten Daten (vgl. Sun, 2017). Durch die erforderliche Annotation dieser Daten wird der Aufwand zusätzlich erhöht. Daher ist es sinnvoll, Strategien zu erarbeiten, welche die Anzahl der benötigten annotierten Daten verringern oder die Annotation vereinfachen. Solche Algorithmen werden dem Gebiet des Aktiven Lernens zugeordnet.

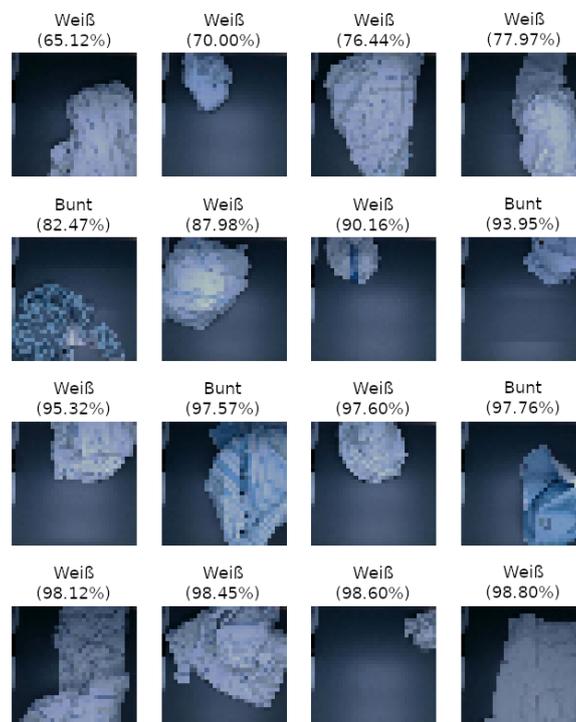


Abbildung 4: Exemplarische Klassifikation und zugehörige Sicherheit für 16 Wäschestücke während eines Trainings eines neuronalen Netzes. Die Sicherheit ist hier die Ausgabe des Neurons, welches mit der wahrscheinlichsten Klasse assoziiert ist.

Ein grundlegender Ansatz dabei ist es, zuerst einen kleinen Teil der Daten zu annotieren und damit das Training des neuronalen Netzes zu beginnen. Ist dieses abgeschlossen, werden neue, noch ungesehene Daten durch das trainierte Netz propagiert. In gewissen Fällen kann die Ausgabe des Netzes als Wahrscheinlichkeit für die jeweiligen Klassen interpretiert werden (siehe Abbildung 4). Daher ist es möglich, anhand der Verteilung der Klassenwahrscheinlichkeiten die Unsicherheit des Netzes bei der Klassifizierung einer Eingabe zu bestimmen, d.h. welche Daten sicher und welche unsicher vom Netz klassifiziert wurden. Zur Bestimmung der Sicherheit/Unsicherheit des Netzes ist die Bestimmung von Auswahlkriterien notwendig. Ein einfacher Ansatz dafür wäre, die geringste der maximalen Wahrscheinlichkeiten aller Vorhersagen als die unsicherste Vorhersage zu identifizieren.

Für die unsicher klassifizierte Daten wäre es sinnvoll, von Experten annotiert und für das nächste Training verwendet zu werden. Falls außerdem eine bestimmte Klasse die Menge der unsicher klassifizierten Daten dominiert, ist es des Weiteren sinnvoll, bei der Aufnahme neuer Daten vermehrt Daten dieser Klassen zu verwenden. Die vom Netz sicher klassifizierte Daten können ebenso sinnvoll verwendet werden.

Dies nutzen die Autoren des CEAL-Algorithmus (vgl. Wang & Zhang, 2016) (kosteneffektives Aktives Lernen) und schlagen vor, die sicheren Klassifizierungen, welche durch das neuronale Netz erhalten wurden, zu übernehmen und zusätzlich zu den bereits annotierten Daten für ein erneutes Training der gleichen KI zu verwenden. Das Übernehmen von Annotationen eines Algorithmus ohne diese zu überprüfen wird auch als Pseudo-Annotation bezeichnet. Neben den Pseudo-Annotationen werden des Weiteren auch die Bilder verwendet, welche vom Algorithmus am unsichersten klassifiziert wurden. Diese bieten vermutlich die meisten Informationen für den Algorithmus, benötigen jedoch eine menschliche Annotation. Anschließend wird das neuronale Netz mit den alten und neuen händischen sowie Pseudo-Annotationen trainiert. Dies wird wiederholt, bis entweder eine zufriedenstellende Erkennungsleistung des neuronalen Netzes erreicht wurde oder der komplette Datensatz annotiert ist. Dieser Prozess ist in Abbildung 5 dargestellt.

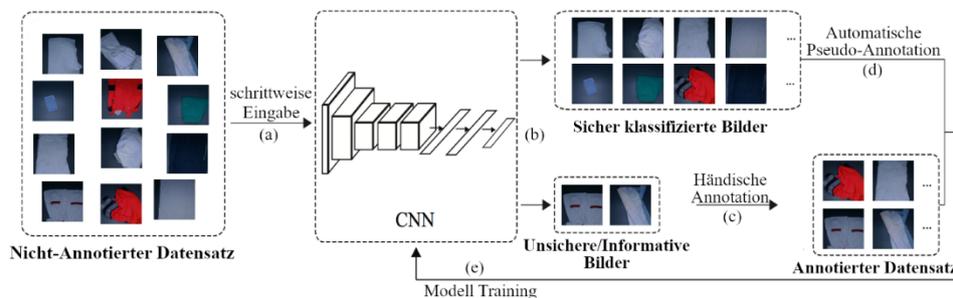


Abbildung 5: Illustration des von Wang et al. vorgeschlagenen CEAL-Paradigmas. Das CEAL speist nach und nach Stichproben aus dem nicht annotierten Datensatz in ein neuronales Netz ein (a). Dann werden die Auswahlkriterien für die eindeutig klassifizierte/sicheren Bilder und die informativsten/unsicheren Bilder auf die Klassifizierungsergebnisse des neuronalen Netzes angewandt (b). Nach dem Hinzufügen von benutzerannotierten, unsicheren Bildern in den annotierten Datensatz (c) und den Pseudo-Annotationen von nicht benutzerannotierten, sicheren Bildern (d) wird das Modell weiter aktualisiert/trainiert (e).

2.2.3 Erklärbare künstliche Intelligenz

Wie bereits erwähnt, ist der Entscheidungsweg tiefer neuronaler Netze schwer nachvollziehbar. Künstliche neuronale Netze enthalten üblicherweise Millionen von Parametern und Verknüpfungen, weswegen es sehr schwierig für einen Menschen ist, den Überblick zu behalten. Daher werden CNNs auch als Black-Box-Modelle bezeichnet.

Methoden der Explainable AI (XAI) versuchen, diese Black-Box zu durchleuchten, so dass die Entscheidungen des neuronalen Netzes nachvollzogen werden können. Populäre Methoden sind beispielsweise kontrafaktische Methoden und die Layerwise Relevance Propagation (LRP).

Bei kontrafaktischen Methoden wird die Eingabe wiederholt in verschiedenen Regionen verändert (z.B. durch einen schwarzen Balken) und dabei die Klassifizierung des neuronalen Netzes betrachtet. Die Gebiete, bei der die größte Änderung in der Klassifikation durch die Störung resultiert, können als maßgeblich und Grund für die Klassifizierung gesehen werden.

Layerwise Relevance Propagation (vgl. Lapuschkin & Binder, 2015) betrachtet ebenso die Ein- und Ausgaben des Netzes. Intuitiv gesehen verwendet LRP die Netzwerkgewichte und die Aktivierungen der Neurone, die durch den Vorwärtspass (Berechnungen aller Netzwerkschichten) erzeugt werden, um die Ausgabe zurück durch das Netzwerk bis zur Eingabeschicht zu propagieren. Dadurch erhält man in der Eingabeschicht die Relevanz jedes einzelnen Pixels für die Ausgabe. Daraus lässt sich ableiten, welche Pixel das Netz für die jeweilige Klassifikation als positiven, neutral oder negativen Indikator nutzt/einordnet.

Durch die Verwendung dieser Methode wurde zum Beispiel ein nicht einwandfreier Lösungsweg in dem InceptionNet von Google gefunden, welches eines der gewinnenden Modelle des Wettbewerbs ILSVRC 2014³ war. Die Klassen „Husky“ und „Wolf“ wurden abhängig von der Anwesenheit von Schnee im Hintergrund und nicht anhand von Merkmalen der Tiere unterschieden.

Einige LRP-Erklärungen im Kontext dieses Projektes sind in der Abbildung 11 zu sehen. Die Erklärungen bilden ab, welche Pixel nach LRP einen neutralen (grün), leicht positiven (gelb) und stark positiven (rot) Einfluss für die Klassifikation der Farbe, des Typs und der Verschmutzung haben.

Betrachtet man die Erklärung der Farbe, erkennt man, dass dieses Modell wider Erwarten hauptsächlich Pixel am Rand des Bildes betrachtet und nicht das Wäschestück selbst. Hier lässt sich vermuten, dass dieses Netz gelernt hat, dass weiße Wäschestücke dieses Datensatzes meistens groß sind und die schwarzen Blöcke am Rand verdecken. Die bunten sind im Gegensatz häufig deutlich kleiner und verdecken die Blöcke nicht.

³ <https://image-net.org/challenges/LSVRC/2014/results>

Daher wäre es am einfachsten, die zwei Klassen zu separieren, indem man lediglich feststellt, ob der schwarze Block zu sehen ist. Anhand dieser Technik ist eine Erkennungsgenauigkeit von 96% erreicht worden. Die logische Konsequenz ist daher, dass solche falschen Erkenntnisse verhindert werden müssen, da diese in der Praxis zu Fehlern führen können und das Modell an dem Lernen einer zulässigen Lösung hindern. Dies kann beispielsweise durch die Segmentierung der Wäschestücke erfolgen, welche im Kapitel 2.2.4 besprochen wird.

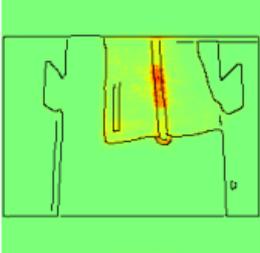
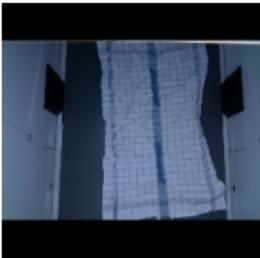
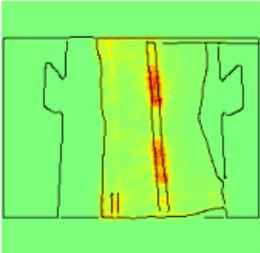
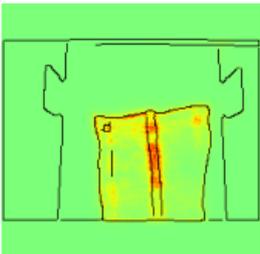
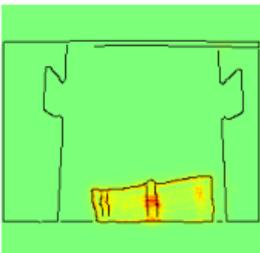
Bild	Klassifikation des Typs	Erklärung der Typklassifikation
	<p>Geschirrtuch: 0.8820</p> <p>Ess-Schürze: 0.0089</p> <p>Inkontinenzunterlage: 0.0072</p> <p>Unterwäsche: 0.0065</p> <p>Shirt: 0.0062</p>	
	<p>Geschirrtuch: 0.9169</p> <p>Ess-Schürze: 0.0081</p> <p>Schürze: 0.0052</p> <p>Inkontinenzunterlage: 0.0051</p> <p>Pullover: 0.0046</p>	
	<p>Geschirrtuch: 0.9300</p> <p>Pullover: 0.0044</p> <p>Shirt: 0.0042</p> <p>Schürze: 0.0042</p> <p>Inkontinenzunterlage: 0.0041</p>	
	<p>Geschirrtuch: 0.9311</p> <p>Ess-Schürze: 0.0050</p> <p>Inkontinenzunterlage: 0.0050</p> <p>Kittel: 0.0050</p> <p>Spezialkleidung: 0.0036</p>	

Abbildung 6: Originalbilder mit Ausgabe der wahrscheinlichsten Klassen eines trainierten neuronalen Netzes. Hierbei ist die wahrscheinlichste Klasse gelb unterlegt. Zudem wird mittels LRP die Relevanz der einzelnen Pixel für die Klassifikation des neuronalen Netzes berechnet und in Form einer Heatmap dargestellt.

Durch die Verwendung des Aktiven Lernens und der Explainable AI erhoffen wir uns sowohl das schnellere und einfachere Training in den Wäschereien, die Vereinfachung der Mensch-Maschine-Interaktion, wie auch die Erhöhung der Akzeptanz und des Verständnisses von KI in den Wäschereien.

2.2.4 Segmentierung

Unter der Segmentierung versteht man in der Bildverarbeitung das Unterteilen eines Bildes in inhaltlich zusammenhängende Regionen. Diese Regionen sind in diesem konkreten Fall die des Wäschestücks und die des Hintergrunds.

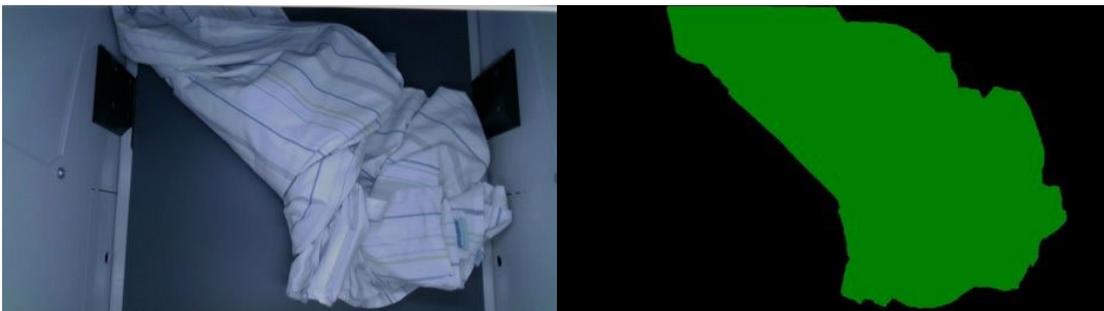


Abbildung 7: Originalbild (links) und pixelweise Segmentierung (rechts). Die Pixel, welche zum Wäschestück gehören sind grün eingefärbt und die, welche zum Hintergrund gehören, schwarz.

Naheliegende Techniken für die Segmentierung im konkreten Anwendungsfall sind die Nutzung der Tiefeninformation der Kamera oder die Differenz zu einem Bild des Förderbands ohne Wäschestück, da der Hintergrund statisch ist. Beide Ansätze wurden verfolgt und führten jedoch zu keinen zufriedenstellenden Ergebnissen, da die Farbbilder und insbesondere die Tiefenbilder sehr verrauscht waren. Daher wurden Deep Learning Methoden zur Segmentierung exploriert.

Da der Aufwand der pixelweisen Segmentierung, um überwacht Modelle trainieren zu können, nochmal deutlich höher als bei der Klassifikation ist, wurde zunächst mit dem PerturbGAN eine unüberwachte Methode umgesetzt. Unüberwachte Methoden verwenden beim Training im Gegensatz zu den überwachten Methoden keine korrekten Ausgaben. Daher ist eine geschickte Umformulierung des Problems notwendig.

Die Autoren des PerturbGAN (vgl. Bielski & Favaro, 2019) versuchen zunächst anhand eines Generative Adversarial Networks (GAN) (vgl. Goodfellow, 2014) künstliche Bilder anhand eines Datensatzes zu erzeugen. Dabei verwenden sie zwei Generatoren, wobei einer den Hintergrund erzeugen soll und der andere das Objekt von Interesse und dadurch ebenso die entsprechende Segmentierung. Da der Hintergrund statisch ist, war es hier möglich, den einen Generator durch ein Durchschnittsbild des Förderbands ohne Wäschestück zu ersetzen. Der andere Generator versucht nun möglichst realistische Wäschestücke zu erzeugen, sodass der Diskriminator die generierten Wäschestücke auf dem statischen Hintergrund nicht von den Originalen unterscheiden kann. Anhand der Merkmale, durch die der Diskriminator die beiden Datenquellen unterscheiden kann,

lernt der Generator dazu. Schrittweise verbessert sich der Generator dadurch, bis die generierten Bilder ununterscheidbar von den Originalen sind. Die beiden Netzwerke arbeiten demnach den ganzen Prozess über gegeneinander.

Als zweiter Schritt wird ein Enkodier-Netzwerk trainiert, welches zu den Originalbildern eine Kodierung berechnet, anhand derer der Generator des GANs möglichst verlustfrei die Eingabe rekonstruieren soll. Da der Generator ebenso die Segmentierung erstellt, erhält man je nach Rekonstruktionsleistung des Gesamtnetzwerks die pixelweise Segmentierung eines Bildes.

Des Weiteren wurde zu Vergleichszwecken ebenso ein Netzwerk überwacht trainiert. Dazu wurden 5000 Bilder händisch segmentiert. Um diese Arbeit zu vereinfachen, wurde ein Programm entwickelt, welches die pixelweise Segmentierung erheblich vereinfacht. Zunächst zieht man ein Rechteck um das Wäschestück. Dieses wird mittels GrabCut (vgl. Rother, 2004) im Idealfall lassoartig um das Wäschestück zusammengezogen, so dass Wäschestück und Hintergrund optimal voneinander getrennt sind. Sollte dies nicht zutreffen, ist es möglich, die Auswahl zu verfeinern, indem zusätzliche Bereiche markiert werden, welche eindeutig zum Wäschestück gehören bzw. nicht gehören. Anhand dessen passt GrabCut die Segmentierung an. Dies wird wiederholt bis das Bild optimal segmentiert wurde. Je nach Farbe, Form und Muster des Wäschestücks lässt sich die Segmentierung innerhalb von 10-120 Sekunden erstellen. Durchschnittlich benötigt eine Person für eine Segmentierung 55 Sekunden.

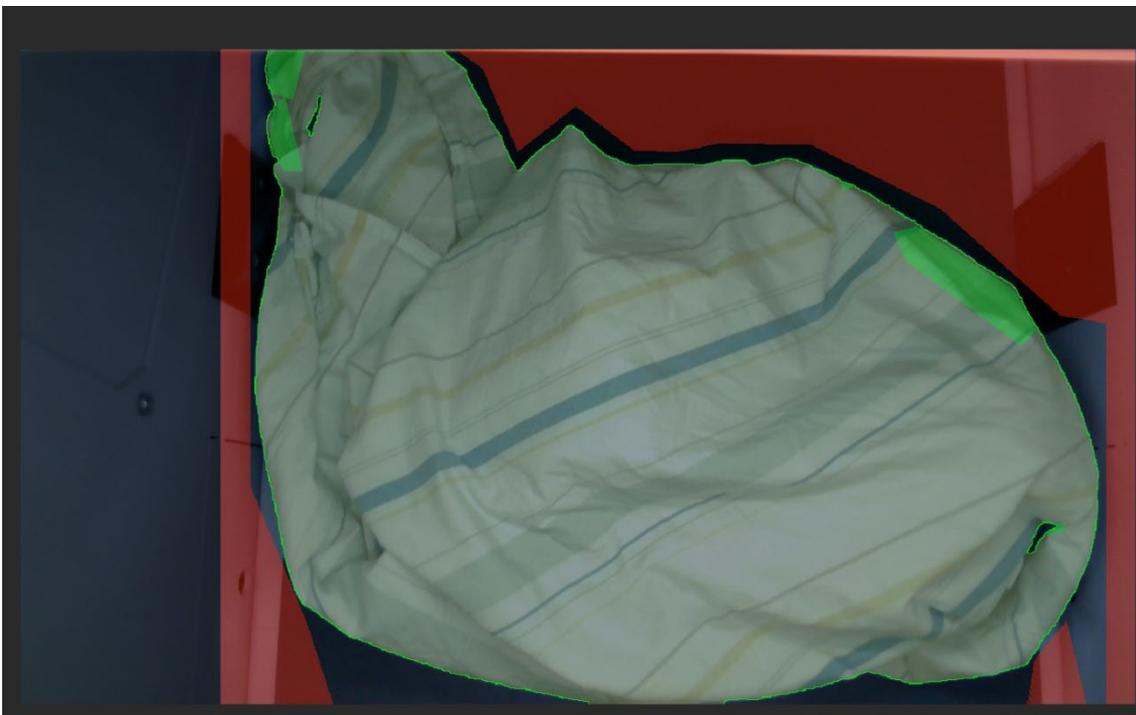


Abbildung 8: Erstellung einer Segmentierung mit Hilfe des GrabCut Algorithmus. Nutzer:innen definieren mit der Maus Bereiche, welche zum Wäschestück (grün) und welche zum Hintergrund (rot) gehören. Die Software ermittelt anhand dieser Informationen eine pixelgenaue Segmentierung.

In diesem Anwendungsfall verwenden wir zur überwachten Segmentierung das U-Net (vgl. Ronneberger, 2015), welches eine der bekanntesten Netzwerk-Architekturen zur Bildsegmentierung ist.

Das U-Net ist ein Netzwerk zur semantischen Segmentierung, auch bekannt als pixelweise Segmentierung, und ordnet somit jedem Pixel eine Klasse zu. Seine Architektur kann man sich grob als ein kontrahierendes Enkodier-Netzwerk vorstellen, gefolgt von einem expandierenden Dekodier-Netzwerk. Die Hauptidee besteht darin, ein gewöhnliches kontrahierendes Netz durch aufeinanderfolgende Schichten zu ergänzen, in denen Pooling-Operationen durch Upsampling-Operationen ersetzt werden. Diese Schichten erhöhen also die Auflösung der Ausgabe und anstatt die Pixelwerte dabei lediglich zu wiederholen wird die Erhöhung der Auflösung gelernt. Diese Faltungsschicht kann dann lernen, eine präzise Ausgabe auf der Grundlage der Informationen des Enkodier-Netzwerks zusammenzustellen. Dabei lernt das Enkodier-Netzwerk optimalerweise die Bereiche der unterschiedlichen Regionen zu finden. Der Name ergibt sich dadurch, dass der kontrahierende Pfad mehr oder weniger symmetrisch zum expandierenden Teil ist, womit sich eine U-förmige Architektur ergibt.

3 Ergebnisse

Anhand des Versuchsstandes wurde anfangs ein initialer Datensatz aus 9405 Bildern erstellt. Zu jedem der Bilder existieren Information über die Farbe, den Typ und zumeist auch über die Kontaminierung, die Schäden und das Material. Da für diesen Datensatz die Segmentierung essenziell für die Klassifizierung ist, betrachten wir diese Ergebnisse im Folgenden zuerst.

Segmentierung

Zunächst sind 5000 der Bilder händisch segmentiert worden. Im Anschluss wurde ein U-Net mittels dieser 5000 Bilder trainiert und die Segmentierungen für die restlichen 4405 durch dieses Netz vorhergesagt. Die Segmentierungen des Netzes wurden dann händisch angenommen bzw. abgelehnt. 1000 der abgelehnten Bilder wurden darauffolgend händisch segmentiert und zusammen mit den vorherigen 5000 und den angenommenen Segmentierungen wurde ein neues U-Net trainiert. Anhand dieses Netzes wurden dann Segmentierungen für die restlichen Bilder berechnet und erneut händisch akzeptiert bzw. abgelehnt. Für die Bilder mit abgelehnten Segmentierungen wurden die Segmentierungen abschließend erneut händisch erstellt. Durch diesen am aktiven Lernen orientierten Prozess konnten 2091 Segmentierungen automatisch erstellt werden. Ohne Abzug des Aufwandes für das Akzeptieren/Ablehnen konnten dadurch etwa 32 Stunden Arbeitszeit eingespart werden.

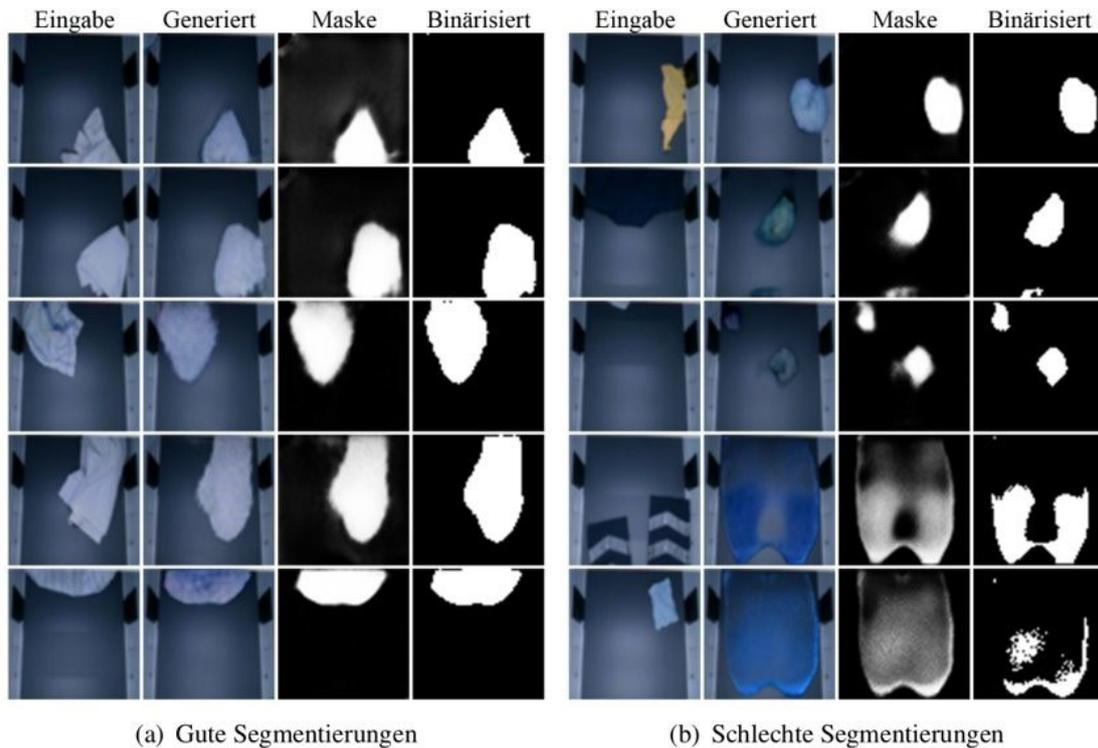
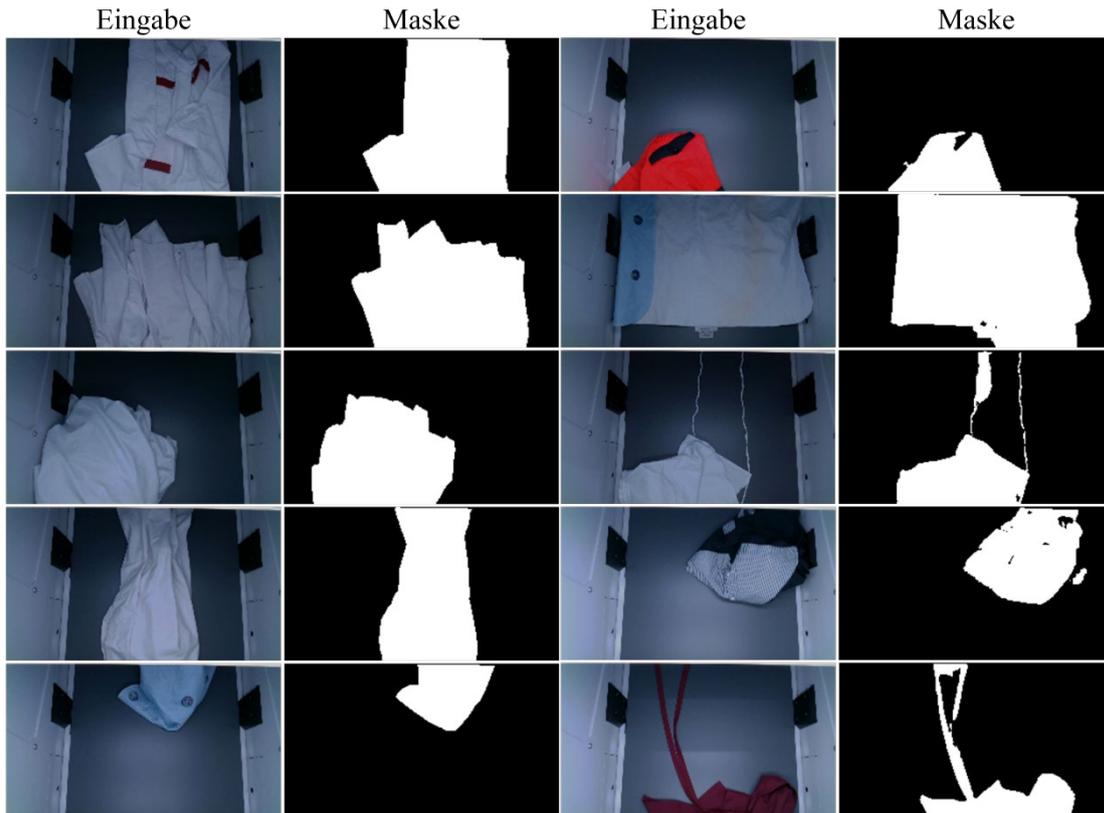


Abbildung 9: Originalbilder (Eingabe), Rekonstruktion des Originalbilds durch den Enkodierer und das GAN (Generiert), Segmentierungsausgabe des GANs (Maske) und die binärisierte Version der Maske anhand eines Schwellwerts. Auf der linken Seite sind zufriedenstellende und auf der rechten Seite schlechte Segmentierungen des PerturbGAN zu sehen.

Die Ergebnisse der Segmentierung sind in Tabelle 1 zu sehen. Um die Genauigkeit (Anteil korrekt klassifizierter Pixel der Vorhersage) und die IoU (Verhältnis von Schnitt zur Vereinigung der Vorhersage und korrekten Ausgabe) einordnen zu können, wurden diese neben den beiden neuronalen Netzen ebenso für die Segmentierung mittels Differenz zu einem Bild eines leeren Förderbands und der Ausgabe des GrabCut-Algorithmus ohne Nachbesserungen berechnet.

	Differenz	GrabCut	PerturbGAN	U-Net
Genauigkeit	0.8971	0.9245	0.8864	0.9993
IoU	0.6989	0.7994	0.6106	0.9947

Tabelle 1: Genauigkeit und IoU der verwendeten Methoden auf einem Validierungsdatensatz.



(a) Gute Segmentierungen

(b) Schlechte Segmentierungen

Abbildung 10: Originalbilder (Eingabe) und die Segmentierungsausgabe des U-Nets (Maske). Auf der linken Seite sind zufriedenstellende und auf der rechten Seite schlechte Segmentierungen des U-Nets zu sehen.

Klassifikation

Wie zuvor erwähnt sind 6855 Bilder (72.89%) von Wäschestücken, welche aus Pflegeheimen stammen und 2550 Bilder (27.11%) von Arbeitsbekleidung. Da bei jedem Durchlauf eines Wäschestücks fünf Bilder gespeichert werden und jedes Wäschestück drei Mal von unterschiedlichen Startpositionen aus über das Förderband fährt, sind insgesamt 1881 Durchläufe mit 627 verschiedenen Wäschestücken getätigt worden.

Zu jedem dieser Bilder sind zudem Informationen über die Farbe, den Typen, die Verschmutzung, das Material, mechanische/chemische Schäden und die Temperatur, bei der das abgebildete Wäschestück gewaschen werden darf, gegeben. Jedoch ist aufgrund von Unsicherheiten bei der Annotierung die Klasse Unsicher für alle Kategorien eingeführt worden. Da lediglich für die Farbe, den Typ und die Verschmutzung zu über 90% die Klassen sicher bestimmt werden konnten und der Datensatz schon generell eher klein ist, wurde der Fokus zunächst auf diese drei Kategorien gelegt.

Die Tabelle 2 zeigt, wie die Klassen der drei Kategorien jeweils verteilt sind. Dabei fällt auf, dass die Klassen der Kategorien Typ und Kontamination sehr unbalanciert verteilt sind.

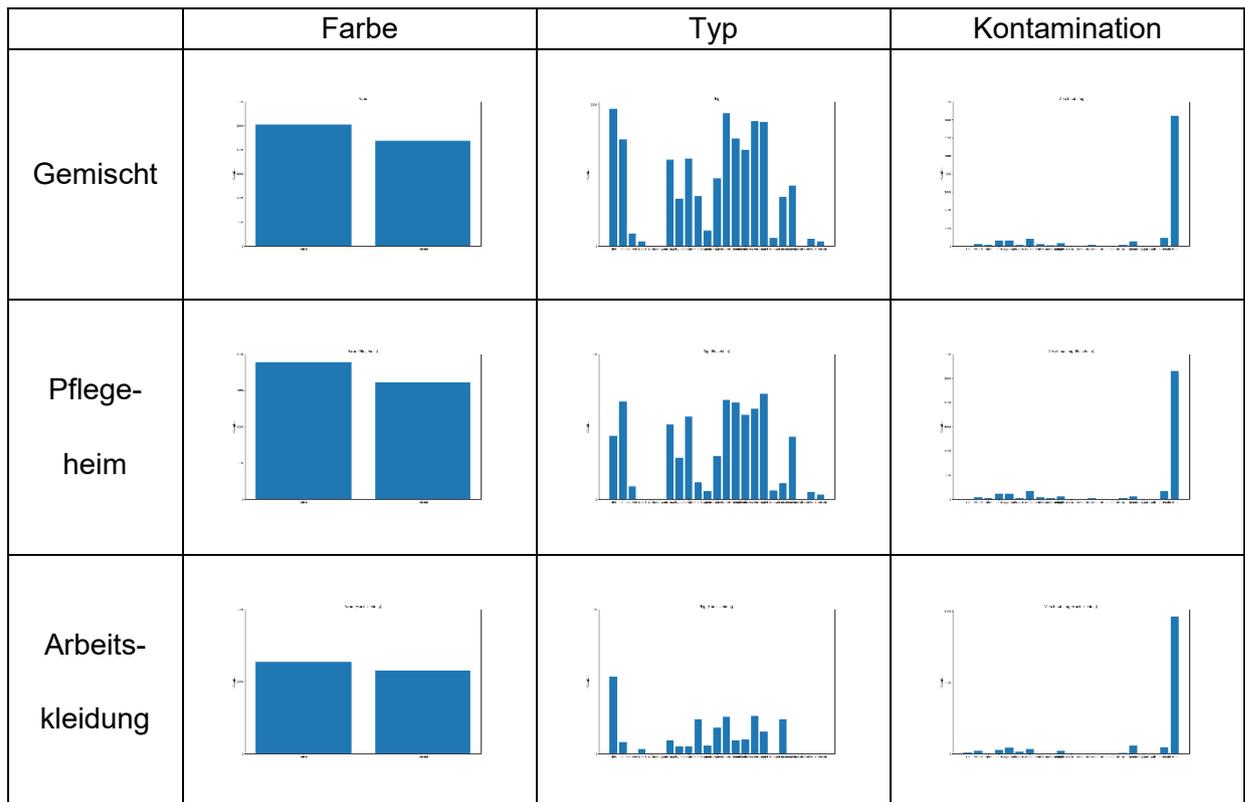


Tabelle 2: Verteilung der Klassen für die Kategorien Farbe, Typ und Verschmutzung des gesamten Datensatzes und der Subdatensätze Pflegeheim und Arbeitskleidung.

Für beide CNN-Architekturen wurde für jede Kategorie kombiniert mit jedem Datensatz (unverarbeitet und segmentiert) eine Hyperparameteroptimierung durchgeführt. Dabei haben wir uns zunächst auf die Anzahl der Bilder, die pro Optimierung gemeinsam betrachtet werden (Batch Size), die Lernrate und auf die Verwendung eines vortrainierten Netzes beschränkt. Die Ergebnisse sind in Tabelle 3 zu sehen. Dabei ist die beste Genauigkeit auf einem gemischten Validierungsdatensatz für die Farbe 96.49%, für den Typ 79.89% und für die Verschmutzung 81.00%. Auf dem Pflegeheim-Datensatz für die Farbe 95.44%, für den Typ 84.05% und für die Verschmutzung 84.30% und auf dem Arbeitskleidung-Datensatz für die Farbe 100%, für den Typ 82.31% und für die Verschmutzung 85.03%.

Modell	Parameter	Datensatz		Farbe	Typ	Verschm.
VGGNet	134.3 Mio.	Gemischt	orig.	0.9649	0.7989	0.7841
			seg.	0.9539	0.7952	0.8100
		Pflegeheim	orig.	0.9518	0.8227	0.7569
			seg.	0.9215	0.8253	0.8430
		Arbeitskleidung	orig.	1.0000	0.7823	0.8095
			seg.	1.0000	0.7755	0.8095

DenseNet	5.1 Mio.	Gemischt	orig.	0.9631	0.7048	0.7620
			seg.	0.9539	0.7841	0.7878
		Pflegeheim	orig.	0.9544	0.8000	0.7797
			seg.	0.8810	0.8405	0.7721
		Arbeits- kleidung	orig.	1.0000	0.8231	0.8095
			seg.	1.0000	0.8231	0.8503

Tabelle 3: Genauigkeit der zwei unterschiedlichen Modelle für die Erkennung der Farbe, des Typs und der Verschmutzung. Die Modelle wurden dabei jeweils einmal auf dem unveränderten originalen Datensatz (orig.) und einmal auf dem segmentierten Datensatz (seg.) trainiert.

Erklärbare künstliche Intelligenz

Zudem wurde der LRP-Algorithmus implementiert und kann für jede beliebige Netzarchitektur angewandt werden. Ein Schaubild für die Erklärungen der Farbklassifikation mitsamt der Sicherheit des jeweiligen neuronalen Netzes ist in Abbildung 11 zu sehen.

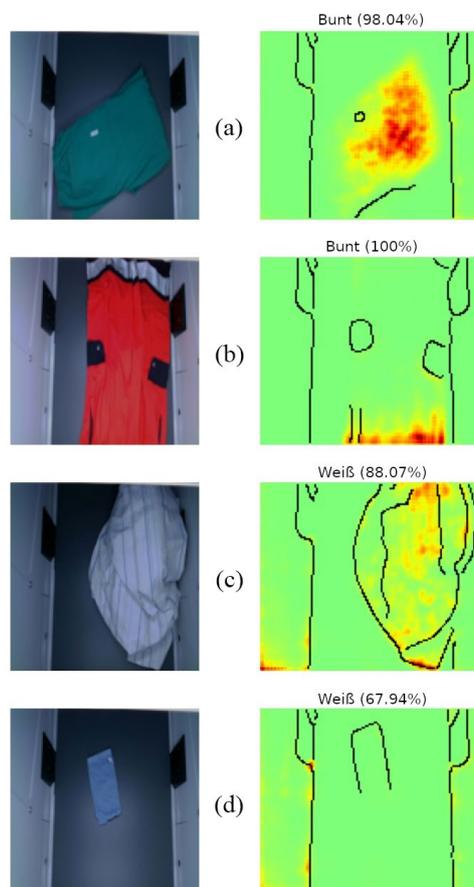


Abbildung 11: Originalbilder (links) und die Klassifikation mitsamt der LRP-Erklärung (rechts).

Betrachtet man die Klassifikation mit hoher Sicherheit (a-b, >98%) fällt auf, dass lediglich Bereiche des Wäschestücks für die Klassifikation genutzt wurden. Im weniger sicheren Bereich (c, 80-90%) wurde größtenteils das Wäschestück genutzt, aber auch Bereiche am Rand. Bei einer relativ unsicheren Klassifikation (d, <70%) wurde das Wäschestück offensichtlich nicht erkannt und nur Randpixel sowie Pixel am Übergang zwischen Förderband und Wand genutzt. Mögliche Konsequenzen, um solch einen Fall zu verhindern wäre eine Segmentierung oder eine Anpassung des Moduls, sodass der Hintergrund gleichfarben ist.

In Abbildung 12 ist die finale Darstellung der gesamten Ausgabe mitsamt Visualisierungen für ein zufällig gewähltes Wäschestück zu sehen.

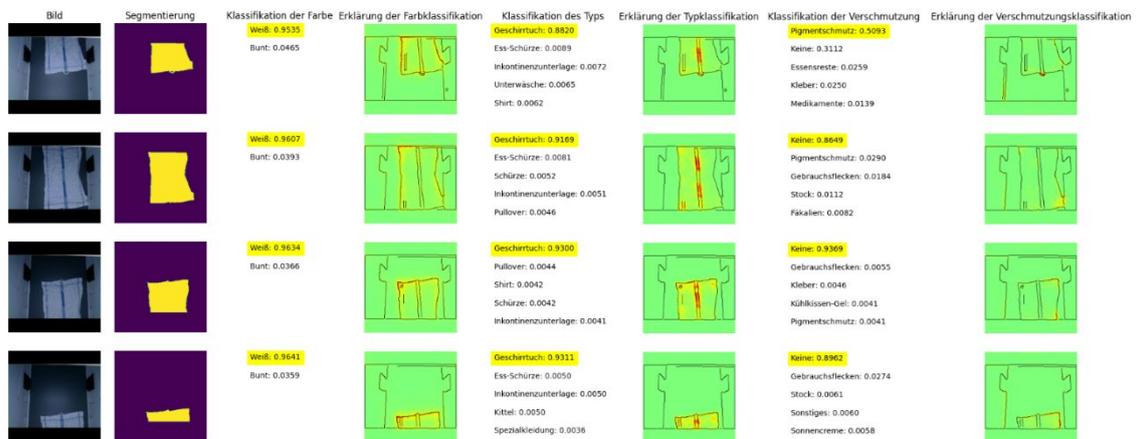


Abbildung 12: Originalbilder mit den Segmentierungen des trainierten U-Net, den Ausgaben der wahrscheinlichsten Klassen eines trainierten neuronalen Netzes pro Kategorie. Hierbei ist die wahrscheinlichste Klasse gelb unterlegt. Zudem wird mittels LRP die Relevanz der einzelnen Pixel für die Klassifikation des neuronalen Netzes berechnet und in Form einer Heatmap dargestellt.

4 Diskussion

Durch die beiden neuronalen Netze konnten gute Erkennungsgenauigkeiten für alle Kategorien, insbesondere der Farbe, erzielt werden. Hierbei muss jedoch beachtet werden, dass die Datensätze und dementsprechend vor allem die Validierungsdatensätze klein sind. Es ist unbedingt notwendig den Datensatz zu vergrößern. Dabei sollte höchste Priorität sein, die Aufnahme und Annotation in einer Wäscherei zu verlagern um möglichst automatisiert und möglichst schnell Daten aufnehmen und annotieren zu können. Dafür wollen wir das Aufnahmemodul in einer Wäscherei einsetzen, die sortenrein wäscht und das Modul zwischen der Sortierung und den Waschmaschinen platzieren. Da die Wäsche zu diesem Zeitpunkt bereits sortiert ist, können wir für eine große Menge an Daten die gleiche Annotation annehmen und müssen lediglich auf den Wechsel achten. Sofern möglich sollte ebenso auf die Balance der verschiedenen Klassen geachtet werden. Sollten weiterhin einige Klassen unterrepräsentiert sein, wie beispielsweise seltene Verschmutzungen, könnten diese händisch auf einigen Wäschestücken aufgetragen werden und so die Verteilung künstlich ausbalanciert werden.

Die Segmentierungsleistung des U-Nets ist sehr gut. Die Verwendung von Segmentierungen verbessert nach Tabelle 3 die Klassifikationsleistung zwar nicht in jedem Fall, hindert das Netz aber daran, Artefakte in den Bildern wie die schwarzen Blöcke zu nutzen (siehe Abbildung 11), um einen zulässigen Lösungsweg zu umgehen. In diesem Fall sind die umgesetzten Methoden zur Erklärung der KI bereits hilfreich gewesen und haben dazu beigetragen bereits in der Entwicklung Fehler zu erkennen und zu beseitigen. Einen weiteren Aufschluss könnten die Erklärungen der besten neuronalen Netze geben, welche auf dem originalen und segmentierten Datensatz trainiert wurden, indem man die Erklärungen der beiden Netze, welche sich für dasselbe Bild am meisten unterscheiden, vergleicht.

Zudem ist bei der Ausstellung des Moduls an einem anderen Standort die Segmentierung und somit auch die Klassifikation aufgrund unterschiedlicher Lichtverhältnisse teils massiv beeinträchtigt worden. Dies, die Erklärungen und die Beobachtungen, die uns beim Annotations- und Segmentierungsprozess aufgefallen sind, dienen als Grundlage um einige Optimierungen am Aufnahmemodul vorzunehmen. Ziel ist es, sowohl die Störobjekte wie auch Varianz/Rauschen durch z.B. unterschiedliche Lichtverhältnisse zu eliminieren. Der Innenraum sollte möglichst abgeschirmt von außen sein, einen homogenen/gleichfarbigen Hintergrund haben und gut ausgeleuchtet sein. Da uns Wäschestücke unterschiedlicher Länge vorliegen und die Tiefendaten keine Verbesserung bei der Segmentierung und Klassifikation erbracht haben, ist die Verwendung einer Zeilenkamera eine sinnvolle Anpassung. Durch diese wird das Bild nicht wie durch eine normale Kamera zu einem festen Zeitpunkt aufgenommen, sondern Zeile für Zeile über die Zeit hinweg synchron zur Förderbandgeschwindigkeit. Somit ist es möglich Bilder von Wäschestücken beliebiger Länge aufzunehmen. Die Zeile der Kamera wird so ausgerichtet, dass diese keine Störobjekte wie die schwarzen Blöcke oder Löcher beinhaltet. Zudem müssen dann lediglich im Bereich der Zeile die Wände entsprechend der Förderbandfarbe lackiert werden und die Ausleuchtung angepasst werden.

Ausblick

Der Fokus wird auf die automatisierte Aufnahme in einer Wäscherei mit Vorsortierung gelegt, um kontinuierlich an Trainingsdaten zu gelangen. Dazu werden zeitnah die beschriebenen Anpassungen am Aufnahmemodul wie der Wechsel auf eine Zeilenkamera durchgeführt. Ebenso ist es langfristig erstrebenswert für die Erkennung des Materials eine Hyperspektralkamera zu verwenden (vgl. Blanch, 2019).

Parallel zu dem Aufnahmeprozess wollen wir die Methoden des Aktiven Lernens für die Segmentierung und Klassifikation einsetzen, um die Erkenntnisse für den neuen Datensatz zu nutzen. Hierbei ist besonders die Segmentierung interessant, da hier pro Pixel eine Klassifikation durchgeführt wird und dadurch verschiedene Methoden zur Bestimmung der Sicherheit betrachtet werden können. Beispielsweise die durchschnittliche

Sicherheit oder Varianz der Randpixel oder die Anzahl der zusammenhängenden Vordergrundobjekte.

Zudem ist die Umsetzung weiterer Methoden der erklärbaren künstlichen Intelligenz geplant. Um die Effektivität und Akzeptanz dieser in Wäschereien evaluieren zu können, planen wir eine Studie, in der wir messen wollen, wie Mitarbeiter:innen mit der KI und auch verschiedenen Stufen von Erklärungen der KI umgehen. Daraus wollen wir ableiten, wie die KI und auch die Erklärung der KI die Entscheidungen der Mitarbeiter:innen beeinflusst.

5 Literaturverzeichnis

- Bielski, Adam & Favaro, Paolo. (2019). Emergence of Object Segmentation in Perturbed Generative Models.
- Blanch, Carolina & Saeys, Wouter & Lambrechts, Andy. (2019). Hyperspectral imaging for textile sorting in the visible–near infrared range. *Journal of Spectral Imaging*. 8.10.1255/jsi.2019.a17.
- Goodfellow, Ian & Pouget-Abadie, Jean & Mirza, Mehdi & Xu, Bing & Warde-Farley, David & Ozair, Sherjil & Courville, Aaron & Bengio, Yoshua. (2014). Generative Adversarial Networks. *Advances in Neural Information Processing Systems*. 3. 10.1145/3422622.
- He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.
- Huang, Gao & Liu, Zhuang & van der Maaten, Laurens & Weinberger, Kilian. (2017). Densely Connected Convolutional Networks. 10.1109/CVPR.2017.243.
- Lapuschkin, Sebastian & Binder, Alexander & Montavon, Grégoire & Klauschen, Frederick & Müller, Klaus-Robert & Samek, Wojciech. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE*. 10.e0130140.10.1371/journal.pone.0130140.
- LeCun, Yann & Boser, Bernhard & Denker, John & Henderson, Donnie & Howard, Richard & Hubbard, Wayne & Jackel, Lawrence (1989). Handwritten Digit Recognition with a Back-Propagation Network
- Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *LNCS*. 9351. 234-241. 10.1007/978-3-319-24574-4_28.
- Rother, Carsten & Kolmogorov, Vladimir & Blake, Andrew. (2004). GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Trans. Graph.* 23. 309-314. 10.1145/1186562.1015720.
- Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* 1409.1556.
- Sun, Chen & Shrivastava, Abhinav & Singh, Saurabh & Gupta, Abhinav. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era.
- Szegedy, Christian & Liu, Wei & Jia, Yangqing & Sermanet, Pierre & Reed, Scott & Anguelov, Dragomir & Erhan, Dumitru & Vanhoucke, Vincent & Rabinovich, Andrew. (2015). Going deeper with convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1-9. 10.1109/CVPR.2015.7298594.
- Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
- Wang, Keze & Zhang, Dongyu & Li, Ya & Zhang, Ruimao & Lin, Liang. (2016). Cost-Effective Active Learning for Deep Image Classification. *IEEE Transactions on Circuits and Systems for Video Technology*. 27. 1-1. 10.1109/TCSVT.2016.2589879.

IMPRESSUM

Verantwortlich für den Inhalt

Nico Rabethge, Forschungsinstitut für Kognition und Robotik (CoR-Lab), Universität Bielefeld
Kurt-Georg Ciesinger, DAA Deutsche Angestellten-Akademie GmbH

Fotos/Abbildungen

Titel: Herbert Kannegiesser GmbH & Co. KG

S. 2: it's OWL Clustermanagement, DAA

S. 6, 8, 9, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23: Universität Bielefeld

Gestaltung & Redation

Salome Leßmann

it's OWL Clustermanagement GmbH

Empfohlene Zitierweise

Rabethge, N.; Ciesinger, K.-G. (2023): KI in der Schmutzwäschesortierung - Entwicklung einer intelligenten Mensch-Maschine-Interaktion zur Identifikation von Waschkategorien. Working-Paper-Reihe des Kompetenzzentrums Arbeitswelt.Plus, Paderborn, Nr. 4, <https://doi.org/10.55594/WGCT6835>

Erscheinung

01/2023



Möchten Sie mehr über die Forschungsarbeit im Kompetenzzentrum Arbeitswelt.Plus erfahren? Auf unserer Website finden Sie detaillierte Informationen zu allen Forschungsschwerpunkten.

Kompetenzzentrum Arbeitswelt.Plus

c/o it's OWL Clustermanagement GmbH

Zukunftsmeile 2

33012 Paderborn

www.arbeitswelt.plus



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Dieses Forschungs- und Entwicklungsprojekt wird durch das Bundesministerium für Bildung und Forschung (BMBF) im Programm „Zukunft der Wertschöpfung – Forschung zu Produktion, Dienstleistung und Arbeit“ gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin / beim Autor.