

Vergleich von Methoden zur Erzeugung synthetischer Datensätze für medizinische Eingriffsräume am Beispiel der Erkennung medizinischer Kleidung

Zweck

Die Verfügbarkeit von echten Daten aus Bereichen mit hohen Datenschutzerfordernungen, wie dem medizinischen Interventionsraum, ist gering und die Erfassung ist in Bezug auf den Datenschutz komplex. Um die Forschung für Assistenzsysteme im medizinischen Interventionsraum zu ermöglichen, müssen neue Methoden zur Datengenerierung für diese Bereiche erforscht werden. Daher präsentiert diese Arbeit einen Weg, um einen synthetischen Datensatz für den medizinischen Kontext zu erstellen, wobei die medizinische Kleidungserkennung als Beispiel dient. Das Ziel besteht darin, die Kluft zwischen synthetischen und echten Daten zu schließen.

Methoden

Es werden Methoden zur 3D-gescannten Kleidung und entworfenen Kleidung in einem Szenario der Domänen-Randomisierung und strukturierten Domänen-Randomisierung miteinander verglichen, wobei zwei verschiedene Rendering-Engines verwendet werden. Zusätzlich wurde ein Mixed-Reality-Datensatz vor einem grünen Hintergrund und ein Ziel-Domänen-Datensatz verwendet, wobei letzterer zur Bewertung der verschiedenen Datensätze dient. Die durchgeführten Experimente sollen zeigen, ob gescannte Kleidung oder entworfene Kleidung bessere Ergebnisse in der Domänen-Randomisierung und strukturierten Domänen-Randomisierung liefern. Ebenso wird eine Baseline mit den Mixed-Reality-Daten erstellt. In einem weiteren Experiment wird untersucht, ob die Kombination von realen, synthetischen und Mixed-Reality-Bilddaten die Genauigkeit im Vergleich zu nur realen Daten verbessert.

Ergebnisse

Unsere Experimente zeigen, dass die strukturierte Domänen-Randomisierung von entworfener Kleidung zusammen mit Mixed-Reality-Daten eine Basislinie erreicht, die auf dem Testdatensatz der klinischen Ziel-Domäne eine mittlere durchschnittliche Präzision (mAP) von 72,0% erzielt. Wenn zusätzlich 15% (99 Bilder) der verfügbaren Trainingsdaten der Ziel-Domäne verwendet werden, konnte die Kluft zu 100% (660 Bilder) der Trainingsdaten der Ziel-Domäne nahezu geschlossen werden, mit einer mAP von 80,05% (81,95% mAP). Schließlich zeigen wir, dass bei Verwendung von 100% der Trainingsdaten der Ziel-Domäne die Genauigkeit auf 83,35% mAP gesteigert werden kann.

Schluss

Zusammenfassend kann festgestellt werden, dass das vorgestellte Modellieren von Gesundheitsfachkräften eine vielversprechende Methodik darstellt, um das Problem fehlender Datensätze aus medizinischen Interventionsräumen anzugehen. Wir werden es weiterhin in verschiedenen Aufgaben, wie Assistenzsystemen im medizinischen Bereich, untersuchen.

1 Einführung

Für Herausforderungen im Bereich der Computer Vision in medizinischen Interventionsräumen, wie der Objekterkennung, der Personenerkennung oder komplexeren Aufgaben wie der Aktivitätserkennung, existieren nur wenige Datensätze [1, 19]. Diese Datensätze konzentrieren sich auf die 2D- und 3D-Erkennung der menschlichen Körperhaltung. Obwohl in den Datensätzen mehrere Kameras verwendet werden, sind nur etwa 700 Frames pro Kamera annotiert. Darüber hinaus sind die Datensätze außerhalb der oben genannten nur institutionell gebunden und nicht öffentlich verfügbar, möglicherweise aufgrund von Datenschutzbestimmungen und ethischen Anforderungen [18, 28]. Ebenso können die Datensätze möglicherweise nicht über die notwendige Varianz verfügen, um auf andere Orte übertragen werden zu können. Darüber hinaus gibt es noch andere Anwendungsfälle neben der Erkennung der menschlichen Körperhaltung. Dazu gehören KI-basierte Erkennung von Sterilität bei Gesundheitsfachkräften, die Lokalisierung bestimmter medizinischer Geräte oder die Erkennung von Handlungen von Gesundheitsfachkräften [14]. Soweit uns bekannt ist, stehen keine anderen öffentlich verfügbaren Datensätze für die genannten Szenarien zur Verfügung, die kamerabasierte Systeme während echter Interventionen verwenden.

Die Erfolge des Deep Learnings in den letzten Jahren sind unter anderem auf die Verfügbarkeit großer Datensätze wie Imagenet [16] für die Bildklassifizierung oder MS COCO [8] für die Objektbegrenzungsdetektion zurückzuführen. Zusätzlich sind die Forschung neuer Methoden oder Architekturen wie [4, 9, 21, 2] sowie die Forschung an leistungsstarker Hardware für Parallelrechnungen zu erwähnen. Die Autoren von [20, 7] analysieren das Thema Hardware eingehend. In dieser Arbeit liegt jedoch besonderer Fokus auf der Verfügbarkeit von Datensätzen und Methoden zur Datensatzerzeugung, um den notwendigen Umfang an echten Daten aus der Ziel-Domäne zu reduzieren.

Bereits mehrere Arbeiten beschäftigen sich mit der Erzeugung synthetischer Daten und dem Ziel, die Kluft zwischen synthetischen und echten Daten zu verringern. Dazu gehören Arbeiten zu Domain Randomization (DR) und Structured Domain Randomization (SDR) [24, 25, 12]. Neben anderen Arbeiten wurde bereits gezeigt, dass der Einsatz von synthetischen Bilddaten den Bedarf an echten Daten verringern kann [25]. Ebenso existieren synthetische Daten von Personen [5]. Die Herausforderung liegt jedoch in den spezifischen Merkmalen eines medizinischen Interventionsraums. Gesundheitsfachkräfte tragen spezielle Kleidung mit manchmal mehreren Schichten, sterile Handschuhe, Masken und Haarnetze. Die Unterschiede zwischen konventionellen menschlichen Daten und Daten aus dem medizinischen Bereich sind groß. Dennoch scheinen DR-Techniken vielversprechend für den Einsatz in Forschungsfragen rund um medizinische Interventionen zu sein.

Diese Arbeit präsentiert einen Vergleich hinsichtlich der Erkennungsgenauigkeit und der Verallgemeinerungsfähigkeit verschiedener Methoden zur Erzeugung synthetischer Kleidung, die entweder aus 3D-Kleiderscans (SCANS) oder entworfenen CAD-Kleidungen (CAD) mit dem Skinned Multi-Person Liner Model (SMPL) [10] bestehen. Der Vergleich wird am Beispiel der Erkennung von medizinischer Kleidung durchgeführt. Um synthetische Trainingsdaten zu erzeugen, werden beide Methoden (SCANS, CAD) in eine DR-Umgebung namens NVIDIA Deep Learning Dataset Synthesizer (NDDS) [23] und eine SDR-Umgebung, die auf [22] basiert und in Unity implementiert ist, integriert. Ebenso zielt die vorgestellte Methodik darauf ab, eine

Pipeline zur Generierung synthetischer Daten für den medizinischen Bereich zu erkunden, um weitere Forschungsfragen aus dem Interventionsraum untersuchen zu können. Neben den synthetischen Daten werden echte Daten von verschiedenen Personen vor einem grünen Bildschirm aufgenommen, um die Kluft zwischen synthetischen Daten und der Ziel-Domäne zu verringern. Zur Bewertung der verschiedenen Ansätze wird ein Datensatz aus der Ziel-Domäne des Krankenhauses aufgezeichnet. Alle Daten werden in Trainings-, Validierungs- und Testsets aufgeteilt, wobei der klinische Datensatz als Testset für alle Methoden dient.

2 Damit verbundene Arbeit

Mit dem Aufkommen von Methoden zur Generierung synthetischer Daten, wie zum Beispiel DR [24], wurde bereits gezeigt, dass synthetische Daten den Bedarf an echten Daten verringern können [25]. Ein Schwerpunkt der Forschung liegt jedoch darauf, die Kluft zwischen den synthetischen Daten und der Ziel-Domäne zu verringern.

Hier hat sich das oben erwähnte DR als eine Möglichkeit zur Reduzierung der Kluft herausgestellt. Eine Idee des DR ist, dass wenn ausreichend Varianz in den synthetischen Daten erzeugt werden kann, die Realität eine weitere Variante der Ziel-Domäne darstellt [24].

Die Arbeiten von Tobin et al. [24] und Tremblay et al. [25] zeigten, dass ein Objekterkennungsnetzwerk für das Greifen von Robotern oder die Erkennung von Autos allein aus synthetischen Bildern mit zufälliger Positionierung, zufälliger Beleuchtung, zufälligen Hintergründen, Ablenkungsobjekten und nicht realistischen Texturen trainiert werden kann. Darüber hinaus zeigte die Arbeit von Tremblay et al., dass die notwendige Menge an echten Daten aus der Ziel-Domäne reduziert werden kann, während eine ausreichende Genauigkeit erhalten bleibt, wenn mit DR-generierten Bildern vorab trainiert wird.

Auch die Arbeit von Borkman et al. [3] zeigte, dass bei Verwendung von Unity Perception für die Generierung synthetischer Daten die Menge an realen Daten auf 10% reduziert werden konnte, wenn sie zusammen mit den synthetischen Daten verwendet wurden, während eine bessere durchschnittliche Präzision (AP) im Vergleich zu allein verwendeten realen Daten erreicht wurde.

DR wurde bereits erfolgreich in verschiedenen Bereichen angewendet. Neben den genannten Bereichen der Autodetektion und des Greifens von Robotern können die Arbeiten von Sadeghi et al. [17] zum Fliegen eines Quadrocopters durch Innenräume, Zhang et al. [31] für eine Objekterreichungsaufgabe auf einer Tischplatte durch Unordnung oder James et al. [6] zum Greifen eines Würfels und Platzieren in einem Korb genannt werden.

Dies lässt uns glauben, dass DR ein geeigneter Ansatz für den Bereich medizinischer Interventionen ist, in dem nur wenige echte Daten verfügbar sind und der Zugang zu dieser Domäne stark eingeschränkt ist.

Ablationsstudien von [25] und [24] zeigten, dass hochauflösende Texturen und eine größere Anzahl einzigartiger Texturen in der Szene die Leistung verbessern. Auch [31] kommt nach dem Testen ihrer

Hypothese zu dem Schluss, dass die Verwendung komplexer Texturen eine bessere Leistung bringt als die Verwendung von zufälligen Farben.

Im Gegensatz zum DR-Ansatz steht die fotorealistische Darstellung der Szene und der Objekte. In den letzten Jahren wurden eine Reihe von Datensätzen zu diesem Zweck erstellt. Hier sind die Arbeiten von [26, 5, 29] oder [27] zu nennen. Einige dieser Arbeiten kombinieren reale Bilddaten mit DR und fotorealistisch gerenderten Bilddaten.

In [26] wurde ein fotorealistisch gerendeter Datensatz für 21 Objekte des YCB-Datensatzes erstellt. Hier werden die Objekte in verschiedenen Szenen mit Kollisions-Eigenschaften gerendert, wenn sie herunterfallen. Der Datensatz soll den Fortschritt im Bereich der Objekterkennung und Pose-Schätzung beschleunigen.

In [27] wird DR mit fotorealistisch gerenderten Bilddaten kombiniert, um das Greifen von Haushaltsgegenständen durch Roboter zu ermöglichen. Mit den auf diese Weise generierten Daten haben es die Autoren geschafft, ein Echtzeitsystem für die Objekterkennung und das Greifen von Robotern mit ausreichender Genauigkeit zu erkunden. Sie zeigten auch, dass die Kombination beider Domänen die Leistung im Vergleich zu nur einer allein verbesserte.

Im Bereich der Schätzung der menschlichen Körperhaltung müssen die Arbeiten von [5] und [29] erwähnt werden. Beide Arbeiten konnten zeigen, dass die Leistung von Netzwerken durch Verwendung synthetischer bzw. animierter Personen erhöht werden kann.

Die Arbeit von [29] generiert fotorealistische synthetische Bilddaten und deren Ground-Truth für die Klassifikation von Körperteilen.

In [5] werden animierte Personen in Mixed-Reality-Umgebungen integriert. Die Bewegungen wurden von Schauspielern in einem Motion-Capture-Szenario aufgezeichnet und auf 3D-gescannte Meshes übertragen. In ihren Experimenten konnten sie eine 20%ige Leistungssteigerung im Vergleich zum größten verfügbaren Trainingsdatensatz in diesem Bereich erreichen.

State-of-the-Art-Modelle für realistische menschliche Körperformen sind die von [10] eingeführten SMPL-Modelle, die von STAR in [11] verbessert wurden. Laut den Autoren ist das SMPL-Modell ein skinned vertex-basiertes Modell, das menschliche Körperformen in einer breiten Vielfalt darstellt. In ihrer Arbeit lernen sie männliche und weibliche Körperformen aus dem CAESAR-Datensatz [13]. Ihr Modell ist mit einer Vielzahl von Rendering-Engines wie Unity oder Unreal kompatibel und daher sehr gut geeignet für die Verwendung bei der Generierung synthetischer Daten für Menschen. Es existieren auch Erweiterungen des SMPL-Modells wie MANO und SMPL-H, die ein deformierbares Handmodell in das Framework einführen. MANO [15] wird aus 1000 hochauflösenden 3D-Scans verschiedener Handhaltungen gelernt.

3 Methoden

Wie zuvor erwähnt, ist die Sammlung von Echtwelt-Daten in medizinischen Interventionsräumen komplex, kostspielig und erfordert die Genehmigung eines Ethik-Gremiums und der beteiligten Personen. Wie zuvor gezeigt wurde, kann DR/SDR dazu beitragen, ein Objekterkennungsnetzwerk mit ausreichender Leistung in realen Anwendungen zu trainieren.

Eine Herausforderung bei der Datensatzerzeugung für den medizinischen Interventionsraum ist jedoch die domänenspezifische Kleidung. Wir argumentieren, dass durch Zufallsvariation der Kleidungstexturen mit zufälligen Texturen die Erkennungsrate der Kleidungsarten verbessert werden könnte. Wenn jedoch diese Technik in realen Anwendungen angewendet wird, wäre zum Beispiel ein farbiges T-Shirt nicht von der gezielten blau gefärbten spezifischen Bereichskleidung zu unterscheiden. Für die allgemeine Erkennung von Autos wie in [25] macht die Randomisierungstechnik Sinn, aber für den hier vorgestellten domänenspezifischen Anwendungsfall sollte unserer Meinung nach etwas anderes verwendet werden.

Die Fragen, die wir in dieser Arbeit zu adressieren versuchen, lauten:

1. Wie können Gesundheitsfachkräfte für die Generierung synthetischer Daten modelliert werden?
2. Welche Techniken eignen sich am besten für die SDR/DR-Kleidungserzeugung?
3. Können wir die Kluft zur Realität weiter schließen, indem wir Greenscreen-Daten (Mixed Reality, MR) einbeziehen?
4. Kann die erforderliche Menge an echten Daten durch den Einsatz von SDR/DR/MR reduziert werden?
5. Kann die Genauigkeit verbessert werden, wenn reale und synthetische Daten kombiniert werden?

Für Punkt (1) argumentieren wir, ein deformierbares Modell für die menschliche Form wie die SMPL-Modelle zu verwenden. Dies bietet ausreichende Varianz für verschiedene menschliche Formen und Größen. Für Punkt (2) untersuchen wir zwei verschiedene Methoden zur Kleidungserzeugung. Erstens scannen wir verschiedene Personen mit medizinischer Kleidung in 3D ein und erstellen eine Datenbank verschiedener medizinischer Kleiderscans für jeden Kleidungsstyp, die wir SCANS nennen. Zweitens beauftragen wir einen professionellen Grafikdesigner, Assets auf Basis der Bereichskleidung zu erstellen, die wir CAD nennen. Bezüglich Punkt (3) nehmen wir Bilder vor einem Greenscreen von verschiedenen Personen auf, die medizinische Kleidung tragen, die wir manuell beschriften. Für Punkt (4) untersuchen wir, ob die erforderliche Menge an echten Daten mit konstanter Genauigkeit durch Mischen von echten und synthetischen Daten reduziert werden kann. Schließlich untersuchen wir in Punkt (5), ob die Kombination aus synthetischen Bilddaten und einem Prozentsatz an realen Daten die Genauigkeit von nur realen Daten verbessert.

Um die genannten Fragen weiter zu behandeln, haben wir Experimente durchgeführt, bei denen wir mit Hilfe des Scaled Yolov4-Objektdetektors [30] die folgenden Klassen erkennen möchten.

Die zu erkennenden Klassen sind:

Menschen

Bereichskleidung Hemd

Bereichskleidung Hose

Sterile Kittel

Medizinische Gesichtsmaske

Medizinische Haube

Medizinische Handschuhe.

Beispiele für die medizinische Kleidung sind in Abbildung 1 dargestellt.

3.1 Erschaffung des Charakters

Die medizinischen Charaktere, die wir in SDR/DR verwenden, werden durch eine Kombination von SMPL-Körpermodellen, Texturen, Animationen und Kleiderassets erstellt. In dem folgenden Abschnitt werden jede der verwendeten Komponenten präsentiert und erklärt, warum sie verwendet werden.

Ein Körpermodell ist für die Erstellung synthetischer Menschen erforderlich. Als Grundlage für unsere Charaktere verwenden wir das männliche und weibliche Modell des SMPL+H-Modells von [15]. Die Modelle decken eine Vielzahl realistischer menschlicher Formen ab, die durch zehn Blendformen zufällig variiert werden können. Wir entscheiden uns für das erweiterte SMPL+H-Modell anstelle des ursprünglichen SMPL-Modells [10]. Dies liegt daran, dass eines unserer Kleidungsstücke Handschuhe sind und durch das Hand-Rig des SMPL+H-Modells mehr Deformationen des Handschuh-Assets erstellt werden können.

Die SMPL-Modelle allein sind Oberflächenmodelle ohne Textur. Für die Erzeugung von Menschen wird eine menschliche Textur benötigt. Um mehr Variation und Realismus in das Erscheinungsbild der Charaktere zu bringen, werden die Texturkarten aus [29] verwendet. Von den 930 Texturen wurden nur 138 (69 von jedem Geschlecht) verwendet. Dies liegt daran, dass wir unsere eigenen Kleiderassets erstellt haben und nur die Texturen von Personen in Unterwäsche relevant waren. Diese Texturkarten wurden aus 3D-Körperscans aus dem CAESAR-Datensatz [13] erstellt und decken eine Vielzahl von Hautfarben und Identitäten ab, jedoch wurden alle Gesichter anonymisiert [29].

Bei der Arbeit mit synthetischen Menschen in Rendering-Engines muss die menschliche Pose modifiziert werden, um eine Vielzahl realistischer Körperhaltungen zu erzeugen. Die Modelle wurden daher durch Motion Capture (MoCap)-Daten animiert, die in unserem Labor erfasst wurden. Wir verfolgen die Bewegung von 74 Gelenken bis zu den Fingerspitzen. Wir verwenden ein eigenes Motion Capture-System mit dem Hand Gloves Add-on namens Perception Neuron Studio. Um den Datensatz einfach zu halten, haben wir in unseren Experimenten nur eine Animation verwendet. Die Möglichkeit, weitere unterschiedliche Animationen hinzuzufügen, besteht jedoch.

Nachdem das Körpermodell, die Körpertexturen und die Körperhaltungen definiert wurden, wird die medizinische Kleidung benötigt. Hier werden zwei verschiedene Ansätze untersucht. Einer ist die Generierung von medizinischer Kleidung mit einem 3D-Scanner und der andere ist die Erzeugung von entworfener

Kleidung durch einen Grafikdesigner. Die mit dem 3D-Scanner erstellten 3D-gescannten Kleidungsassets, die wir SCANS nennen, werden mit einem 3D-Scanner namens Artec Leo erstellt. Eine 3D-Auflösung von 0,2 mm wurde verwendet, um die medizinischen Kleider zu erfassen. Für unseren synthetischen Trainingsdatensatz haben wir Kleiderscans von 4 männlichen und 4 weiblichen Modellen verwendet. Auf diese Weise werden Variationen der Echtwelt-Texturen, einschließlich Reflexionen, Falten, Farben und Oberflächentexturinformationen, gesammelt. Nachdem ein erstes Modell aus dem 3D-Scanner erstellt wurde, passen wir die Kleidung an den Standard-Männlichen- und Weiblichen-SMPL+H-Charakter mit 3D-Modellierungstechniken an. Nach unseren Recherchen gibt es in der medizinischen Kleidung üblicherweise die Farben Blau, Grün und Hellrosa. Um diese Variation in unserem Datensatz abzudecken, haben wir die Texturkarten ergänzt. Beispiele für die gescannten und geriggeden Kleidungsassets sind in Abbildung 2 zu sehen.

Um die Leistung der 3D-gescannten Kleidungsassets zu bewerten, vergleichen wir sie mit von Hand entworfenen Kleidungsassets, die wir CAD nennen. Dafür haben wir einen Designer auf Fiver gebeten, die Kleidung zu modellieren. Beispiele für diese Assets sind in Abbildung 3 zu sehen. Zunächst haben wir geprüft, inwieweit frei verfügbare Assets aus den Asset-Stores zu diesem Zweck verwendet werden können. Es gibt jedoch keine Assets, die unsere spezifische Kleidung vollständig abdecken. Daher haben wir uns entschieden, die Assets entwerfen zu lassen. Die entworfenen Assets wurden genauso wie unsere gescannten Assets verarbeitet. Sie sind ebenfalls verformbar und sind an dasselbe Rig gebunden.

Die Erstellung der synthetischen Personen erfolgt mithilfe der Rendering-Engines Unreal Engine 4 und Unity. Das NDDS-Plugin für die Unreal Engine wird verwendet, um die DR-Bilddaten zu generieren, und ein Unity-Plugin wird verwendet, um die SDR-Bilddaten zu generieren.

Für die synthetische Datenerzeugung von DR wird ein Unreal Engine 4-Plugin namens NDDS [23] verwendet. Dies ermöglicht die Erzeugung von RGB-Bildern mit Geschwindigkeiten ähnlich echter Kameras sowie Tiefenbildern und Segmentierungsmasken der Szene innerhalb von Unreal Engine 4. Das Plugin erstellt auch Bounding-Box-Etikettierungsdaten für jedes Objekt in der Szene in 2D und 3D. Das Tool wurde speziell für DR entwickelt und bietet daher Werkzeuge für die zufällige Gestaltung der Szene, wie die Position von Objekten oder Kameras, Beleuchtung und Ablenkungsobjekte, unter anderem. Mit einem modularen Charakter-Blueprint ermöglicht NDDS die Erzeugung synthetischer Datensätze für sterile Kleidung, wobei 3D-gescannte Kleidung oder entworfene Kleidung verwendet wird. Beispielbilder sind in Abbildung 4 in der oberen Reihe gezeigt. Wir erstellen zwei separate Datensätze, einen mit SCANS-Assets und einen anderen mit CAD-Assets für DR. Ein Aktivitätsdiagramm, das den Blueprint für die modulare Charaktererstellung in NDDS darstellt, ist in Abbildung 5 gegeben.

Für die Datensatzerzeugung mit SDR haben wir ein Unity-Plugin namens ML-ImageSynthesis [22] als Basis verwendet und es an das Universal Rendering Pipeline (URP) angepasst, um die Qualität zu verbessern. Unter Verwendung von Unity 2020.3.32f1 wurden zusätzliche Komponenten hinzugefügt, um einen Export zusätzlicher Metadaten für jedes generierte Bild zu ermöglichen, wie z.B. Kameraparameter, Bounding-Boxen und Weltposition. SDR wird durch die Verwendung einer Vielzahl von maßgeschneiderten Komponenten ermöglicht, die die Zufallsvariation von Parametern wie Beleuchtung, Material, Textur und Position

ermöglichen. Das Plugin ProBuilder, das von Unity bereitgestellt wird, wurde verwendet, um einen Interventionsraum basierend auf der Ziel-Domäne des realen Datensatzes (Klinikum) zu erstellen. Die Zufallsgenerierung der Szene wird durch die Verwendung der zuvor genannten Zufallskomponenten erreicht. Ein Aktivitätsdiagramm, das den Blueprint für die modulare Charaktererstellung in Unity darstellt, ist in Abbildung 6 gegeben.

3.2 Datasets

Um potenzielle Genauigkeitsunterschiede zwischen SCANS, CAD und der Kombination mit echten Daten zu untersuchen, wurden verschiedene Datensätze generiert.

Zunächst wurden synthetische Datensätze von DR und SDR für sowohl SCANS als auch CAD-Kleidung mithilfe der vorgestellten Pipelines in der Unreal Engine und Unity generiert. Diese Datensätze werden verwendet, um zu experimentieren und herauszufinden, ob gescannte Kleidung oder entworfene Kleidung bessere Ergebnisse liefern.

Zweitens wurde ein Datensatz vor einem Greenscreen gesammelt, den wir Mixed-Reality (MR) nennen. Er besteht aus 8 Personen im Trainingsdatensatz und 2 Personen im Validierungsdatensatz. Die aufgezeichneten Personen bewegen sich vor dem Greenscreen mit einer bestimmten Greifbewegung, die auch als Bewegungsanimation für die synthetischen Daten verwendet wird. Dieser Datensatz zielt darauf ab, die Lücke zwischen den synthetischen Bilddaten und den echten Daten weiter zu schließen, indem reale Daten in einem Mixed-Reality-Szenario eingeführt werden, ohne Daten in der Ziel-Domäne aufzeichnen zu müssen.

Schließlich wurde ein Datensatz der Ziel-Domäne aufgezeichnet, den wir Klinikum nennen. Er dient als Baseline-Vergleich für alle Modelle und stellt auch die Testdaten dar. Dies ergibt 331 markierte Testdaten. Um eine ausreichende Menge an Testdaten aus den verfügbaren zu erhalten, haben wir uns hier entschieden, ein anderes Split im Vergleich zu den anderen Datensätzen zu verwenden.

In den folgenden Abschnitten stellen die Zeilen Klinikum (100) oder Klinikum (15) alle verfügbaren Klinikum-Trainingsdaten dar, bzw. 15 % zufällig ausgewählte Trainingsdaten. Die Zeilen real (100) und real (15) bedeuten dasselbe.

Alle Datensätze sind in Trainings- und Validierungsdaten aufgeteilt.

Beispiele für echte Daten vor dem Greenscreen mit ausgetauschtem Hintergrund sind in Abbildung 7 zu sehen. Beispiele für die synthetischen Daten sind in Abbildung 4 zu sehen, und schließlich sind Beispiele aus den klinischen Testdaten in Abbildung 8 zu sehen.

Tabelle 1 gibt eine Aufschlüsselung der Größen und Verteilungen der Datensätze.

4 Experimente

Es wurden Experimente durchgeführt, um zu untersuchen, ob und wie gut SCANS im Vergleich zur CAD-Kleidung zur Erkennung in der medizinischen Umgebung geeignet sind. Zusätzlich wurden Experimente durchgeführt, um festzustellen, ob eine Kombination aus echten Daten und synthetischen Daten eine ausreichende Genauigkeit erreichen oder sogar echte Daten allein übertreffen kann. Schließlich wurden MR-Daten in die Experimente einbezogen, um festzustellen, ob sie die Kluft zur Realität weiter schließen können.

Für unsere Experimente haben wir die Implementierung von Scaled-Yolov4 [30] von GitHub verwendet. Zunächst wurden 6 verschiedene Basisnetzwerke trainiert, um einen grundlegenden Vergleich der verschiedenen Methoden zu zeigen und festzustellen, ob SCANS oder CAD-Kleidung bessere Ergebnisse liefern. Diese Basismodelle umfassen Trainings mit synthetischen Daten (DRscans, DRcad, SDRscans, SDRcad), Mixed-Reality-Daten (MR-DR) und echten Daten aus dem klinischen Bereich (Klinikum train).

Das Training wurde mit den YOLOv4-p5-Gewichten und den Standard-Feinabstimmungsparametern durchgeführt, die vom Scaled Yolo-V4 GitHub-Repository bereitgestellt wurden. Nur die Mosaik-Augmentationsverhältnis-Parameter α und β wurden von 8,0 auf 20,0 erhöht. Dadurch wird der Effekt der Bildüberlagerung in der verwendeten Implementierung geschwächt. Zusätzlich wurde eine Green-Channel-Augmentation verwendet, wenn MR-Daten im Trainingsatz vorhanden waren, um den Greenscreen-Effekt zu reduzieren, mit dem wir in einigen Klassen Probleme hatten. Hierbei versuchen wir, eine Basislinie für die MR-DR-Daten zu etablieren. Experimentell haben wir festgestellt, dass die Verwendung der Green-Channel-Augmentation die Genauigkeit verbessert.

Alle Netzwerke wurden für 300 Epochen trainiert und erreichten Konvergenz. Alle trainierten Modelle wurden mit einem IoU-Schwellenwert von 0,5 und einem Konfidenz-Schwellenwert von 0,2 auf dem Klinikum-Testdatensatz getestet. Das verwendete Yolov4-Netzwerk war yolov4-p5, die Bildgröße wurde sowohl im Training als auch im Test auf 896 festgelegt und die bereitgestellten vortrainierten Gewichte wurden verwendet. Die Ergebnisse der Basismodelle sind in Tabelle 2 dargestellt.

Die Ergebnisse zeigen, dass CAD-basierte synthetische Daten im Allgemeinen bessere Ergebnisse liefern als auf SCAN basierende Daten in diesem Experiment. Deshalb verwenden wir den SDRcad-Datensatz für alle weiteren Experimente.

Um zu untersuchen, inwieweit die Menge an echten Daten reduziert werden kann, wenn sie zusammen mit synthetischen oder MR-Daten verwendet werden, um eine ausreichende Genauigkeit zu gewährleisten, wurden Experimente mit einer prozentualen Verteilung der Klinikum-Trainingsdaten durchgeführt. Hier ist unser Hauptziel herauszufinden, ob die Verwendung von synthetischen Daten zusammen mit MR-Daten und einem Anteil an echten Daten die Genauigkeit von echten Daten allein überschreitet. Wir haben uns für 15% echte Daten entschieden, da dies zu 99 verbleibenden Trainingsbildern führt, die unserer Meinung nach eine ausreichende Menge an Bildmaterial darstellen, das von Hand beschriftet werden kann. Wir

haben uns entschieden, die Mosaik-Augmentation auch während dieser Experimente zu verwenden und alle Datensätze als Trainingsdaten zu verwenden, anstatt ein Feinabstimmungsexperiment durchzuführen. Wir argumentieren, dass das Netzwerk relevante Merkmale besser lernen kann, während es gleichzeitig die Vorteile der zusätzlichen synthetischen Daten beibehält, wenn es eine Variation aller verwendeten Datensätze mit Mosaik-Augmentation sieht, im Vergleich zu nur einer Feinabstimmung. Während dieser Experimente haben wir uns entschieden, die oben genannte Green-Channel-Augmentation in allen Trainings einzubeziehen. Zusätzlich wurden die Echt-Datenläufe mit der gleichen Anzahl von Optimierungsschritten trainiert, um sicherzustellen, dass das Modell konvergiert, während weniger Trainingsdaten verwendet werden.

Die Ergebnisse der Folgeexperimente mit SDRcad, MR-DR sowie echten Daten sind in Tabelle 3 dargestellt.

5 Ergebnisse

Die Ergebnisse des ersten Experiments sind in Tabelle 2 zu sehen. Beim Vergleich von SCANS-Kleidung und CAD-Kleidung in den DR- und SDR-Szenarien zeigten beide Male die Datensätze mit CAD-Kleidung bessere Ergebnisse. Dies hat uns zu diesem Zeitpunkt überrascht. Die möglichen Gründe dafür werden im Diskussionskapitel erörtert.

Ebenso zeigt sich aus den Ergebnissen der einzelnen Klassen in Tabelle 4, dass die CAD-Kleidung mit Ausnahme der Maskenklasse in jedem Fall bessere Ergebnisse liefert als die SCANS-Kleidung.

Es ist auch klar aus den Ergebnissen ersichtlich, dass SDR DR überlegen ist. Dies war angesichts früherer Arbeiten in diesem Bereich zu erwarten, da in den präsentierten SDR-Experimenten die Umgebung mit den in der Klinik vorhandenen Objekten angereichert wird und das Netzwerk dadurch besser auf Ablenkungen eingestellt ist.

Während die MR-DR-Ergebnisse in vielen Klassen der SDR unterlegen sind, sind sie im Vergleich zu DR mit Ausnahme der Gown-Klasse überlegen. Die Gründe für die schlechte Leistung der Gown-Klasse in den SDR-Experimenten wurden bereits im Kapitel 3.1 erwähnt. Hier war der Greenscreen-Effekt besonders negativ, weshalb die zusätzliche Green-Channel-Augmentation angewendet wurde.

Die verwendete Bewertungsmetrik ist die durchschnittliche Präzision (mAP) mit 2 verschiedenen Intersection over Union (IoU) Schwellenwerten. Die beiden Schwellenwerte sind 0,5:0,95 für mAP und 0,5 für mAP50, wie sie in der Scaled-Yolov4-Implementierung [30] verwendet werden.

Für die Hosenklasse erzielte MR-DR in diesem Experiment die besten Ergebnisse für mAP. Für mAP50 hingegen ist die Situation bei den anderen Klassen gleich. Dieser Unterschied kann wahrscheinlich auf die ungenaue Begrenzung der Hosen unter dem Hemd zurückgeführt werden. Hier erklärt sich der Unterschied

durch eine möglicherweise unterschiedliche Kennzeichnung der realen Bilddaten (Klinikum, MR-DR) im Vergleich zur automatischen Kennzeichnung mit den synthetischen Bilddaten.

Im Allgemeinen liefern die Klassen Maske und Handschuh die schlechtesten Ergebnisse. Dies gilt auch für die Kategorie mAP50 im Vergleich zur Wirkung, die für die Hosenklasse beschrieben wurde. Dies kann auf die relativ geringe Größe dieser Klassen zurückgeführt werden. Im Testdatensatz sind schwierige Fälle enthalten, in denen Personen von der Seite zu sehen sind. In diesen Fällen sind die Maske oder die Handschuhe nur seitlich sichtbar, und das begrenzte Bounding-Box-Gebiet umfasst nur wenige Pixel. Dieser Effekt tritt auch in den realen Daten auf. Hier erreicht der Klinikum-Trainingsdatensatz eine Genauigkeit von 53,66% bei mAP, während mAP50 wieder bei 95,85% liegt. Dieser Effekt ist auch in den anderen Trainingsdatensätzen zu sehen, ist aber weniger stark ausgeprägt.

Die Ergebnisse des Folgeexperiments, das den Vergleich von synthetischen Bilddaten zusammen mit Mixed-Reality-Daten und einer prozentualen Verteilung von echten Daten untersucht, sind in Tabelle 3 dargestellt.

Hier verbessert der gemeinsame Trainingsdatensatz aus SDR+MR die Genauigkeit der beiden einzelnen Datensätze aus dem ersten Experiment. Dennoch besteht immer noch ein Unterschied im Vergleich zu 100% echten Daten und sogar 15% echten Daten bei mAP und ist kleiner bei mAP50. Trotzdem ist dieses Ergebnis von großem Interesse für zukünftige Arbeiten und Experimente, da es einen Weg zeigt, komplett auf echte Daten aus der Ziel-Domäne zu verzichten. Die Möglichkeiten von Mixed-Reality zusammen mit synthetischen Daten sollten daher weiter untersucht werden.

Des Weiteren ist zu erkennen, dass die Hinzunahme von SDR+MR-Daten zu 15% und 100% echten Daten die Genauigkeit der Erkennungen im Vergleich zu echten Daten allein erhöht. Bei 15% echten Daten liegt dies bei einer Steigerung von 2,53% und bei 100% echten Daten bei einer Steigerung von 1,4% für mAP.

Betrachtet man die Ergebnisse der einzelnen Klassen, die in Tabelle 5 gezeigt werden, liefert der Datensatz mit SDR+MR+Klinikum(100) die besten Ergebnisse für alle Klassen außer Masken.

In Bezug auf die Klassen Maske und Handschuh, die im ersten Experiment die schlechtesten Ergebnisse lieferten, kann die Genauigkeit durch die Zusammenführung von SDR+MR-Daten verbessert werden. Dies ist ein weiterer Hinweis auf das Potenzial von synthetischen und Mixed-Reality-Daten für Anwendungen im medizinischen Bereich, um den Bedarf an echten Daten erheblich zu reduzieren.

Bei Betrachtung der Ergebnisse der Hosenklasse, die im ersten Experiment bei SDR 55,32% in mAP erreichte, kann diese durch Kombination mit MR-Daten auf 80,22% verbessert werden. Ebenso ist der Einfluss der MR-Daten mit dem Greenscreen-Effekt bei der Gown-Klasse zu sehen. Die Kombination von SDR+MR-Daten beseitigt weitgehend den negativen Einfluss der MR-Daten aus dem ersten Experiment. Dies ist ein

Indikator für die spürbar niedrige Genauigkeit dieser Klasse, möglicherweise aufgrund des Greenscreen-Effekts.

Die Inferenz-Ergebnisbilder des Trainings mit SDR+MR+real(100)-Daten sind in Abbildung 9 zu sehen. Nur zur Visualisierung der Bildergebnisse wurde hier eine leicht höhere Konfidenzschwelle von 0,4 verwendet im Vergleich zu den in allen Tabellen präsentierten Ergebnissen (0,2).

6 Schlussfolgerung

Wir konnten zeigen, dass die Verwendung von SMPL-Modellen in Kombination mit gescannter oder entworfener medizinischer Kleidung eine geeignete Methode zur Modellierung von medizinischem Personal für KI-Fragen im Interventionsbereich ist, basierend auf dem Beispiel der Erkennung von medizinischer Kleidung. Während unserer Experimente stellten wir fest, dass die entworfene Kleidung im Allgemeinen bessere Leistungen auf unserem Testdatensatz erzielte als die per 3D-Scan erfassten Kleidungsstücke. Dieses Ergebnis hat uns überrascht, da wir erwartet hatten, dass die potenziell genaueren Texturen des 3D-Scans sich positiv auf die Erkennungsraten auswirken würden. Gemäß den Ergebnissen kann jedoch nicht ausgeschlossen werden, dass Artefakte in der Render-Pipeline oder Preprocessing-Pipeline, die wir nicht erkannt haben, darauf Einfluss haben könnten. Zusätzlich kann in weiteren Arbeiten untersucht werden, ob gescannte Kleidung möglicherweise so gestaltet werden sollte, dass sie besser verformbar ist, um den Vorteil gescannter Texturen mit realistischer Bewegung des Gewebes zu kombinieren. Um eine endgültige Aussage über das Potenzial von 3D-gescannter Kleidung für die Modellierung von medizinischem Personal zu treffen, sollten weitere Experimente durchgeführt werden. Die Verwendung von Mixed-Reality-Daten zusammen mit den synthetischen Daten hat die Kluft weiter geschlossen, und obwohl der Unterschied gering ist, konnten wir zeigen, dass die Verwendung von synthetischen, Mixed-Reality- und 15% echten Daten die verbleibende Lücke zu 100% echten Daten nahezu schließen kann. Generell konnten wir zeigen, dass die Verwendung von synthetischen und Mixed-Reality-Daten zusammen mit einem Prozentsatz an echten Daten die Verwendung von nur echten Daten übertrifft.

Dies ist ein gutes Zeichen für das Potenzial von synthetischen und Mixed-Reality-Daten bei Fragen rund um medizinische Interventionen, da sie ausreichend Informationen enthalten, um die Kluft zur Realität zu schließen. Ein Weg mit mehreren prozentualen Verteilungen von echten Daten zusammen mit SDR+MR-Daten ist für einen größeren Testdatensatz mit mehreren medizinischen Fachkräften interessant und Gegenstand zukünftiger Arbeiten. In den gezeigten Ergebnissen wurde bereits gezeigt, dass die Fusion von SDR+MR-Daten zusammen mit den echten Daten die Genauigkeit verbessert.

Für Fragen im Interventionsbereich ermöglicht insbesondere Mixed-Reality den Erwerb von Daten außerhalb der Ziel-Domäne, um Datenschutz-Herausforderungen zu minimieren. In zukünftigen Arbeiten sollten Methoden erforscht werden, um den Greenscreen-Spill-Effekt während der Datenerzeugung zu reduzieren und die resultierenden Daten in komplexeren Szenen ähnlich wie SDR zu visualisieren. Hierbei ist die Verwendung von Deep-Learning-Netzwerken zur Bildverbesserung interessant, um diese Möglichkeit zu untersuchen.

Zusammenfassend ist die vorgestellte Modellierung von medizinischem Personal eine vielversprechende Methode, um das Problem fehlender Datensätze aus medizinischen Interventionsräumen zu lösen. Wir werden es weiter für verschiedene Aufgaben im medizinischen Bereich untersuchen.

Verfügbarkeit von Daten und Materialien

Zusätzlich zu den generierten synthetischen Bilddaten bestehen die verwendeten Datensätze auch aus echten Bilddaten, die personenbezogene Daten enthalten. Dies erlaubt uns nicht, die Daten online zur Verfügung zu stellen. Es ist jedoch möglich, die Daten vom Autor nach einer gerechtfertigten Anfrage durch einen bilateral vereinbarten Datenschutzvertrag zu erhalten.

Anmerkung

1. Perception Neuron Studio suite and gloves addon: <https://neuronmocap.com/perception-neuron-studio-system>.
2. 3D Scanner Artec Leo: <https://www.artec3d.com/portable-3d-scanners/artec-leo>.
3. azeemdesigns: [https://www.fiverr.com/azeemdesigns?](https://www.fiverr.com/azeemdesigns?source=order_page_user_message_link)
source=order_page_user_message_link.
4. <https://github.com/WongKinYiu/ScaledYOLOv4>.

Veröffentlichungen Thema KI & Co.

PERSONALENTLASTUNG – MIT HILFE DES HYGIENE KI – COPILOTEN

LERNEN SIE IN UNSEREM WORKSHOP, KI-KONZEPTE FÜR IHRE ARZTPRAXIS & CO. ZU ERSTELLEN

KÜNSTLICHE INTELLIGENZ IN DER HYGIENE: EINE REVOLUTION AUS SICHT EINES HYGIENEINSPEKTORS

ALLE REDEN VON CHATGPT – WELCHEN NUTZEN HAT IHRE PRAXIS?