Einleitung

**Die Fähigkeit** zur Pose Estimation, also das bestimmen der Position und Orientierung eines Gerätes, in Gebäuden **verschafft** im Bauwesen eine Reihe von **Anwendungen** wie z.B. die automatische Baufortschritterfassung oder Facility-Management und Navigation über Augmented Reality.

**Es gibt Lokalisierungsansätze** in Gebäuden, die sich an **Technologien** wie z.B. **Bluetooth Beacons** oder Wireless Access Points usw. bedienen, worin ein **visuelle Lokalisierungsverfahren** über eine mobile Kamera die kostengünstigste darstellt, da zum einen heute jeder ein Smartphone bei sich führt und zum anderen eine **flächendeckende Hardware Installation** erspart bleibt.

**Visuelle Lokalisierungsansätze** wie z.B. die **visuelle Odometrie** oder **SLAM** sind **relativ zum Ausgangspunkt**. Eine Lösung für die absolute Bestimmung der Pose über einen Anfragebild ist durch das Suchen eines korrespondierenden Bildes in einer Bildergalerie oder Pose Regression über Bild-Features möglich. Allerdings **benötigen** solche Verfahren **Ground-Truth Daten** von jeder Ecke, die Zeit**- und kostenintensiv zu erheben sind**.

Dafür versuchen Ansätze über **künstliche neuronale Netze** wie z.B. PoseNet eine Lösung zu finden, worin das Ermitteln **der Ground-Truth-Daten über SfM-Methoden** genügt, sprich man nimmt nur ein Video der Zone auf, worin man es quasi überquert. Allerdings sind SfM-Methoden sehr **fehleranfällig**, sprich dreht man sich zu schnell um die Ecke wird die Verbindung zum vorherigen Frame verloren etc. und können sehr viel Zeit in Anspruch nehmen.

Grundlagen

Acharya et al. versuchten daher, **statt über SfM-Methoden** Trainingsdaten aus der Simulation eines Korridors zu gewinnen. Dabei erzeugten Sie synth. Daten die sich von der **Realitätstreue** von Karikaturistischem bis hin zum Photorealistischem unterscheideten.

**Sie konnten dabei über PoseNet** durch das Trainieren mit den unterschiedlichen synth. Datentypen die Pose mit einer Akkuratesse von ca. 5m in der Position und 20° in der Orientierung bestimmen. Durch die Gradientenbildern der synth. Daten konnten Sie mit einer besseren Akkuratesse von ca. 2m zu 7 Grad konnten die Pose bestimmen.

**Allerdings** verlief deren Strecke **überwiegen in einer Richtung,** auf **einer Ebene und** in einem **kleinen Korridor**. Deshalb war **es Ziel** meiner Bachelorarbeit, deren Ansatz über Gradientenbildern in einem **größeren Gebäude** auf längeren Strecken zu untersuchen, die einerseits auf mehreren Richtungen verlaufen und anderseits auf mehreren Etagenebenen erstrecken.

Methodik

In dieser Abbildung ist noch einmal der **Ansatz als Pipeline** dargestellt. PoseNet wird mit den Gradientenbilder der synth. Daten trainiert und anschließend mit den Gradientenbilder der realen Daten evaluiert. Vorerst wird PoseNet mit den Gewichten eines Models initialisiert, welches auf der GoogLeNet Architektur mit den Places Datensatz trainiert wurde. Die Position und die Orientierung als Quaternion ist das Output des Netzwerks.

**Angelehnt an Acharya et al.** wurden **3 synthetischen Datentypen** erstellt, auf die ich bei der Verarbeitung der Daten eingehen werde. Und insgesamt wurde dieser Ansatz **auf 4 Strecken in 2 Gebäuden** untersucht, die ich im **Unterkapitel Datensätze** näher beschreiben möchte.

**Die Trainingsparameter** wurden weitestgehend übernommen oder gleichermaßen bestimmt, **bis auf eins**, welches mit Grid-Search Verfahren bestimmt wurde und die dabei ermittelten Werte sollen auch später als Referenzwerte dienen. Nun möchte ich jetzt mit den **zwei markierten Punkten fortfahren**.

## Verarbeitung der Daten

Ziel war es mit Gradientenbildern das Netzwerk zu trainieren und zu evaluieren. Also wurden die Gradientenbilder der realen sowie synth. Bilder berechnet. Dabei **entstanden bei den synth. Datensätzen** durch die **künstlichen Lichter** **Artefakte**, welche mit einem **Treshold-Verfahren** unterdruckt werden konnten.

Hier sieht man die **erwähnten Artefakte** und hier sieht man **Beispielbilder von jedem Datentyp**. Ebenso sieht die **dazugehörigen Gradientenbildern**. Die Datensätze **cartoon** und **photoreal** **unterscheiden** sich nur bei den **Render-Engines**. Wie man sich denken kann, wurde im photoreal-Datensatz ein **komplexeres Render-Engine** verwendet. Beim edge-Datensatz wurden die **Objektkanten** über die **Simulationssoftware** markanter dargestellt und eine **konstante Beleuchtung** verschaffen.

## Datensätze

Ziel war es den Ansatz von Acharya **auf längeren und in mehreren Richtungen verlaufende** sowie auf **mehrere Etagenebenen erstreckende** Strecken zu untersuchen. Es wurden Daten aus der nördlichen Hälfte des 6. Stockwerks des IC-Gebäudes und der Seminargebäude der Hochschule Bochum erhoben.

Die Gebäudesimulationen unterschieden sich im Detail. Während die Simulation der IC-Gebäude **minimalistisch** war, enthielt die Simulation von der Hochschule Objekte der **technischen Gebäudeausrüstung** wie z.B. Feuermelder usw. Weiterhin bestand die IC-Simulation aus sich **wiederholenden Strukturen**. Hingegen hatte die HS-Simulationen **ein eindeutiges Treppenhaus**, weshalb ein Datensatz aus dem IC erhoben wurde und 3 aus der Hochschule.

Der aus dem IC-Gebäude erhobene Datensatz wurde IC-loop benannt und bildet wie man sehen kann eine **geschlossene Scheife**. Die HS-Gamma Strecke, ausgehend aus einem Zimmer, verläuft in eine Schlaufe und übergeht **zu einem Flur.** Die Strecken HS-stairs-up und down erstrecken sich **ab – und aufwärts** auf einer Treppe **naheliegend der Schlaufe** im HS-gamma.

Die Strecken, die in den **Simulationen rekonstruiert** wurde, sind grün gekennzeichnet. Die Blauen Strecke geben die erhobenen realen Daten an. Wie man sehen kann, wiesen die realen Daten **eine Drift** auf. Die **schwarzen Umrisse** der Kameras geben die **Aufnahmerichtung an.**

Ergebnisse

Wie **anfangs erwähnt** besteht das Output des PoseNet aus **zwei Komponenten**. Einmal die Positionsfehler und einmal aus dem Orientierungsfehler. Ein Ergebnis wird anhand seines **Positionsfehler verglichen**. Die Akkuratesse der Netzwerke gibt **den Median der Evaluationsergebnisse an**.

Außerdem wurden die **zuvor mit den synth. Daten** trainierten Netzwerke **zweimal evaluiert.** Die erste Evaluation folgte mit den korrespondierenden synth. Evaluationsdaten. Die zweite Evaluation folgte mit den realen Evaluationsdaten. **Ferner sollten** die bei der Bestimmung des Beta Hyperparameters **ermitteln Akkuratessen** als **Referenzwerte** dienen.

## IC-loop

Auf der IC-loop Strecke konnte eine durchschnittliche Akkuratesse von 1.80m, 8.05° mit den synth. Daten erzielt werden. Hingegen konnte mit realen Daten 16.68, 73.25 erreicht werden. **Betrachtet man** nun das beste Netzwerk mit realen Daten, stellt man fest, dass dieses Netzwerk alle Evaluationsdaten in einem ca. 5x30m großen **Teilbereich bestimmt hat**. Ebenso wurde die Orientierung als die Aufnahme Richtung **der unteren horizontalen Strecke** bestimmt.

## HS-gamma

Auf der HS-gamma Strecke konnte eine durchschnittliche Akkuratesse von 0.85m, 8.07° mit synth. Daten erreicht werden. Mit den realen Daten konnte durchschnittlich 9.67m, 32.10° erzielt werden. **Bei genauem Hinsehen des Netzwerks** mit den besten Ergebnissen bei realen Daten stellt man fest, dass das Netzwerk die Evaluationsdaten in einem ca. 5x20m großen Teilbereich lokalisiert hat, ebenso wurde die **Orientierung als die Aufnahmerichtung der horizontalen Strecken** bestimmt.

## HS-stairs-up

Auf der HS-stairs-up Strecke konnte mit den synth. Daten eine durchschnittliche Akkuratesse von 0.85, 8.07° erreicht werden. Mit den realen Daten hingegen nur 4.75m, 56.15° **und beim genauen Betrachten der Evaluation des grad-edge Netzwerks**, sieht man das die Evaluationsdaten zwischen dem oberen und unteren Treppenlauf lokalisiert wurden. Es gibt **abwechselnd immer größere Positionsfehler**. **Die Größten sind auf dem Treppenpodest** zu erkennen. Ebenso konnte man auch erkennen, dass **die Orientierung hin und wiedermal in der entgegensetzten**, sprich 180°, falsch bestimmt wurden.

## HS-stairs-down

**Die Ergebnisse von HS-stairs-down sind den Ergebnissen von HS-stairs-up ähnlich**. Es konnte mit synth. Daten eine Akkuratesse von 0.93m, 8.03° erreicht werden, und mit den realen Daten 5.01m, 49.29°. Hierbei wurden die Evaluationsdaten **ebenso zwischen** dem **oberen und unteren Treppenlauf** lokalisiert. **Abwechselnde Positionsfehler sind hierbei sichtbarer, vor allem die der entgegengesetzten Orientierungen.**

Diskussion

Nach dem die **Ergebnisse sachlich vorgestellt** wurden, möchte ich diese nun disktutieren. Die Diskussion meine Bachelorarbeit ist in 3 Abschnitte geteilt. Zuerst wurden die **Defizite** der **angewandten Methodik** auflisten, s.d. man diese im Hinterkopf behalten sollte. Danach wurden die Ergebnisse interpretiert und mit den Ergebnissen der **Literatur verglichen**. Im Anschluss wurden Empfehlungen für **weiterführende Forschungen** gegeben. Ich werde mich jetzt nur auf die Diskussion der Ergebnisse fokussieren.

## D. der angewandten Methodik

Die in dieser Arbeit verwendetet Intel Realsense T265 versprach bei Bestkonditionen eine Drift von 1%. Wir konnten die Bestkonditionen nicht erfüllen s.d. eine Abweichung bis zu 5% entstanden ist. Trotzdem konnte eine Korrespondenz der Ground-Truth-Daten zwischen der Erhobenen und von der Simulation verschafft werden. Allerdings ist ein negativer Einfluss auf die Akkuratesse der domänenübergreifenden Evaluationen nicht auszuschließen bzw. denkbar.

Ein Weiterer Punkt ist, dass die Akkuratesse eines KNNS in dieser Arbeit durch das stochastische Gradientenabstiegsverfahren im Trainingsprozess vom Zufall abhängt. Die bestmögliche Akkuratesse zu finden würde den Rahmen einer Bachelorarbeit sprengen. Deshalb wurden je Datentyp die Trainingsprozesse 5-mal wiederholt. Daher könnte bei weiteren Trainingsprozesse bessere Ergebnisse erzielt werden.

Zudem wurden die Hyperparameter von Acharya et al. übernommen und könnten auf deren Datensatz optimiert sein bzw. mit deren Datensatz besser harmonieren. Infolgedessen könnten bei einer anderen Belegung der Hyperparameter ebenso bessere Ergebnisse erzielt werden.

## Diskussion der Ergebnisse

In dieser Arbeit konnte auf die erhobenen Datensätze bei einer **domänenübergreifenden Evaluation** eine durchschnittliche Akkuratesse von 10.95m, 49.69° erzielt werden. **Ferner** sollten die bei der Hyperparameterbestimmung erhaltenen Akkuratesse von 1.17m, 7.34° als **Referenzwerte** dienen. Dieses konnte die durch die Evaluationen **mit Daten der gleichen**

Domäne mit 1.19m, 8.35° angenähert werden. Allerdings wurde ein domänenübergreifender Ansatz **gestrebt** und diese ist bei den betroffenen Gebäuden mit einer **Positionsakkuratesse** von ca. 11m **nicht denkbar**.

Bei den HS-stairs-down und up Strecken **konnte keine Generalisierungsfähigkeit** der Netzwerke erkannt werden, vielmehr konnte eine **zufällige Zuordnung interpretiert** werden. Naja, so eine Treppe weist eben von Natur aus **wiederholenden Strukturen.** Deshalb könnten die Unfähigkeiten der Netzwerke mit **perceptual-aliasing begründet werden**. Perceptual-aliasing ist **eine der größten Herausforderungen von Lokalisierungsverfahren** und wird in diesem Kontext als das **gleiche Erscheinen zweier verschiedenen Stellen** eines Gebäudes definiert.

Weiterhin wurden **die Evaluationsdaten** der Strecken IC-loop und HS-gamma in einem

ca. 5m breiten und 20m bis 30m langen **Teilbereich lokalisiert**. Ebenso konnte man auch erkennen, dass die Netzwerke **nur eine Richtung erlernt haben**. **Interessanterweise** stellten diese Ergebnisse **Parallelen** zudem Datensatz bzw. Ergebnissen von Acharya et al. her.

Während das **Lokalisieren** in **einem Teilbereich** bei der IC-Loop Strecke auf **perceptual-aliasing geführt werden** könnte, da sowohl **die vertikalen als auch die horizontalen Strecken optisch gleich aussehen,** unterscheiden die Evaluationsdaten der vertikalen und entlang der Schlaufe Strecke sich von den horizontalen Strecken. **Nichtdestotrotz wurden diese nur in einem Teilbereich der linken horizontalen Strecke lokalisiert**.

Die schlechten Ergebnisse auf den HS Datensätzen **könnten auf das hohe Level-of-Detail** der Simulationen zurückgeführt werden, da Achaya et al. **überraschenderweise feststellten**, dass die **Zunahme des** Level-of-Details zu einer **Abnahme der Akkuratesse führte**.

**Widersprüchlicher weise** empfahlen die Autoren für eine besseres Akkuratesse ein hohes Level-of-Detail der Simulationen. Naja, die Ergebnisse der HS-gamma Strecke **waren weit über einem vorstellbaren Einfluss** des hohen Level-of-Detail. Es sollte auch nicht außer

Acht gelassen werden, dass die Ergebnisse von HS-Gamma Gemeinsamkeiten zu den Ergebnissen von IC-loop und den Ergebnissen von Acharya et al. aufweisen.

**Was so eigentlich schon fraglich ist**. PoseNet ist grundsätzlich nicht begrenzt auf eine Teilzone von ca. 5m breite und 30m lange**. In dieser Arbeit konnte mit Daten der gleichen**

**Domäne** eine Positionsakkuratesse von ca. 1m erreicht werden. Ferner konnten Wlach et al. mit PoseNet auf einen größeren Datensatz die Pose mit ca. 2m Positionsgenauigkeit bestimmen. **Daher liegt die Schlussfolgerung nahe**, dass PoseNet bei **domänenübergreifender Anwendung** nur in einem kleinen Bereich in einer Richtung trainiert werden kann.

Fazit

So, um jetzt alles nochmal abzurunden und abzuschließen**. Ziel dieser Bachelorarbeit** war es den Ansatz von Acharya in **größeren Gebäuden auf längeren Strecken** zu untersuchen. **Insgesamt** wurde dieser Ansatz mit **2 Gebäuden auf 4 Strecken untersucht**.

**Zusammenfassend** konnte festgestellt werden, dass ein **Lokalisierungsverfahren durch eine domänenübergreifende Anwendung von PoseNet** auf die hier **erhobenen Datensätze** nicht möglich gewesen ist.

Es konnte **bei Daten der gleichen Domäne** eine Akkuratesse von ca. 1m in der Position und 8° in der Orientierung erzielt werden. Bei der domänenübergreifenden Anwendung konnte eine Akkuratesse von 10.95m in der Position und 49.69° in der Orientierung erreicht werden. Hierbei konnte PoseNet **nur in einem Teilbereich und in einer Richtung trainiert** werden. **Angesichts der Parallelen** zu den Acharya et al. Datensätzen lag **die Schlussfolgerung nahe**, dass der Ansatz auf **die erwähnten Eigenschaften begrenzt** ist.

Ein **Lokalisierungsverfahren** mit einer Positionsakkuratesse von **ca. 11m ist undenkbar**. Sogar die **potenziell mögliche Akkuratesse von 1m** wäre **im** **direkten Gebrauch** für z.B. Anwendungen mit Augemented Reality nicht brauchbar, allerdings **durch Kaskadeneffekt** korrigierbar, s.d. man die Genauigkeit verbessern kann.

**Jetzt nochmal im Allgemeinen.**

Die Unzureichende Akkuratesse bei domänenübergreifender Anwendung von künstlichen neuronalen Netzwerken zur Pose Estimation **liegt** den **Simulationsdefiziten**, wie das Fehlen von Objekten, oder **domänenspezifsiche Artefakten,** wie z.B. durch Licht oder Reflektieren von Gläsern etc. **zugrunde**. Daher ist es l**ohnenswert** in diesem **Zusammenhang** zu untersuchen, ob **die Diskrepanz-Minimierung** zwischen der

synth. Daten und der realen Daten z.B. durch **generative adverserial networks** zu besseren Ergebnissen führt. Ferner könnte man die möglichen Posen im Trainingsprozesses so einschränken, dass diese immer im **begehbaren Flächen** **oder tatsächlich möglichen Bereichen** liegen. Die **dafür nötigen Informationen kann** man ja durch die Simulation bzw. BIM der Gebäuden gewinnen.