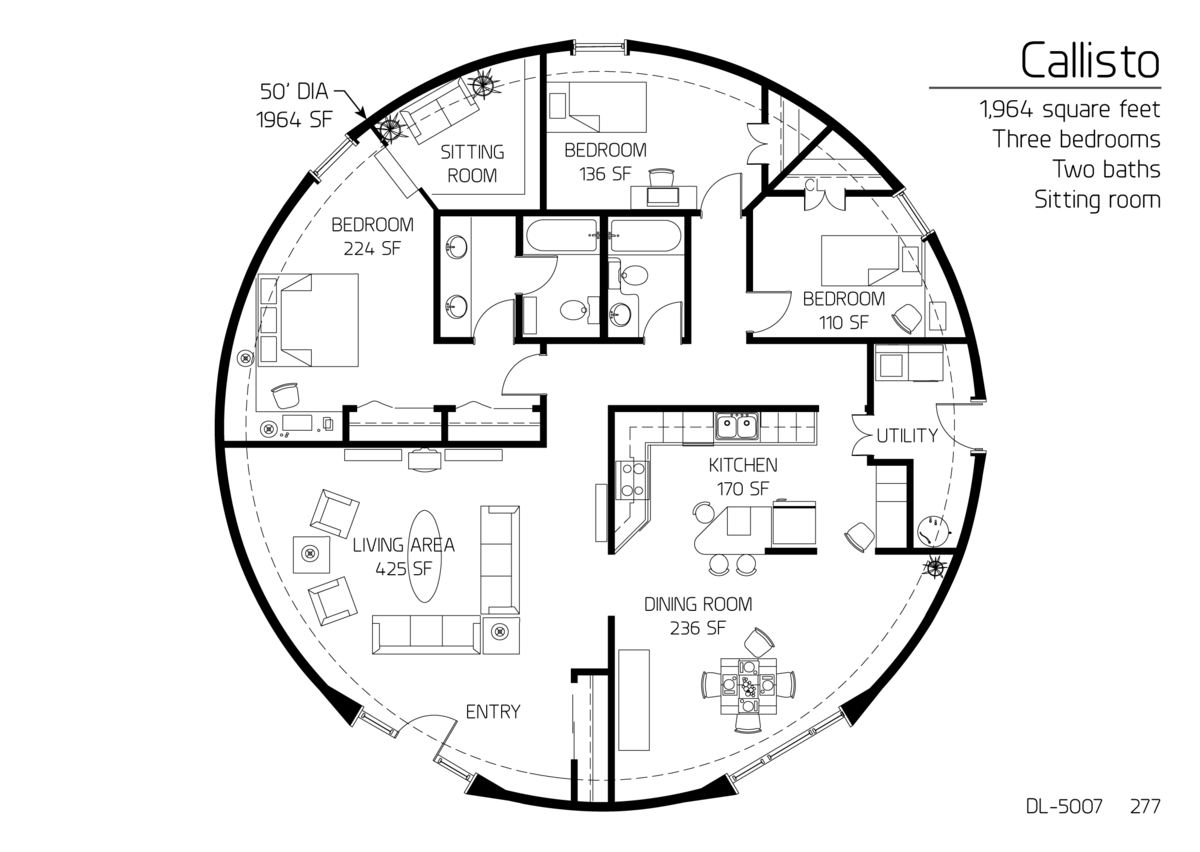
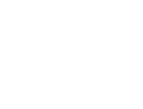




**Deep Learning zur**

**Objekterkennung in Grundriss-Bildern**





Abschlussbericht der Kooperationsphase 2019/20

Durchgeführt am Deutschen Forschungszentrum für Künstliche Intelligenz

Betreuer: Dr. Stefan Agne, Christoph Balada

Betreuer am Hector-Seminar: Dr. Joachim Götz, Ingmar Oehme

Annika Nassal, PF14

nassalan@hector-seminar.de

Inhaltsverzeichnis

[1 Einleitung 1](#_Toc42697633)

[2 Durchführung 2](#_Toc42697634)

[2.1 Ansatz 2](#_Toc42697635)

[2.2 Datenvorverarbeitung und -vervielfältigung 2](#_Toc42697636)

[2.3 Grundaufbau des Netzes 2](#_Toc42697637)

[2.4 Optimierung 2](#_Toc42697638)

[3 Ergebnisse 3](#_Toc42697639)

[3.1 Variation in der Anzahl der Layers 3](#_Toc42697640)

[4 Fehleranalyse und Deutung der Ergebnisse 4](#_Toc42697641)

[5 Zusammenfassung und Ausblick 5](#_Toc42697642)

[5.1 Erweiterte Fragestellungen 5](#_Toc42697643)

[5.2 Anwendung und Nutzen des Projekts 5](#_Toc42697644)

[6 Danksagung 6](#_Toc42697645)

[7 Quellenverzeichnis 7](#_Toc42697646)

[8 Selbständigkeitserklärung 8](#_Toc42697647)

[9 Anhang 9](#_Toc42697648)

[10 weitere Hinweise 10](#_Toc42697649)

Abbildungsverzeichnis

[Abb. 1: Beispiel einer Abbildung 1](#_Toc21887860)

Tabellenverzeichnis

[Tab. 1: Beispiel einer Tabelle 3](#_Toc21882774)

# Einleitung

In unserem heutigen Alltag sind Computer allgegenwärtig und sie werden immer intelligenter. Sie lernen, Sprache zu verarbeiten und Antworten zu geben und können uns bei der Navigation ohne große Schwierigkeiten den schnellsten Weg zu unserem Ziel unter Einbezug der Verkehrsdaten zeigen. So übernehmen sie immer öfter Aufgaben, die sonst Menschen übernehmen müssten, welche dafür deutlich länger brauchen. Doch es gibt eine Aufgabe, in deren Bewältigung Computer noch weit hinter uns Menschen liegen – die Interpretation visueller Daten. Während Menschen ohne Probleme Informationen aus einem Bild extrahieren können und wissen, was dargestellt ist, sind Bilder für den Computer zunächst einfach nur Daten in Form von Zahlen. Dies wird zu einem Problem, wenn Computer Aufgaben für uns durchführen sollen, die auf Bilddokumenten beruhen. Ein bekanntes Beispiel hierfür ist die Klassifikation von Bildern anhand des dargestellten Tiers – Katze oder Hund. Dieses Problem wurde mit einem relativ neuen, doch immer häufiger erfolgreich eingesetzten informatischen Konzept gelöst – dem neuronalen Netz. Neuronale Netze gehören zum Bereich der künstlichen Intelligenz und simulieren ein virtuelles Netz aus Neuronen ähnlich dem menschlichen Gehirn. Im hier vorgestellten Projekt wird mithilfe eines solchen Neuronalen Netzes konkret das Problem der Analyse von Grundrissplänen mit eingezeichneter Möblierung bearbeitet. Dabei sollen die Pläne eingelesen und danach mit entsprechenden Objektbezeichnungen an allen Möbelstücken wieder ausgegeben werden. Schwierigkeiten, die hierbei auftreten können, sind beispielsweise eine sehr variierende Symbolwahl in Plänen unterschiedlichen Ursprungs und die Feststellung der Position der Objekte auf dem Plan. Für die Lösung der Aufgabe soll ein passendes neuronales Netz trainiert werden, bis es die höchstmögliche Effizienz erreicht hat. Für das Training wird ein Datensatz aus einem Wissenschaftswettbewerb zum Einsatz kommen.

**Fragestellung:**

Wie muss ein neuronales Netz trainiert und in Vor- und Nachverarbeitung der Daten eingebettet werden, um die höchstmögliche Genauigkeit bei der Objekterkennung in Grundrissplänen zu erreichen?

# Durchführung

## Ansatz

Um das Ziel zu erreichen, auf einem Grundrissbild alle Möbelstücke und Einrichtungsobjekte zu erkennen und deren Position zu bestimmen, wurde ein mehrteiliger Ansatz gewählt. Dieser beinhaltet für die Objektklassifizierung ein Faltungsnetz, das zwischen den verschiedenen Möbelkategorien unterscheiden kann. Da ein solches Netz eine feste Eingabegröße hat und zur Positionsbestimmung nicht geeignet ist, muss vorher und nachher ein Datenverarbeitungsschritt eingefügt werden. Bevor die Daten zur Klassifizierung dem Netz übergeben werden, ist es daher notwendig, den Plan in Ausschnitte fester Größe zu teilen. Um dabei zu gewährleisten, dass kein Objekt nur am Rand eines Ausschnitts zu finden ist, wird das Bild in überlappende Quadrate geteilt. Dabei weicht jedes Bild vom vorherigen nur genau eine Pixelreihe oder eine Pixelspalte ab. Aufgrund einer Größe der Teilstücke von 100\*100 Pixeln ist jeder Pixel außer den ersten 99 Randpixeln an allen Rändern 10.000 Mal in der Netzeingabe zu finden. Dabei befindet er sich auf jedem Teilstück an einer anderen Position. Das Netz verarbeitet dann nacheinander alle Teilstücke und gibt für jedes von diesen einen Vektor mit einer Wahrscheinlichkeit für die verschiedenen Möbelklassen aus. Diese Vektoren werden zugeordnet zu den entsprechenden Pixeln abgespeichert und am Ende werden mithilfe von Schwellwerten aus den Wahrscheinlichkeiten die tatsächlichen Positionen und Objekttypen extrahiert.

## Datenvorverarbeitung und -vervielfältigung

Der Grundrissplan-Datensatz des Wettbewerbes „ICDAR2019-ORF“ der „International Conference on Document Analysis and Recognition 2019“ (engl.: Internationale Konferenz zur Dokumentenanalyse und -erkennung 2019; zweijährlich stattfindende wissenschaftliche Konferenz) beinhaltet knapp 250 Pläne, auf denen mehr als 7000 Objekte aus 12 Kategorien zu finden sind. Diese Kategorien sind Toilette, Dusche, Badewanne, Waschbecken, Bidet, Tisch, Stuhl, Sofa, Sessel, Nachttisch, Bett und Herd. Die Bilder stehen als PNG-Dateien zur Verfügung und die Metadaten sind in einer JSON-Datei festgehalten.

### Aufbereiten des Datensatzes

Schon früh während der Arbeit mit dem Datensatz konnten Fehler und Probleme mit den Daten festgestellt werden. Zunächst fiel auf, dass die Metadaten Informationen zu Grundrissbildern enthielten, die im Download nicht mitgeliefert wurden. Um diesen Fehler zu beheben, wurde ein Programm geschrieben, das die Informationen, die keinem Bild zugeordnet werden konnten, aus den Metadaten entfernt. Das selbe Programm entfernte auch Fehlinformationen, die in den Metadaten enthalten waren. So waren beispielsweise Angaben zu Positionen von Symbolen außerhalb der Größe des Grundrissplans, was nicht möglich ist. Des Weiteren hatten die Symbole auf den Grundrissplänen sehr variierende Größen, also unterschiedliche Seitenlängen in Pixeln, was ein Problem darstellte, da die Eingabegröße des Neuronalen Netzes, das zur Klassifizierung verwendet werden sollte, fest ist. Die Größe des größten Symbols war dafür nicht geeignet, da auf anderen Grundrissbildern die Symbole sehr klein waren. Um einen Überblick zu erhalten, wie viele Symbole mit welcher Größe vorhanden sind, wurden die Metadaten mit einem Programm ausgewertet, das diese Information als Histogramm darstellt. Darauf ist zu erkennen, dass viele Symbole kleiner als 100\*100 Pixel sind. Also wurde das Projekt limitiert auf die Erkennung von Symbolen mit einer Größe von 100\*100 Pixeln und kleiner. Diejenigen Symbole, die größer waren, wurden aus den Metadaten gelöscht, sodass das Netz beim Training im Abgleich mit den korrekten Lösungen keine verfälschten Werte ausgab.

### Zuschneiden und Vervielfältigen der Trainingssymbole

## Aufbau und Training des Netzes

## Optimierung mithilfe von Metriken

## Vor- und Nachverarbeitung von Anwendungsdaten

# Ergebnisse

## Hyperparameter

## Metriken

# Fehleranalyse und Deutung der Ergebnisse

# Zusammenfassung und Ausblick

## Erweiterte Fragestellungen

## Anwendung und Nutzen des Projekts

# Danksagung

# Quellenverzeichnis

# Selbständigkeitserklärung

Ich versichere, dass ich diese schriftliche Prüfungsarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass ich alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, durch Angabe der Quellen als Entlehnung kenntlich gemacht habe.

Königsbach-Stein, den [Datum]

Unterschrift

# 9 Anhang