October University for Modern

Sciences and Art

Faculty of Computer Science

Graduation Project

Title

Supervisor: ……..

Name: ……..

ID: …….

Abstract

Table of Contents

Table of Figures

Table of Tables

Chapter 1: Introduction

1.1 Introduction

- state the general topic and give some background (4 : 5 Lines)
- define the terms and scope of the topic (4 : 5 Lines)
- outline the current situation (4 : 5 Lines) + Figure
- evaluate the current situation (advantages/ disadvantages) and identify the gap (4:5 Lines)

Eg:

It is supposed to understand a wide range of living room and bedroom image scenes with several home styles by using a single image of an indoor scene. In this thesis, a fully automatic system is presented to deal, detect and segment image objects with high quality using different methods and techniques of deep learning nets. The project will present methods to distinguish between the main different home furniture (chairs, beds, tables, nightstands, and sofas). Depending on subsequent researches in computer vision field that have many limitations to overcome. Lately, understanding the indoor scenes and their included furniture has a significant usage in interior home design and remodeling websites for different users [1].

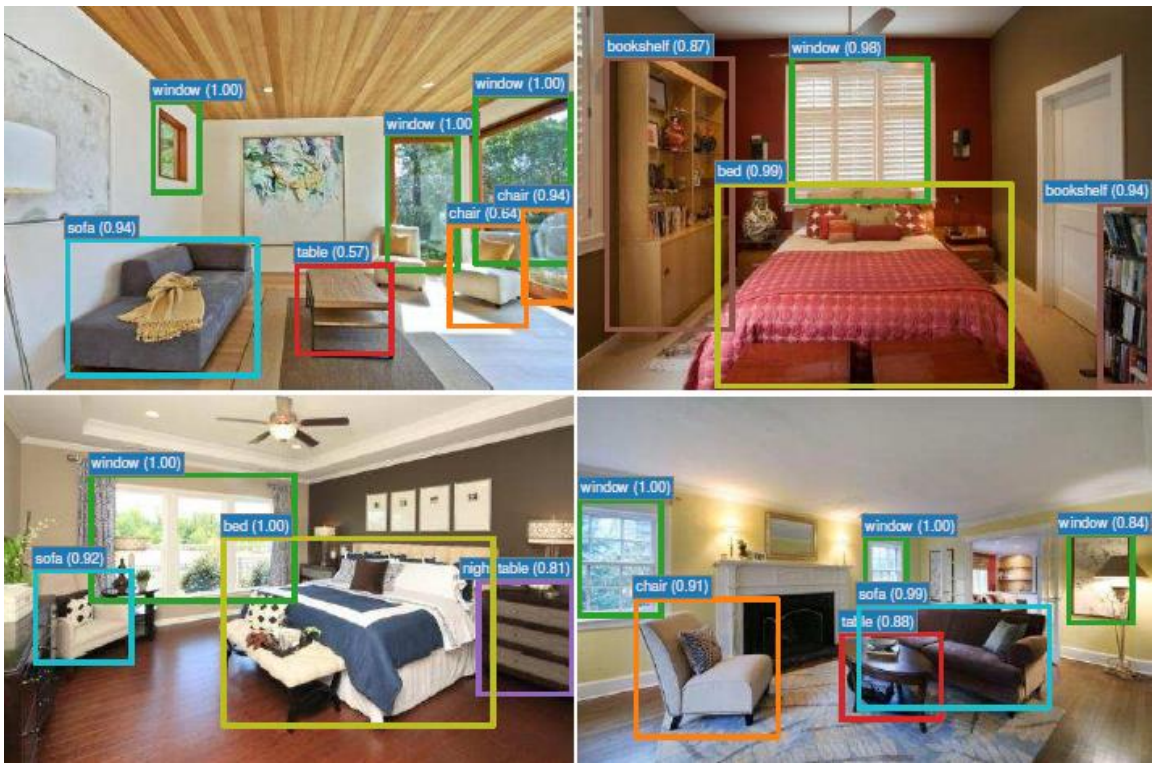


Figure 1: Object detection result on sample images

1.2 Problem statement

- State what is the problem to be solved in this area…… (3 : 6 Lines)

  Eg:

  One of the computer vision key problems is *scene understanding* and *segmenting objects* because it is a necessary precondition to many high-level tasks of autonomous intelligent machine operating in real-world environments. Recently, there are many promising attempts in this area of supervised learning which need enormous quantity of supervised data to improve the performance of the machine learning processes [3].

1.3 Objective

- state the research aims and/or research objectives (when we reach something ……) [4 : 6 lines]

  Eg:

  This project should present a fully automatic system that understands and segment objects in the image with accurate annotation of an indoor scene from a single photograph. It is supposed to provide approaches for room understanding which is the key of the success of our system.

1.4 Motivation

- Motivation for others
  - (identify the importance of the proposed research)
  - (identify the need of the proposed research) [3 lines]
- Motivation for the researcher (you) [3 lines]

  Eg:

  Semantic understanding of indoor scenes from single image that contains a variety of furniture objects is a particular interest for several attractive applications that need dealing and interaction with these objects. Individual object detectors and scene classifiers are powerful; however, they are inaccurate in case of complicated real-world scenes variety and different viewpoints [4].

1.5 Thesis layout

- Chapter 2 will be ……..
- Chapter 3 will be …… [3 lines]

Eg:

In this thesis, the first chapter will provide an introduction about the project and its aim. Then, the second chapter will provide a literature review and a background of the previous work in the same area of research.

Chapter 2: Background and Literature Review

2.1 Background

– Mention different approaches & algorithms & give some examples (don't mention other researches , don't mention results , just explain) [about 4: 5 pages + figures]

Eg:

A number of industries are adapting technology such as furniture industry and most people are becoming more technology orientated. Now furniture designers are not considering only how comfortable a chair is, but also how easy it is to use technology. That is why using image processing, deep learning and computer vision techniques are used in order to improve the usability of technology in furniture interior design field.

Figure 2: Annotation interface of objects in the image

2.1.1 Algorithm X

……..

2.1.2 Algorithm Y

……..

2.1.3 Algorithm Z

……..

Eg:

2.1.1 Machine Learning

Machine learning “ML” is a subset of Artificial Intelligence “AI”, but the difference aspect is the ability of machine learning to adapt itself over time when feeding the machine with more data about its surrounding environment and interactions.

2.1.2 Neural Network and Deep Learning

Neural network is a set of algorithms that are designed to mimic the human brain to be able to recognize patterns and interpret data in terms of machine perception, classifying, or clustering this raw input data.

2.2 Previous Work

- We will Analyzing & evaluating literature review sources…..

2.2.1 Research 1

2.2.1.1 Strategy & Structure

- Strategy → [What are the key concepts and how are they defined ?] (3 : 6 Lines for each research)
- Structure → [What are the key theories, models and methods? Does the research use established frameworks or take an innovative approach?] (3 : 6 Lines for each research + figure(s))

Eg:

The major limitation for some models is that these models has the ability of semantic segmentation but unable to identify the instance segmentation of the same class. So, in this research, the researchers have an attempt to build a model can perform semantic and instance segmentation using a Coverage Loss as a high-order loss function. By performing the hierarchical segmentation of the image, the researchers use heuristics to restrict search space.

2.2.1.2 Data

- [what is the nature of the used data ?]
- [how is its size?]
- [how it is collected?]

(3 : 6 Lines for each research)

Eg:

The model is applied on NYU Depth V2 dataset to evaluate its performance in instance segmentation because the other datasets which are used in the segmentation task customarily do not provide instance labels or provide instance labels for few objects [16].

2.2.1.3 Method Evaluation

– [What are the strengths and weaknesses of the research?]

Eg:

Using the CNN architecture to extract additional feature moreover the segmentation tree, improves the performance of the model at the instance segmentation task as a qualitative result. However, the model struggles in case of larger numbers of objects in the same scene. So, the perfect coverage score has not accomplished yet [16].

2.2.1.4 Results Evaluation

– [What are the results and conclusions of the study ?] (3 Lines for each research)

Eg:

The model with extracted CNN features is trained on segmentation tree biased by the ground truth as a label and achieved 62.8 accuracy results.

Chapter 3: Material and Methods

3.1 Materials

- In this section we will explain the used materials int this project which varied between data & tools & environments.

3.1.1 Data

- Source: collected or downloaded
- Format
- Size
- Preprocessing
- Subsampling

Eg:

This dataset is created to be evaluated in the academic researches of deep-learning algorithms for object recognition. The images consist of 65 different categories of office and home settings. These images are collected through a python web-crawler and then are verified and filtered to ensure that the desired objects exist with a certain number of images [17].

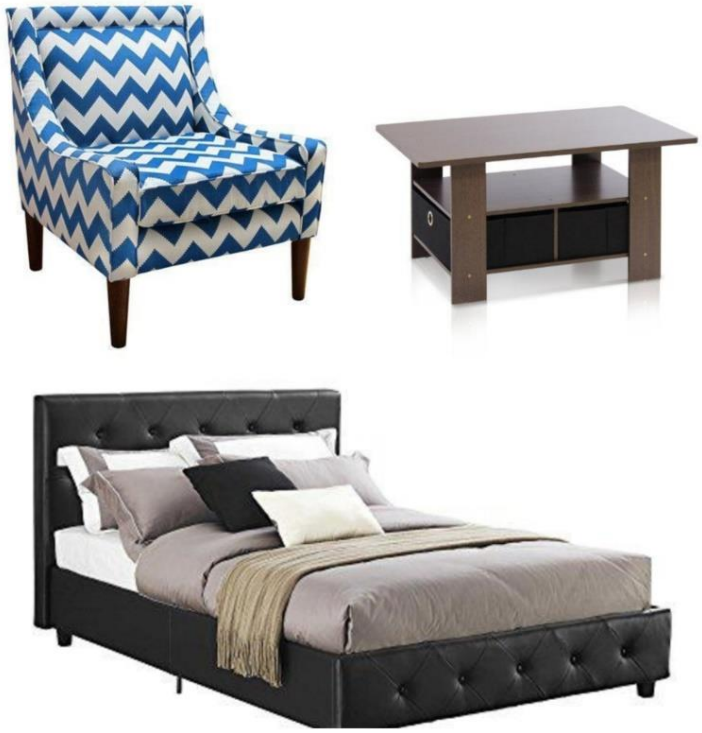


Figure 3: Office-Home Dataset samples

3.1.2 Tools

- Software
- Library

Eg:

  - Anaconda: is a platform for efficient developing and applying AI and machine learning models.
  - Python 3: An open-source programming language that enables developers to work and integrate their systems quickly and effectively.

3.1.3 Environment

- GPU OR CPU + why
- cloud or local + why
- single mod or multi + why

Eg:

  - Local CPU, intel i7 processor with 6 cores.
  - Google Colaboratory GPU

3.2 Methods

- In this section we will mention & explain our solution methodology + used approach(algorithms)

3.2.1 System architecture Overview

Eg:

First of all, the office-home dataset is used after filtration as an input to the model. As a preprocessing step, the data is loaded to our model and a fixed size is used by resizing or padding the images.

.

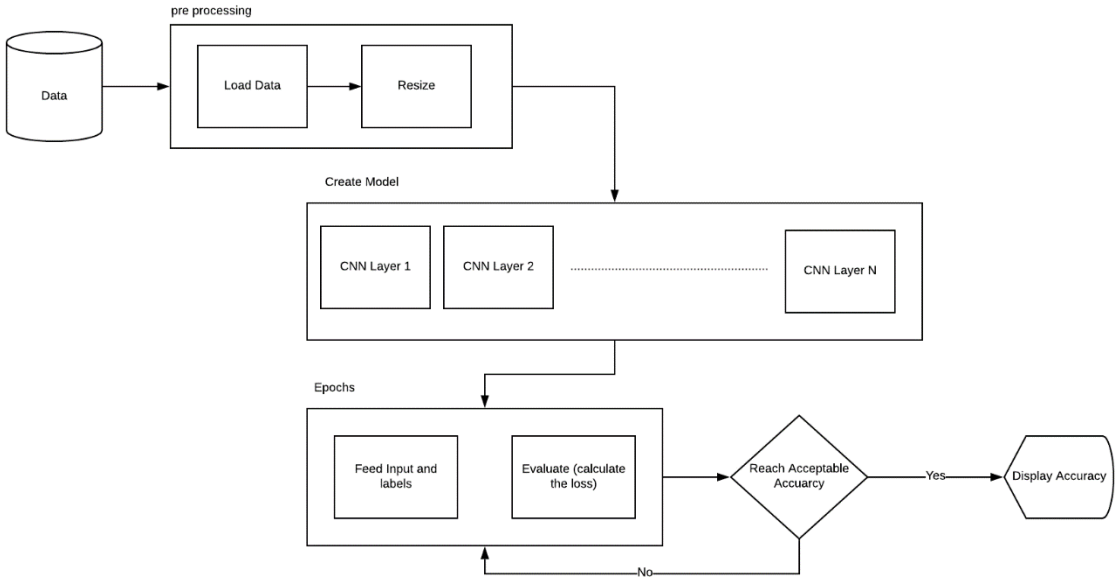


Figure 4: CNN system architecture overview

Chapter 4: System Implementation

4.1 System Development **(2 pages + figures)**

- The system was developed on different phases, at first we implement a small model to check that the result of partX will give a good results or not.
- After that we make a helper tool to prepare the data (or convert its format – or extract subset of it …..) .
- We then deveope partY to check the algorithmZ.
- Finally, we integrate partX & partW which is the enhanced version of partY , all of those made the whole final proposed system.

- during the system development we tried several software (or platforms)

[unity – opencv – matlab - tensorflow – keras …]

[spyder – jupyter – colab]

(don't mention here the problems you faced in each plateform, you will mention it in chapter 6 ["discussion & conclusion"]

Eg:

The first model was developed to check the ability of the machine to classify the whole image consists of only chair, bed, or table by using Office-Home dataset after filtration. The dataset images were not in a fixed size. So, the first solution is resizing the input images by using OpenCV library after reading images and their labels.

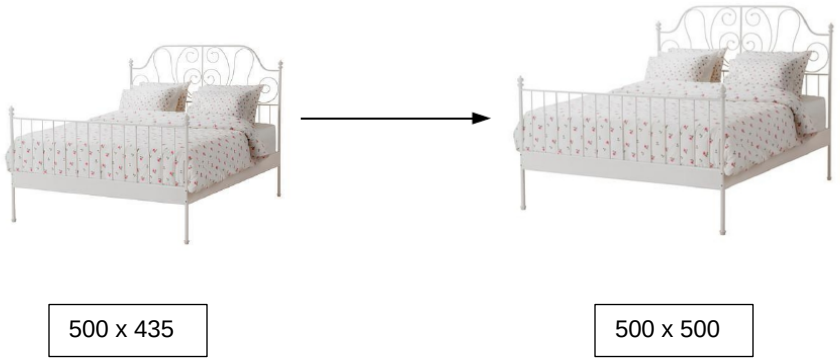


Figure 5: Input image before and after resizing

By using Keras APIs a simple classification model of CNN architecture was developed to ensure that the model is able to recognize the difference between the three categories and give the right label for each class which is 0 for chairs, 1 for beds, 2 for tables.

4.2 System Structure (5~ 6 lines)

- Here we will explain on the coming subsections the overview of the final system we developed by explaining the flow between its components , and in the second subsection we will illustrate the class diagram for the used classes in this system.

4.2.1 System Overview (2~ 4 pages)

- The system composed of N main stages, at the first we ……

  Eg:

  This system is composed of 3 essential stages. The first stage is responsible for the task of data preparing and preprocessing. The first step in this stage is dataset loading and reading. The second step is data resizing to get a fixed size and data augmentation to increase the number of it and avoid overfitting.

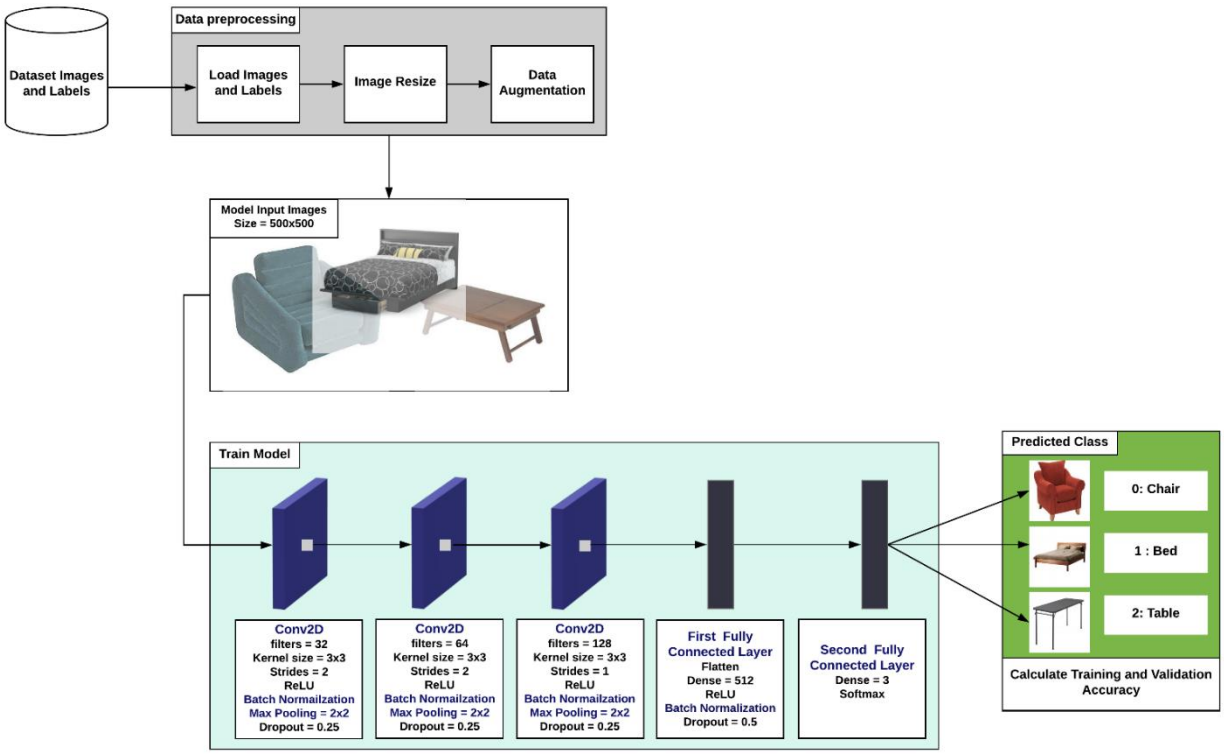


Figure 6: Classification CNN architecture overview

4.2.2 Class Diagram or TensorBoard (5 ~ 7 pages)

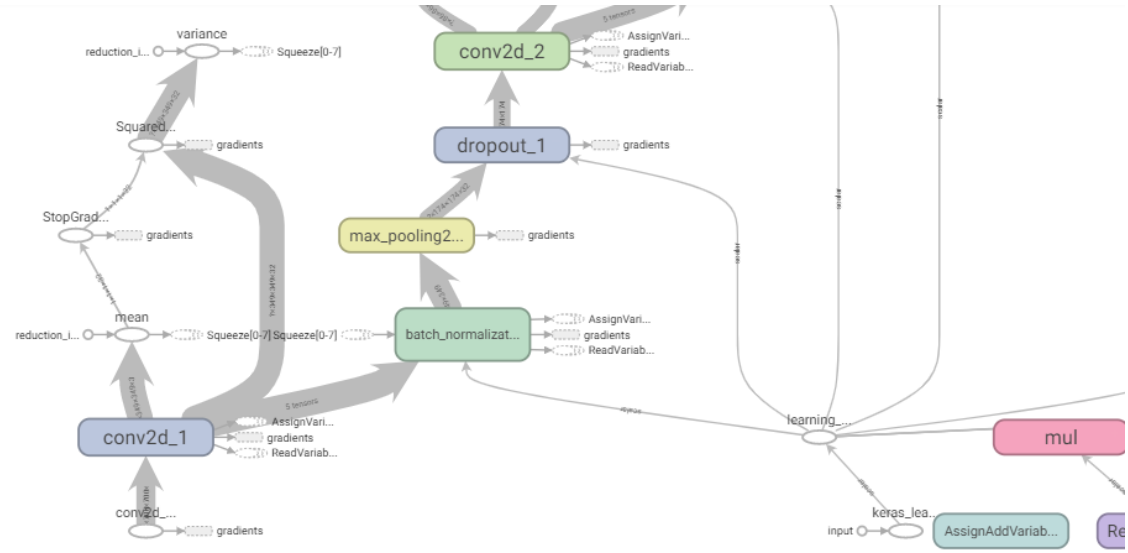
- In this section we will explain the class diagram for the developed classes in this system. The system finally composed of N classes as the following:

For each class:

- Mention its Name, its responsibility, its relation with other classes.
- In the level of (attributes and methods) mention its Name , its responsibility , its relation with other attributes
- In the level of (attributes and methods), don't explain all of them just select the important of them.

Eg:

The following figure is the TensorBoard about the whole main components of the CNN architecture for the classification task. It describes its convolutional layers, max-pooling layers, batch normalization layers, the drop out, the flatten layers with dense, and the activation functions.



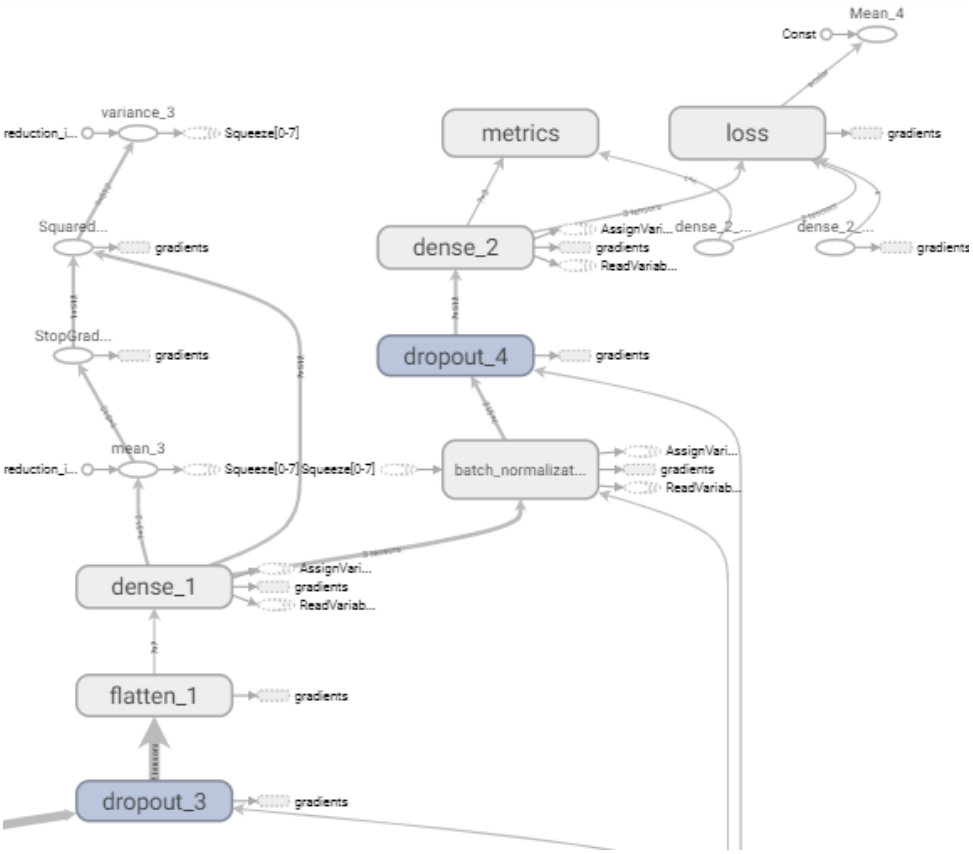


Figure 7: TensorBoard Convolutional layers and Flattens

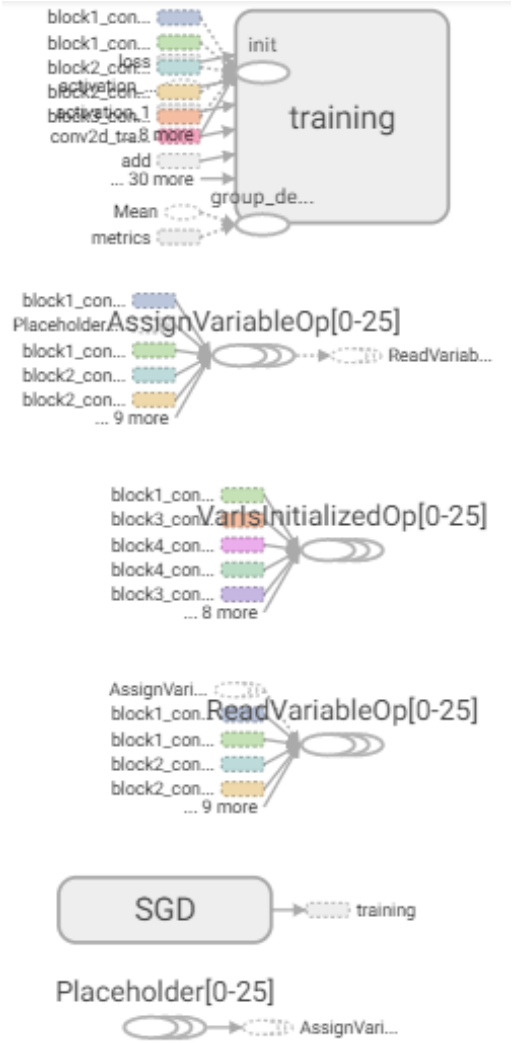


Figure 8: FCN TensorBoard Auxiliary Nodes

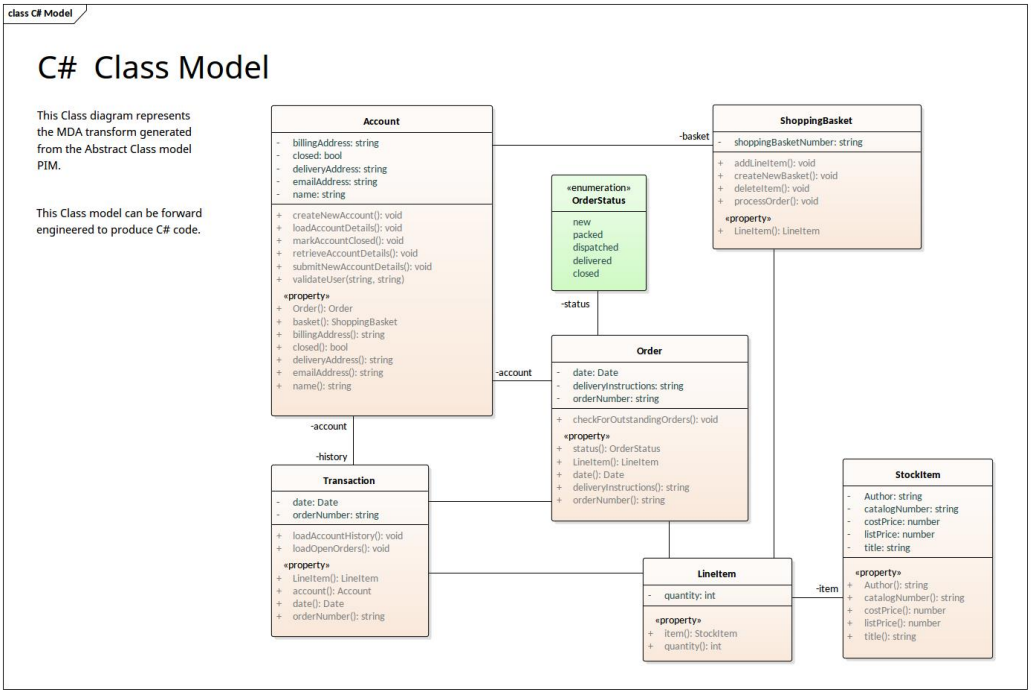


Figure 9: Class Diagram Example

4.3 System Running (5 ~ 7 pages)

– In this section we will demonstrate the input-output for each component in our final system

4.3.1 Component A

– You have to explain what is the input (type , size , …) and what is the output .

– [don't evaluate (don't say why some results good & some are not good) the output of your components in this section, you will in chapter 5 "Results & Evaluation"].

Eg:

The input of this function is the original image dataset, and the output is 15 different images generated by data augmentation function to increase the small number of the dataset images and avoid model overfitting.

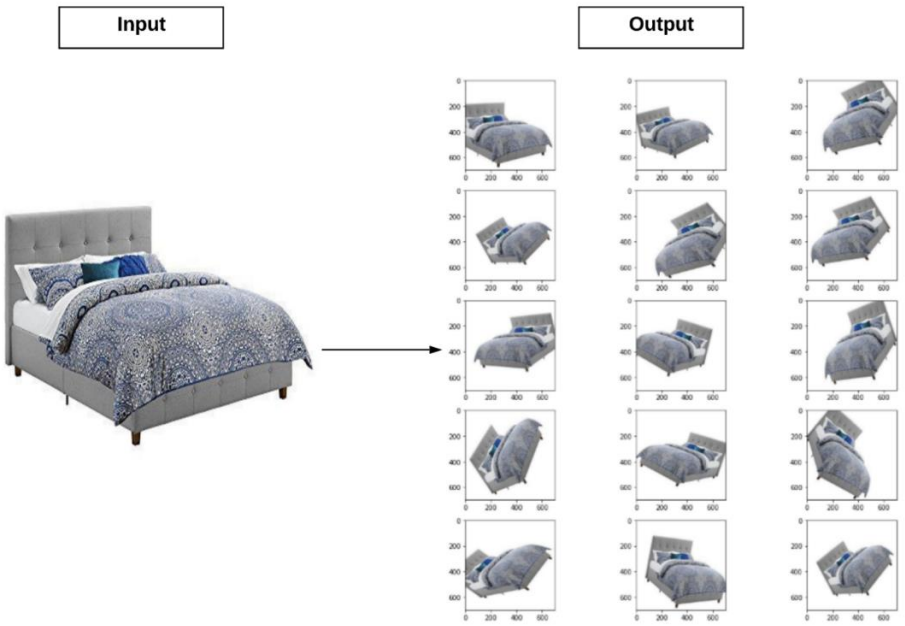


Figure 10: Input and output of CNN data augmentation

Chapter 5: Results and Evaluation

5.1 Testing Methodology

Explain your testing methodology

- Testing the whole system, or each component
- Loss
- Validation or training-testing separation
- Selection of Batch size

Eg:

There are different methodologies to assess how well the model performs in prediction of specific classes from specific dataset. And the used methodology in the classification CNN model for chairs, beds, and tables will be demonstrated in detail through this subsection. First of all, the dataset was split into two subsets of training and validation and the feeding of the dataset to the model is done according to the assigned value of the batch size. Each of these methods are evaluated and tested separately according to the accuracy and the loss to decide the preferable preprocessing method to be used for the existing classification problem.

5.2 Results

Case study for 3 cases

- 1 case for perfect result
- 1 case for acceptable result
- 1 worst case.

Comment on each

Limitations

- Data filtration or Augmentation or preprocessing
- Comment on the results with & without filtration.
- Tools or platforms (openCV , matlab , tensor , keras …)

Comment why?

Eg:

5.2.1 Best Results Cases

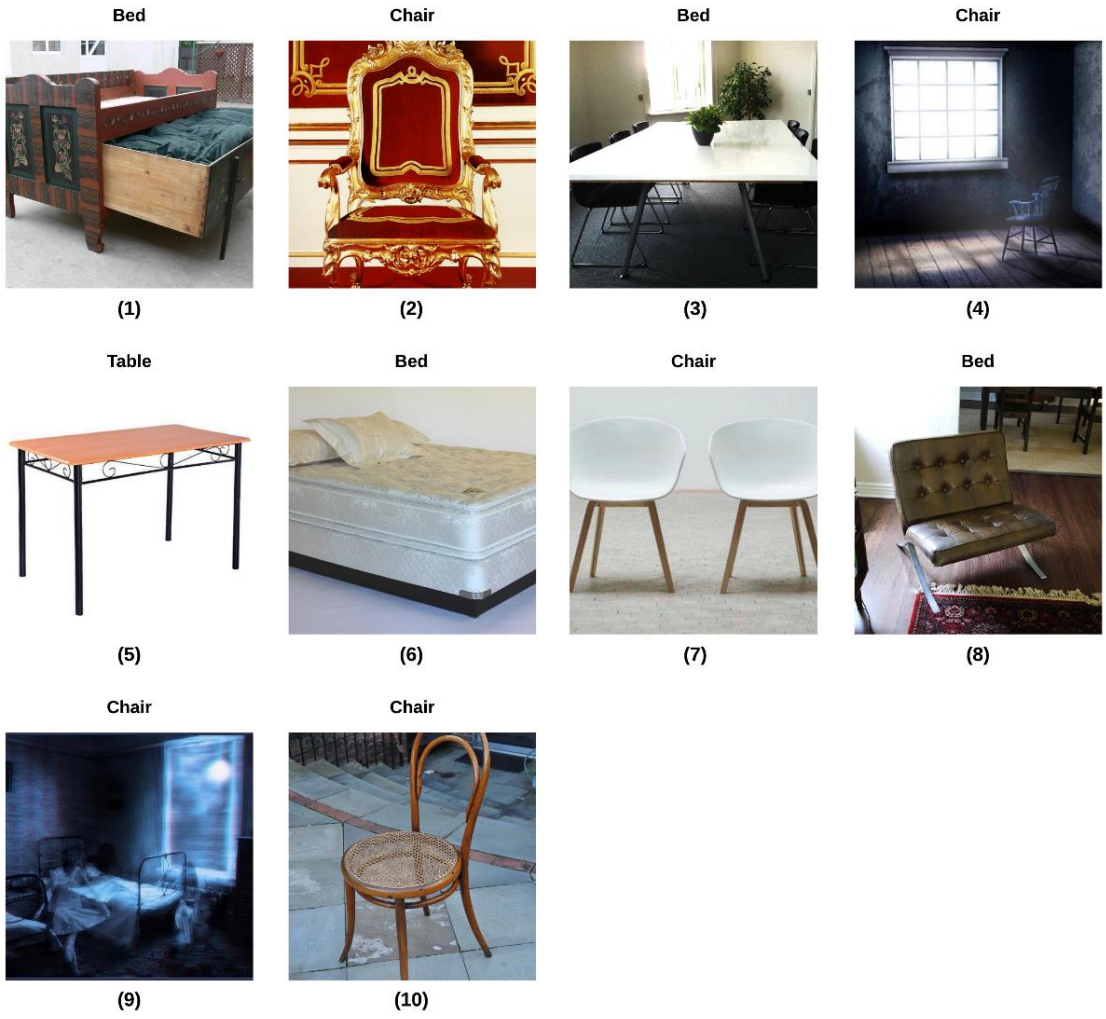


Figure 11: Random Results from the Best CNN Classification Model

The first best result was generated by the following architecture model. The model consists of three convolutional layers with ReLU activation function and 3×3 kernel size. The first two layers with 2×2 strides and without padding but the third layer without strides and with padding. The model was trained for 30 epoch and 15 batch size. The previous figure shows some random results of this model. If these results were observed, it will be recognized that the results are perfect because most of the predicted labels are right even the images that are hard to be recognized such as in

the images 1 and 2 and the wrong labels are misleading because of the light or the orientation of the image such in result 3 and 9.

5.2.2 Limitations

The selected dataset was filtered to become products and real-world images of chairs, beds, and tables. After this selection, the number of images was inadequate to train the CNN model and it was difficult to find another dataset consists of only one class in each image. So, the solution of this problem was the data augmentation which was better for the results to avoid the overfitting. The limitation that could not be solved is the quality of the images because many of them has poor lighting or the angle of the images lose many details and features about the existing object. And if these problems were solved, the model could achieve better results.

5.3 Evaluation

- Accuracy
  Under different conditions (parameters)

- Space
  Degree of scalability for data size
  Degree of scalability for architecture

- Performance
  Time (speed)

Eg:

5.3.1 Accuracy Evaluation

Table 1: First Summary of the CNN Classification Models with their Accuracy

Models Number	No. of Conv layers	Kernel size	Strides	Padding	Activation Function	Batch Size	Training Accuracy	Validation Accuracy
Model 1	3	3×3	1×1	valid	ReLU	15	71%	51%
Model 2	5	3×3	1×1	valid	ReLU	10	69.33%	60.7%
Model 3	3	5×5	1×1	valid	ReLU	10	68.94%	64.91%
Model 4	3	3×3	1×1	valid	ReLU	10	73.18%	68.44%
Model 5	3	3×3	1×1	valid	Sigmoid	15	56.25%	50.88%
Model 6	5	3×3	1×1	valid	Sigmoid	15	66.73%	56.14%

Model 7	3	3×3	1×1	valid	Swish	15	76.56%	55.38%
Model 8	3	5×5	1×1	same	ReLU	10	77.92%	59.62%
Model 9	3	3×3	2×2	same	ReLU	15	80.27%	58.33%
Model 10	3	3×3	2×2	valid	ReLU	15	79.96%	70.23%

As shown in the previous table, different implementation to the classification CNN architecture were tried with different permutations of parameters such as the activation function, kernel size, padding, strides, and batch size until the best training and validation accuracy were accomplished.

The first graph represents the training and validation loss and it could be noticed that the training loss is 0.5276 and the validation loss is 0.9055, and the second graphs represents the training and validation accuracy.

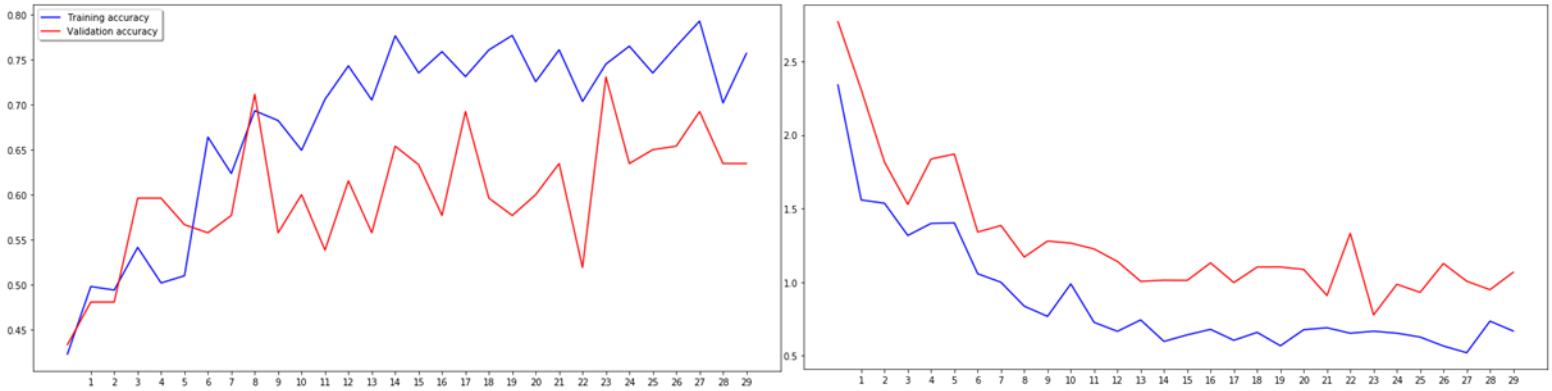


Figure 12: Graphs of the best CNN model loss and Accuracy

In the figure of the activation functions, the training and validation accuracy and loss of three different activation functions are mentioned. The training accuracy of the sigmoid activation function it the worst unlike the ReLU and swish activation functions.

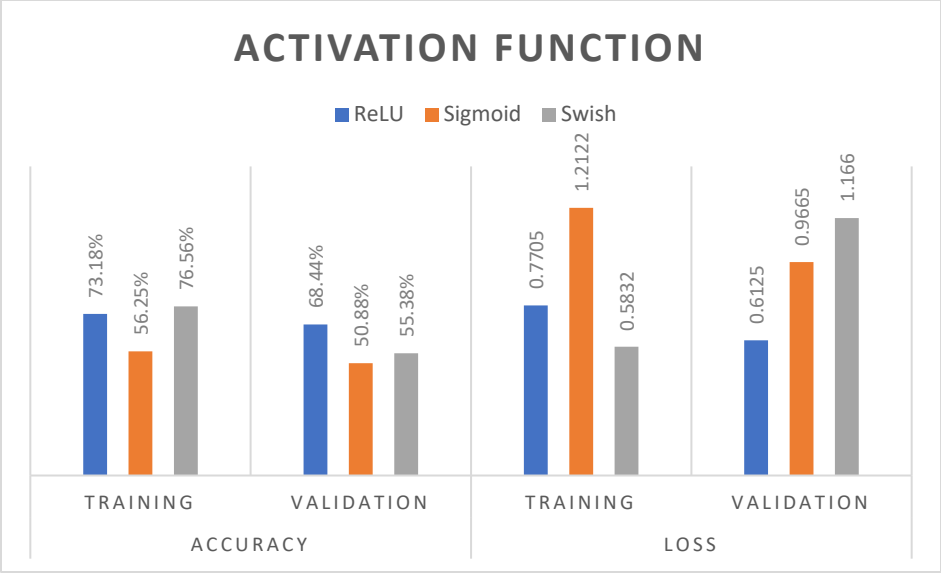


Figure 13: Graph of the CNN results of different Activation Functions

5.3.2 Time Performance

Regarding the time, it was a significant advantage for using Google Colab GPU resources rather than using Spyder environments and Keras APIs on the local PC CPU. If both are compared, the time taken for the same architectures models on the local PC CPU was 76s per epoch and 4s/step, and on Google Colab GPU was 18s per epoch and 1s/step.

Chapter 6: Conclusion and Future Work

6.1 Conclusion

State what you've done and what you've found

- We develop a complete System for …….. ,
- starting with investigation of the different approaches …….,
- then the try several algorithm …….,
- after that we test our system under different conditions , finally we found that the architecture with bla bla is the most promising one that give us a good results ……
- we think that if the data was more homogenous , or more large in size , we think that this approach will give more accurate results

Summarize contributions (achievements and impact)

- we think that system could commercialize to beneficiate the persons like ……..

Eg:

To sum up, in this thesis a complete automatic system was developed to understand and segment a wide range of indoor scene furniture. After searching about several approaches to determine the most auspicious approach that has the ability to solve this target problem, it was the decision of heading for the deep learning approach and its various algorithms to build up a successful system that is considered as an appropriate system for different home styles.

First, it was the start by implementing a classification algorithm to distinguish between images of one object of the main furniture which are chairs, beds, and tables. And the selected algorithm for this task was the Convolutional Neural Network "CNN" algorithm which is a well-known algorithm for the similar classification tasks. This algorithm has proven its capability to classify between these objects.

6.2 Problem Issues

6.2.1 Technical issues:

- System crash with large data , && how you solve or tackle it
- GPU not working && how you solve or tickle it
- Debugging very difficul && we couldn't solve this problem

  …..

Eg:

- The system consumes a huge amount of computational power and time on the local PC CPU and with the large architectures it was crashed.

6.2.2 Scientific issues:

- Change batch size
- RNN algorithm not give, so we go to CNN …..
- Pipeline between subsystems

  …….

Eg:

- To include as much as possible of images for model training and validation, a large fixed size of images padding was selected which was 700×700 but this affected the batch size because more than 5 batch size caused a system out of resources crash.

6.3 Future Work

Outline open issues/directions for future work

Eg:

In the future work, deeper pre-trained models will be used such as ResNet269, and ResNet101 in order to increase the MIoU value and the accuracy. In addition to that more classes will be added to be segmented using the FCN model by utilizing more resources of a cloud service provider such as AWS and Azur.

References

1. Izadinia, H., & Seitz, S.M. (2016). IM 2 CAD.
2. Potdar, K., Pai, C.D., & Akolkar, S. (2018). A Convolutional Neural Network based Live Object Recognition System as Blind Aid. ArXiv, abs/1811.10399.