
# FACULTY OF COGNITIVE SCIENCES AND HUMAN DEVELOPMENT

## FYP Presentation

## Design and Development of a Scene Recognition System using Neural Network models

Presented by: Ong Hui Xin (62322)

Supervised by: Assoc. Prof. Dr Teh Chee Siong

# Outline of presentation

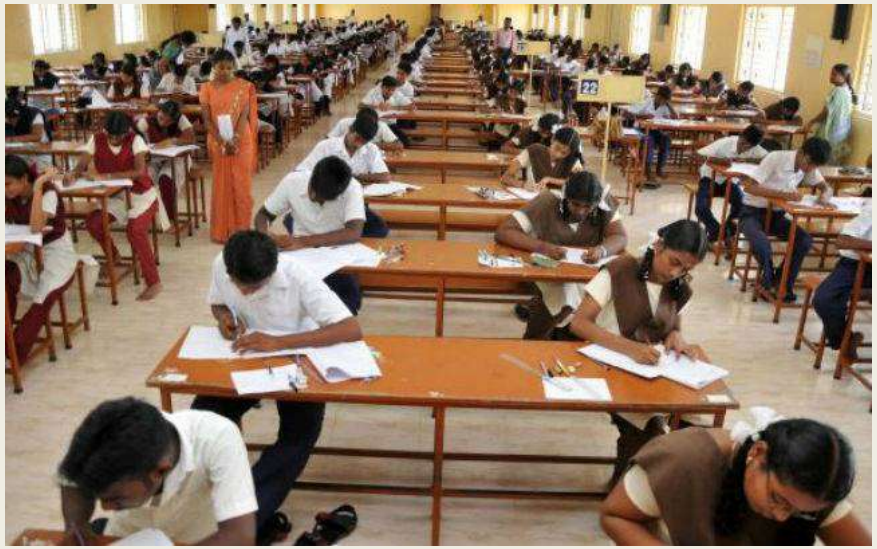
- Introduction
  - Background
  - Problem statement
  - Objectives
  - Scope of study
- Literature review
- Methodology
- Feature extraction
- Scene classification
- Results and Discussion
- Conclusion

# INTRODUCTION

Background, Problem statement, Objectives, Scope of study

# BACKGROUND

- There are many tables and chairs lined up in a hall. People are sitting on the chairs facing down at their tables. There are stationeries on the tables.

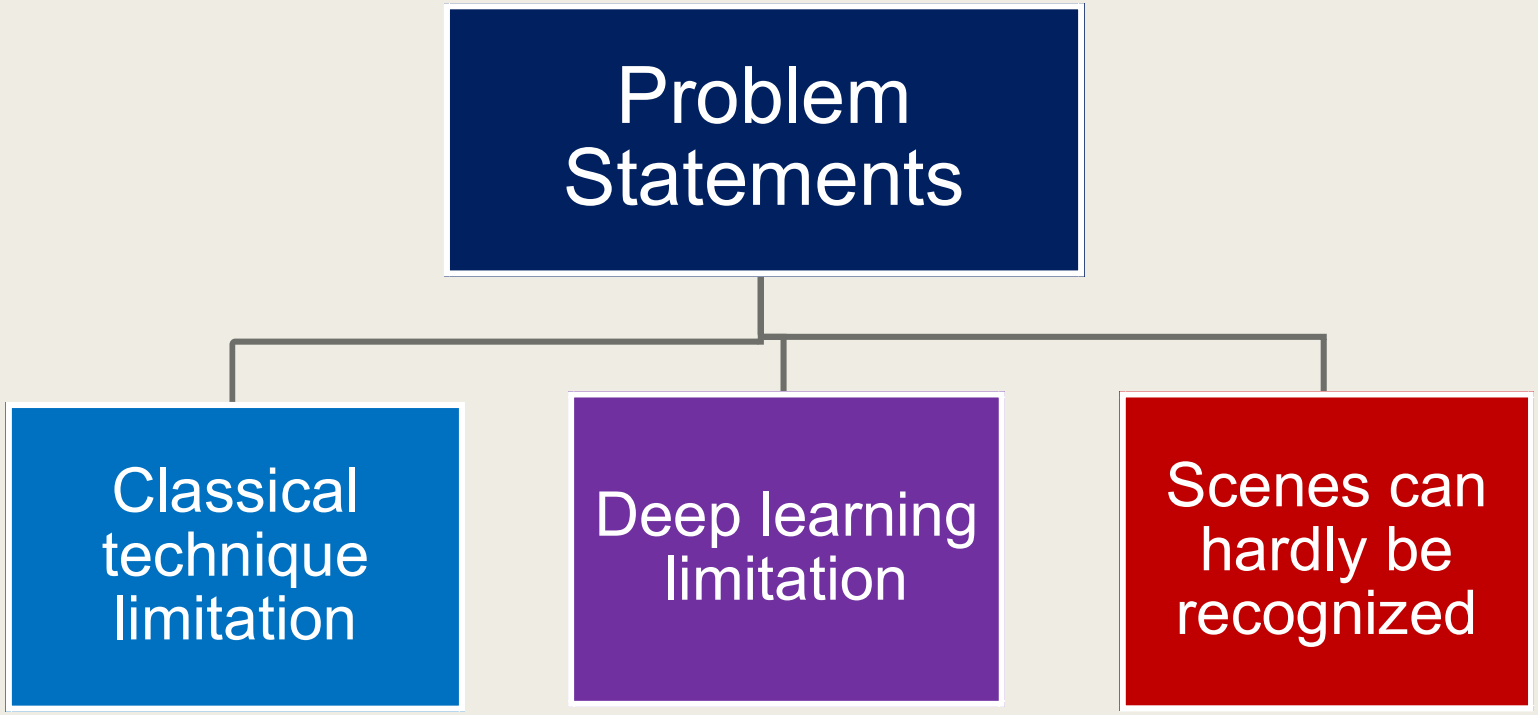


# BACKGROUND

## Scene Recognition

- Top-challenging research field in computer vision – may be due to the ambiguity between classes (López-Cifuentes et al., 2020).
- Belongs to visual classification: means to determine the scene category of an image by analysing its contents (Jiafa et al., 2019).
- Needed in many fields of research: remote sensing, aerial scene classification, computer vision and robotics (Jiafa et al., 2019).

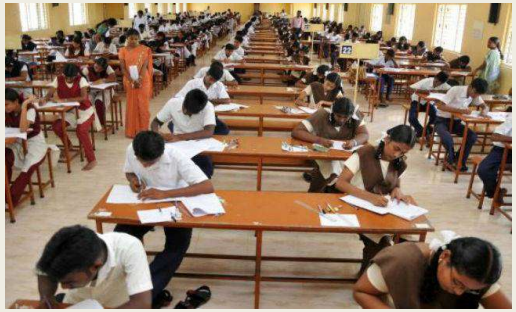
# PROBLEM STATEMENT



# PROBLEM STATEMENT:
## Classical technique limitation

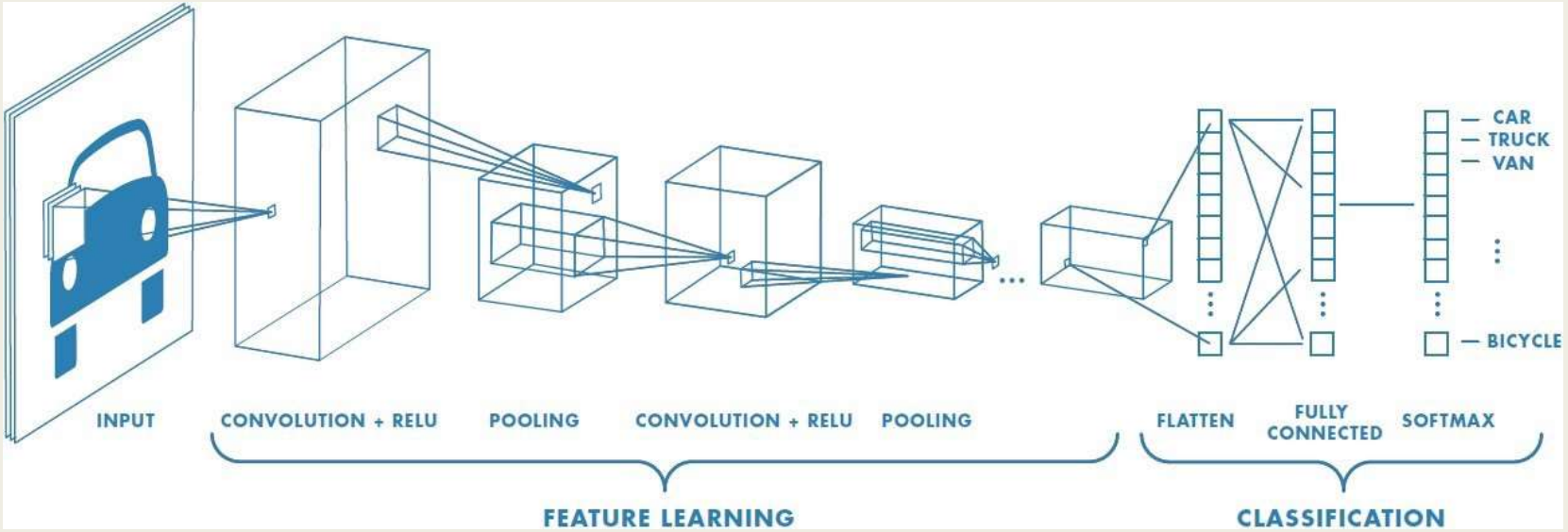
- Pattern recognition in Computer Vision:
  - Classical technique – feature recognition
  - Involves many image pre-processing like image segmentation & enhancement (Balan P & Prof Sunny, 2018).

- Object recognition:
  - two eyes on top of a mouth = a face
- Scene recognition:
  - Needs a lot of features and data to categorize places.
  - Time consuming
  - Computational expensive



# PROBLEM STATEMENT:
## Deep learning limitation

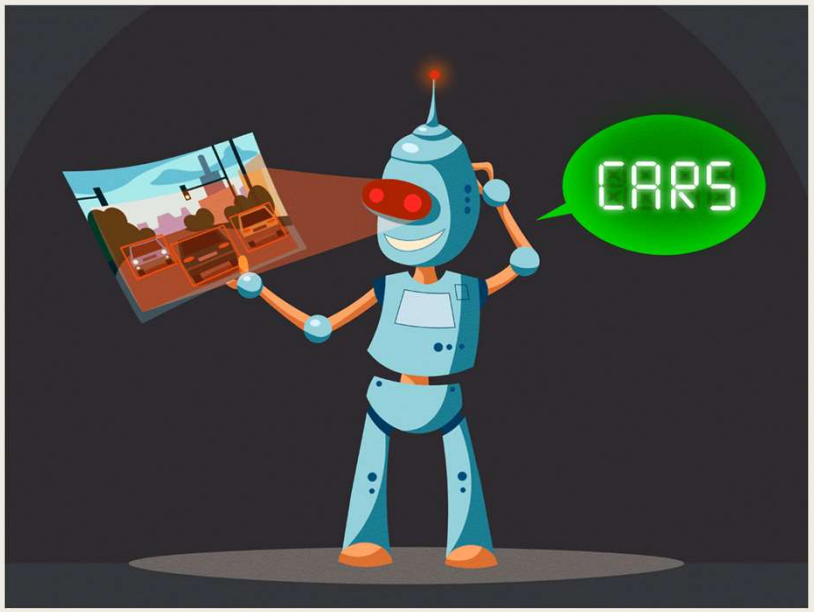
- Deep learning network to recognize image: Convolutional Neural Network
- Input: image data as a whole to train model (object and scene recognition)
- Limitation: needs a lot of image database



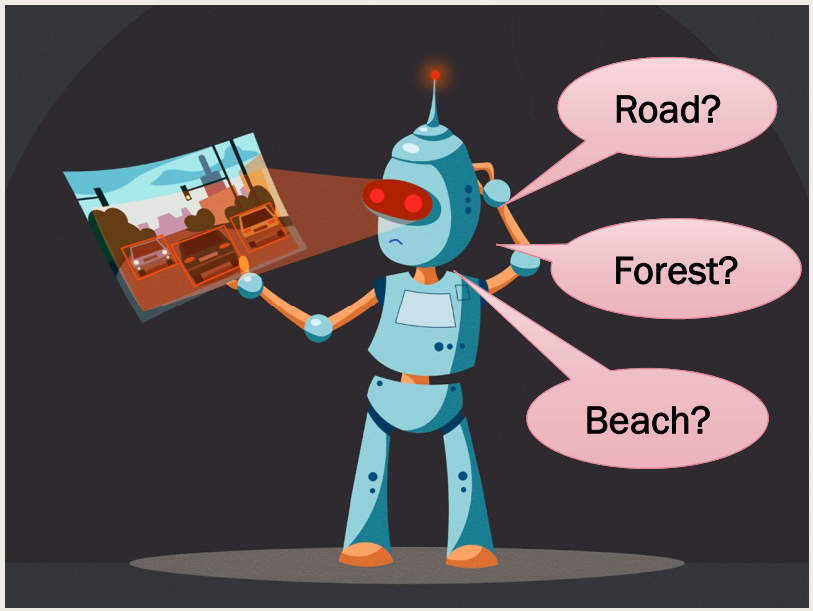
# PROBLEM STATEMENT:

## Scenes can hardly be recognized.

Today, objects can be recognized easily. (VERY SUCESSFULLY)




However when it comes to recognizing scenes…



# Objectives

Generally, the project is to conduct experimental studies by implementing neural network models in and redesign a scene recognition system and evaluating its performance by using a well-known evaluation method.

- To study in depth the currently available deep learning algorithms for scene recognition.
- To implement feature extraction using CNN.
- To design and develop a new scene recognition system.
- To evaluate the performance of the system by using a well-known evaluation method.

# Scope of study

- Database:
  - **Google** image database
  - Reason: Easy to be reached, rich dataset
  - Constraints: Irrelevant images act as noise and need to be filtered out.
  - Method: 2 or 3 scenes that are common are selectively chosen from the database.

- Object detection model:
  - **YOLO-v3**
  - Reason: Recent state-of-art, fast, pre-trained object detection, transferred learning
  - Constraints: Less number of labelled pre-trained object detection, only 80 classes.

- Scene classification model:
  - Multi-layer perceptron (**MLP**) using **scikit-learn** library model
  - Reason: Pre-defined model
  - Constraints: Needs a lot of hyper-parameter tuning

- Computational device:
  - Computational device: Personal laptop (with CPU)
  - Reason: Cheap and available

# LITERATURE REVIEW

# Image recognition

- Artificial Intelligence + Computer Vision
- Pattern recognition = process of description, grouping, and classification of patterns.
- Recognize patterns by training models

# Object recognition

- Given
  - an image containing one or more objects of interest (and background)
  - a set of labels corresponding to a set of models known
  - ✓ Object recognition systems should assign correct labels to regions/ a set of regions, in the image.
- Closely tied to the segmentation problem:
  - No partial recognition of objects -> no segmentation -> object recognition is impossible

# Scene recognition

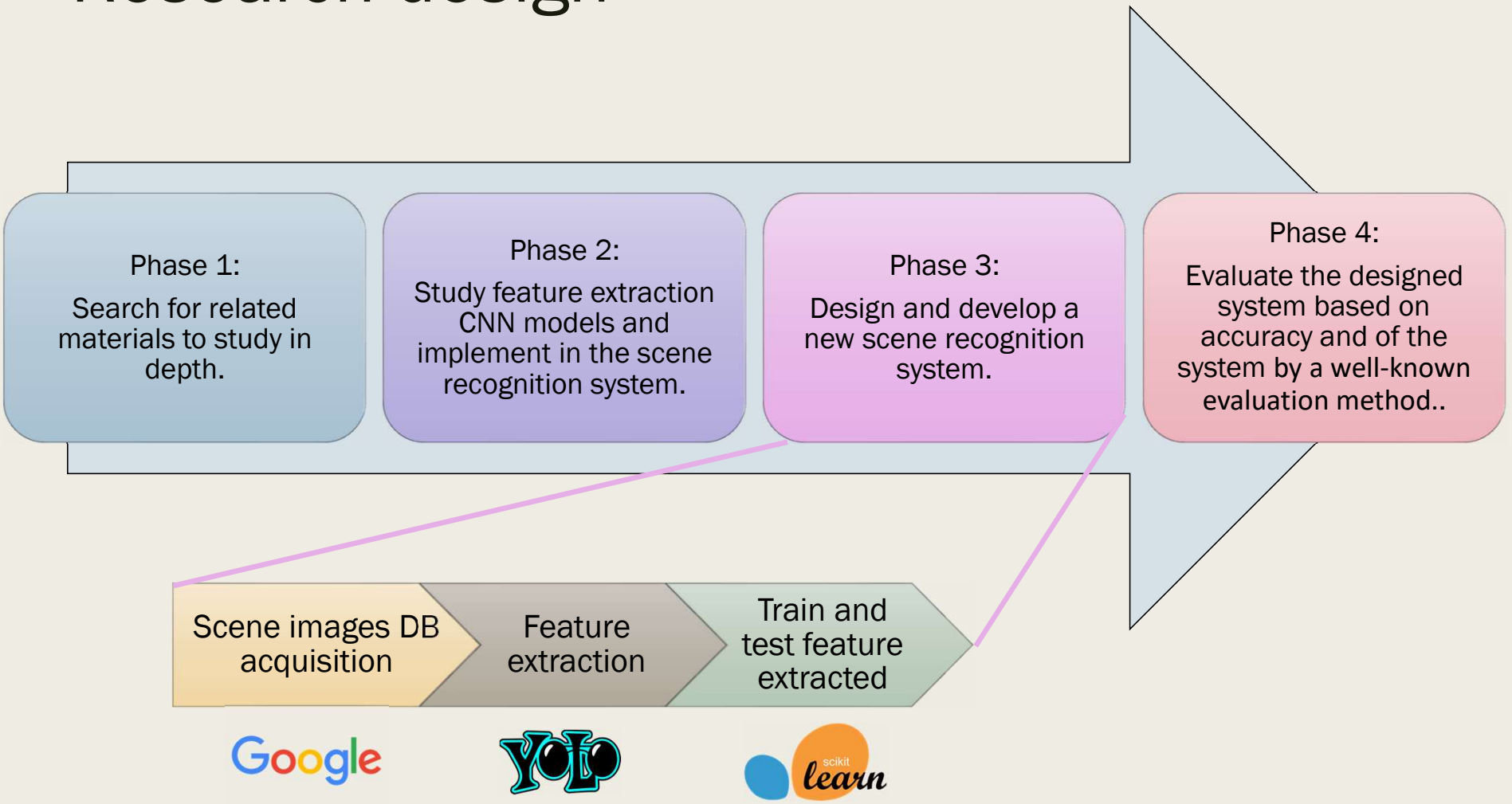
- Identify the place in which the objects seat (Zhou et al., 2017)
- Today, trained by CNN
- A lot of images are passed through the model as input.
- Needs a lot of scene image dataset
- Needs powerful computational device
- Parameters are tuned to identify the best model

- Performance of well-consolidated network architectures classifying images from the challenging Places-365 Standard dataset (Zhou et al., 2018).:

| CNN Network Architectures | Performance accuracy |
|---|---|
| AlexNet | 53.17% |
| GoogleNet | 53.63% |
| VGG-16 | 55.24% |
| ResNet-152 | 54.74% |
| DenseNet-161 | 56.10% |

# METHODOLOGY

# Research design



# FEATURE EXTRACTION

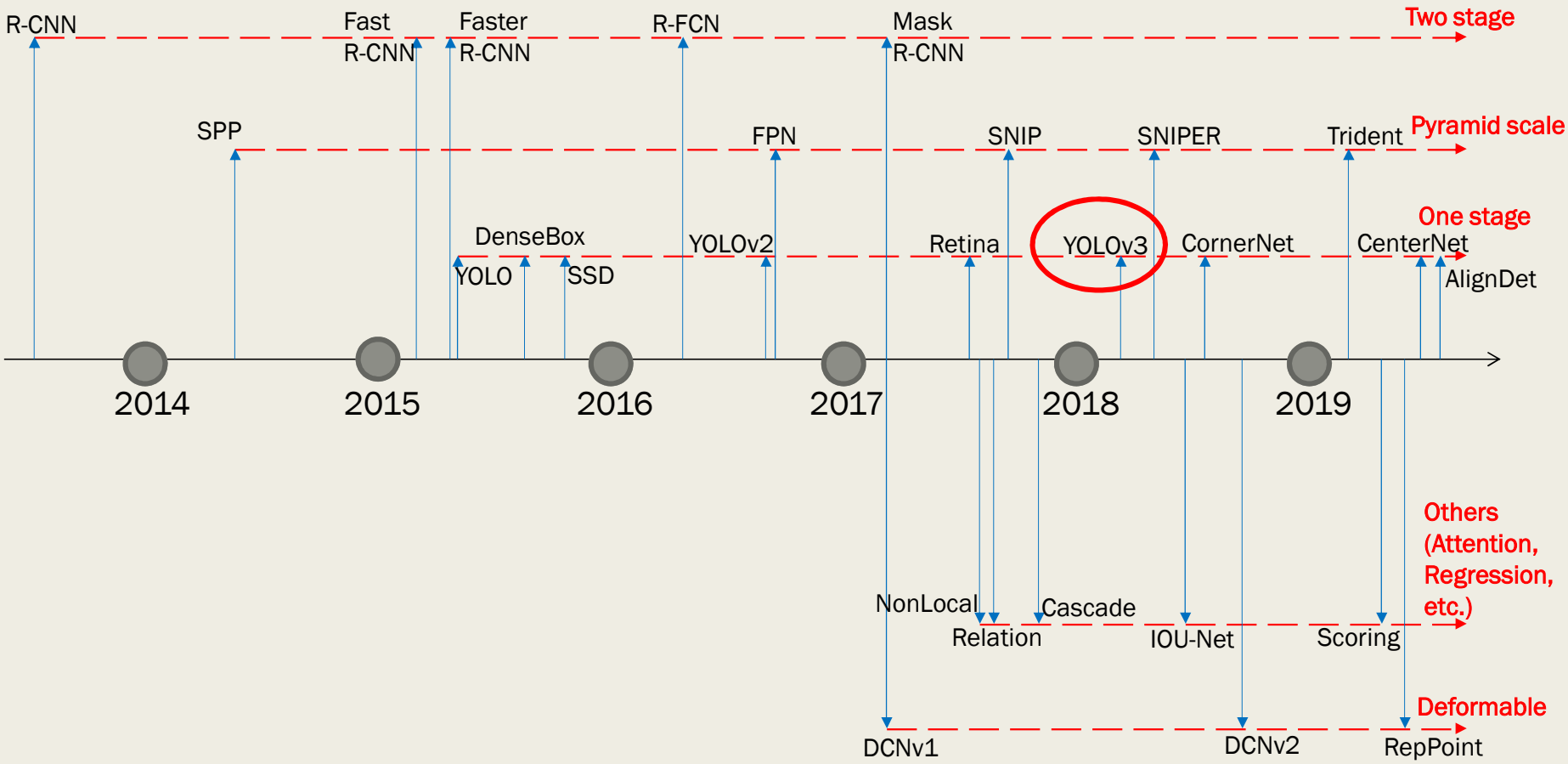

# FEATURE EXTRACTION

Looking at an image, how do we know where the scene in the image is?

# Feature Extraction

# Feature Extraction

## Detection System evolution

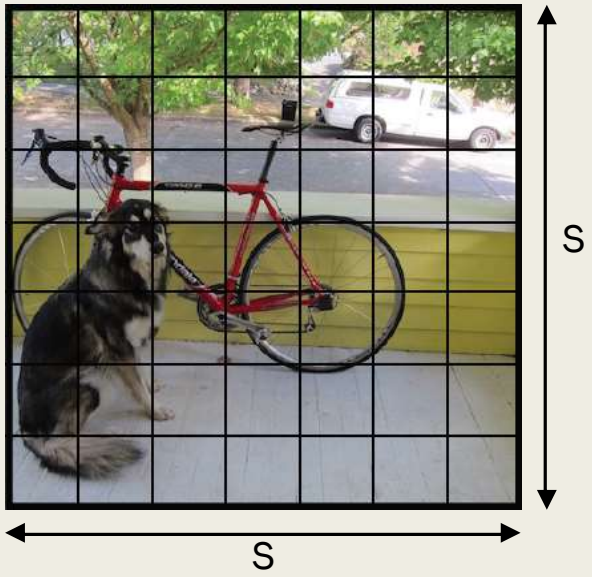


# Feature Extraction

## What is YOLO?

- YOLO takes a completely different approach.
- It is not a traditional classifier that is repurposed to be an object detector.
- YOLO actually looks at the image just once (hence its name: You Only Look Once) but in a clever way:

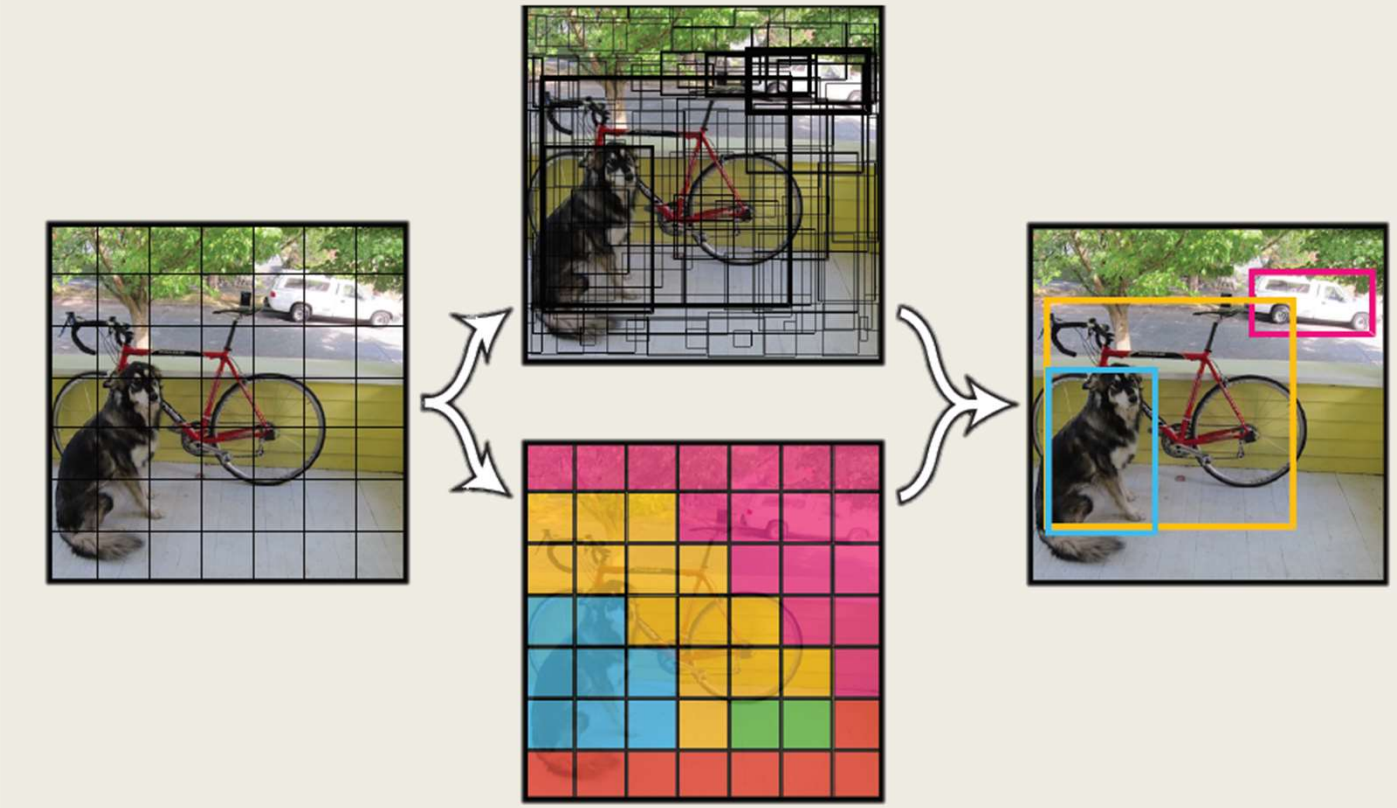
1. YOLO divides up the image into a grid of $S$ by $S$ grid cells.
2. Each grid cell predicts $B$ bounding boxes.
3. A bounding box describes the rectangle that encloses an object.
4. YOLO also output a confidence score reflecting how confident it is that the box contains an object.



# Feature Extraction
## What is YOLO?

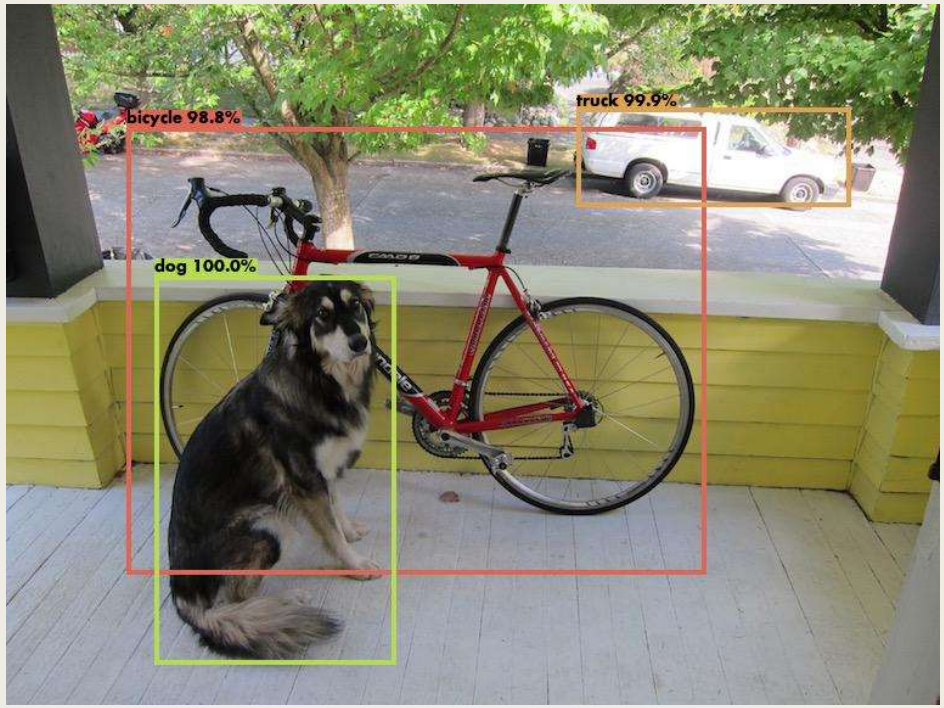

5. Each bounding box consists of 5 predictions: $x$, $y$, $w$, $h$ and confidence.
6. Higher confidence score -> thicker bounding box.
7. Each grid cell predicts $C$ conditional class probabilities, $\Pr(Class_i|Object)$, that are conditioned on the grid cell containing an object.
8. Confidence score + Conditional class probability -> **Class-specific confidence score**

# Feature Extraction
## What is YOLO?



# Feature Extraction

## What is YOLO?



# Feature Extraction


**Object count:**

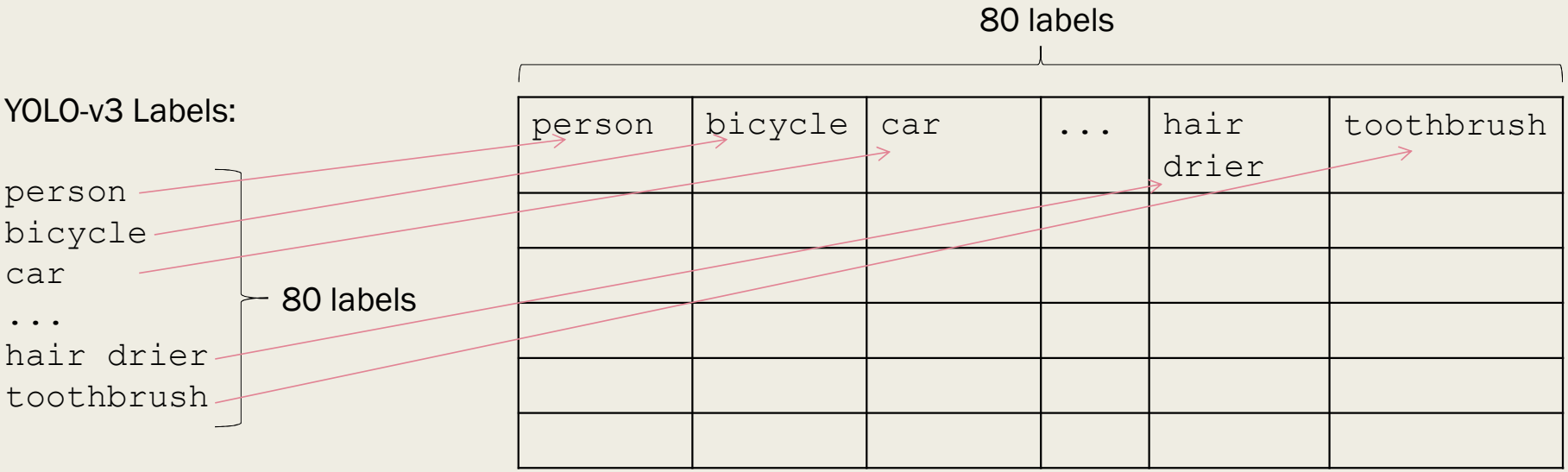
person: 2

chair: 8

pottedplant: 2

# Feature Extraction

## Storing of feature data



# Feature Extraction

## Storing of feature data


**Object count:**

person: 2

chair: 8

pottedplant: 2

| person | bicycle | car | ... | hair drier | toothbrush |
|---|---|---|---|---|---|
| 2 | 0 | 0 | ... | 0 | 0 |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

# SCENE CLASSIFICATION

# Scene Classification

## Storing of scene data

**Object count:**

person: 2

chair: 8

pottedplant: 2

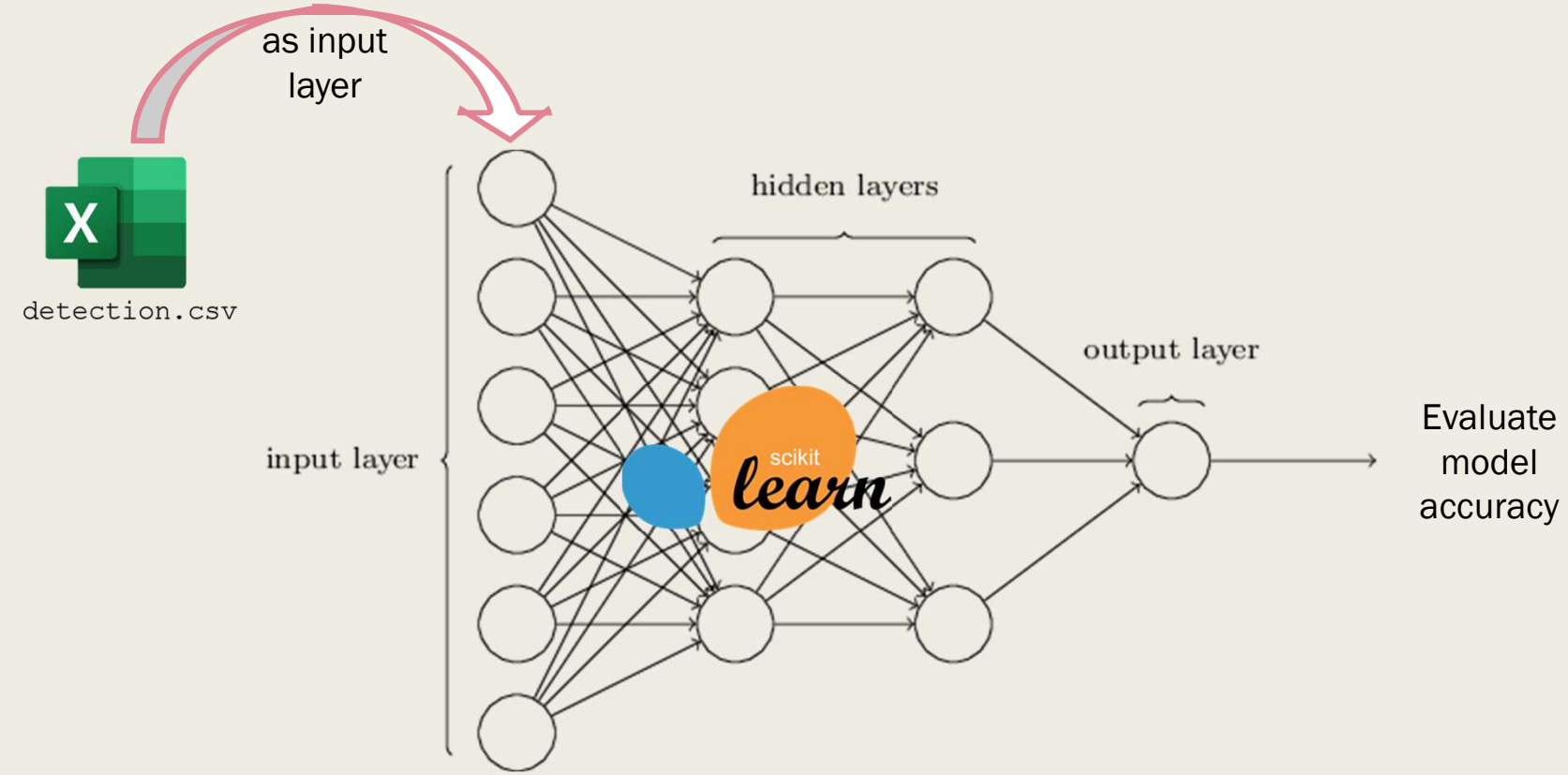
**Scene: cafe**

| person | bicycle | car | ... | hair drier | toothbrush | Scene |
|---|---|---|---|---|---|---|
| 2 | 0 | 0 | ... | 0 | 0 | cafe |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

detection.csv

# Scene Classification

## Training & testing using MLP



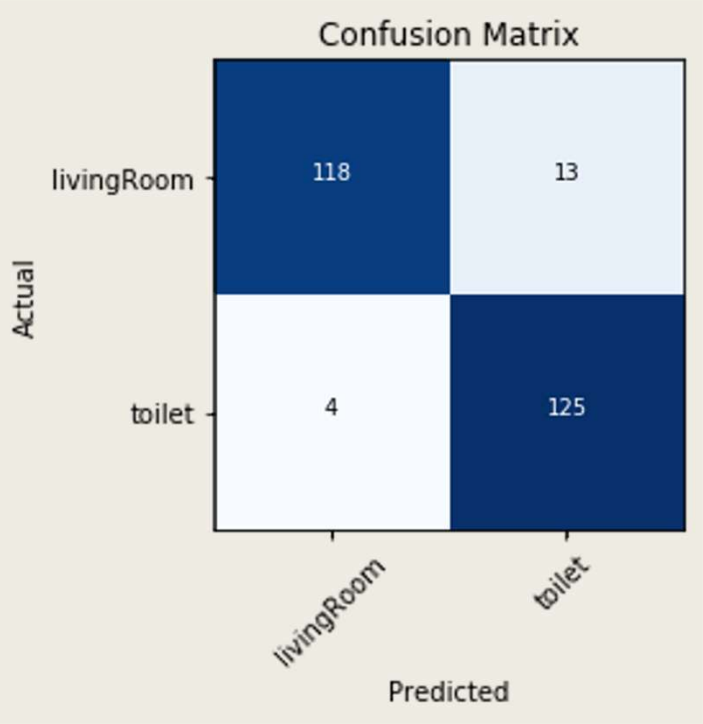
# RESULT & DISCUSSION

# Result and Discussion

## Experiment 1: Toilet VS LivingRoom

| | |
|---|---|
| Toilet images used: | 520 |
| LivingRoom images used: | 520 |
| K-fold: | 300 fold cross validation |
| Mean accuracy: | 0.9711 |

- Very accurate
- Toilet and living room are quite distinct classes

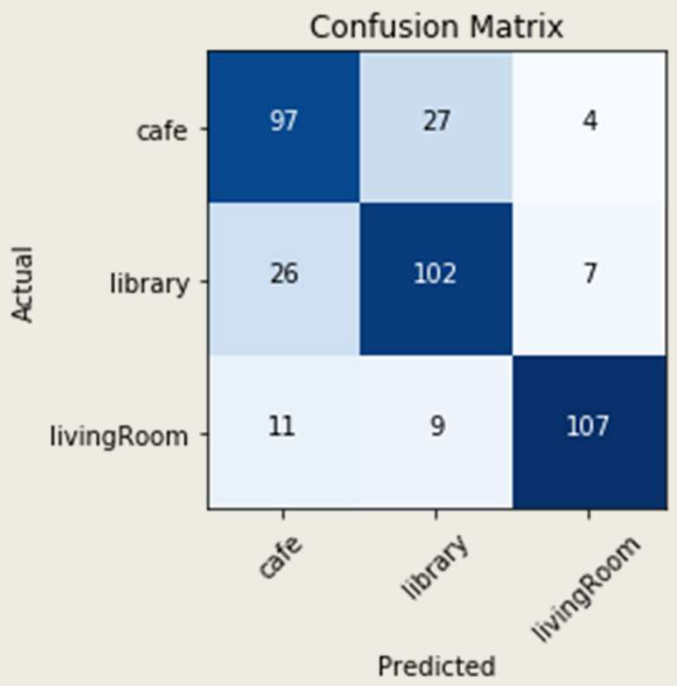


# Result and Discussion

## Experiment 2: Cafe VS Library VS LivingRoom

| | |
|---|---|
| Cafe images used: | 520 |
| Library images used: | 520 |
| LivingRoom images used: | 520 |
| K-fold: | 300 fold cross validation |
| Mean accuracy: | 0.8022 |

- Quite accurate
- Café, library and living room are not so distinct classes



# CONCLUSION

# Conclusion

All objectives of the study have been fully achieved.

- Both experiments using the scene recognition system developed show good results (0.9711 and 0.8022 accuracy on Toilet VS LivingRoom and Café VS Library VS LivingRoom respectively)
- Classification of more distinct classes shows higher accuracy level

# Future Work

- Future work can be done to improve the performance of this scene detection algorithm
- Train and fit in more labels for feature extraction
- Filter out noisy /irrelevant images
- Use different SVM

Thank you.