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

ONG HUI XIN (62322)

Supervised by: Assoc. Prof. Dr Teh Chee Siong

Introduction

Scene recognition is currently one of the top-challenging research fields in computer vision. This may be due to the ambiguity between classes: images of several scene classes may share similar objects, which causes confusion among them (López-Cifuentes et al., 2020). Scene recognition belongs to visual classification, as it means to determine the scene category of an image by analysing its contents. It is very needed in many fields of research, such as remote sensing, aerial scene classification, computer vision and robotics. This research aims to design and develop a scene recognition system using different neural network models.

Problem Statements

In scene recognition, it is necessary to recognize and extract patterns in a scene. **Statistical pattern recognition technique** is a classical method applied in pattern recognition, and it involves a lot of image pre-processing like image segmentation and enhancement (Balan & Sunny, 2018). Hence, when dealing with scene recognition with a lot of objects in a scene, it is very time consuming and computational expensive.

Shifting to **deep learning method**, image data are collected and input as a whole to train the model (Sultana et al., 2018). Hence, there is a need of a lot of image data to be used in deep learning neural network models, which makes it a hard part in this research.

When the enormous recent research in object recognition tasks keeps developing because of the large datasets available like ImageNet and the advancements made in CNN especially for high level feature learning, **performance shown in scene recognition has not achieved the same success level**.

Objectives

The objectives of the study are:

- 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 system performance by a well-known evaluation method.

Literature Reviews

Image pattern recognition and computer vision

Pattern recognition is a process of description, grouping, and classification of patterns. A pattern is an entity that could be associated with a name. In a real-world pattern recognition, problems in diversified forms are ubiquitous and are critical in most human decision-making tasks (Chen et al., 2001).

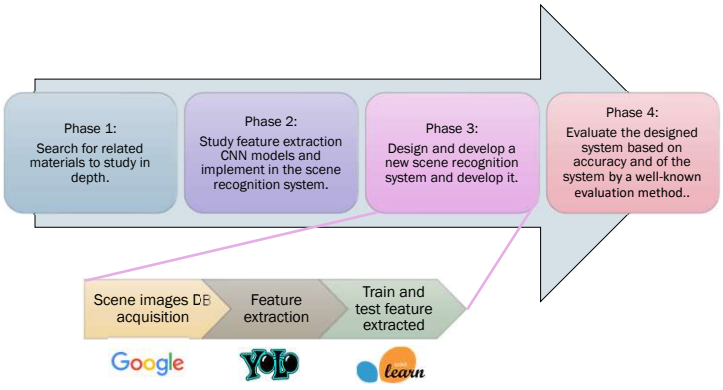
Object recognition

Formally, when an object recognition system is given an image containing one or more objects of interest (and background) and a set of labels corresponding to a set of models known, it should assign correct labels to regions, or a set of regions, in the image. The object recognition problem is closely tied to the segmentation problem: without at least a partial recognition of objects, segmentation cannot be done, and without segmentation, object recognition is not possible (Jain, Kasturi & Schunck, 1995).

Scene Recognition

Well-consolidated network architectures, such as AlexNet, GoogLeNet, VGG-16, ResNet-152 and DenseNet-161 have reported accuracies accu of 53.17%, 53.63%, 55.24%, 54.74% and 56.10% respectively when classifying images from the challenging Places-365 Standard dataset (Zhou et al., 2018).

Methodology



Feature Extraction

Features of an image is extracted using YOLOv3 model and is stored in a csv file.

80 labels | Class

person	bicycle	...	hair drier	toothbrush	SCENE CLASS
2	0	...	0	0	cafe

detection.csv

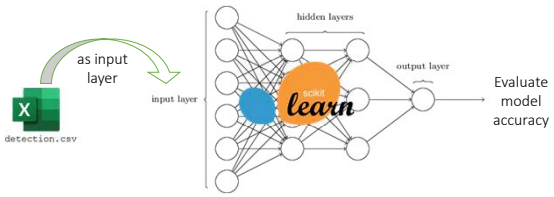
Scene Classification

MLP model provided by scikit-learn library is used to train and test data of scene images. Hyper-parameters of the model are fine-tuned to achieve best results.

Number of images per class: 520

K-fold: 300-fold cross validation

Multi-layer perceptron (MLP) model

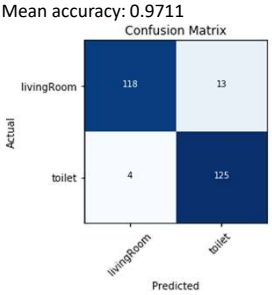


Results and Discussion

Experiment 1:

Toilet VS LivingRoom

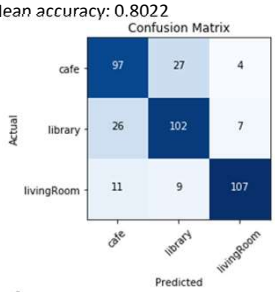
Mean accuracy: 0.9711



Experiment 2:

Cafe VS Library VS LivingRoom

Mean accuracy: 0.8022



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.

In future study, more labels can be trained to fit into YOLOv3 for feature extraction purpose. Besides, noisy or irrelevant data images should be filtered out to prevent outliers. Lastly, different support vector machine can be used to find out the most suitable model when doing scene classification.

References

Balan, S. & Sunny, L. E. (2018). Survey on feature extraction techniques in image processing. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 6*(3), 217-222.

Chen, K., Kvasnicka, V., Kanen, P. C. & Haykin, S. (2001). Supervised and unsupervised pattern recognition: Feature extraction and computational intelligence [Book Review]. In *IEEE Transactions on Neural Networks, 12*(3), 644-647.

Jain, R., Kasturi, R. & Schunck, B. G. (1995). *Machine vision*. US: McGraw-Hill.

López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J., & García-Martín, Á. (2020). Semantic-aware scene recognition. *Pattern Recognition, 102,* 107256.

Sultana, F., Sufian, A. & Dutta, P. (2018). Advancements in image classification using convolutional neural network. In *Forth International Conference on Research in Computational Intelligence and Communication Networks.* Kolkata, India: RCCIIT.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. & Torralba, A. (2018). Places: A 10 million image database for scene recognition. *Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452-1464.