THE UNIVERSITY OF HONG KONG

DEPARTMENT OF COMPUTER SCIENCE

COMP4801 Final Year Project

Object Recognition by Deep Learning Neural Networks

Project Plan

Du Haiyang (3035087124)

Wang Shunqi (3035085736)

Supervised by Dr. KP Chan

October 1st, 2017

Abstract

Deep Learning Neural Networks have been commonly used in the field of object recognition. This project plan intends to give a detailed overview on the Final Year Project "Object Recognition by Deep Learning Neural Networks". The ultimate objectives of this project are to: 1) reproduce R-CNN on Python; and 2) replace original classifier with LDA classifier to improve on accuracy rate. To achieve the goal, project team will utilize ImageNet ILSVC 2012 and PASCAL VOC 2007 datasets to train the model and leverage PASCAL VOC 2012 dataset to evaluate the performance.

This project plan will start from an introduction and provide some related researches that have been conducted in the object recognition field of studies. It will then move on to discuss the scope, prerequisites, methodology and deliverable of this project. Also, to offer more realistic insight into the project, analysis on possible challenges will be provided. Eventually, it will outline a detailed time schedule and conclusion for this project.

Table of Contents

Introduction

Human beings are capable of recognizing objects through their eyes, even if the objects merely varied from each other in the slightest. However, object recognition, a technology to capture and recognize objects in an image or a series of images by utilizing computer vision algorithm, is far from the efficiency of human's eyes. Extensive research on object recognition has been conducted over the past decade, where deep learning neural networks, a methodology for machines to understand and interpret data representations without specific tasks being enforced, were widely utilized to conduct the researches.

Regional Convolutional Neural Network (R-CNN) represents the usage of deep learning neural networks on identification of target object [1]. The key objectives of this Final Year Project are to implement R-CNN through Python and to leverage Latent Dirichlet Allocation (LDA), a graphical model for topic discovery, as the new classifier to improve performance.

Related Studies

R-CNN operates three modules, starting from selective search on image, moving on to extract features by CNN and eventually identify objects through Support Vector Machine (SVM) classifier [1]. The above-mentioned process is illustrated as Figure 1 below.

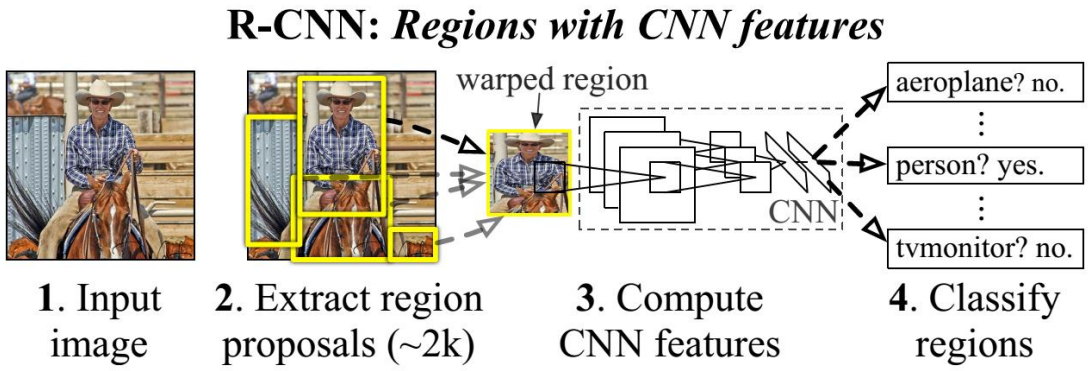

Figure 1: R-CNN process. From *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation* by Girshick R, Donahue J, Darrell T, Malik J.

Since the introduction of R-CNN, research on improving the algorithm performance based on original version has attracted uprising interest in this field of study. Main objectives of these researches concentrate on accelerating training time and raising mAP accuracy rate.

Take Fast R-CNN as an example, which proposed Regional Proposal Network (RPN) to accelerate processing time of R-CNN, and at the same time adopted to replace SVM classifier to Softmax classifier in order to increase accuracy [2]. The result yielded a great improvement in the performance, where Fast R-CNN is capable to achieve up to 66% mAP when evaluating by PASCAL VOC 2012, compared to record high accuracy of 62% with R-CNN [2].

In this project, baseline framework has been chosen to be R-CNN and re-implementation on Python will be introduced. Focusing solely on improving accuracy rate, LDA classifier will

be introduced to replace SVM classifier in this project, which has already been used in the research study of other image classifications. Han Bing and Yang Chen has applied a LDA based algorithm to classify and analysis aurora images by treating the images as documents [3]. There are also researches making comparisons to LDA and SVM showing that LDA based algorithms are more robust and outperforms SVM when trained by either large scale or small-scale datasets [4].

Scope

Given thorough consideration on the complexity in conducting this research, the scope is temporarily limited to the following four parts and is subjected to changes that may occur during actual research progress:

- ***RCNN Focused***

  Although improved versions of R-CNN, such as Fast R-CNN and Faster R-CNN, have emerged, this FYP project will focus solely on R-CNN, considering the better extensibility of R-CNN than other two improved versions.
- ***Single Controlled Variable of Classifier***

  Maintaining other variables to stay the same, this project aims to explore by how much extent could LDA classifier help improving RCNN's performance.
- ***Three Datasets for Research***

  Among numerous datasets available and take into consideration of the time limit, three datasets will be leveraged in this project. Selected two datasets, ImageNet ILSVC 2012 and PASCAL VOC 2007, will be used for training the model. Another dataset, PASCAL VOC 2012, will be utilized to evaluate the performance for all three versions of implementation.
- ***Code Base Only***

  Considering the project nature, no UI will be developed. Instead, a code base will be provided to implement the algorithm.

Prerequisites

1. **Caffe** [5] and Caffe's prerequisites with Python layer and pycaffe framework.
2. **Matlab** to build R-CNN source code.
3. **Python 3** to reimplement R-CNN.
4. **R-CNN** source code which based on Matlab.

Methodology

In order to micro-manage the project process, the methodology has been divided into three different research phrases that are listed as follows:

- ***Matlab Version Source Code Implementation***

  RCNN source code is originally implemented through Matlab. This version will be used as benchmark for the project, which will also be trained through ImageNet ILSVC 2012 and PASCAL VOC 2007 and evaluated by Pascal VOC 2012, to keep all training and evaluating environment stable.
- ***Python Version Implementation***

  RCNN will be implemented through python based on Matlab version. A same training and evaluating environment will be provided for this replicated version. The evaluation result will be recorded as well to compare with Matlab version's result and also set as another benchmark for the project.
- ***LDA Classifier Version Implementation***

  The original SVM classifier will be replaced by LDA classifier, while other parts of the algorithm will remain the same as python version. Trained and evaluated by the exact same PASCOL database, the result will be put under comparison with Matlab version and python version.

Deliverable

This project will deliver an improved object recognition code base that is implemented through Python and based on Matlab version R-CNN. The code base will be trained by ImageNet ILSVC 2012 and PASCAL VOC 2007 datasets and is expected to achieve higher accuracy rate, evaluating by PASCAL VOC 2012 dataset. Project progress can be checked on the project website: http://i.cs.hku.hk/fyp/2017/fyp17015/ .

Challenges and Mitigations

The main challenges for this project are hardware constraint and uncertainty in Python version performance. Detailed obstacles that may occur and corresponding mitigations can be found as follow:

- ***Hardware Constraint***

  RCNN Matlab version requires sophisticated hardware support, such as high-performance GPU, large disk spaces (around 200G) to cache image and feature vector. The requirements may vary when we implement it through python. As a consequence, potential delay to original project schedule and lower-than-expected performance shall occur.
  Mitigation:

  1) *Current hardware will be tested under Matlab version RCNN to check if they are satisfactory enough. New disk and GPU will be bought using the research funding if current hardware could not meet the demand.*
  2) *More flexible project schedule will be adopted.*

- ***Uncertainty in Python Version Performance***

Due to the change in language environment, selective search may be realized in a different way than the original version. Furthermore, python version may lead to worse performance than original Matlab version, such as slower processing time and less accuracy rate.
Mitigation:

1) *Only compare performance between python version and LDA version, which controls all other variables and focus solely on whether LDA classifier is able to improve the RCNN performance.*

Time Schedule

Time Frame	**Task**
Sep. 2017	● Detailed project plan submission ● Project webpage go live ● Initial meeting with Dr. KP Chan to confirm research topic and setup regular meeting time slots ● Related research paper reading
Oct. 2017	● Environment Setup ● Download RCNN Matlab version source code and get familiar with the algorithm ● Train and evaluate Matlab version's performance
Nov. 2017 – Dec. 2017	● Reproduction of RCNN through python based on Matlab version source code ● Check hardware status and decide on whether higher standard hardware should be used ● Train and evaluate the python version's performance ● Interim report submission
Jan. 2018 – Feb. 2018	● Replace classifier from SVM to LDA on python version, while maintaining other parts exactly the same ● Train and evaluate the LDA classifier version's performance

Mar. 2018	● Comparison among the performances of three different version of RCNN implements ● Caliber the parameter of LDA classifier version to improve accuracy rate
Apr. 2018	● Final report submission ● Final project presentation ● Final presentation poster design
May 2018	● Final project exhibition

Conclusion

This project is based on R-CNN, a cutting-edge technology used for identification of objects in the field of computer vision. Reproduction from source version to Python version will be commenced, further modification in classifier will be introduced and consistent training through ImageNet ILSVC 2012 and PASCAL VOC 2007 datasets will be enforced. In conclusion, an increase in mAP accuracy rate shall be expected from this project evaluating by Pascal VOC 2012 dataset.

References

[1] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in IEEE conference on computer vision and pattern recognition (pp. 580-587), 2014.

[2] R. Girshick, "Fast R-CNN," in In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448), 2015.

[3] H. Bing and G. X. B. Yang Chen, "Aurora image classification based on LDA combining with saliency information.," Ruan Jian Xue Bao/Journal of Software, vol. 24, no. 11, pp. 2758-2766, 2013.

[4] T. Rubin, A. Chambers, P. Smyth and M. Steyvers, "Statistical Topic Models for Multi-Label Document Classification," Machine learning, vol. 88, no. 1, pp. 157-208, 2012.

[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.