香港大學計算機科學系

THE UNIVERSITY OF HONG KONG

DEPARTMENT OF COMPUTER SCIENCE

COMP4801 Final Year Project

Modified R-CNN:

Object Recognition by Deep Learning Neural Networks

Interim Report

Du Haiyang (3035087124)

Wang Shunqi (3035085736)

Supervisor: Dr. KP Chan

Submitted on January 21st, 2018

ABSTRACT

Deep Learning Neural Networks have been commonly used in the field of object recognition. This draft interim report 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 Latent Dirichlet Allocation classifier to improve accuracy. In order to achieve the goal, project team will utilize public datasets to train and evaluate the algorithm. At current stage, Matlab version R-CNN has been implemented with a 49.6% mAP and selective search has been reproduced in Python version R-CNN. No particular difficulties were encountered at this stage since the project is still in the early phase. It is expected that ultimate deliverable will be able to achieve higher accuracy rate than Python implemented R-CNN.

ACKNOWLEDGEMENT

Progress of this project and featured interim report would not have been possible without the kind support and help of two individuals.

We would like to express our gratitude to Dr. KP Chan, our project supervisor, for providing useful guidance, valuable hardware resources and inspiring advices throughout the project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

CNN -- Convolutional Neural Networks

R-CNN – Regional Convolutional Neural Networks

LDA – Latent Dirichlet Allocation

mAP – Mean Average Precision

1. INTRODUCTION

Human beings are capable of identifying objects through their eyes with little effort, even if the objects merely vary from each other in the slightest. Moreover, biological visual system of human beings can recognize the objects from different viewpoints, even when the objects are partially obstructed. However, the same kinds of task are extremely difficult for computer system to imitate, despite the fact that computers nowadays can surpass human easily in many ways given the powerful calculation speed and enormous memory size. The technology to capture and recognize objects in an image or a series of images by utilizing computer vision algorithm is called object recognition and extensive research in this field of study has been conducted over the past decades.

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 algorithm in Python language environment and to improve object recognition accuracy rate by leveraging advanced classifier. The project will utilize Latent Dirichlet Allocation (LDA) classifier, a topic model for text mining, in replacement of original classifier in the purposing of raising algorithm performance.

This interim report intends to present a comprehensive illustration of the final year project. It starts by examining previous related studies in the field of object recognition using deep learning neural networks. Then, it outlines the scope and prerequisites of the final year project. After that, it discusses in detail the approach applied to conduct the project and report the current progress of this project. Eventually, it proposes some potential challenges that may be expected and corresponding mitigations.

2. BACKGROUND

Related studies on R-CNN and topic model will be covered in this section.

- ***2.1 R-CNN***

  R-CNN operates three modules, selective search on image, CNN feature extraction and object identification through Support Vector Machine (SVM) classifier [1]. This process is illustrated in Figure 1 below.

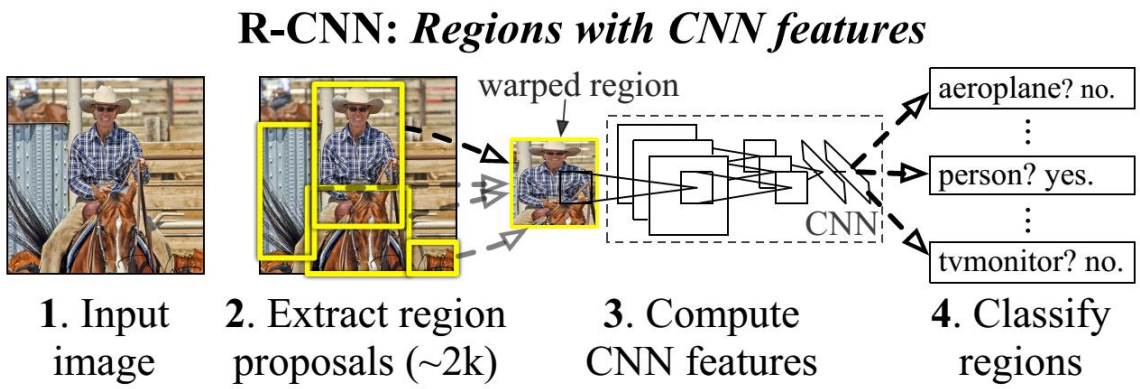

Figure 1: R-CNN Process [1]

In Figure 1, R-CNN first input the image from its data source and then extract region proposals that may contain objects. Then it computes the CNN features of each region and eventually classify each region according its specific characteristic.

- ***2.2 Topic Model and Latent Dirichlet Allocation***

  Topic model is a machine learning model originally used in natural language processing. It is a statistical model to identify the abstract topics in a document. There exist several different topic models based on the statistical distribution they are using. One of the most popular topic models now is Latent Dirichlet Allocation (LDA) [10], which is developed by David Blei, Andrew Ng, and Michael I. Jordan in 2002. This model is using Dirichlet distribution and the name of the model comes from this distribution.

  Topic model technology are broadly used in document classification and reading recommendation. By applying topic modeling to a reader's reading history, content providers can learn which topic the reader is more interested in and will then offer similar documents to this reader. Detail algorithms and why LDA is chosen in this project to replace SVM classifier will be explained later.

Although sophisticated algorithm such as Fast R-CNN and Faster R-CNN has been introduced, this project will base its framework on R-CNN given its originality and better extensibility compare to other advanced algorithms. Considering the technological constraint and time limit, this project cannot perform substantial improvement to R-CNN algorithm as what Fast R-CNN and Faster R-CNN do. The scope of the research that will be covered in this project will be introduced in the next section.

3. SCOPE

Given thorough consideration on the complexity in conducting this research, the scope of this project is currently limited to the following four parts and will subject to changes that may occur during actual research progress:

- ***3.1 R-CNN 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.

- ***3.2 Datasets for Research***

  Among numerous public datasets available for object recognition study, two datasets will be leveraged in this project. PASCOL VOC 2007 will be used for training the algorithm and PASCOL VOC 2012 will be utilized to evaluate the performance for implementations.

- ***3.3 Code Base Only***

  Given the project nature, no user interface will be developed. Instead, a code base will be provided for implementation.

In short, the scope of this project will be limited to training and evaluating algorithm's performance by two pre-selected public datasets, with ultimate deliverable to be a code based algorithm that will generate better mAP accuracy rate. In light of the challenging aspects from both hardware and code perspective, several high standard prerequisites will be required, listing in the next section.

4. METHODOLOGY

In order to successfully conduct this project, research process is separated into three different phases as follows:

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

  R-CNN source code is originally implemented in Matlab environment. This version will be used as benchmark for the project, which will be trained by Pascal VOC 2007 and evaluated by Pascal VOC 2012 to keep all training and evaluating environment on accord.

- ***4.2 Python Version Implementation***

  R-CNN will be reprogrammed to implement in Python environment based on the Matlab version. It is generally recognized that Python offers a better adaption to all programmers, since that Python is more commonly used compare to Matlab. Moreover, Python environment offers more extensibility so that further study based on this research project can be easily conducted. The 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. It is expected to achieve relatively the same training speed and accuracy rate as Matlab version. However, a 5% fluctuation will be allowed given the transition in language environment. Noted that deviation from the expectation may be encountered, and with the progress of research, the project schedule will subject to changes due to this reason.

- ***4.3 LDA Classifier Version Implementation***

  Based on the Python version R-CNN obtained, research on improving accuracy rate will be conducted by replacing SVM classifier with LDA classifier while other parts of the algorithm remain the same. LDA classifier will be adapted to fit into R-CNN algorithm, and is expected to provide more accurate identification on objects compare to original SVM classifier.

  As introduced before, Latent Dirichlet Allocation is an example of topic model and is originally used in text modeling, analysis, and classification [10]. LDA is a model that treats documents as a collection of words and there is no sequence order between words. It claims that a document is consist of several topics and every word is generated by one of the topics. This concept also applies to object recognition, because images can also be treated as a combination of topics. The popular concept of tagging images is quite like the idea of topic in LDA. Additionally, compared to SVM, LDA is an unsupervised learning algorithm and does not require manual work to prepare pre-training data. This will save memory and speed the system up. Since R-CNN is using a modular design, a replacement in this section won't have great effects on other parts of the model. Only the input datatype format should be slightly modified since the model is using a new classifier.

LDA Classifier version R-CNN will be trained and evaluated by the exact same PASCOL datasets, in order to keep apple-to-apple principle. Based on the result of output, specific parameter will be calipered for achieving optimal performance.

To summarize, the research will be conducted in 3 phrases, starting with Matlab version R-CNN implementation, then with Python version R-CNN reproduction, and eventually with LDA Classifier version R-CNN creation. The performance of LDA Classifier version R-CNN is expected to exceed that of Python version R-CNN. Progress of the final year project will be reported in the next section.

5. PROGRESS

- ***5.1 Current Stage of Work***

  Given the time limit, only three parts were finished at this stage: environment setup, Matlab version R-CNN implementation and selective search implementation in Python version R-CNN.

  - *Project Website*

    A webpage has been set up for the project to give the audience a brief introduction of the project. Documents, project progress and contacts of project members are posted on the webpage and it will be updated regularly. The URL of the webpage is: http://i.cs.hku.hk/fyp/2017/fyp17015/ and below is a screenshot of the project website.

  - *Environment Setup*

    A computer embedded with Intel i7 and 32G RAM was provided. 332G disk space was granted to this final year project to store image cache. In addition, GeForce GTX GPU was also setup for training, which will be used in this project.

    On top of the hardware, Python and Caffe were installed in the system.

  - *Selective Search Implementation in Python Version R-CNN*

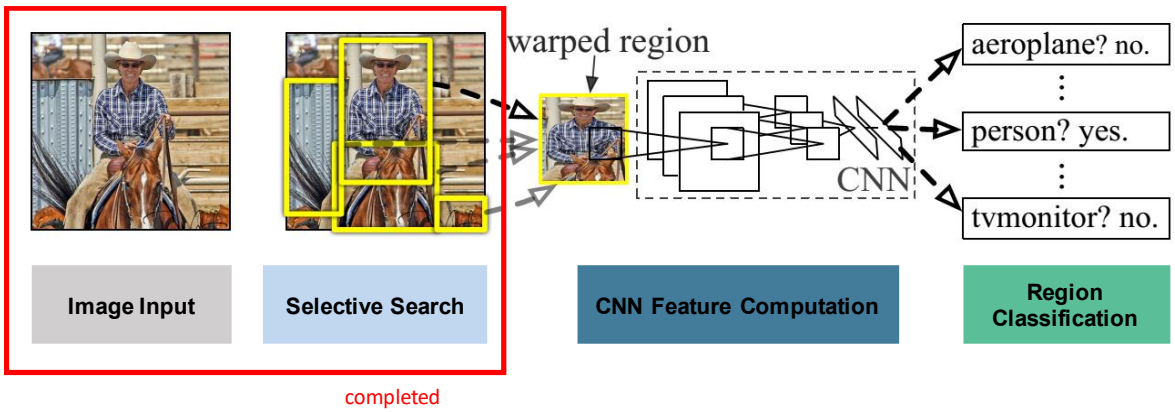

Figure 2: Current Stage of Work [1] (Figure was reproduced on top of original graph in paper)

Figure 2 illustrates the current stage of Python version R-CNN implementation, which is highlighted in the red square. The first part was to set up environment and prepared PASCOL VOC 2007 and PASCOL VOC 2012 datasets for image input. The second part was to implement selective search algorithm in Python.

- **5.2 Interim Result**

The interim result is that Python Version R-CNN is now able to process image input and apply selective search algorithm to segment minimum regions with the highest possibility to contain objects.

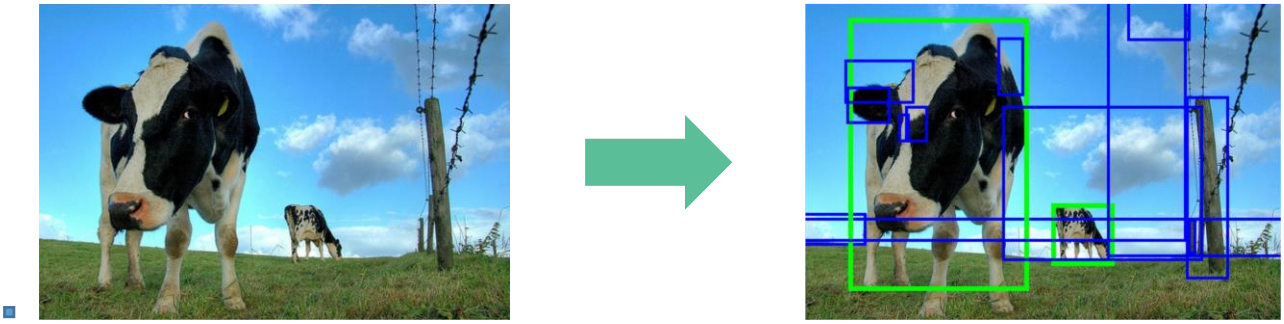

Figure 3: Interim Result Example [9]

Figure 3 gives a visual example on interim result. In the figure, an input image with two milk cows and fence was applied to selective search. The proposed regions with the highest possibility to contain object are circled in green. The less possible regions were circled in blue for CNN computation use, which would be implemented in later phases.

- **5.3 Timeframe**

  Future work will focus on completed implementation of R-CNN algorithm in Python. For the next phase, Python version R-CNN will be tested and unexpected technical issues raised by the transition of language environment will be dealt with. Eventually, LDA classifier version R-CNN will be adapted into the R-CNN algorithm in replacement of SVM classifier in the purpose of achieving higher mAP rate.

  Table 1 below gives a detailed overview on the time schedule of this project:

Time Frame	**Task**
Sep. 2017 **(Completed)**	● Detailed project plan submission ● Project webpage creation ● Initial meeting with Dr. KP Chan to confirm research topic and setup regular meeting time slots ● Related research paper reading
Oct. 2017 **(Completed)**	● Download R-CNN Matlab version source code and familiarize with the algorithm ● Train and evaluate Matlab version R-CNN
Nov. 2017 – Jan. 2017 **(In Progress)**	● Python version R-CNN implementation ● Check hardware status and decide on whether higher standard hardware will be used ● Train and evaluate Python version R-CNN ● Interim report submission
Jan. 2018 – Feb. 2018 **(To Be Completed)**	● Replace classifier from SVM to LDA based on Python version R-CNN ● Train and evaluate LDA classifier version R-CNN
	● Caliber the parameter of LDA classifier to improve accuracy rate

Mar. 2018 **(To Be Completed)**	
Apr. 2018 **(To Be Completed)**	• Final report submission • Final project presentation
May 2018 **(To Be Completed)**	• Final project exhibition

Table 1: Project Schedule (Completed, In Progress and To Be Completed)

6. DELIVERABLES

This project will deliver an improved R-CNN algorithm using in object recognition study, in the form of code base. The algorithm will be implemented in Python environment, and in expectation to achieve a mAP accuracy rate higher than 49.6% evaluating by PASCOL VOC 2012 dataset. Project progress can be checked on: http://i.cs.hku.hk/fyp/2017/fyp17015/

7. POTENTIAL CHALLENGES AND MITIGATIONS

Potential challenges for this project are hardware constraint, uncertainty in Python version performance and uncertainty in LDA Classifier's improvement. Detailed obstacles that may occur and corresponding mitigations can be found as follow:

- ***7.1 Hardware Constraint***

  R-CNN 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 in Python environment, and as a consequence, potential delay to original project schedule and lower-than-expected performance may occur.

  Mitigation:

1) *Current hardware provided by supervisor will be tested on Matlab version R-CNN to check if they are satisfactory enough. New disk and GPU may be bought using the research funding if current hardware is not able to meet the demand.*

2) *Flexible project schedule will be adopted.*

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

  Due to the change in language environment, Python version R-CNN may perform worse than original Matlab version. Slower processing time and less accuracy rate may be encountered.

  Mitigation:

  1) *Only compare the performance between Python version R-CNN and LDA Classifier version R-CNN, which controls all other variables and focus solely on whether LDA classifier is able to raise R-CNN algorithm's accuracy rate in identifying objects.*

- ***7.3 Modifications on Selective Search***

  Selective search algorithm used in the final year project generates around 2000 segments with different sizes and shapes. Nonetheless, the CNN requires all the input segmentations to have exactly same size. To resolve this conflict, two possible measures can be taken.

  Mitigation:

  1) *Wrap all the images to a certain size after generating the images. However, this may cause some misunderstanding for CNN. Because the size of the object has been modified, a small little tree may be recognized and treated as a big wooden door after resizing.*

  2) *Pad the picture with some background color. This may also cause confusion since the neural network is taking the whole picture into consideration.*

8. CONCLUSION

This project researches on the heated object recognition topic, targeting to improve the performance of widely used R-CNN algorithm and aiming to provide a better algorithm that is capable of achieving higher object identification rate in the form of code base. Currently, the project has finished three parts: environment setup, Matlab version R-CNN implementation and selective search implementation in Python version R-CNN.

Following phases of research will focus on completion of Python version R-CNN and corresponding evaluation, as well as LDA Classifier version R-CNN implementation.

In light of the upside benefit of potential improvement on object identification accuracy, this project is regarded as scientifically meaningful. It is believed that the ultimate deliverable can inspire researchers in the field of object recognition by deep learning neural networks.

REFERENCE

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

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

[3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

[4] Artificial Intelligence. Retrived from https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/object_localization_and_detection.html

[5] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678). ACM.

[6] Ludwig, T. (2004). Research trends in high performance parallel input/output for cluster environments.

[7] Gu, S., Tan, Y., & He, X. (2010). Discriminant analysis via support vectors. *Neurocomputing*, *73*(10), 1669-1675.

[8] Bellingegni, A. D., Gruppioni, E., Colazzo, G., Davalli, A., Sacchetti, R., Guglielmelli, E., & Zollo, L. (2017). NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. *Journal of neuroengineering and rehabilitation*, *14*(1), 82.

[9] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2), 154-171.

[10] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993-1022, 2003.