An application of Residual Network and Faster - RCNN for Medico: Multimedia Task at MediaEval 2018

Trung-Hieu Hoang[1], Hai-Dang Nguyen[2], Thanh-An Nguyen[1],
Vinh-Tiep Nguyen[3], Minh-Triet Tran[1],
[1] Faculty of Information Technology, University of Science, VNU-HCM, Vietnam [2] Eurecom, France
[3] University of Information Technology, VNU-HCM, Vietnam
{hthieu,ntan}@selab.hcmus.edu.vn,nguyenhd@eurecom.fr,tiepnv@uit.edu.vn,tmtriet@fit.hcmus.edu.vn

ABSTRACT

The Medico: Multimedia Task focuses on developing an efficient framework for predicting and classifying abnormalities in endoscopic images of gastrointestinal (GI) tract. We present the HCMUS Team's approach, which employs a combination of Residual Neural Network and Faster R - CNN model to classify endoscopic images. We submit multiple runs with different modifications of the parameters in our combined model. Our methods show potential results through experiments.

1 INTRODUCTION

Medico: Multimedia Task at MediaEval 2018 challenge [4] aims to bring new achievements in computer vision, image processing and machine learning to the next level of computer and multimedia assisted diagnosis. The goal of the challenge is to predict abnormalities and diseases in an efficient way with as less training data as possible [5]. The task organizers also provide a priority list for the classes in other to accommodate with the single-class classification challenge. Thus, this leads to some modifications of our model, which are meticulously described in section 3.

In our approach, we introduce a stacked model consisting of two deep networks, a Residual Neural Network (Resnet) [2] followed by a Faster Region-based Convolutional Neural Network (Faster R-CNN) [7]. Since Resnet mostly focuses on deep global features of image, it fails to classify images that symptoms of abnormal diseases or instruments appear as small objects on diversity backgrounds. Therefore, this is the reason of using Faster R-CNN to re-classify the images of some classes that Resnet usually mis-classify.

2 RELATED WORK

In the field of medical image processing, deep neural networks have been used in order to solve several problems related to endoscopic images of the gastrointestinal (GI) tract. Particularly, to localize and identify polyps within real-time constraint, deep CNNs has recently shown an impressive potential when achieving up to 96.4% accuracy - published in 2018 by Urban G et al. [9]. Another interesting article of Satoki Shichijo et al. [8] also applies multiple deep CNNs to diagnose Helicobacter pylori gastritis based on endoscopic images. Further, gastrointestinal bleeding detection using deep CNNs on endoscopic images has been successfully done and published by Xiao Jia et al. [3].

MediaEval'18, 29-31 October 2018, Sophia Antipolis, France

3 APPROACH

3.1 Dataset Preparation

3.1.1 Disease region localization. In order for the Faster R-CNN model to be trained, objects in the image have to be tagged with bounding boxes and passed to the model as input. We annotate the signal of disease in all images of the following classes: *dyed-resection-margins, dyed-lifted-polyps, instruments* and *polyps*.

3.1.2 Re-labeling Medico development dataset. After training with the development set, we find some training samples with inappropriate labels according to the priority list. Therefore, in order for our model to learn with the least confusing, we apply the new labels, predicted by the trained model, to these images.

3.1.3 Instruments dataset augmentation. Instruments - the second highest priority class has only 36 images with the limitation of background context in the development set. In order to maintain the balancing between all of the classes and also improve the diversity of the *instruments* images, we generate more images for the *instruments* based on the current given development set by placing the instruments on the foreground of other diseases backgrounds.

Among the 36 *instruments* images, we carefully select 24 of them and crop the instruments along their edges. Then, we randomly select 20% of the images from *dyed-lifted-polyps, dyed-resection-margins*, *ulcerative-colitis* classes, and use them as the background of the cropped instruments. By applying this method, we are able to generate more than 800 images for the *instruments* class.

3.2 Method

3.2.1 Fine-tuning deep neural network for medical images. In our approach, both Residual Network with 101 layers and Faster R-CNN [1] (both pre-trained on ImageNet) are fine-tuned by using our modified development dataset. In term of using convolution neural network for medical images, knowledge transferring from natural images to medical images is possible, even though there is a large difference between the source and target databases. It is especially useful in the case of small dataset of images provided [6]. Our experiment results also support this idea. Fine-tuning on the ImageNet pre-trained model significantly improves the efficient of classification model.

3.2.2 First run. Residual network with 101 layers model are fine-tuned on the original development set provided by the task organizers along with our instruments increased dataset. After passed through Resnet101, output images classified as special classes become the input of Faster R-CNN network, which is trained for detecting instruments in images.

- First case: Images predicted as *instruments* by Resnet101 are double-checked. In case instruments are not detected by Faster R-CNN in those images, they are re-labeled as the class of their second highest score proposed by Resnet101.
- Second case: Images predicted as *dyed-lifted-polyps, dyed-resection-margins, ulcerative colitis* by Resnet101 are fed forward through Faster R-CNN network to detect instruments. They are classified as *instruments* if detected or keep the original prediction otherwise.

3.2.3 Second run. Feeding forward a large number of images in the three classes through Faster R-CNN causes a bottle-neck of inference time, as Faster R-CNN has high time complexity. Therefore, in this second run, we limited the images passed through Faster R-CNN by only performing the first case of the first run.

3.2.4 Third run. The configuration of the third run is as same as the second run. Instead of using the original training set mentioned in the first run, we train our model on the re-labeled development set combined with the augmented instrument set.

3.2.5 Forth run. In this run, we reduce the number of images used for training by selecting randomly 75% images of each class in the same training set as the third run. Other processing steps are also configured in the same way.

3.2.6 Fifth run. Throughout our experiments, *normal-z-line* and *esophagitis* are the top most confusing classes not only for Resnet101 but also for human to distinguish them. In the priority list, *esophagitis* has a higher rank than *normal-z-line's*. Thus, after several times evaluating our model on the development dataset, we propose a condition for these two classes when they are predicted by Resnet101. As Resnet101 provides a probability distribution over the 16 classes for each image, whenever the *normal-z-line* appears to be the highest class, we add a small bias 0.3 to the probability of the *esophagitis*. Hence, the model is more likely to emit the *esophagitis* class. This intuitively means that our model prefers *esophagitis* to *normal-z-line* when it is confused between these classes.

4 RESULTS

Table 1: Official evaluation result for both sub-tasks (provided by the organizers) and speed (fps) on Tesla K80 GPU

RunID	PREC	REC	ACC	F1	MCC	RK	FPS
Run01	94.245	94.245	99.281	94.245	93.861	93.590	6.589
Run02	93.959	93.959	99.245	93.959	93.556	93.273	**23.191**
Run03	**94.600**	**94.600**	**99.325**	**94.600**	**94.240**	**93.987**	23.148
Run04	93.043	93.043	99.130	93.043	92.579	92.257	22.654
Run05	94.508	94.508	99.314	94.508	94.142	93.884	21.413

There is a trade-off between speed and accuracy when comparing the result of Run01 and Run02. In Run02, we reduce a large number of images passing through Faster R-CNN for the sake of time, so its performance seems to be relatively worse than Run01's.

As we mentioned earlier in section 3, data pre-processing takes an important role in building a deep-neural network model. Through our experiments, in the case of less training data, the augmented dataset helps us improve the performance of deep-neural network model. Run03 and Run05 show impressive results comparing to

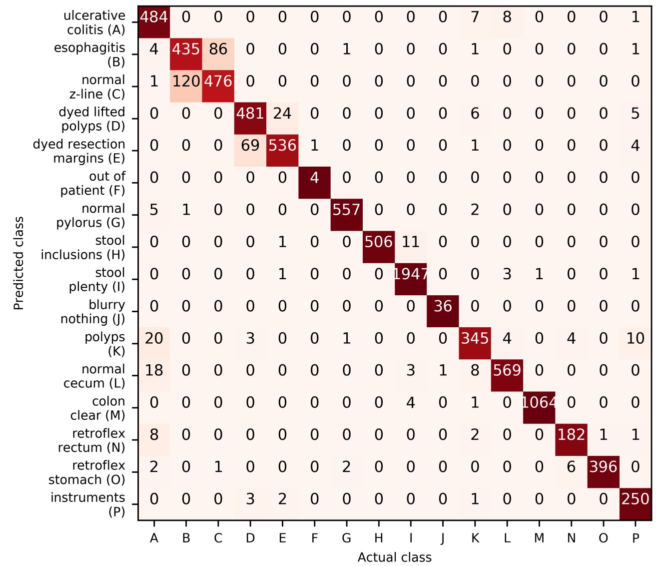

Figure 1: Confusion matrix of our best run - *Run03*

the first two runs. This implies that training on our re-labeled development set provides better models.

On the other hand, using the Residual neural network cannot classify efficiently the two classes *esophagitis* and *normal-z-line*. The same problem also occurs between the *dyed-resection-margins* and *dyed-lifted-polyps* classes. It can be observed in the confusion matrices of the two pairs (Figure 1). Therefore, these are the two main reasons which mainly bring negative impact to our results.

Additionally, as we mentioned in section 3, the configuration of Run05 intuitively prefers *esophagitis* to *normal-z-line*, which may leads to an increasing of the false-positive cases in the result.

By comparison to the others, Run04 has the lowest precision since it uses 75% of training data. Decreasing the amount of training samples of course affects the performance in deep-learning models. Nevertheless, the result is still acceptable when it decreases only a few percentages and its configuration is as same as Run03. This is an evidence that we are even able to reduce up to 50% of data when the less training time is preferred over the accuracy.

5 CONCLUSION AND FUTURE WORKS

Medico image classification is a challenging problem because of the fine-grained images, less training data and require high accuracy. In our current approach, we focus on training a combination of Residual Neural Network and Faster R-CNN with different modifications of the training set. Additionally, object detection method is applied to detect small symptoms of diseases, which are useful evidences for the classification task. Accuracy and inference time that we reach is acceptable and appropriate for real-time constraint. However, for future works, we need a more robust approach to exploit the distinction between easy-confused classes, e.g, *esophgitis* and *normal-z-line*, or *dyed-lifted-polyps* and *dyed-resection-margins*.

REFERENCES

[1] Xinlei Chen and Abhinav Gupta. 2017. An Implementation of Faster RCNN with Study for Region Sampling. *arXiv preprint arXiv:1702.02138* (2017).

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Jun 2016). https://doi.org/10.1109/cvpr.2016.90

[3] Xiao Jia and Max Q.-H. Meng. 2016. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2016). https://doi.org/10.1109/embc.2016.7590783

[4] PËŽal Halvorsen Thomas de Lange Kristin Ranheim Randel Duc-Tien Dang-Nguyen Mathias Lux Olga Ostroukhova Konstantin Pogorelov, Michael Riegler. 2018. Medico Multimedia Task at MediaEval 2018. *Media Eval' 2018* (2018).

[5] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17)*. ACM, New York, NY, USA, 164–169. https://doi.org/10.1145/3083187.3083212

[6] Adnan Qayyum, Syed Anwar, Muhammad Majid, Muhammad Awais, and Majdi Alnowami. 2017. Medical Image Analysis using Convolutional Neural Networks: A Review. 42 (09 2017).

[7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., 91–99.

[8] Aoyama Kazuharu Nishikawa Yoshitaka Miura Motoi Shinagawa Takahide Takiyama Hirotoshi Tanimoto Tetsuya Ishihara Soichiro Matsuo Keigo Tada Tomohiro Shichijo Satoki, Nomura Shuhei. 2017. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. *EBioMedicine* 25 (01 Nov 2017), 106–111. https://doi.org/10.1016/j.ebiom.2017.10.014

[9] Gregor Urban, Priyam Tripathi, Talal Alkayali, Mohit Mittal, Farid Jalali, William Karnes, and Pierre Baldi. 2018. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* 155, 4 (2018). https://doi.org/10.1053/j.gastro.2018.06.037