Inception-ResNet [2] for Detecting Cardiomegaly on Chest X-rays

于治維
Department of Biomedical Engineering of
National Cheng Kung University
Tainan, Taiwan, ROC
c54076055@gmail.com

張荷昕
F94071021@gs.ncku.edu.tw

林雨農
F94076102@gs.ncku.edu.tw

卓庭萱
C44071053@gs.ncku.edu.tw

陳柏霖
F94071186@gs.ncku.edu.tw

1. Introduction of the paper

1.1. Motivation

The researchers in this paper [2] would like to know the result between different versions of Inception networks. Also, the previous studies showed the benefits of combining Inception architectures with residual connections.

So, the researchers want to put these implementations into practice and evaluate the performance of all models.

1.2. Objective

In this report, the researchers will compare the pure Inception variants, Inception-v3 and v4, with similarly expensive hybrid Inception-ResNet versions.

They also try more uniform simplified architecture and more inception modules.

1.3. Contribution of the paper

1. Compare different versions of Inception.

2. Change the first convolution layer of residual connection to 1×1.

3. Scaling down the residuals stop the network from dying when the number of filters exceeded 1000.

2. System framework

2.1. Deep learning architecture

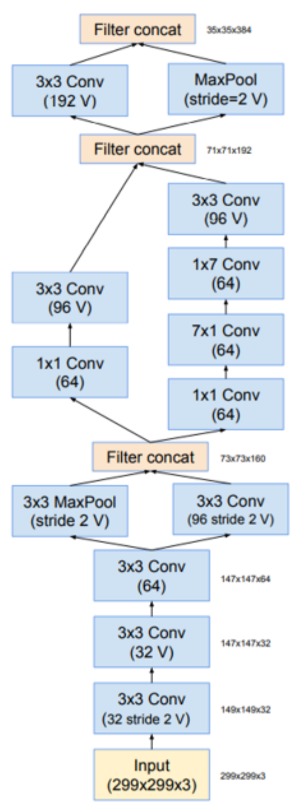

Figure 3. The schema for stem of the pure Inception-v4 and Inception-ResNet-v2 networks. This is the input part of those networks. Cf. Figures 9 and 15

Figure 1. Stem structure

The structure we used is an implementation of the paper and it is combined with the structure above. [2]

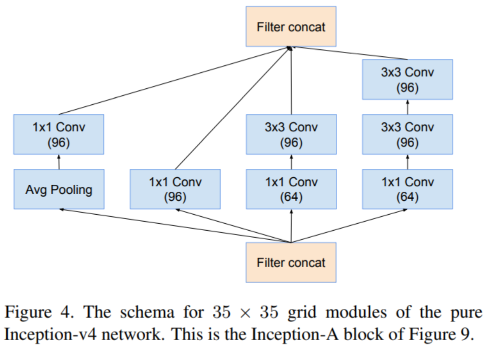

Figure 2. 35*35 grid structure

2.2. Dataset

The dataset we used is "NIH Chest X-ray Dataset" [3] the version which is public on Kaggle.

There are 112,120 X-Ray images in PNG format over 30,000 patients and classified into 15 categories (14 disease & normal). In this report, we are aimed to detect do the subject in the X-ray image got cardiomegaly or not.

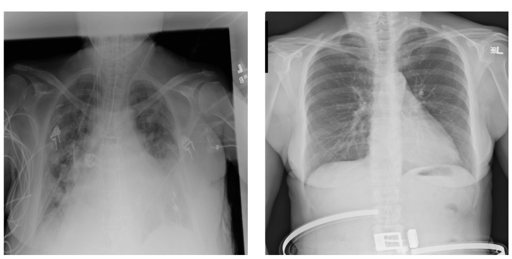

Figure 3. The image above shows the difference between cardiomegaly and normal subjects, the image on the right-hand side is a normal subject, and on the left-hand side is a subject with cardiomegaly

2.3. Training process

When we try to implement the model, we found out that the result wasn't so good. We try to add a batch normalize layer into the structure, even the paper didn't do this. And we see a huge improvement after doing it. But it also gets costly on both memory and training time.

3. Result analysis

3.1. Result

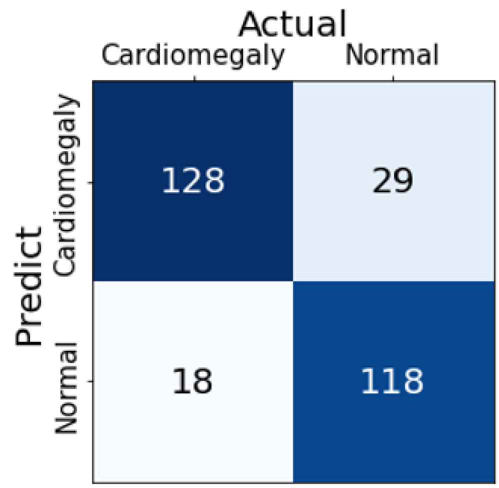

Figure 4. Confusion matrix of the result

In the result, we got 0.89 in sensitivity, 0.80 in specificity, and 0.85 in accuracy. We will try to explain what happened in the model.

3.2. Why correct?

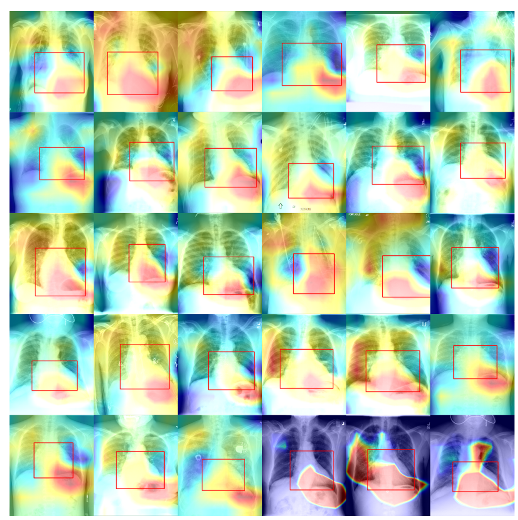

Figure 5. true positive

We implement the grad-CAM in this paper [1] , and apply it to our model. What we see is the attention hot spot when the model is predicting is around the bounding box, which makes it do the right classification.

3.3. Why wrong?

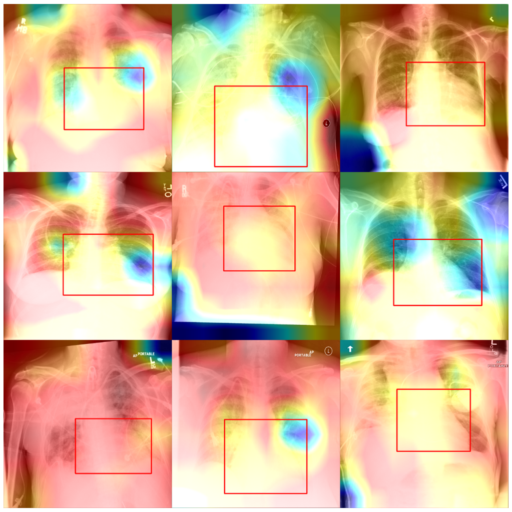

Figure 6. false negative

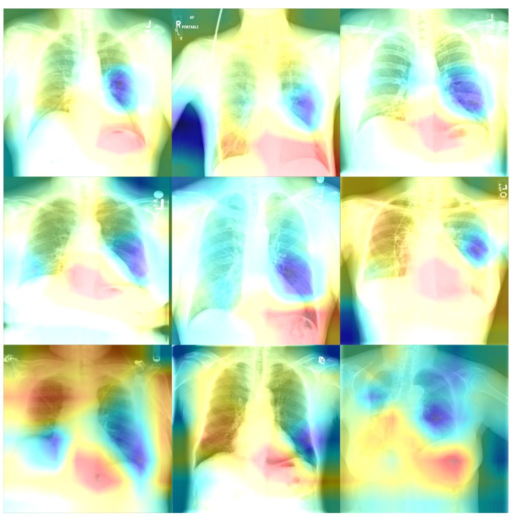

Figure 7. false positive

The model doesn't seem to focus on the right point. It can be seen more clearly in the false positive condition.

References

[1] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, pages 618–626, 2017. 2

[2] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence, 2017. 1

[3] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3462–3471, 2017. 2