COVID-19 Radiography Analysis by DL methods

Peng Deng*, **Chaoyue Duan***, **Haoyu Kang***, **Yuxuan Wang***, **Mingyu Xu***
School of Data Science
The Chinese University of Hong Kong, Shenzhen
{pengdeng,chaoyueduan,haoyukang,yuxuanwang,mingyuxu}@link.cuhk.edu.cn

Abstract

With the development of big data technology in the information age, artificial intelligence has been used in more and more fields. Among them, the application of artificial intelligence or deep learning in the medical imaging industry has broad prospects and potential. In this article, we design a deep learning-based COVID-19 diagnosis system, which aims to help doctors make better diagnosis. Firstly, we train an image classifiers based on the image recognition network ResNet18, pre-trained ResNet18, CBAM conbined ResNet18 respectively. The dataset we used is the "COVID-19 Radiography Database" dataset, which contains 3616 COVID-19 positive cases and 10,192 Normal cases. The dataset was split into training dataset and validation dataset with ratio 7:3. The validation accuracy are 87.86% (vanilla ResNet18), 89.29% (pre-trained ResNet18), 88.78% (CBAM-ResNet18). Next, we used Grad-CAM to handle explicability, which can figure out the parts by which the network classified the image. The result of Grad-CAM shows that CBAM really helps. Then, we use U-Net and DeepLabV3 to do segmentation work, and the validation mean IOU score are 94.9% (U-Net) and 93% (DeepLabV3). Although DeepLabV3's mean IOU is lower, it can get a much more clean prediction result. Finally, we designed a software to facilitate the use of doctors. Users only need to load a COVID-19 radiography image, then the software will give the corresponding diagnosis results.

1 Introduction

Since January 2020, COVID-19 has swept the world, bringing a huge disaster to the health and economy of the people of the world, our group hopes to make a certain contribution to the diagnosis of the disease based on the deep learning knowledge. We use the Kaggle public dataset (COVID-19 Radiography Database[1]), this is a database of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images. We used this dataset to do image classification and image semantic segmentation respectively. For image classification, we used ResNet18[9] combined with Convolutional Block Attention Module (CBAM[22]) method to improve the effect of traditional ResNet. The results show that the new method can be more effective identified the X-ray chest radiographs of patients with confirmed COVID-19, and we leveraged Grad-CAM to enhance the applicability of the model. At the same time, we have performed image semantic segmentation tasks on the existing chest radiographs of patients with COVID-19 pneumonia and other types of pneumonia patients. We used the traditional U-Net model and the DeepLabV3 model for comparison. In model deployment, we deploy the classification model and segmentation model on a software. It can help doctors easily get access to use our model on diagnosing COVID-19.

*Equal contribution

Preprint. Under review.

2 Related Work

2.1 Image Classification

Traditional image classification algorithms generally include several stages, such as feature extraction, feature encoding, spatial constraints and image classification. But with the development of artificial intelligence, the model of image classification task has transitioned from traditional algorithms to deep learning algorithms.

The CNN model proposed by Alex Krizhevsky[11] in ILSVRC in 2012 achieved a historic breakthrough, the results greatly surpassed the traditional method, and won the ILSVRC2012 championship. The model is called AlexNet. This is also the first time deep learning has been used in large-scale image classification. Since AlexNet, a series of CNN models have emerged, which continuously refresh the results on ImageNet[7]. As the model becomes deeper and more sophisticated, the error rate of Top-5 is getting lower and lower. The model proposed by the VGG (Visual Geometry Group) group of Oxford University in ILSVRC in 2014 is called the VGG model[18]. Compared with previous models, this model further widens and deepens the network structure. Its core is five groups of convolution operations, and Max-Pooling space dimensionality reduction is performed between each two groups. The convolution is followed by two fully connected layers, followed by a classification layer. The ILSVRC image classification top-5 error rate of VGG model is 7.3%. Also in 2014, GoogLeNet[20] won the ILSVRC championship with top-5 error rate 6.7%, and the GoogLeNet model consists of multiple groups of Inception modules. In addition, at the end of the network, the traditional multi-layer fully connected layer is not used, but the mean pooling layer is used like the NIN (Network in Network)[13] network. In 2015, the ResNet[9] launched by Kaiming He swept all the players in ISLVRC and COCO and won the championship. ResNet has made great innovations in the network structure, rather than simply stacking layers. ResNet's new ideas in convolutional neural networks are definitely a milestone in the development of deep learning. ResNet successfully trained a 152-layer deep neural network by using Residual Unit, and won the championship in the ILSVRC2015 competition, with a top-5 error rate as 3.57%, while the number of parameters is much lower than that of VGGNet. The problem of degradation in deep networks is successfully solved. On the same ImageNet data set, the recognition error rate of the human eye is about 5.1%, that is, the recognition ability of the current deep learning model has surpassed that of the human eye.

2.2 Medical Image Recognition

There is a big difference between the medical image analysis of different imaging principles and the natural image analysis in the field of computer vision. So far, domestic and foreign scholars have carried out a series of deep learning research work mainly on medical image analysis tasks with different imaging principles such as MRI, CT, X-ray, ultrasound, PET, and pathological optical microscopy. Therefore, this section mainly briefly describes the analysis of these types of medical images.

Medical image classification can be divided into image screening and target or lesion classification. Image screening is one of the earliest applications of deep learning in the field of medical image analysis. It refers to taking one or more examination images as input, predicting it through a trained model, and outputting a disease or severity. Graded diagnostic variables. Image screening is an image-level classification, and deep learning models used to solve this task initially focused on SAE, DBN, and DBM networks and unsupervised pre-training methods. Research has mainly focused on the analysis of neuroimaging, such as the diagnosis of Alzheimer's disease (AD) or mild cognitive impairment (MCI) by neuroimaging. These algorithms usually use multimodal images as input to extract complementary feature information in modalities such as MRI and PET. At present, CNN is gradually becoming a standard technique in image screening and classification, and its application is very extensive. For example, Arevalo et al proposed a feature learning framework for breast cancer diagnosis, using CNN to automatically learn discriminative features to classify mammogram lesions [3]. Kooi et al. compared manual design and automatic CNN feature extraction methods in traditional CAD, both of which were trained on a large dataset of about 45,000 mammogram images, and the results showed that CNN outperformed traditional CAD system methods at low sensitivity, and in Both are equivalent at high sensitivity [10]. Spampinato et al applied deep CNN to automatically assess skeletal bone age [19]. In addition, there are some works that combine CNN

and RNN. For example, Gao et al. used CNN to extract low-level local feature information in slit lamp images, and combined RNN to further extract high-level features to classify nuclear cataract [8].

2.3 Image segmentation

Image segmentation is a key topic in image processing and computer vision with applications such as scene understanding, medical image analysis, robotic perception, video surveillance, augmented reality, and image compression, among many others. Various algorithms for image segmentation have been developed in the literature. The tasks processed in the field of medical image segmentation are of the following types: liver and liver-tumor segmentation, brain and brain-tumor segmentation, optic disc segmentation, cell segmentation, lung segmentation and pulmonary nodules[12]. Recently, due to the success of deep learning models in a wide range of vision applications, there has been a substantial amount of works aimed at developing image segmentation approaches using deep learning models. Currently commonly used deep learning models are 1) Fully convolutional networks 2) Convolutional models with graphical models 3) Multi-scale and pyramid network based models 4) Dilated convolutional models and DeepLab family. Long et al.[15] proposed one of the first deep learning works for semantic image segmentation, using a fully convolutional network (FCN). An FCN includes only convolutional layers, which enables it to take an image of arbitrary size and produce a segmentation map of the same size. However, FCN ignores potentially useful scene-level semantic context. To integrate more context, several approaches incorporate probabilistic graphical models, such as Conditional Random Fields (CRFs) and Markov Random Field (MRFs), into DL architectures. Chen et al.[4] proposed a semantic segmentation algorithm based on the combination of CNNs and fully connected CRFs. Next are multi-scale and pyramid network based models. Multi-scale analysis, a rather old idea in image processing, has been deployed in various neural network architectures. One of the most prominent models of this sort is the Feature Pyramid Network (FPN) proposed by Lin et al.[14], which was developed mainly for object detection but was then also applied to segmentation. Moreover, Dilated convolution (a.k.a. atrous convolution) introduces another parameter to convolutional layers, the dilation rate. Some of most important include the DeepLab family[5], multiscale context aggregation[23], dense upsampling convolution and hybrid dilatedconvolution (DUC-HDC)[21].

3 Methodology

3.1 ResNet

As the depth of the network increases, there may be a degradation problem, that is, as the depth of the network increases, the accuracy reaches saturation, and the depth continues to increase, resulting in a rapid decline in accuracy. The degradation problem shows that not all network models can be easily optimized. Kaiming[9] argues that at least deep networks should not perform worse than shallow networks, the reason is that we can set up a network that stacks identity mapping layers in a shallow network to increase the network depth. These identity mapping layers Just keep the input and output the same. The residual block is as Figure 1. So the author designed a residual learning

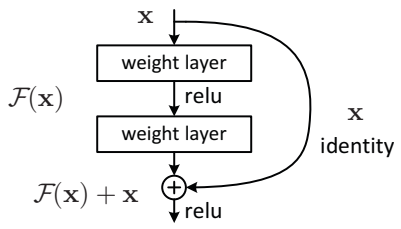


Figure 1: Residual learning: a building block.

module, assuming that the basic mapping of data expectations is $H(x)$, and let the stacked nonlinear layers fit the residual mapping $F(x) = H(x) - x$ instead of directly fitting $H(x)$.

Through experiments, it is found that the deep residual network is easier to optimize than the ordinary network. As the depth of the network increases, the performance of the residual network will also improve and achieve much better results than the previous network.

3.2 Pre-train for CNN

Deep convolutional neural networks revolutionized computer vision arguably due to the discovery that feature representations learned on a pre-training task can transfer useful information to target tasks. In recent years, a well-established paradigm has been to pre-train models using large-scale data (e.g., ImageNet`https://image-net.org/`) and then to fine-tune the models on target tasks that often have less training data. Pre-training has enabled state-of-the-art results on many tasks, including object detection, image segmentation, and action recognition. In the field medical image identify, as the problem data scarcity, pre-train can be useful to improve the performance.

3.3 Convolutional Block Attention Module

Attention is a method that focus on the area that matters for the identify task going on. This is helpful for medical image classification cause the lesion area of the image is relatively small scale as the whole image. Model will understand which part to focus on through attention module. In our experiment, we explore the CBAM[22] module represents Convelutional Block Attention Module. The module including two parts, a channel base attention module and a spacial base attention module.

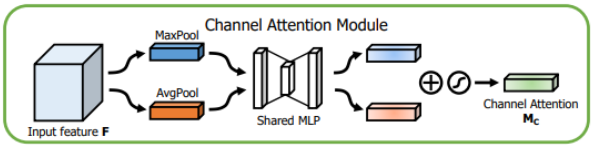


Figure 2: Channel base attention block

The CAM part pass the input feature map F ($H \times W \times C$) through global max pooling and global average pooling based on width and height respectively, to obtain two $1 \times 1 \times C$ features Figure, and then send them into a two-layer neural network (MLP), the number of neurons in the first layer is C/r (r is the reduction rate), the activation function is `Relu`. And then go through a two-layer neural network which is shared. Then, the features output by the `MLP` are added based on element-wise, and then the `sigmoid` activation operation is performed to generate the final channel attention feature, that is, M_c. Finally, the element-wise multiplication operation is performed on M_c and the input feature map F to generate the input features required by the Spatial attention module.

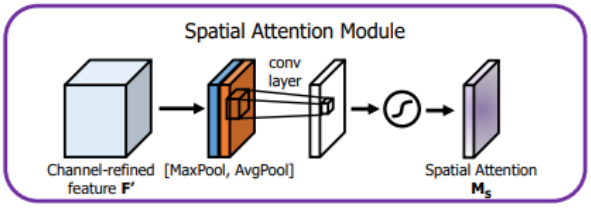


Figure 3: Spatial base attention block

For the SAM part, first do a channel-based global max pooling and global average pooling to obtain two $H \times W \times 1$ feature maps, and then perform the concat operation (channel splicing) based on the channel. Then after a 7×7 convolution (7×7 is better than 3×3) operation, the dimension is reduced to 1 channel, that is, $H \times W \times 1$. Then through `sigmoid` to generate spatial attention feature, namely M_s. Finally, the feature is multiplied by the input feature of the module to obtain the final generated feature.

3.4 Grad-CAM

For commonly used deep learning networks (such as CNN), it is generally believed that a black box is not very interpretable. We do not know why it makes such predictions and where it focuses. For the classification task of medical images, if we can visualize the attention regions where the network makes decisions, it will help doctors make better diagnoses. Here we use grad-CAM[17] technology.

First, the network performs forward propagation to obtain the feature layer A (generally refers to the output of the last convolutional layer) and the network prediction value y (note that this refers to the value before softmax activation). Suppose we want to see the network's region of interest for the "Tiger Cat" category, and the network's predicted value for the "Tiger Cat" category is y^c. Then back-propagate y^c to get the gradient information A' on the feature layer A. The importance of each channel of the feature layer A is obtained by calculation, and then the weighted summation is done through ReLU, and the final result is Grad-CAM.

Grad-CAM can be summed up with the following formula:

$$L^c_{\text{Grad}-\text{CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{1}$$

In the formula, A represents a feature layer, which generally refers to the feature layer output by the last convolutional layer in the paper. k represents the k^{th} channel in the feature layer A. c stands for category c. A^k represents the data of channel k in feature layer A. α^c_k represents the weight for A^k. The formula for calculating α^c_k is as follows:

$$\alpha^c_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{2}$$

Among them, y^c represents the score predicted by the network for category c, note that there is no softmax activation here. A^k_{ij} represents the data of feature layer A in channel k, at the coordinate position ij. Z is equal to the width times the height of the feature layer.

3.5 U-Net

The U-Net[16] network structure is a U-shaped structure. The left half is the Encoder and the right half is the Decoder, which is a milestone turning point in medical image segmentation based on deep learning. The architecture of U-Net we used in this work can be seen as Figure 4[2], which is a little different from the original paper.

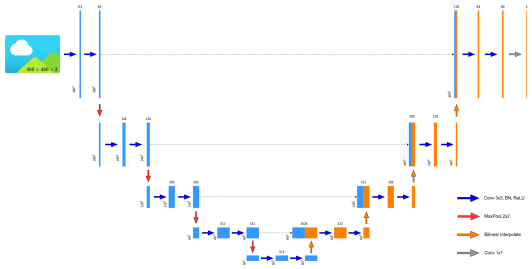


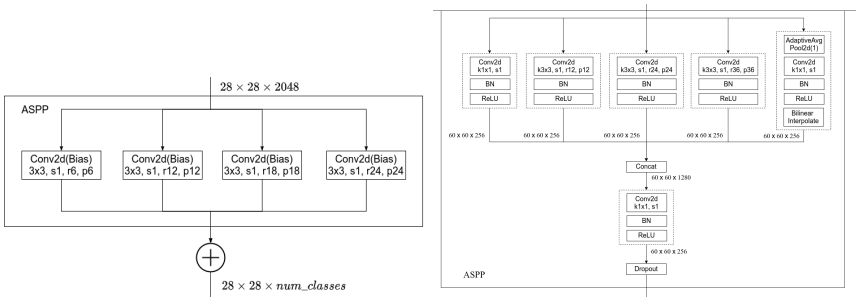
Figure 4: The architecture of U-Net

3.6 DeepLab[4, 5, 6]

DeepLabV1: There are two technical hurdles in the application of DCNNs to image segmentation tasks: signal downsampling, and spatial "insensitivity"(invariance). For the first problem, the author use dilated convolution, and for the second problem, they use use fully-connected CRF(Conditional Random Field). Compared with the previous segmentation network, DeepLabv1 is faster and more accurate, and the model is very simple, mainly composed of DCNN and CRF cascade.

DeepLabV2: DeepLabV2 is an upgrade of V1, its backbone changes from VGG16 to ResNet and it still use dilated convolution. Besides, the model use ASPP(atrous spatial pyramid pooling) to deal with multi-scale and it changes fully-connected CRF to fully-connected pairwise CRF.

DeepLabV3: Compared with DeepLabV2, DeepLabV3 has three changes: 1) Multi-grid is introduced, 2) ASPP structure is improved, and 3) CRFs post-processing is removed.



(a) (b)

Figure 5: The modification to ASPP[2]

4 Experiments

4.1 Dataset

The data set we used is "COVID-19 Radiography Database"[1], which is a public database. The database including chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images and corresponding lung masks. Figure 6 is an example plot of some samples in the dataset.

(a) The COVID image (b) The Normal image (c) The ground truth

Figure 6: The dataset

As of now, the database contains 3616 COVID-19 positive cases along with 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection), and 1345 Viral Pneumonia images and corresponding lung masks. Obviously, the number of chest X-rays of normal people is much larger than that of COVID-19 patients, which is also in line with common sense. We will only use COVID-19 positive cases and Normal cases in this work.

4.2 Results

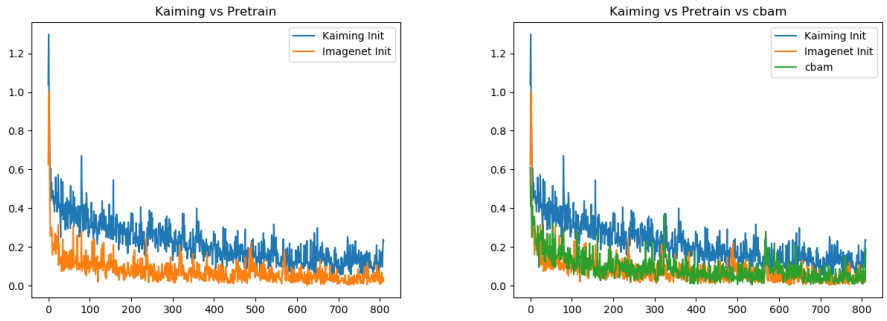
4.2.1 Classification

Pretrained Residual Network: In this part of experiment, we mainly focus on explore how pretrain benifits the model. Our baseline model is ResNet18 due to the dataset we use is relatively simple. The baseline cold start model using Kaiming initialization for Conv parameters and random

initialization for Linear layers. The pre-trained models are download from `https://download.pytorch.org/models/resnet*.pth` which is pre-trained from ImageNet. We set the origin input size 224×224, the learning rate 10^{-3}, weight decay 10^{-5} and bachsize 60, training epoch 10. We can see form Figure 7(a) that that the pre-trained model result in a faster speed to converge and remain a lower loss than Kaiming initialized model after epochs. Also the best accuracy jumps from 87.86% to 89.29%. In conclusion, pre-train parameters give a better initialization for training as it saved common experience from ImageNet.

CBAM: In the CBAM experiment, we mainly focus on explore the how the CBAM module affects our model. Our baseline model is also Resnet18. we did not put the CBAM block directly in the residual architecture, instead after the first 7×7 Conv block and the final output of the residual block for the efficient of the model. For connecting the module and ResNet, we use max pooling and average pooling.

We set the origin input size 224×224, the learning rate 10^{-3}, weight decay 10^{-5} and bachsize 60, training epochs 10. We can see from Figure 7(b) that with the attention method the model convergence rate is much similar to the pretrained model. Also the model performance best accuracy is 88.87% which is a little lower than the pretrained model. But we can further explore the attention performance using Grad-CAM method to see if the CBAM can really increase the ability to focus more on the abnormal area to help medical diagnose.



(a) Training loss (b) Training loss

Figure 7: Classification training results

4.2.2 Grad-CAM

As we mentioned before, Grad-CAM can visualize the attention regions where the network makes decisions, and it will help doctors make better diagnoses. Besides, CBAM can help model understand which part to focus on through attention module. As a result, we will compare the Grad-CAM visualization result of vanilla ResNet18 and CBAM combined ResNet18. The result is as Figrue 8. As we can see, CBAM exactly help the model to pay attention to the correct parts of the images to do prediction, as a result, we can get a more interpretable and accurate visualization result by Grad-CAM.

4.2.3 Segmentation

Implementation: We use U-Net and DeepLabV3 to do medical image segmentation work respectively and then compare their performance. In U-Net, we use dice loss and the cycling learning rate (LR) scheduler is implemented. For both model, we use the same data augmentation methods. The data augmentation includes `RandomResize`, `RandomHorizontalFlip`, `RandomVerticalFlip`, `RandomCrop`, `Normalize`. For detail, please refer to `tranforms.py`.

Training result: Due to the large dataset and our limited computing resources, we only trained one epoch for each method. For evaluation, we use mean IOU as the metrics. As we can see from Figure 9, both U-Net and DeepLabV3 can achieve over 90% validation mean IOU, U-Net achieves 94.9% and DeepLabV3 achieves 93%. It seems that the U-Net model is more powerful from the mean IOU

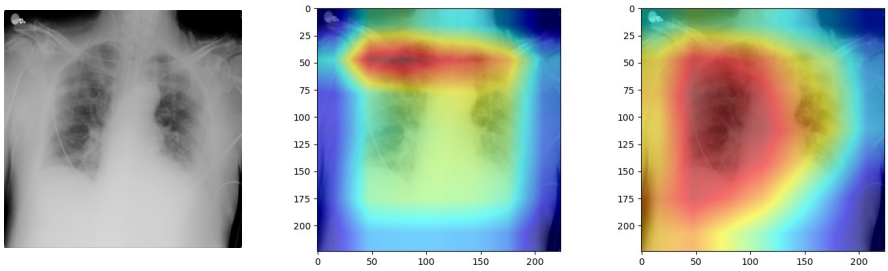
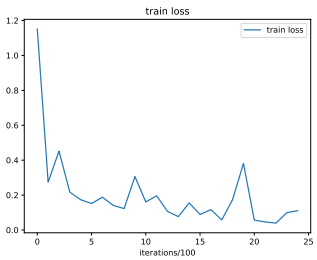
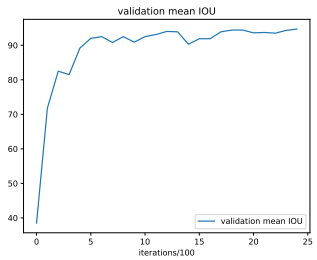


Figure 8: The grad-CAM visualization result; original image(left), ResNet18(mid), ResNet18-CBAM(right)

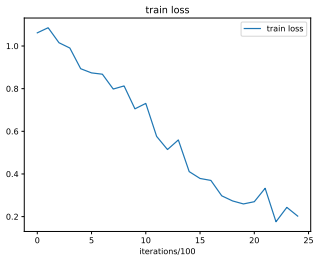
score, more specifically, we should refer the prediction result to see which method is better in this task.Moreover, we can see that U-Net can converge more quickly but is not stable, which means the higher mean IOU score can be unstable.



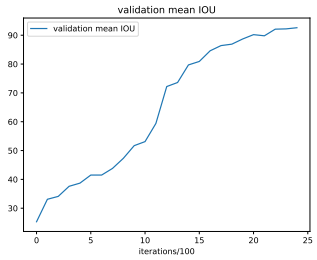
(a) Training loss (U-Net)



(b) Validation mean IOU (U-Net)



(c) Training loss (DeepLab)



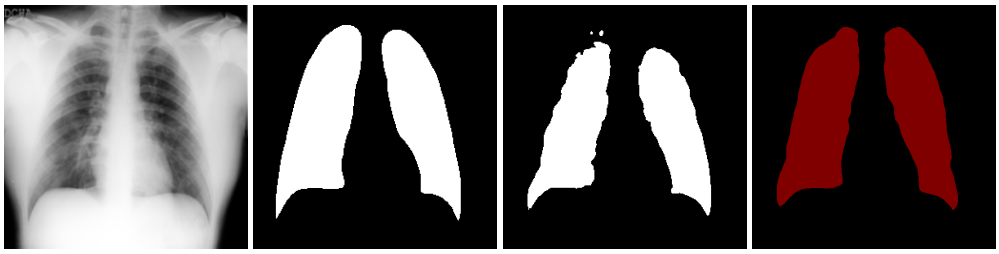
(d) Validation mean IOU (DeepLab)

Figure 9: Segmentation training results

Prediction result: We choose file `"COVID-16.png"` in the validation dataset as the original image to show the prediction results of our model. The result is as Figure 10. As we can see, the prediction result of DeepLabV3 is much more clean and accurate than the result of U-Net, despite its mean IOU score is lower than U-Net. Thus, in the software phase, we will choose DeepLabV3 as our backend segmentation model.

4.3 Software

We build a software to help intuitively determine whether getting covid-19 from chest X-ray images, and to help us better discover which areas have potential pathological conditions. Besides, we show you segmented area of pathology. The whole software can be partitioned into two blocks: The frontend and backend. The former is development by tool called PYQT5 which is a set of frameworks for Python binding Digia QT5 applications. PyQt5 is a Python language implementation based on the graphical programming framework Qt5, which consists of a set of Python modules. It is also a



(a) The original image (b) The ground truth (c) Result of U-Net (d) Reuslt of DeepLab

Figure 10: The prediction results

cross-platform toolkit that can run on all major operating systems including UNIX, Windows, Mac OS. pyqt5 is dual licensed. The modules of PYQT5 we apply contains: QtGui, QtCore, QtWidgets, QFileDialog. Then we connect front-end and back-end. We trigger the operation of back-end mainly by clicking the button like Select the image, Recognition, Lung Segmentation. As for procedure of how back-end works, it contains image preprocessing like converting and scaling, function of model prediction, grad-cam and segmentation, and sent back final pictures to front-end. The front-end interface is displayed as Figure 11 in the Appendix.

5 Conclusion

In this work, we mainly do image classification and image segmentation to "COVID-19 Radiography Database" dataset. In the classification part, we tried ResNet18 and pre-trained ResNet18, besides, we also used CBAM to help the model to pay attention to the correct parts. We used Grad-CAM to do visualization work to help doctors make better diagnosis. In the segmentation parts, we tried U-Net and DeepLabV3, and the result shows that DeepLabV3 has a much better clean segmentation result. Finally, we integrated all the modules of our work into a software to facilitate the use of doctors which can help doctors make better diagnosis.

Appendix

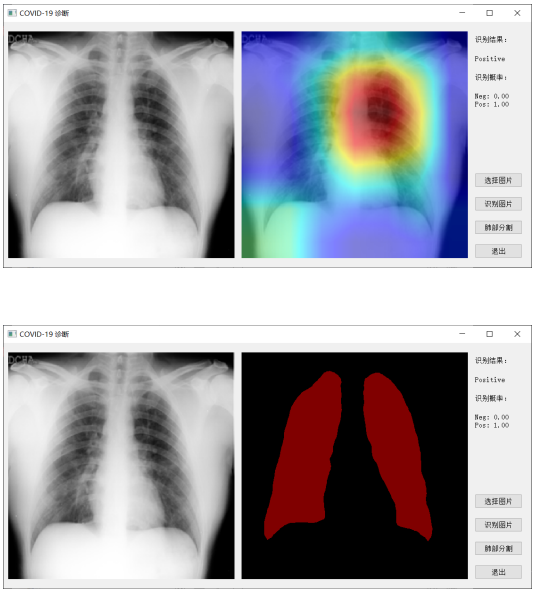


Figure 11: The UI of the software

References

[1] https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/metadata.

[2] https://github.com/WZMIAOMIAO/deep-learning-for-image-processing.

[3] John Arevalo, Fabio A. González, Raúl Ramos-Pollán, José Luís Oliveira, and Miguel Ángel Guevara-López. Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer methods and programs in biomedicine*, 127:248–57, 2016.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[6] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[8] Xinting Gao, Steve Lin, and Tien Yin Wong. Automatic feature learning to grade nuclear cataracts based on deep learning. In *Asian Conference on Computer Vision (ACCV)*, November 2014.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Thijs Kooi, Geert Litjens, Bram van Ginneken, Albert Gubern-Mérida, Clara I Sánchez, Ritse Mann, Ard den Heeten, and Nico Karssemeijer. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, 35:303312, January 2017.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[12] Tao Lei, Risheng Wang, Yong Wan, Xiaogang Du, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. 2020.

[13] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[14] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[17] 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.

[18] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[19] C Spampinato, S Palazzo, D Giordano, M Aldinucci, and R Leonardi. Deep learning for automated skeletal bone age assessment in x-ray images. *Medical image analysis*, 36:4151, February 2017.

[20] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[21] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. Ieee, 2018.

[22] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[23] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.