一般口演 | 画像情報・生体信号処理
一般口演11
画像情報・生体信号処理
2019年11月23日(土) 09:00 〜 11:00 E会場 (国際会議場 3階中会議室301)

[3-E-1-04] 胸部単純レントゲン写真の肺野のセグメンテーション

○王 博文[1]（1. 大阪大学大学院医学系研究科医療情報）

キーワード：deeplearning, segementation, x-ray

胸部単純レントゲン写真（ ChestXp）は、高頻度で撮られるが、多くの臓器が重なって映し出されるため、その読影はむしろ難しい。一方、放射線科医は、撮像される画像検査数に対して相対的に不足である。深層学習の技術を応用して ChestXpの自動診断ができると、一般医の読影による見落とし等が防止できる可能性がある。深層学習を応用するためには、大量の学習データが必要となるが、過去画像に対して異常陰影に印を付ける方法で学習データを作成することは放射線科医の負担となり、律速段階となる。そこで、 ChestXpの診断レポートから、Information Extractionの方法で、部位と所見の種類を抽出し、これを教師として自動診断のための学習する方向で本課題に取り組む方針とした。
本研究は、レポートから抽出した部位に相当する領域をセグメンテーションが可能かを肺の領域について調べることを目的としている。
セグメンテーションタスクのために、深層学習モデル U-netを使用した。セグメンテーションの領域設定は、CITECの規格に従って作成し、左右肺の肺尖部、上肺野、中肺野、下肺野の8つの部分に分けた。画像データは、DICOM形式から256 x 256サイズの PNGに圧縮変換した。250枚の学習および評価用のデータを手作業で作成した。学習データについて4倍のデータ増強を行った。
テストの結果、全体ピクセル精度は0.951であった。肺尖部、上肺野、中肺野、下肺野の IOU(Intersection over Union)値はそれぞれ0.866、0.872、0.846、0.812であり、かなりの精度でセグメンテーションが可能であることが確認できた。
本法のセグメンテーションで得られた mask情報に従って、オリジナル画像から高解像度のデータを取得し、これを識別 CNNモデルのデータとして、次の学習ステップに進める計画である。

Regional Segmentation in Chest Radiographs with Fully Convolutional Architectures

Bowen Wang[1] Yasushi Matsumura[1] Toshihiro Takeda[1] Shirou Manabe[1]
Kento Sugimoto[1]Syo-ya Wada[1] Asuka Yamahata[1]
Department of Medical Information, Osaka University Hospital[1]

Abstract: The deep learning method has achieved good results in the diagnosis of chest X-ray (CXR). Existing research is mostly based on the Chest X-ray 14 Dataset. It contains the CXR images with multi-label for the whole image, but the type and specific location of the disease is also recorded in the report. For using such data, regional segmentation for CXR is necessary. In order to be consistent with the records in the report, we divide the lung into four areas. And then, a full convolution network used to achieve semantic segmentation of lung regions. Our best performing model, trained with the loss function based on the Focal Loss, reached mean intersection-over-union (IoU) of 0.850.

Keywords: Lung regional segmentation, fully convolutional network, chest X-ray

---

1. Introduction

Chest X-ray is a basic procedure in radiology for lung disease prediction and diagnosis. However, such diagnosis is very time consuming and not objective. By the development of AI, deep learning has demonstrated remarkable strength in a variety tasks of medical image analyses. With the recent availability of the Chest X-ray 14 Dataset [1], many deep learning approaches have been proposed to automatically diagnose the thoracic diseases in CXR images. But these studies do not give good results for diseases which do not have obvious symptom (e. g., nodule). Wang et al. [1] evaluated several deep convolutional neural network (DCNN) architectures, reporting an area under the ROC curve (AUC) of 0.75 on average. In his work, the high resolution CXR image is compressed before it can be used as a model input. Some other studies [2,3] provide higher resolution input for the model by segmentation or other methods. Through their results, higher resolution does help improve the accuracy of prediction.

In X-ray 14 Dataset, every CXR image has its multi-label which tells the kinds of disease for the whole image. This means that we have to use an entire map as input of the model. However, CXR has a very high resolution and image compression will lose some information. In fact, there is a more detailed record in the report. In general, CXR report records disease information according to the lung field. (e. g., “There is a nodule in the left upper lung”) This shows that the disease in the report is recorded by type and field of the lung. The definition of lung field is in accordance with CITEC [4] standards. You can see the relevant content in Sugimoto et al.’s [5] research. If we segment each part according to the field, DCNN will have a higher resolution input. Therefore, we want to make a model for regional segmentation in CXR image.

Because there is no public dataset that is segmented according to the fields of the lung. In this paper, we made a dataset names Regional Segmentation in Chest Radiographs (RSCR) basically according to CITEC standards. The whole lung is divided into four parts: the tip lung field, the upper lung field, the middle lung field and the lower lung field (see Fig. 1). And then, we used a full convolutional network structure DU-Net for the segmentation. DU-Net has the structure similar to U-Net and used the dense block which proposed in DenseNet [6]. Since the number of pixels belonging to each category of the four lung fields is not uniform

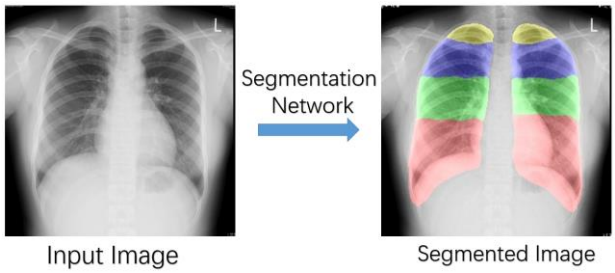

Fig. 1 Segmentation process: from input CXR to mask of four lung fields. Tip lung field, the upper lung field, the middle lung field and the lower lung field were given the color yellow, blue, green, red respectively.

(especially for background), and we also want to focus on the samples which are difficult to train, Focal Loss [7] was used as loss function in our work. We use the mean IOU and Pixel Accuracy (PA) to evaluate the results. We also implemented some comparative experiment on the frequently used methods. Finally, we also designed a method for obtaining the lung fields from the high-resolution image based on the predicted results. The segmented image will be facilitated to the subsequent research.

2. Related Work

For segmentation of medical images, Greenspan et al. [8] made a summary of the recent state-of-the-art works in the field. The semantic segmentation typically builds upon a vast set of training data (e.g., Pascal VOC-2012 [9]). Such large datasets are not typical for the medical domain. This means most current approaches unfeasible, thus calling for a finely tailored strategy.

Long et al. [10] introduced the concept of the Fully Convolutional Net (FCN). Different from the classic CNN that using the convolutional layer to obtain fixed-length feature vectors for classification, the FCN can accept input images of any size, and use the deconvolution layer to restore the feature map of the last convolutional layer. Upsampling it to the same size as the input image produces a prediction for each pixel ,preserving the spatial information of the original input image at the same time. Pixel-by-pixel classification is performed on the upsampled feature map to solve the semantic segmentation problem. Ronneberger et al. [11] guide the resolution of feature extraction, and thereby control the relationship of features. The proposed U-Net consists of contraction and expansion parts: in the contraction part abstract features are extracted by consecutive application of pairs of convolutional and pooling layers, while in the expansion part the upsampling abstract features are merged with the features from the contractive part. U-Net is often used in medical image segmentation tasks.

SJSRT (Japanese Society of Radiological Technology) image set often used for study of CXR segmentation. It contains manual segmentations for total lung fields, heart and clavicles. Alexey et al. [12] reached mean Jaccard overlap scores of 95.0% for lungs, 86.8% for clavicles and 88.2% for heart by using the JSRT. Although his research is multi-class, he did not follow the CITEC format for the lung field. Wang et al. [13] used U-Net to get the segmentation of total lung and made a CNN model to determine whether CXR image is normal or not. His experiments prove that using the segmented CXR image can improve the accuracy of the judgment. Different from their approach, we want to make a more specific segmentation in the lung field.

Sugimoto et al. used Natural Language Processing (NLP) to obtain disease information from CXR reports. In his research, the information in the report was extracted and formed into the structure according to the CITEC. In order to use such data, we design a regional segmentation model for the CXR image.

3. Materials and Method

1 RSCR

Our data set for the division of the lung field, basically in accordance with the standard of CITEC.

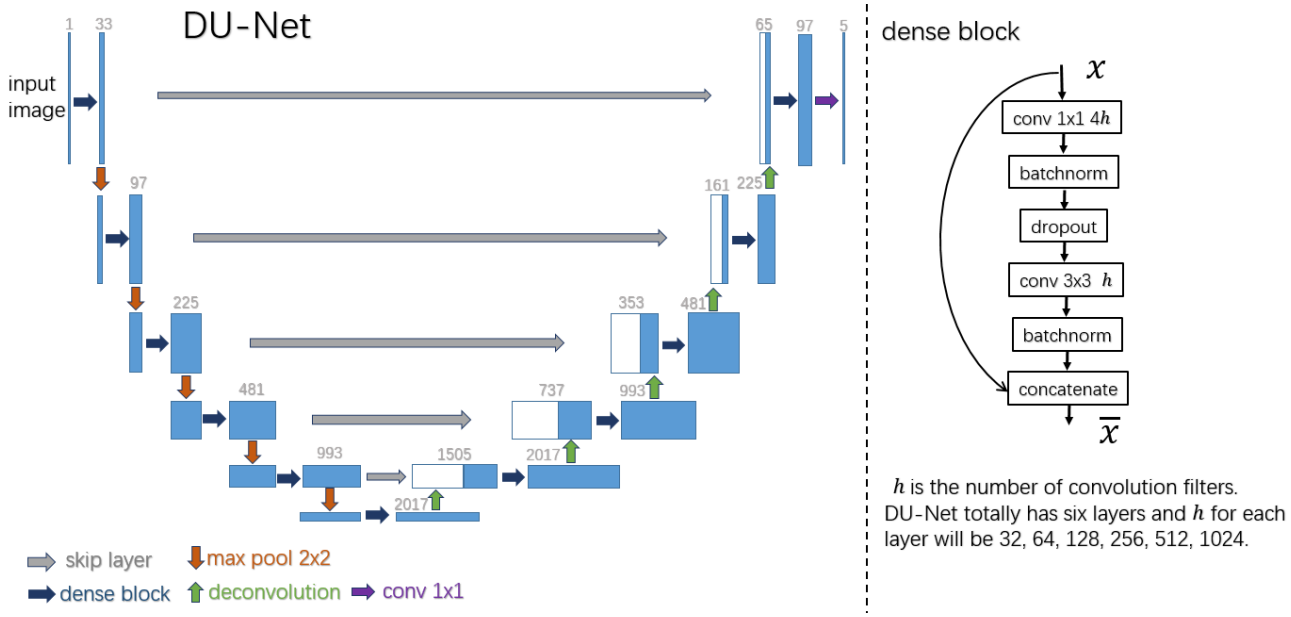

Fig.2 Structure of DU-Net

However, only the tip lung field, the upper lung field, the middle lung field and the lower lung field were recorded in the study by Sugimoto et al., and there is no separate record for the clavicle. So we divided the clavicle into the area of the upper lung field. For the part where the organs overlap, when making the mask, restore the shape of the lungs as far as the naked eye is visible (there may be lesions in these areas). In addition, pacemaker will be considered as the background in this dataset. We used Labelme [14] as the data annotation software. When labeling, we classified left and right upper lung fields as the same class, and the other lung fields have the same practice. PSCR contains 250 anonymized CXR images taken from Osaka University Hospital. The dicom format images were converted to PNG format and retained the original resolution. All labels are manually made and converted to binary image format. 80% of the images are used for training, 10% of images are used for validation, and 10% of images are used for test.

2 DU-Net

U-Net is often used in the study of CXR semantic segmentation. In this paper, we used the improved DU-Net. The structure is shown in Figure 2.

We used a six-layer U-shaped structure that would allow for higher resolution input. Skip layer in the original U-Net helps upsampling to combines the various layers of information in downsampling, in order to restore detail information and progressively restore image accuracy. In addition to retaining skip layer, DU-Net also replaced the original convolution layer with the dense block. It could superimpose the feature map of current layer with the previous layer and pass it to the next layer, which increases the fitting ability of the model. In the process of upsampling, we used deconvolution (start from the bottom layer, the number of filters for deconvolution will be 512, 256, 128, 64, 32). Besides, relu combine with batchnorm were added after convolution, and L2 regularization is used on the global parameters.

3 Loss Function

In the semantic segmentation study of the lung, cross entropy or dice is often used as a loss function. But our target is to divide the lung into multi-class. Considering the number of pixels belonging to each class is not balanced, we used Focal Loss in the model. It has the function as follow for multi-class task.

$$L_{focal\_loss}(y, y_{hat}) = -\frac{1}{n}$$

$$\sum_{i=1}^{n} y_{hat_i} \times \alpha_i \times (1 - softmax(y_i))^{\gamma} \log(softmax(y_i))$$

γ makes it possible to reduce the loss of easily categorized samples. In other words, it will focus on the difficult and misclassified samples. ($\gamma = 2$ is recommended in the paper). In addition, the balance factor α is added to balance the uneven proportion of the different class samples. This need to be artificially given.

4 Segmented Image Acquisition

In this section, we used the prediction results of the model to obtain the segmentation mask image for each lung field from the original resolution CXR. It was done as the process of Fig. 3.

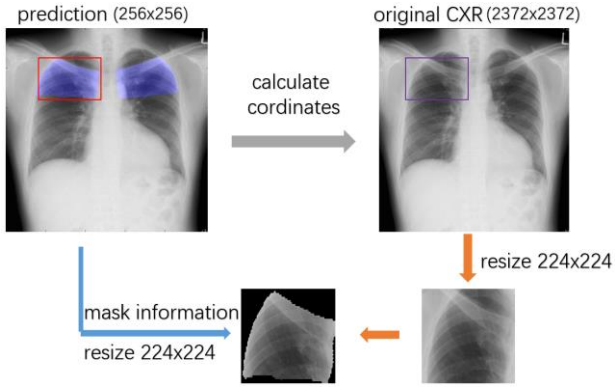

Fig. 3 Process of segmented image acquisition.

Here we use the upper right lung to illustrate. First we calculated the smallest bounding box which contain all the pixels predicted as right upper lung. And then, project the bounding box onto the original CXR image proportionally based on the resolution information. We used the projected coordinates to get the higher resolution segmentation images from original CXR. Finally, the mask information of the prediction result was merged with the segmentation image to obtain the outcome.

5 Training Details

The proposed algorithm was implemented using Keras [14] with Tensorflow [15] backend in Python 3.6. Input data were resized into two kinds of resolution 256x256 and 384x384. They also have the augmentation with a random translation maximum range of 30 and 50 pixel for two different size. We set the epoch with 300 iterations (batch size is 8) and totally have 20 epoches. We adopted the Adam and set the base learning rate to 0.001, and then reduced it a factor of ten at 10 and 15 epoch. All weights were randomly initialized, and the regularization coefficient is 1e-4. For Focal Loss, we used recommended $\gamma = 2$, and α for background 0.5, for other classes 1. Our code is available in Github[17].

4. Results

1 Multi-class Segmentation

In the experiment we also used other models for comparative experiments. Different loss functions and different sized of inputs were adopted. Mean IOU and Pixel Accuracy were used to evaluate the outcome. You can see the results from Table 1.

For FCN, original U-Net and SegNet, we used cross entropy as the loss function. All methods were implemented with the same epoch and learning rate. It can be seen from the table that DU-Net using focal loss has obtained the best results. From the IOU results for each class, the forecast score of the upper lung field is the highest. This may be due to the clearer outline of the upper lung field, while other parts are not easy to distinguish. The tip lung field has the clavicle as the dividing line with the upper lung field, so it has the second best result. However, the judgment in the middle lung field is rather vague.

Table 1

Method	tip	upper	middle	lower	mean IOU	pixel accuracy
FCN[10]	0.757	0.816	0.681	0.772	0.757	0.903
U-Net[11]	0.825	0.830	0.811	0.833	0.826	0.937
SegNet[13]	0.799	0.806	0.785	0.781	0.792	0.920
DU-Net+dice	0.762	0.782	0.679	0.761	0.746	0.893
DU-Net+Cross Entropy	0.856	0.871	0.820	0.838	0.846	0.947
DU-Net+Focal Loss	0.852	0.875	0.828	0.846	0.850	0.951
DU-Net+Focal Loss (384x384)	0.847	0.869	0.815	0.836	0.842	0.943

We also plotted the pixel accuracy variation of the DU-Net on the validation set during the entire training process for different loss functions and input sizes. The study speed of Focal Loss is lower than cross entropy at the beginning, but it has a higher precision in the end. It seems that higher size input did not work in this experiment. Dice as a loss function did not get good result. See Fig.4

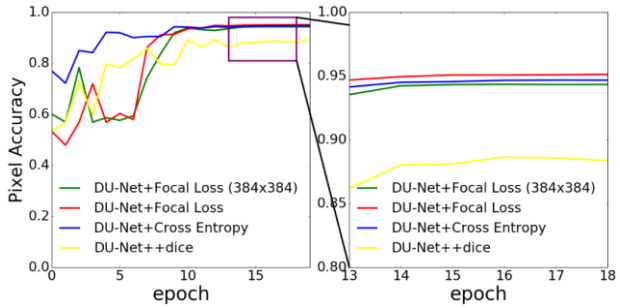

Fig.4 Pixel accuracy in validation set when training.

2 Segmented Image Acquisition

According to the mask information, we separated each lung field from the original resolution CXR. The segmented image is unified to a resolution of 224x224, which is generally the input of the classification model. The outcome of segmentation is shown in Fig.5.

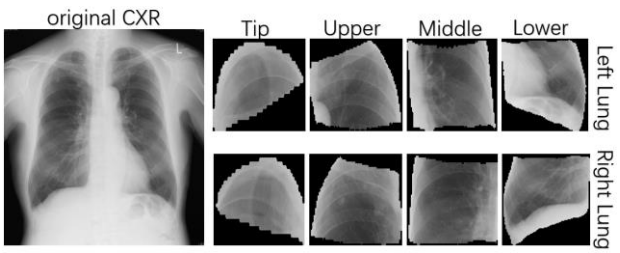

Fig.5 Segmented lung fields

5. Conclusion

In this study, we have designed an end-to-end regional segmentation method for lung field based on deep learning. We introduced and evaluated our model with some common methods. We also compared different loss functions to prove that focal loss worked best in this experiment. We believe that with more training data and transfer learnt features from other CXR-related segmentation tasks, it is possible to further improve the scores.

Then, we used the model prediction to get the mask information, and segmented each lung field from the original high-resolution CXR image. The obtained lung field segmentation image is consistent with the records in the report. We will use these data in the future classification tasks.

References

[1] Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[2] Yao L, Prosky J, Poblenz E, et al. Weakly supervised medical diagnosis and localization from multiple resolutions[J]. arXiv preprint arXiv:1803.07703, 2018.

[3] Liu H, Wang L, Nan Y, et al. SDFN: Segmentation-based Deep Fusion Network for Thoracic Disease Classification in Chest X-ray Images[J]. Computerized Medical Imaging and Graphics, 2019.

[4] http://citec.kenkyuukai.jp/special/index.asp?id=25698

[5] Sugimoto et al. Extracting structured data from diagnostic imaging reports Proceedings of the 38th Medical Informatics Conference. 2018; 718-20.

[6] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected

convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[7] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[8] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1153–1159, May 2016.

[9] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” arXiv e-prints, arXiv:1412.7062v4 [cs.CV], vol.

[10] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks forsemantic segmentation,” in Computer Vision and Pattern Recognition(CVPR 2015), 2015.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional Networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI 2015). Springer, 2015, pp. 234–241.

[12] Novikov, Alexey A., et al. "Fully convolutional architectures for multiclass segmentation in chest radiographs." IEEE transactions on medical imaging 37.8 (2018): 1865-1876.

[13] Wang C, Elazab A, Jia F, et al. Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder[J]. Biomedical engineering online, 2018, 17(1): 63.

[13] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.

[14] http://labelme.csail.mit.edu/Release3.0/index.php

[15] https://github.com/fchollet/keras

[16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin,S. Ghemawat, G. Irving, M. Isardet al., “Tensorflow: A system for large-scale machine learning,” in12th{USENIX}Symposium on OperatingSystems Design and Implementation ({OSDI}16), 2016, pp. 265–283.

[17]https://github.com/wbw520/DU-Net.git