II-FCN FOR SKIN LESION ANALYSIS TOWARDS MELANOMA DETECTION

Hongdiao Wen

School of Electronic Engineering/Center for Robotics
University of Electronic Science and Technology of China (UESTC), Chengdu, China

ABSTRACT

Dermoscopy image detection stays a tough task due to the weak distinguishable property of the object. Although the deep convolution neural network significantly boosted the performance on prevalent computer vision tasks in recent years, there remains a room to explore more robust and precise models to the problem of low contrast image segmentation. Towards the challenge of Lesion Segmentation in ISBI 2017, we built a symmetrical identity inception fully convolution network which is based on only 10 reversible inception blocks, every block composed of four convolution branches with combination of different layer depth and kernel size to extract sundry semantic features. Then we proposed an approximate loss function for Jaccard index metrics to train our model. To overcome the drawbacks of traditional convolution, we adopted the dilation convolution and conditional random field method to rectify our segmentation. We also introduced multiple ways to prevent the problem of overfitting. The experimental results shows that our model achieved Jaccard index of 0.82 and kept learning from epoch to epoch.

Index Terms — Dermoscopy image detection, lesion segmentation, Inception network, fully convolution network, dilation convolution, conditional random field

1. INTRODUCTION

Melanoma, also known as malignant melanoma, is a type of cancer that develops from the pigment-containing cells known as melanocytes. Melanomas typically occur in the skin but may rarely occur in the mouth, intestines, or eye. In women they most commonly occur on the legs, while in men they are most common on the back. Sometimes they develop from a mole with concerning changes including an increase in size, irregular edges, change in color, itchiness, or skin breakdown. Melanoma is the most dangerous type of skin cancer. Globally, in 2012, it occurred in 232,000 people and resulted in 55,000 deaths[1].

Dermoscopy image detection is helpful to assist the diagnosis and prevention of skin lesion like melanoma. However, the melanoma in dermoscopy image is diversiry in color, contrast, and shape, even human beings would have a hard time to recognize. To improve the accuracy and efficiency in melanoma detection, automatically segmentation of melanoma is highly evaluated for medical usage.

Some of former efforts for segmentation is based on hand-crafted features[2, 3]. These hand-crafted features methods are limited and unstable due to the weak non-linear representation ability. In recent years, many works toward object segmentation like melanoma has been based on deep convolution network[4]. Benefiting from the excellent ability to extract abstract features from original image, the model based on dcn is more likely to work well in melanoma segmentation. However, more deep the network, more large should the dataset, otherwise it would take the risk of overfitting.
He, Kaiming, et al [5] proposed the residual network which overcomed overfitting problems of deep network. The residual scheme should be deep enough to acheive complicated non-linearity. The shortcut in residual block is to carry message of features at lower layer to higher layer, if the number of residual blocks is less, then the output would be much influenced by features in lower layer. While many segmentation model are based on fully convolution networks [6] which made up of encoder and decoder, this makes the residual block too heavy for segmentation. Meanwhile, the inception block [12] intergrated with different layer depths and kernel sizes, which is also able to convey diversity semantics to next block. In this paper, we formed a identity inception block, which can fuse different convolution branches together without any zero padding. Our model is basically made up of 10 this kind of blocks, the total size of our model is comparably smaller.

In order to train a deep model with millions of parameters, there exists many models took image augmentation as a basic way to overcome overfitting [11, 12]. We specified our dynamic dataset augmentation method in detail in part 2.1. This made our model feed with new image content from epoch to epoch. Besides, a loss function normalization was adopted to balance the sample number of pixels. Then, we proposed a approximate loss function for jaccard index metrics in addition.

As described in [7], the dilation convolution is best suit for semantic segmentation benefiting from the ability to cover more large receptive field. In our model, we add several dilation convolution at last to fill the black holes. Then we adopted the conditionnal random field [8, 9, 10] to drop the small objects and rectify the boundary.

In this paper, we propose a original model named II-FCN with several training methods and conditional random field technique to segmentate the melanoma from dermoscopy image. Meanwhile, we introduced a identity inception block which can be reversible to build a symmetrical FCN. We proposed an loss function for better training our model. Note, our model size is more small so as to bring a light solution compared with formally FCRN model [13]. The contributions of our work can be summarized in the following:

(1) We proposed a identity inception block with four valid padding convolution branches combined the ability to extract fusion abstract features and stable the training process. Different from residual block, our implement of identity inception block enjoying non-linear feature learning without identity mapping and can be applied with the FCN model with less layer.

(2) Besides, we embedded the dilation convolution at the last of our II-FCN network, which is designed to remove the black holes in object. In order to adjust the boundary according to the texture of original image, we applied the conditional random field as a post processing method to generate the final result.

(3) We also introduced several dataset augmentation policiies and training tricks with our II-FCN model to improve the generalization of our model. The experiments shows that this strategy is helpful to reduce the difference of data distribution between training set and test set. We believe this is generasively effective to many other semantic segmentation tasks.

The remaining content of this paper is organised as follows. We introduced our methods in Section 2. Experiments and results in Section 3. Concludes at Section 4.

2. METHODS

2.1. Dynamic Dataset Augmentation

As introduced above, due to the weak objectiveness of melanoma and nonobvious boundary between melanoma and the background, the melanoma segmentation images are difficult to label. The dataset in this challenge contains only 2 thousand training images, the distribution of training set and test set is highly different. We thought that dataset augmentation is essential to train a model with millions of parameters. We specifies our data augmentation strategies as follows:

(1) First, the original image has possibility of 0.2 to adjust contrast, color, apply gaussian blur and histogram equalization respectively.

(2) Then, the processed image from (1) has 0.2 possibility to flip left and right, flip up and down and rotate a random degree respectively.

(3) Last, the processed image from (2) has possibility of 0.5 to zoom in a random area.

The corresponding segmentation performed the same operation as original image. To conclude, our dataset augmentation method is able to generate different state of original image from epoch to epoch.

2.2. II-FCN Model

In order to build a strong and robust model to better extract features from original image, we propose the Identity Inception block which is able to combine branches with different layer depth and kernel size to generate equal sized feature maps without zero padding. The Identity Inception block can be taken as a basic feature map transform, the forward version was used to make up encoder and backward form is used to make up decoder. Note, this block architecture is a specially designed form of inception block. The details are as Fig.1.

Taking the advantage of indentity inception block as previously described, the symmetrical II-FCN is proposed. Ranging from the numbers of the Identity Inception block, the II-FCN can be adapted to II-FCN8, II-FCN16, II-FCN32 and so on. Our experiments are all based on II-FCN32. We build bridges between forward Identity Inception block and backward Identity Inception block to prevent loss of texture and stable training. The architecture is specified in Fig.2.
Fig. 2. We constructed 5 forward identity inception block and 5 backward identity inception block in our II-FCN segmentation model. The shortcut took a residual scheme to keep texture details. The dilated convolution at last is to drop black holes. The sample segmentation is taken from validation set, best viewed on the color sensitiveness.

2.3. Loss Normalization

Although our dynamic dataset augmentation increased image diversity greatly, the random crop and zooming methods would cause more unbalanced sample. Here, we adopted a reweighting or normalizing method. This makes our segmentation boundary more distinguishable. Though our normalizing technichs, the loss curve could drop from random guessing value of binary classification, which can be used to evaluate whether the model has worked in early time. We formulated our methods as follows.

For a segmentation image \( s \), we write the pixel number of object as \( n \), we created a corresponding reweighting filter \( f \). The \( N \) can be computed through

\[
n = \frac{\sum_{i=1}^{h} \sum_{j=1}^{w} (s[i][j])}{255} \tag{1}
\]

The percentage of object \( p \) is represented as

\[
p = \frac{n}{w \times h} \tag{2}
\]

where \( w \) and \( h \) is the image width and height respectively.

Then we filled the corresponding percent of object and background in filter \( f \) as follows

\[
f[i][j] = \begin{cases} 
p & \text{if } s[i][j] = 0 \\ 
1 - p & \text{otherwise} 
\end{cases} \tag{3}
\]

At last, we normalized our filter by

\[
f[i][j] = \frac{f[i][j]}{2 \times p \times (1 - p)} \tag{4}
\]

Our filter was used to normalize the final loss and reweight the gradients of back propagation, thus achieved balanced training between melanoma and background.

2.4. Jaccard Index Loss

In this challenge, the evaluation metrics including sensitivity, specificity, pixel accuracy, dice coefficient, and jaccard index. The jaccard index is the final metrics to test the model. We put forward a method to approximate the jaccard index as a loss in our network. We specifies it at bellow.

We took softmax output of an image as \( A \) with probability of being the melanoma, the label as \( B \) with value 1 represent object and 0 represent background. First we computed the corresponding category number

\[
A = \begin{cases} 
\max(A - 0.5, 0) & \text{if } A \neq 0.5 \\
1 & \text{otherwise}
\end{cases} \tag{5}
\]

this maps the probability over 0.5 to label 1 and below 0.5 to label 0.

Then we approximate the jaccard index as the following

\[
J = \begin{cases} 
\lg\left(k \times \frac{\sum A + \sum B - \sum |A - B|}{\sum A + \sum B} + \frac{1}{k}\right) & \text{if } A \neq 0 \text{ and } B \neq 0 \\
0 & \text{otherwise}
\end{cases} \tag{6}
\]

where we took \( k \) as 1.1, then our jaccard index loss is between 0 and 1.004, which is comparable with the binary entropy loss. Then we took entropy loss with jaccard index loss together as our final optimization function.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup

Transfer learning is proven to be an efficient method to ease the training process of deep convolution models, which implies that the feature distribution of a large dataset may cover a small dataset more or less. However, we were going to train from scratch so as to prove the effectiveness of applying a bunch of image augmentation strategies with our II-FCN model.
We appended the multi-scale training method apart from our dynamic dataset augmentation. We started from small scale training with image height and width of 252 and 380, then training at middle scale with image height and width of 444 and 688, finally, training at large scale with image height and width of 636 and 956. We set the batch size of training at different scale at 8, 4, 2, respectively. We used Adam optimizer in all scale with same setting of learning rate = 0.0001, \( \beta_1 = 0.9 \), \( \beta_2 = 0.9999 \), \( \epsilon = 1e^{-8} \). We performed our experiments with a 8GB nvidia GPU k5200. Our model is based on TensorFlow.

### 3.2. Results

We split 200 images from labeled data as validation set. The Fig. 3 shows the jaccard index metrics from epoch 5 to 25. We masked from the softmax output with a probability larger than 0.8 as a baseline. Then we performed conditional random field method to rectify the segmentation. We found that after 10 epoch the jaccard index was above 0.8, the CRF method improved the baseline by a mean of about 0.005.

The figure also shows that our model kept learning from epoch to epoch, this demonstrated the effectiveness of our dataset augmentation method.

![Evaluation of post processing methods](image.png)

**Fig. 3.** Comparison of post processing methods

### 4. CONCLUDES

We proposed an II-FCN model for melanoma segmentation with several training methods, including augmentation, loss normalization, jaccard index loss optimizing, dilation convolution, and CRF processing. Our results showed that our model can automatically segment melanoma at high jaccard index without overfitting.

### 5. REFERENCES

[1] Melanoma, https://en.wikipedia.org/wiki/Melanoma.

[2] Dima, A., Scholz, M., Obermayer, K.: "Automatic Segmentation and Skeletonization of Neurons from Confocal Microscopy Images Based on the 3D Wavelet Transform," TIP. 2002.

[3] Roberto Rigamonti and Vincent Lepetit, "Accurate and Efficient Linear Structure Segmentation by Leveraging Ad Hoc Features with Learned Filters," Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2012.

[4] Wen Li, Fucang Jia, Qingmao Hu, "Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks," Journal of Computer and Communications. 2015.

[5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[6] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

[7] Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." arXiv preprint arXiv:1511.07122 (2015).

[8] Shotton, Jamie, et al. "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation." European conference on computer vision. Springer Berlin Heidelberg, 2006.

[9] Chen, Liang-Chieh, et al. "Semantic image segmentation with deep convolutional nets and fully connected crfs," arXiv preprint arXiv:1412.7062 (2014).

[10] Philipp Krhenbhl and Vladlen Koltun."Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials". NIPS. 2011

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

[12] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[13] Lequan, Y. U., et al. "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks." IEEE Transactions on Medical Imaging (2016).