ISSN: 1813-4890

Fovea detection based on Resnet_vd

Tingyu Li

College of Computer Science and Engineering, Northeastern University, Shenyang, 110819, China

20203465@stu.neu.edu.cn

Abstract

In this paper, we propose a center fovea detection method based on Resnet_vd, aiming to automatically identify and locate the foveal region in retinal images. Resnet_vd is a deep residual network architecture that has shown remarkable performance in image classification and detection tasks. We train the network on a large-scale dataset of retinal images to achieve accurate fovea detection. We conduct experimental evaluations on publicly available datasets and compare the proposed method with traditional approaches. The experimental results demonstrate the superiority of the proposed Resnet_vd-based fovea detection method in terms of accuracy and robustness, exhibiting excellent performance in fovea localization tasks.

Keywords

Fovea detection, Deep learning, Resnet_vd, Retinal images

1. Introduction

1.1. Significance of research

The fovea is an important area of the eye, which has an important impact on eye health. Fovea detection has many medical benefits. First of all, the health status of fovea is closely related to vision. Through fovea detection, we can evaluate whether vision is normal and provide appropriate treatment plan. Secondly, fovea detection is very beneficial for early screening of retinal diseases, which often affect fovea area at the earliest. Regular detection can find signs of diseases in time and take early treatment measures. In addition, fovea detection also plays an important role in making treatment plans. Doctors can track and evaluate the changes of fovea area through detection, so as to guide and adjust treatment plans. Advantages of fovea detection include non-invasiveness, speed, efficiency, accuracy and providing objective data. It is analyzed by automated image processing technology, without surgery or traumatic operation, and provides a comfortable and safe detection method for patients. In addition, fovea detection has high accuracy, which can accurately locate the fovea area, provide doctors with objective data and quantitative indicators, and help evaluate patients' eye health. To sum up, fovea has an important impact on eye health, and fovea detection has a wide range of applications and benefits in ophthalmology, which can help vision assessment, early disease screening and treatment planning, and provide an important auxiliary tool for eye health management.

In the past, traditional fovea detection methods mainly rely on manually designed features and threshold settings, such as color, texture and geometry. However, these traditional methods have some limitations and challenges, including subjectivity, professional requirements and dependence on image quality. With the development of deep learning, fovea detection method based on deep learning has gradually become a research hotspot and made remarkable progress.

At present, the method based on depth learning has achieved remarkable results in image classification, target detection and segmentation. Resnet_vd, as a depth residual network

ISSN: 1813-4890

architecture, performs well in image processing tasks, and has strong feature extraction and representation capabilities. Therefore, the application of Resnet_vd network in fovea detection is expected to improve the detection accuracy and robustness.

The purpose of this paper is to propose a method of fovea detection based on Resnet_vd to realize automatic fovea recognition and location. By using large-scale retinal image data sets for network training, we will explore the potential of Resnet_vd network in fovea detection tasks.

1.2. Solutions

Firstly, the image is preprocessed and normalized to [-0.5-0.5], then three data enhancement methods are used, and then feature extraction is carried out on the enhanced image, and the list of foveal position coordinates of fundus image is obtained. Finally, it is put into the model for training. The optimizer uses Moment and the loss function uses Smooth L1.

1.3. Value

Deep learning is of great value in fovea detection. Fovea detection is a complex image processing task, which involves the location and analysis of the fovea region in fundus images or retinal images. Traditional algorithms often need to design and extract features manually, and do not perform well for complex image background and large changes. Deep learning can automatically learn the feature representation and related information of images through neural networks, which has better adaptability and generalization ability. Using deep learning for fovea detection can improve accuracy and robustness. Deep neural network can learn the representation from the bottom image features to the high-level semantic information by stacking layer by layer, and can automatically extract and learn the most distinctive features. In addition, deep learning can comprehensively consider more context information, improve the robustness to complex image background and noise, and improve the accuracy and stability of detection.

2. Related work

2.1. Data preprocessing

The training set includes 80 fundus images, and the test set includes 20 fundus images, most of which have a resolution of 2992*2000, and a few have a resolution of 1956*1934. Because the sizes of the images are different, the images are normalized to [-0.5, 0.5]. When cutting, the proportion of fundus images is kept and the middle blocks are intercepted, and the coordinate mapping information is reserved to facilitate mapping back to the original coordinates during prediction.

```
h, w, d = img.shape
dx = (w-h)//2
img = img[:, dx:(dx+h), :]
img = cv2.resize(img, (INPUT_IMAGE_SHAPE, INPUT_IMAGE_SHAPE))
img=img.transpose([2,0,1]).astype(np.float32)/255 -0.5
```

2.2. Data Enhancement

Due to the limited sample data, and each fundus image is affected by many factors such as illumination, diet, age, condition, etc., the data show great diversity, which will affect the nature

ISSN: 1813-4890

and quality of fundus images, and thus have a negative impact on the performance of the model. In this experiment, the following three enhancement operations are carried out to increase the generalization ability of the model.

2.2.1 Color jitte

The luminance offset range is [-0.2, 0.2], the contrast offset range is [0.8, 1.2], the saturation offset range is [0.8, 1.2], and the hue offset range is [-0.1, 0.1].

```
self.color_jitter=paddle.vision.transforms.ColorJitter(
brightness=0.2, contrast=0.2, saturation=0.2,
hue=0.1)
```

2.2.2 Horizontal flip

```
if np.random.uniform(0,1)>0.5:
    img[:,::-1,:] = img
    loc = (w-loc[0], loc[1])
```

2.2.3 Vertical flip

```
If np.random.uniform(0,1)>0.5:
   img[::-1,:,:] = img
   loc = (loc[0], h-loc[1])
```

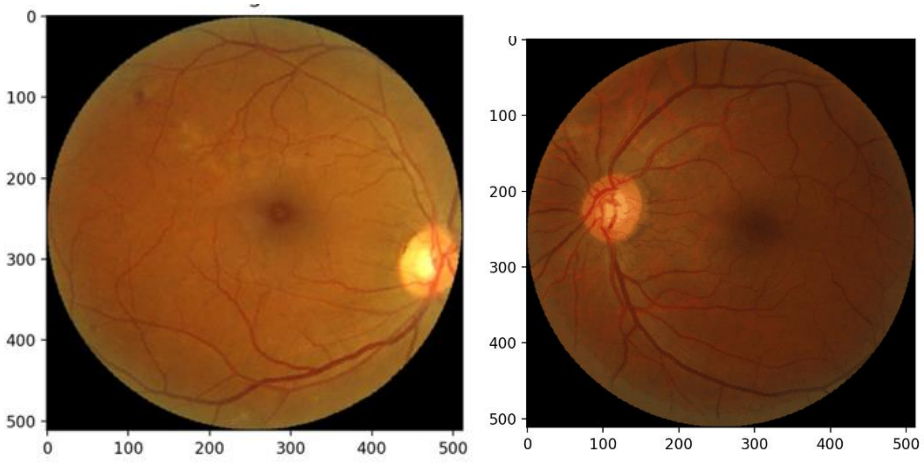

Fig.1. Data Enhancement

2.3. Feature Extraction

Firstly, the pre-trained ResNet50_vd model is selected as the backbone network. Select the feature of the last stage in the output backbone network for subsequent detection tasks. The feature extraction and output part used to detect and locate the fovea in RetinaNet is constructed, which can train and infer the coordinate position task of detecting the fovea.

```
pretrain_url='https://bj.bcebos.com/paddleseg/dygrap
h/resnet50_vd_ssld_v2.tar.gz'
self.backbone=paddleseg.models.backbones.ResNet5
0_vd(pretrained=pretrain_url)
headers=[paddle.nn.Conv2D(self.backbone.feat_chan
nels[i], 2, 1) for i in self.feat_indices]
self.headers=paddle.nn.LayerList(headers)
```

ISSN: 1813-4890

2.4. Calculate concave position coordinates

Firstly, the features of fundus images are extracted, and then the final feature map of the model is used for adaptive average pooling to adjust its size to 1x1. Then, it is converted into a 1-dimensional tensor by a regression header, and the dimension with size 1 is deleted by squeeze operation on each tensor, and finally the list of foveal position coordinates of fundus image is calculated.

```
def forward(self, x):
    feats = self.backbone(x)
    feats = [feats[i] for i in self.feat_indices]
    feats = [paddle.nn.functional.adaptive_avg_pool2d(f,1) for f in feats]
    outs = [paddle.squeeze(header(f), axis=[2,3]) for f, header in zip(feats, self.headers)]
    return outs
```

2.5. Model Testing and Optimization

Maps the predicted values to the specified output range [0.3, 0.7], and then converts the mapped coordinate values back to the coordinate values in the original image coordinate system. Then the measured fundus image is converted into gray image, and then smoothed by Gaussian filtering. Then, a square window is extracted based on the predicted coordinate values in the original image coordinate system and its additional areas, and the point with the smallest gray value is found in the window as the estimation result.

3. Basic theory

3.1. Network model

3.1.1 Resnet_vd

The traditional down-sampling operation in ResNet is realized by using 1 × 1 convolution with step size of 2 in residual structure. The input data will be reduced to half the size of the feature map by 1 × 1 convolution with a step size of 2. 1 × 1 convolution with step size 2 will cause 3/4 of the information of input features to be unused. Therefore, using the structure shown in vd, the down-sampling is moved to 3 × 3 convolution, and the down-sampling in identity mapping is completed by average pooling, which effectively avoids the information loss problem caused by 1 × 1 convolution with step size 2.

Resnet50_vd has good feature extraction ability and spatial alignment ability. In the problem of foveal position detection of fundus image, using Resnet 50_vd as the backbone network can improve the feature representation ability of the model, thus effectively extracting the features of fundus image for foveal coordinate position detection.

ISSN: 1813-4890

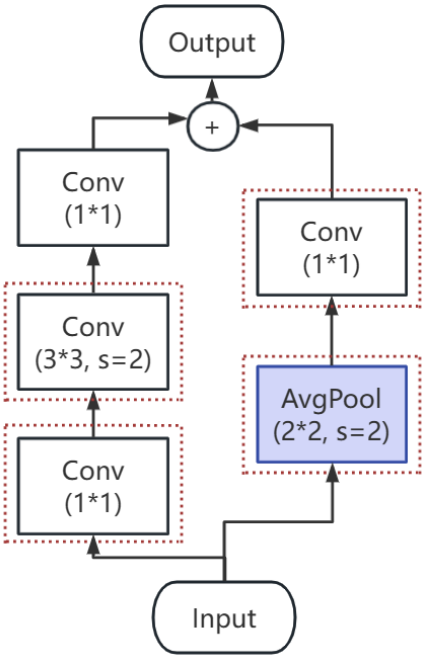

Fig.2. Resnet50_vd

3.1.2 Yolov3

The core idea of traditional YOLO algorithm is to regard target detection as a regression task. The algorithm flow is roughly as follows: the input image is divided into 7x7 grids, and each grid is responsible for predicting the target whose center point is in the grid, the position of the regression center point relative to the grid, the length, width and category of the target.

YoloV3 has three main improvements: (1) using multiple logistic regression classifiers instead of SoftMax classifiers, so that the model can be applied to classification tasks with intersection between categories; (2) Feature pyramid network architecture is introduced, and the deepest feature map is up-sampled twice, which is fused with shallow features respectively. Finally, different anchor points are set on three feature layers to predict targets of different scales; (3) Learning the idea of residual network, Darknet-53 is designed as a new skeleton network.

In the problem of detecting the fovea position of fundus image, the estimation of fovea position is regarded as detecting only one type of target, that is, fovea. Then, for each fundus image, it is intercepted into a local image region containing fovea, which is used as the input of Yolov3 model, and an appropriate target frame size and sample number are set for training and optimizing the model.

3.2. Loss function

Use the smooth_L1_loss function to calculate the error between the label and the predicted value, which is a smooth L1 loss function and is calculated as follows:

$$\text{Smooth L1} = \begin{cases} |x| - 0.5, |x| > 1 \\ 0.5x^2 \ , |x| < 1 \end{cases}$$

L2 loss is between [-1, 1]. To solve the problem that L1 has a break point at 0, and it is L1 loss outside the interval of [-1, 1], and to solve the problem of outlier gradient explosion, the gradient can be limited from the following two aspects:

①When the error between the predicted value and the true value is too large, the gradient value will not be too large.

②When the error between the predicted value and the real value is small, the gradient value is small enough.

ISSN: 1813-4890

The weighted average error value of multi-head output is solved by setting loss weight and customizing loss layer.

```
class MyLoss(paddle.nn.Layer):
    def __init__(self, loss_w=LOSS_WEIGHTS):
        super(MyLoss, self).__init__()
        self.loss_w = loss_w

    def forward(self, preds, label):
        loss = 0
        for pred, w in zip(preds, self.loss_w):
            loss += w*paddle.nn.functional.smooth_l1_loss(pred, label)
        return loss
```

4. Network comparison

4.1. Resnet_vd

Advantages:

It has strong feature extraction ability, and can extract a lot of useful feature information for foveal position estimation while maintaining high classification performance.

Virtual branch down-sampling is adopted, which makes the model less computational and faster in training speed.

Disadvantages:

Due to the lack of fundus image data, Resnet 50_vd model may lack the accuracy of fovea position, and it is easy to have under-fitting or over-fitting problems.

4.2. Yolov3

Advantages:

It has good detection speed and accuracy.

Compared with Resnet 50_vd model, Yolov3 model can directly output the x and y coordinates of the target box if the target box corresponding to the fovea can be obtained.

Disadvantages:

The detection performance is greatly affected by the target selection box. Improper target selection box will lead to the degradation of detection performance, while the features of fovea are small and irregular, and the sensitivity of target detection may be weak.

The training complexity of the model is high, and it needs a large number of samples and fine parameter adjustment to achieve good performance. At the same time, it needs to consider the making of tags and the design of corresponding relations.

4.3. Comparative Selection

This paper attempts to detect the fovea position of fundus image by Yolov3 model, divides the possible area of macular fovea into several grids, first determines which grid the fovea point is in by classification, and then determines the final position by regression. However, the result is not as good as resnet50_vd because the parameters are not set properly, so resnet50_vd is finally selected as the detection model.

5. Parameter tuning

5.1. Learning rate

The learning rate adopts warmup and linear decreasing, and the benchmark learning rate is 1e-3. LinearWarmup () function is used to set up a warm-up interval for PolynomialDecay ()

ISSN: 1813-4890

function in several epoch at the beginning of training, and the learning rate is gradually increased in a linear way, so as to avoid the situation that the learning rate at the beginning of the model is too high, which leads to the failure of convergence or over-fitting of the model. When the learning rate reaches the set maximum value, it begins to decay according to the polynomial decay strategy specified by PolynomialDecay () function until the learning rate reaches the set minimum value. In this way, the training process and learning rate of the model can be better controlled, which is helpful to improve the training speed and performance of the model.

```
lr_scheduler=paddle.optimizer.lr.PolynomialDecay(LR, power=0.9,
decay_steps=batch_per_epoch*TRAIN_EPOCHS,end_lr=0)
lr_warmup=paddle.optimizer.lr.LinearWarmup(lr_scheduler,
batch_per_epoch*WARMUP_EPOCH, 0, LR)
```

6. Experimental results

In the image processing stage, the contrast, saturation and brightness of fundus images are adjusted, hoping to unify the gap between different images and extract fundus features better in the feature extraction stage. However, in the experimental process, it is found that different parameter values have a great influence on fundus images, which is one of the reasons for poor final results.

Try to highlight the dark data in the fundus image, but in the process of data processing, we encounter the problem of image dimension and accuracy explosion. Because the data is not processed well, the final result is not improved.

Insufficient data preprocessing: In the process of data preprocessing, it is necessary to adopt appropriate methods for image enhancement, scaling, cropping and other operations, and improve the training speed while ensuring the data effect.

The final score was 152.6912.

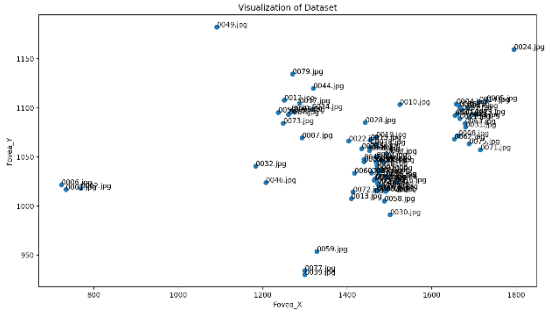

Fig.3. Center concave coordinate distribution in training set

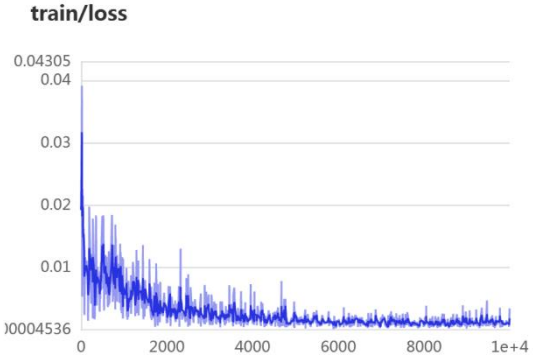

Fig.4. Change of training loss value

ISSN: 1813-4890

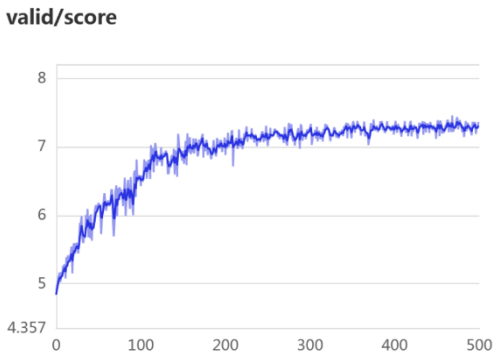

Fig.5. Verify the change of set loss value

References

[1] D · Khosla, R · M · Ullenbroke, Chen Yang, et al.Neuromorphological visual activity classification system and method based on fovea detection and context filtering: CN201980006835.9 [P]. CN111566661A [2023-08-12].

[2] JustincC. Brown, SharonD. Solomon, SusanB. Bressler, et al.Detection of diabetic foveal edema: Comparison of contact lens in vivo microscopy and optical coherence tomography [J]. JAMA Ophthalmology: Chinese Edition, 2004, 16 (4): 6.

[3] Raghu M, Unterthiner T, Kornblith S, et al. Do Vision Transformers See Like Conventional Neural Networks? [C]//2021. DOI: 10.48550/arXiv.2108. 08810.