Seg4Reg Networks for Automated Spinal Curvature Estimation

Yi Lin[1], Hong-Yu Zhou[2(✉)], Kai Ma[2], Xin Yang[1], and Yefeng Zheng[2]

[1] Department of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
[2] Tencent YouTu Lab, Shanghai, China
whuzhouhongyu@gmail.com

Abstract. In this paper, we propose a new pipeline to perform accurate spinal curvature estimation. The framework, named as Seg4Reg, contains two deep neural networks focusing on segmentation and regression, respectively. Based on the results generated by the segmentation model, the regression network directly predicts the cobb angles from segmentation masks. To alleviate the domain shift problem appeared between training and testing sets, we also conduct a domain adaptation module into network structures. Finally, by ensembling the predictions of different models, our method achieves *21.71* SMAPE in the testing set.

Keywords: Spinal curvature estimation · Cobb angle · Deep learning

1 Introduction

Adolescent idiopathic scoliosis (AIS) is the most common form of scoliosis and typically affects children who are at least 10 years old. How to accurately estimate the spinal curvature plays an important role in the treatment planning of AIS. The current clinical standard for AIS assessment relies on doctors' Cobb angle measurement. Such manual intervention process usually makes the operation time-consuming and produces unreliable results. Recently, deep neural networks have got amazing achievements in various image classification tasks. How to apply these deep models to the problem of spinal curvature estimation becomes a hot issue in automated AIS assessment. BoostNet [5] is proposed as a novel framework for automated landmark estimation, which integrates the robust feature extraction capabilities of Convolutional Neural Networks (ConvNet) with statistical methodologies to adapt to the variability in X-ray images. To mitigate the occlusion problem, MVC-Net (Multi-View Correlation Network) [6] and MVE-Net (Multi-View Extrapolation Net) [1] have been developed to make use of features of multi-view X-rays.

Currently, there are two ways to estimate the Cobb angles: (a) predicting landmarks and then angles [5,6] and (b) regressing angle values [1]. The first

The first two authors contributed equally.

Y. Cai et al. (Eds.): CSI 2019, LNCS 11963, pp. 69–74, 2020.
https://doi.org/10.1007/978-3-030-39752-4_7

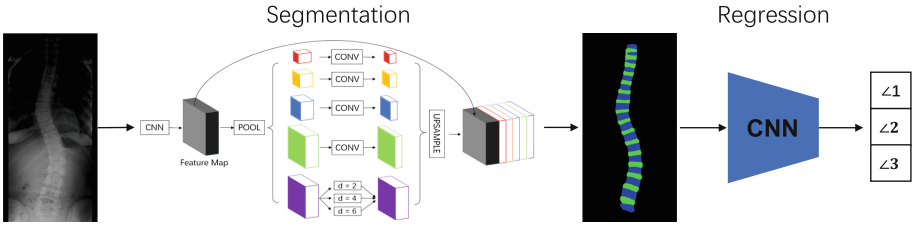


Fig. 1. An overview of our pipeline. We first process the X-ray using a segmentation network. Note that we formalize the groundtruth mask using the provided landmarks. Afterwards, the predicted mask is fed to the regression model to perform angle value prediction.

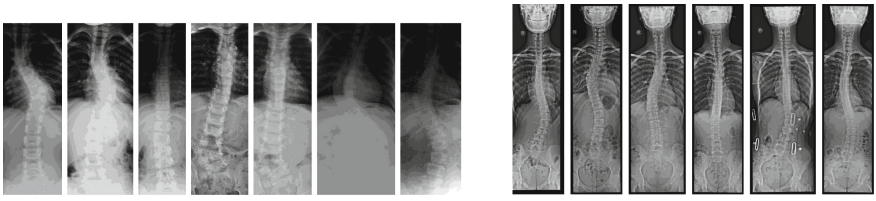
approach is able to produce high-precision angle results but relies heavily on the landmarks predictions, which means a small mistake in coordinates may lead to a big error in angle predictions. On the contrary, angle regression methods are more stable but may lack the ability to generate precise predictions. In this paper, we explore the possibility of aforementioned two directions in MICCAI AASCE 2019 challenge and our experimental results show that the regression strategy outperforms the landmark approach. We will conduct more details in following sections.

2 Proposed Method

We display our pipeline in Fig. 1. The whole process is constructed by two networks: one for segmentation and the other for regression. The architecture of segmentation network is similar to PSPNet [7] while the regression part employs traditional classification models.

2.1 Preprocessing

We observed that there is an obvious domain gap between training and testing sets (as shown in Fig. 2). To mitigate this problem, we first apply histogram



(a) **Training set** (b) **Testing set**

Fig. 2. A comparison of training set and testing set. It is obvious that these two sets have a huge domain gap.

equalization to both sets to make them visually similar. Considering the limited number of testing images, we decided to manually crop these X-rays to remove the skull and keep the spine in the appropriate scope. Besides, we also applied random rescaling (0.85 to 1.25) and random rotation ($-45°$ to $45°$) during the training process. We tried to add gaussian noise to the input images in order to mitigate the overfitting but it did not work.

For the segmentation task, we built the groundtruth masks on top of offered landmarks' coordinates. It is worth noting that we found adding another class "gap between bones" helped the segmentation model perform the best. We argue that such operation may regularize the training process which makes final predictions more precise.

2.2 Network Architecture

We followed the instructions in [7] to design our segmentation network. After the feature extractor, PSPNet [7] utilized different pooling kernels to capture various receptive fields. In order to keep the feature map size, we also append the dilated convolution with different dilation rates to the pooling pyramid. As shown in Fig. 1, we used 2, 4 and 6 as dilation rates while summing their outputs after the convolution operations. For backbone architecture, we simply took ResNet-50 [4] and ResNet-101 as the basic feature extractor (Fig. 3).

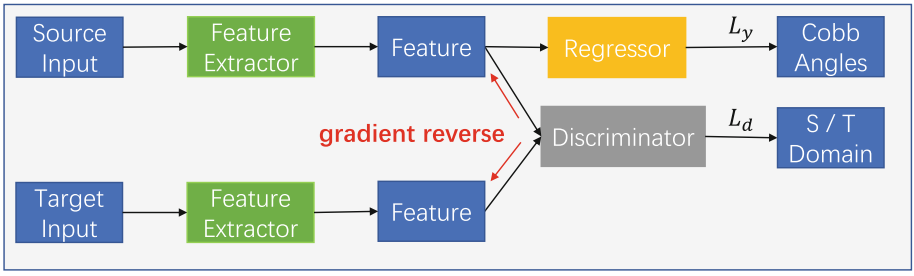


Fig. 3. Our domain adaptation strategy.

As for the classification part, we directly employed recent classification networks to perform the regression task. ImageNet based pretraining was used because we found it helped a lot under limited training samples. Considering the domain gap between training and testing sets, we modified the approach proposed in [3]. The idea is pretty simple which adds a discriminator branch and reverses its gradients during the back propagation so the final loss function can be formalized as:

$$Loss = L_y + \lambda L_d \tag{1}$$

where λ is set to 1 in our experiments.

2.3 Network Training

We used Adam as the default optimizer for both networks where the initial learning rate is 3e−3. β_1 and β_2 are set to 0.9 and 0.999, respectively. We also used weight decay which is 1e−5 and cosine annealing strategy. For the segmentation model, we ran each network for 50 epochs while 90 epochs seem to be a better choice for the regression model. We resized the segmentation input to 1024×512 and regression input to 512×256. The batch size is 32 on 4 NVIDIA P40 GPUs.

Table 1. We made an ablative study on input types and input sizes. It is easy to find that using segmentation mask as input performs the best on the validation set while (512, 256) is the best regression size. The default segmentation backbone is ResNet-50 and regression model is DenseNet-169.

Input type	Input size	Angle1	Angle2	Angle3
Img	(512, 256)	6.0754	7.3386	6.7629
Img + Mask	(512, 256)	5.4489	6.4599	5.8470
Mask	(512, 256)	**4.7128**	**5.7965**	**5.6596**
Mask	(1024, 512)	4.9360	7.2436	6.7321

Table 2. The segmentation performance of PSPNet and DeepLab V3+.

Metric	Ours	PSPNet	DeepLab V3+
mIOU	0.8943	0.8715	0.817

3 Experimental Results

We report our experimental results in both local validation and online testing sets. Note that we did not use cross validation.

3.1 Local Validation

In this part, we report the *L1 distance* between model predictions and groundtruth labels. As shown in Table 1, it is obvious that segmentation mask is the best input type and (512, 256) is the best input size. In Table 2, we compare the performance of our improved version with PSPNet and DeepLab V3+ [2]. We can find that adding a dilation pyramid thus improves the performance of previous PSPNet. It is interesting that PSPNet surpasses DeepLab V3+ by a large margin since they have achieved comparable performance in PASCAL VOC segmentation task. We argue that the failure of DeepLab can be attributed to the limited training data and our parameter tuning strategies.

3.2 Online Testing

We formalize our online testing results into Table 3. We can see that our dilation pyramid improves the online SMAPE by 0.48. Also, it is quite normal to see that EfficientNet-b5 is better than DenseNet-169 considering its higher ImageNet performance. After adding domain adaptation module, the single model performance rises to 26.15. During the model ensemble stage, we assigned different weights to different model outputs considering their validation scores. We mainly ensembled ResNet series, DenseNet series and EfficientNet series. This strategy helped us to improve the SMAPE score to 22.25.

Table 3. The segmentation performance of PSPNet and DeepLab V3+.

Strategies					
PSPNet + DenseNet-169	✓				
Ours + DenseNet-169		✓			
Ours + EfficientNet-b5			✓	✓	
Domain Adaptation				✓	✓
Model Ensemble					✓
SMAPE	28.51	28.03	27.07	26.15	22.25

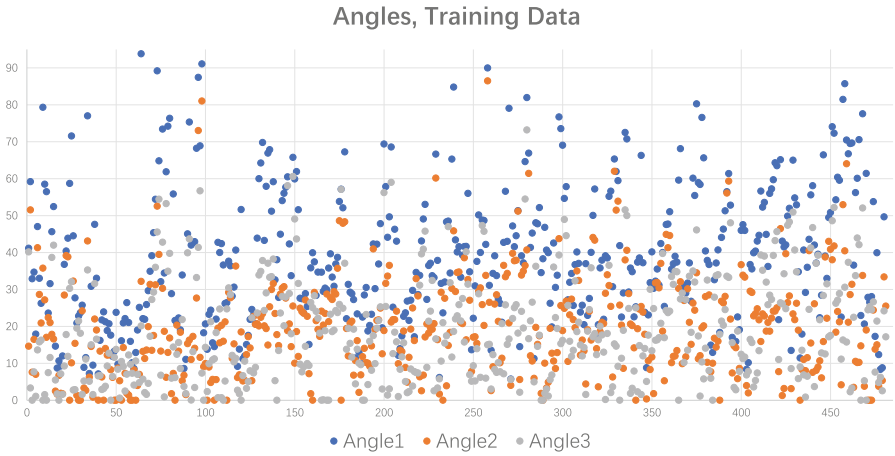


Fig. 4. Distribution of 3 angles in the training set.

During the online testing stage, we revisited the distribution of 3 angles in the training set. From Fig. 4, we can easily find out that Angle2 and Angle3 are much smaller than Angle2. Also, Angle2 has many values which are close to zero. According to such phenomenons, we reduced angles smaller than 4° to zeros which brought us to 21.71 SMAPE.

References

1. Chen, B., Xu, Q., Wang, L., Leung, S., Chung, J., Li, S.: An automated and accurate spine curve analysis system. IEEE Access (2019)
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
3. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. arXiv preprint arXiv:1409.7495 (2014)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
5. Wu, H., Bailey, C., Rasoulinejad, P., Li, S.: Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 127–135. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_15
6. Wu, H., Bailey, C., Rasoulinejad, P., Li, S.: Automated comprehensive Adolescent Idiopathic Scoliosis assessment using MVC-Net. Med. Image Anal. **48**, 1–11 (2018)
7. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)