Automatic Whole Spine Segmentation from Lateral View X-Ray Images using Lightweight Pyramid Attention Network

Liping Zhang\textsuperscript{1} \quad LPZHANG@LINK.CUHK.EDU.HK
Lin Shi\textsuperscript{1} \quad SHILIN@CUHK.EDU.HK
Jack Chun Yiu Cheng\textsuperscript{2} \quad JACKCHENG@CUHK.EDU.HK
Winnie Chiu Wing Chu\textsuperscript{1} \quad WINNIECHU@CUHK.EDU.HK
Simon Chun Ho Yu\textsuperscript{1} \quad SIMONYU@CUHK.EDU.HK

\textsuperscript{1} Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong
\textsuperscript{2} Department of Orthopaedics & Traumatology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong

Editors: Under Review for MIDL 2019

Abstract

Fully automatic whole spine segmentation from lateral view X-ray images is a very challenging task due to the difficulty in locating the thoracic vertebra occluded by the ribcage or overlapped by the upper arm. We propose a novel lightweight pyramid attention network architecture (LPA-Net) that can provide a fully automatic end-to-end system for whole spine segmentation in lateral view X-ray images. The proposed network uses newly designed lightweight network as backbone, embeds global attention refinement to provide the attention weighting mechanism for emphasizing the discriminative features of low-level stages, and incorporates attention-based atrous spatial pyramid pooling to capture the effective contextual information, which allows the network to mitigate the vertebra locating problem. Experimental results on 100 lateral view X-ray images from Adolescent Idiopathic Scoliosis (AIS) patients collected from EOS machine show an impressive segmentation accuracy with 93.07\% Dice coefficient.

Keywords: Convolutional Neural Networks, Whole Spine Segmentation, Lateral View X-ray Images, LPA-Net, Adolescent Idiopathic Scoliosis

1. Introduction

Delineating the detailed shape of the vertebrae can considerably help with the early diagnosis and treatment planning of a number of spinal disorders. However, fully automatic segmentation of whole spine structure from lateral view X-ray images is a very challenging task due to the difficulty in locating the thoracic vertebra occluded by the ribcage or overlapped by the upper arm.

Several previous attempts have been made to fully automatic segment the vertebrae from cervical (Zamora et al., 2003; Roberts et al., 2009) or lumbar (Roberts et al., 2009) images. The design of these hand-crafted features heavily depends on the domain knowledge of spine structures and those low-level features are insufficient to capture the complicated characteristics of whole spine structures. Recently, Al Arif et al. (2018) proposed a deep
learning-based fully automatic framework for segmentation of cervical vertebrae in X-ray images, which achieved excellent results. To the best of our knowledge, however, the fully automatic segmentation of whole spine segmentation in lateral view X-ray images is not exploit yet.

In this paper, we propose a novel lightweight pyramid attention network (LPA-Net) for fast and accurate end-to-end whole spine segmentation in lateral view X-ray images. Our network uses newly designed lightweight network (LW-Net) as backbone, utilizes global attention-based refinement (GAR) module to select the discriminative features of low-level stages, and embeds attention-based atrous spatial pyramid pooling (A-ASPP) module to capture weighted pyramid contextual information for accurate whole spine segmentation.

2. Methods

There are few pre-trained models available in medical imaging field, thus, networks with huge number of parameters are hard to well optimize and more likely to overfit when the model train on a small dataset from scratch. Therefore, we propose lightweight network with a small number of parameters to server as backbone for our LPA-Net, which greatly alleviates the computational burden and improves gradient flows.

The FCNs (Long et al., 2015) encoded different levels of features with consecutive pooling operators to learn more abstract and robust features also resulted in low-resolution feature maps. Different with U-Net (Ronneberger et al., 2015) fused low-level features from encoder directly, our model utilizes GAR module to introduce the prior information from high stages to weight the features of low-stages such that the useful features will be emphasize during the refinement.

PSPNet (Zhao et al., 2017) and Deeplab v3 (Chen et al., 2017) exploited the multi-scale contextual information to improve the accuracy of dense prediction with SPP (He et al., 2015) and ASPP (Chen et al., 2018) respectively. Inspired by recent PANet (Li et al., 2018a), we further enhance ASPP by proposing A-ASPP to enable our network to utilize global contextual prior to provide guidance for emphasizing the important contextual information and lightening those are less.

3. Experiments and Results

We build a dataset consists of 332 lateral view X-ray images from AIS patients collected from EOS machine for evaluating the effectiveness of our proposed LPA-Net in the whole spine segmentation task. The ground truth segmentation was labeled by an experienced clinician and the labeled vertebrae include cervical vertebra except C1, thoracic vertebra, chest vertebra, and sacrum.

Table 1: Segmentation Performance of Different Models in Lateral X-ray Images

<table>
<thead>
<tr>
<th>Model</th>
<th># of params(M)</th>
<th>DC(%)</th>
<th>JC(%)</th>
<th>SE(%)</th>
<th>SP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net</td>
<td>30.90</td>
<td>90.54</td>
<td>83.06</td>
<td>87.15</td>
<td>99.58</td>
</tr>
<tr>
<td>DenseUNet</td>
<td>65.78</td>
<td>92.18</td>
<td>85.65</td>
<td>89.14</td>
<td>99.65</td>
</tr>
<tr>
<td>LPA-Net</td>
<td>2.98</td>
<td>93.07</td>
<td>87.13</td>
<td>92.40</td>
<td>99.48</td>
</tr>
</tbody>
</table>
LPA-Net for Lateral View Whole Spine Segmentation

Figure 1: Examples of whole spine segmentation results from different models. The content of each case consists of a lateral X-ray image, a ground truth segmentation, and predicted segmentation from U-Net, DenseUNet and LPA-Net, respectively.

lumbar vertebra and sacrum. We randomly split the dataset into three groups with 184, 48, and 100 lateral view X-ray images for training, validation and testing respectively.

In order to demonstrate the effectiveness of LPA-Net for whole spine segmentation in lateral X-ray images, we compare the performances of LPA-Net with U-Net (Ronneberger et al., 2015) and DenseUNet (Li et al., 2018b) on our dataset. And the test results regard with several evaluation metrics, including Dice coefficient (DC), Jaccard coefficient (JC), sensitivity (SE) and specificity (SP), are listed in Table 1.

It is obviously that our LPA-Net achieves best accuracy in all metrics and improves the whole spine segmentation accuracy by a large margin over other models with only 2.98 million parameters, which demonstrates that the effectiveness of our LPA-Net to learn discriminative features for accurate whole spine segmentation in lateral view X-ray images. Figure 1 shows two cases of segmentation results obtained from different models. It is clearly that our LPA-Net can well predict each vertebra, even the thoracic vertebrae that other model failed, which further illustrates the effectiveness of our LPA-Net for mitigating the challenges of thoracic vertebrae locating problem, thus, results in accurate whole spine segmentation.

4. Conclusion

In summary, we presented a novel end-to-end lightweight pyramid attention segmentation system to mitigate the challenges of automatic whole spine segmentation in lateral view X-ray images. Compared with U-Net and DenseUNet, our LPA-Net can generate features with high discrimination capability with much less parameters. Extensive comparative experiments conducted on our dataset demonstrated the effectiveness of the proposed LPA-Net. In principle, the proposed GAR module, A-ASPP module and LPA-Net are general and can be easily extended to other segmentation tasks in medical images.
Acknowledgments

The authors would like to thank Dr. Zongshan Hu for his efforts to carefully label and verify the ground truth of the dataset.

References


