Spinal Curve Assessment of Idiopathic Scoliosis with a Small Dataset via a Multi-scale Keypoint Estimation Approach

Tianyu Liu
ty-liu17@mails.tsinghua.edu.cn
Tsinghua University
Haidian, Beijing, China

Yukang Yang
yyk19@mails.tsinghua.edu.cn
Tsinghua University
Haidian, Beijing, China

Yu Wang
wangyuspine@pkufh.com
Peking University First Hospital
Xicheng, Beijing, China

Ming Sun
ming.sun@clin.au.dk
Aarhus University Hospital
Aarhus, Denmark

Wenhui Fan
fanwenhui@tsinghua.edu.cn
Tsinghua University
Haidian, Beijing, China

Cheng Wu
wuc@tsinghua.edu.cn
Tsinghua University
Haidian, Beijing, China

Cody Bunger
codybung@rm.dk
Aarhus University Hospital
Aarhus, Denmark

ABSTRACT

Idiopathic scoliosis (IS) is the most common type of spinal deformity, which leads to severe pain and potential heart and lung damage. The clinical diagnosis and treatment strategies for IS highly depend on the radiographic assessment of spinal curve. With improvements in image recognition via deep learning, learning-based methods can be applied to facilitate clinical decision-making. However, these methods usually require sufficiently large training datasets with precise annotation, which are very laborious and time-consuming especially for medical images. Moreover, the medical images of serious IS always contain the blurry and occlusive parts, which would make the strict annotation of the spinal curve more difficult. To address these challenges, we utilize the dot annotations approach to simply annotate the medical images instead of precise annotation. Then, we design a multi-scale keypoint estimation approach that incorporates Squeeze-and-Excitation(SE) blocks to improve the representational capacity of the model, achieving the assessment of spinal curve without large-size dataset. The proposed approach uses pose estimation framework to detect keypoints of spine with simple annotation and small-size dataset for the first time. Finally, we conduct experiments on a collected clinical dataset, and results illustrate that our approach outperforms the mainstream approaches.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Spinal curve assessment, Keypoint estimation, Small-size dataset

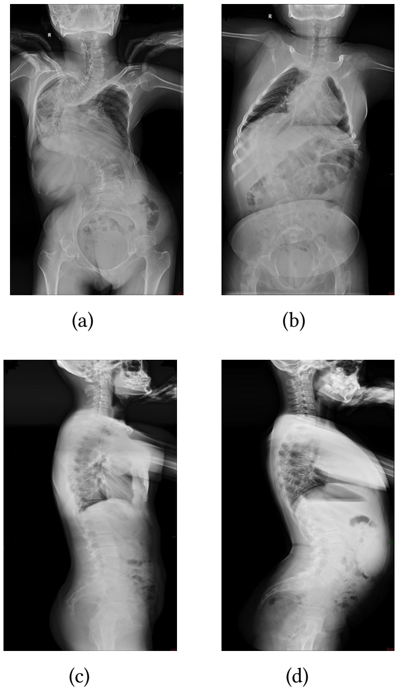
ACM Reference Format:
Tianyu Liu, Yukang Yang, Yu Wang, Ming Sun, Wenhui Fan, Cheng Wu, and Cody Bunger . 2020. Spinal Curve Assessment of Idiopathic Scoliosis with a Small Dataset via a Multi-scale Keypoint Estimation Approach. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12–16, 2020, Virtual Event, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3410530.3414317

1 INTRODUCTION

Idiopathic scoliosis(IS) is a three-dimensional structural spinal deformity that can potentially induce to potential heart damage and respiratory compromise, and also lead to severe psychosocial impairment. The radiographic assessment of spinal curve, which is usually conducted through manual judgment,

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8076-8/20/09…$15.00
https://doi.org/10.1145/3410530.3414317

serves the needs of clinical diagnosis and treatment determination. An automated assessment based on deep learning has the great potential to replace manual method, since the manual one exists large differences and errors due to the subjective judgments. To achieve sufficient assessment quality, the large dataset with well-annotation is the fundamental requirement on learning-based methods, which is often infeasible to prepare. It is a laborious and challenging work for the radiographic experts to annotate the IS-radiographic dataset in the traditional way, especially for the extreme(serious) cases. Collecting sufficiently large data is also onerous for many studies subject to patient privacy and legal considerations. Moreover, the radiography image of IS usually has some inherent limitations just as many vertebrae of the spine are small, occlusive and even invisible(see Fig.1), which are similar to the challenges of pose estimation. Therefore, the automated assessment based on deep learning should overcome three major barriers: the size of training dataset is small, annotations are laborious and vertebrae of extreme cases are always occlusive and invisible.


(a) (b)

(c) (d)

Figure 1: Image examples of full spine with different positions. (a) and (b) are the anterior positions; (c) and (d) are the lateral positions. It is difficult to prepare *strong* annotation since some parts of spine are blurry, occlusive and barely visible.

Recently, deep learning methods have been successfully applied in the computer-aided diagnosis task[1–3]. For the spine analysis, the segmentation-based methods are proposed to achieve accurate estimation of spinal keypoints for IS assessment[4, 5]. Spine-GAN[6] preserved fine-grained information to segment spinal structures by an atrous convolution module. The above approaches have well studied based on the large-size dataset, which are not suitable for the small dataset and extreme case. To deal with small dataset, multi-window conversion was introduced to integrate the information of different channels[7, 8]. The encoder-decoder modules were commonly adopted to augment new images, which can artificially generate training data to alleviate the overfitting[9, 10]. For the annotations strategy, it is essential to consider the task requirement when reducing the cost. Qi et al.[11] used the weak dot annotations to generate confidence maps for the placental lacunae location. The weak annotation was also proposed to train the instance segmentation model without the full annotation[12]. Pose estimation methods are suitable to tackle the problem of invisible keypoint inference by top-down [13] or bottom-up methods[14] . For example, Mask R-CNN[15], the extension of Faster R-CNN[16], added a parallel branch to predict object mask, which is an effective method to detect keypoint. With further development of research, Sun et al.[17] fused multi-resolution feature by the whole process to improve the accuracy of keypoint estimation. Although the above methods can address their special challenge, they are not suitable for our task since these studies are designed without consideration of limited-size dataset.

To solve the aforementioned problems, we propose a multi-scale keypoint estimation approach to assess the spinal curve from a small dataset. The contributions of this work are enumerated as below:

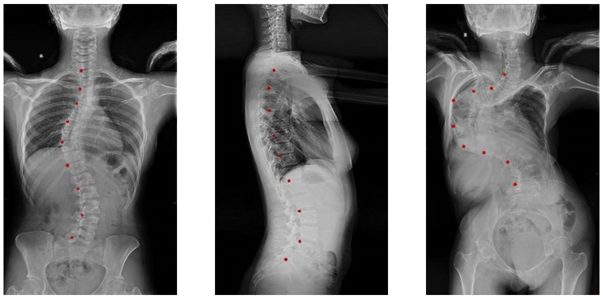
- We formulate the assessment of spinal curve into a keypoint estimation problem, which is more efficient and accurate than the mainstream methods.
- A dot-annotation method rather than the pixel-wise is applied to obviously reduce the annotation cost, especially for the extreme case.
- We incorporate the SE block into a multi-scale architecture to train the models based on a small dataset, and the SE block can improve the performance of the models in accuracy.

The rest of this paper is organized as follows. In Section 2, we present our main contributions: (i) the dot annotation method with higher efficiency(sec.2.1), (ii) our multi-scale keypoint estimation network(sec.2.2), which contains SE blocks and the design of network architecture. Section 3 present the effectiveness of the proposed approach, followed by the conclusions in Section 4.

2 METHODOLOGY

Dot Annotations

To relieve the annotations cost of large-scale radiography images, we propose a *simple* annotation method, which is called dot annotations, to replace the *strong* pixel-wise annotation. The human spine is divided into different regions according to its curves, including cervical, thoracic, lumbar, sacrum and coccyx. Idiopathic scoliosis usually occurs on the thoracic spine(Th) and lumbar spine(L), which consists of 12 and 5 vertebrae, respectively. Assessment of the spinal curve is to determine the maximal curvature degree of spine. Instead of delineating every single vertebra from the radiography image, we develop a new strategy to select representative vertebrae from thoracic and lumbar regions and estimate the spinal curve. To realize simple and high-efficiency annotations without a large loss of accuracy of models, We consecutively number the 17 vertebrae from Th1 to L5 and select every other vertebra as the representative vertebra (9 vertebrae in total). The *cross point* of right lateral edge and top edge of each selected vertebra is marked as a key point and annotated in radiography image (see Fig.2.). The annotation of the *cross point* does not require extensive medical knowledge and clinical experiences. It is noted that we don't select the *central point* as the keypoint because it cannot be accurately assessed by visual observation. Moreover, dot-annotations can also generalize to the task of spinal 3D-reconstruction to reduce the annotation and computation cost[18].



(a) (b) (c)

Figure 2: Image examples of dot-annotation method, in which radiographic experts annotate the truth location of keypoints with the red points. For the occlusive and invisible parts of spine, the annotations resort to educated guess based on medical knowledge and clinical experience, (c) shows the annotation result of Fig.1(a).

Multi-scale Keypoint Estimation Network

As mentioned above, we transform the problem of curve-assessment into that of keypoint estimation. Inspired by the high-quality application of keypoint inference in human pose estimation, we present a multi-scale network for keypoint estimation to assess the spinal curve in the radiography image. However, the size of medical datasets is usually too small to guarantee the representational capacity of deep learning and the accuracy of the results. In this paper, we incorporate an advanced technique called SE blocks into the multi-scale network to improve the efficiency of information extraction and transfer from the small dataset.

Squeeze-and-Excitation. SE blocks, the computation unit proposed in [19], can be fed seamlessly in any deep learning model. It can use global information to perform feature adaptive recalibration so as to emphasize more meaningful features. The variants of SE blocks are introduced to retain spatial information for keypoint estimation, which is called spatial and channel squeeze-and-excitation Block(scSE). For the channel squeeze modules, let $\mathbf{U} = [\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, ..., \mathbf{u}_{H,W}]$ be input feature map and consist of channels $\mathbf{u}_{i,j} \in \mathbb{R}^{1\times1\times C}$, where $i \in \{1, 2, .., W\}$ and $j \in \{1, 2, ..., H\}$. The channel squeeze operation is performed by a convolution $\mathbf{q} = \mathbf{W}_{sq} \star \mathbf{U}$ with kernel weigh $\mathbf{W}_{sq} \in \mathbb{R}^{1\times1\times C\times1}$, generating a attention weight $\mathbf{q} \in \mathbb{R}^{H\times W}$. We recalibrate $\mathbf{U}$ spatially to $\hat{\mathbf{U}}$ through a sigmoid layer $\sigma(\cdot)$, as equation(1).

$$\hat{\mathbf{U}} = [\sigma(q_{1,1})\mathbf{u}_{1,1}, ..., \sigma(q_{i,j})\mathbf{u}_{i,j}, ..., \sigma(q_{H,W})\mathbf{u}_{H,W}]. \quad (1)$$

For the spatial squeeze modules, let $\mathbf{U} = [\mathbf{u}^1, \mathbf{u}^2, ..., \mathbf{u}^C]$ be input feature map and consist of channels $\mathbf{u}^i \in \mathbb{R}^{H\times W}$. The global context information is embedded in a vector $\mathbf{z} \in \mathbb{R}^{H\times W}$ with its k^{th} element, as equation(2).

$$z_k = \frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W}\mathbf{u}_k(i,j). \quad (2)$$

To fully capture channel-wise dependencies, we transform $\mathbf{z}$ to $\bar{\mathbf{z}} = \mathbf{W}_1(\delta(\mathbf{W}_2\mathbf{z}))$, where $\mathbf{W}_1 \in \mathbb{R}^{C\times\frac{C}{2}}$,$\mathbf{W}_2 \in \mathbb{R}^{\frac{C}{2}\times C}$. Then, a resultant vector is used to excite and rescale $\mathbf{U}$ to $\widetilde{\mathbf{U}}$, as follows:

$$\widetilde{\mathbf{U}} = [\sigma(\bar{z_1})\mathbf{u}_1, \sigma(\bar{z_2})\mathbf{u}_2, ..., \sigma(\bar{z_C})\mathbf{u}_C]. \quad (3)$$

where $\sigma(\bar{z_i})$ describes the significance of the i^{th} channel. Fig.3 shows the architecture of SE module.

Network Architecture. Multi-scale feature fusion is crucial for the pose estimation tasks, since the tasks need global and local spatial information to infer the keypoint location. In this paper, we employ High-Resolution Net(HRNet)[17] as a base architecture for the keypoint estimation of radiography image, which is able to learn high-resolution feature pyramids and connect the multi-resolution subnetworks in parallel rather than in series. HRNet is a top-down method and originally proposed for pose estimation based on a relatively large dataset. For our task, we modify HRNet to a bottom-up

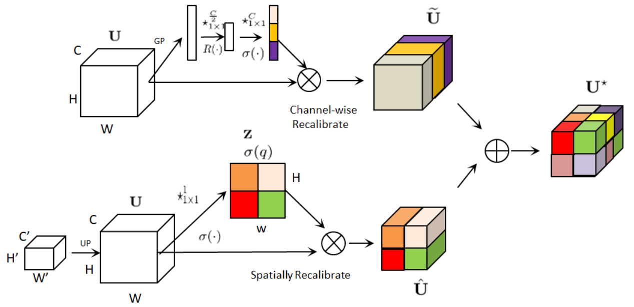


Figure 3: An illustration of SE module, where "UP" represents the upsampling layers, "GP" is the global pooling layers, $R(\cdot)$ is the ReLU activation function and $\sigma(\cdot)$ is the sigmoid function.

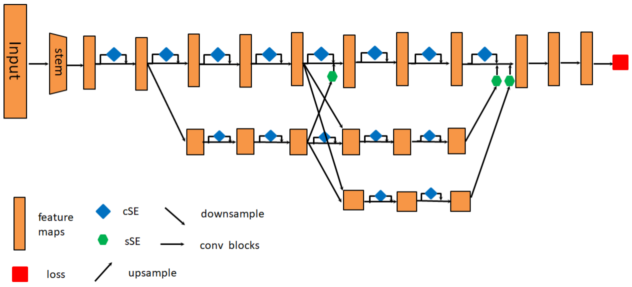


Figure 4: The illustration of multi-scale keypoint estimation network.

method through embedding a 1 × 1 convolution to predict heatmap, which is more suitable for keypoints estimation. With regard to the small dataset, HRNet and other traditional methods will be prone to overfitting and vanishing gradient in the training process. To alleviate this problem, we improve the representational capacity of the HRNet by adding cSE and sSE modules to suppress less meaningful features.

Specifically, we add the cSE modules after each convolution block with the same resolution index, which aim to recalibrate the channel of each feature map for emphasizing the more important channels. For the same stage of network, we add the sSE modules after each upsample block with different resolutions, which squeeze and recalibrate spatial information from subnetwork to provide more useful features. Let $\lambda_{m,n}$ be the mth stage with n resolution index, and input feature map of $\lambda_{m,n}$ is $\mathbf{U}^{\star}_{m,n} = \widetilde{\mathbf{U}}_{m,n} + \hat{\mathbf{U}}_{m,n}$. On the other hand, our high-to-low network only contains three stages with three parallel subnetworks with the aim of further reducing the complexity of architecture. At the first stage, we decrease the resolution to 1/8 by three strided 3×3 convolutions since the radiography image is always *large* in size, and the architecture of each stage is similar to that in [17]. In the prediction stage, we use the deconvolution module to generate new feature maps for keypoint inference, and then design a loss function which is the sum of two parts: focal loss and regularized-L2 loss. The focal loss can deal with hard negative samples for keypoint detection, and regularized-MSE loss is used for location. The network architecture is illustrated in Fig.4.

3 EXPERIMENTS

Data Acquisition and Processing

We conduct experiments to evaluate the proposed method based on a dataset with 200 radiography images, which are collected from Peking University First Hospital. This dataset contains anterior and lateral radiography images of the full spine. Radiographic experts have annotated the truth keypoint locations of full spine by the dot-annotation method. For the dataset details, these images are not unified in size, whose height-width ratios range from 1.85 to 2.16, and the average size is 3560 × 1740 × 3 pixels. The dataset will be divided into two parts, one with 170 images is for training and another with the rest 30 images for testing. We first resize the original dataset to 1024 × 512 × 3 pixels for training. Then, we perform the data augmentation with random scale, random rotation, random crop and random flip to reduce the potential overfitting.

Implementation Details

In this paper, we adopt the Adam optimizer to train the network with the batch size of 10 and patch size of 64 × 64, which is suitable for a small dataset. An initial learning rate was set as 1×10^{-5}, and the decay rate set to 5×10^{-7} after every 100 epochs. The network is trained for a total of 2000 epochs. We use the early stopping strategy to avoid potential overfitting, and the training will be stopped if the validation loss does not decrease by 50 epochs. Besides, the dropout layer with rate 0.5 is added to all convolutional layers with ReLU activate function. We conduct standard 3-fold cross-validation to assess the performance of our models. Without loss of generality, each experiment will be repeated 3 times . The code of our proposed method is implemented by Pytorch on a NVIDIA RTX 2080Ti GPU.

Evaluation Metric. The traditional evaluation metric usually adopts the Object Keypoint Similarity(OKS), which only can assess the visible points of annotations. For our task, the inference of invisible keypoint plays an important part of the method. Therefore, we assess all keypoints based on a modified OKS without subjective effect. The modified OKS

Table 1: Comparison with traditional methods on collected dataset.

Methods	backone	E_d	AP	AP^{50}	AP^{75}
Mask-RCNN	ResNet-101	1.67	60.0	69.3	64.1
CPN	ResNet-101	1.72	45.2	55.2	53.0
CornerNet	ResNet-101	1.58	50.7	62.4	58.1
Hourglass	–	1.32	62.5	70.3	66.5
Our method	HRNet-w32	**0.58**	**72.2**	**82.6**	**76.5**

is written as follows:

$$M_{OKS} = \frac{\sum_i \exp\{-d_i^2/2s^2\sigma_i^2\}\delta(\mu_i = 1)}{\sum_i \delta(\mu_i = 1, 2)} + \frac{\alpha \sum_i \exp\{-d_i^2/2s^2\sigma_i^2\}\delta(\mu_{i=2})}{\sum_i \delta(\mu_i = 1, 2)}, \tag{4}$$

where d_i is the Euclidean distance(E_d) between a predicted keypoint and its ground truth, s is the object scale, σ is the normalizing factor, α is weight coefficient(set as 2), δ is the visibility and its label μ. Following the previous study[15], we report average precision AP: AP^{50}(AP at OKS = 0.5) and AP^{75}(AP at OKS = 0.75).

Results

Baseline Comparisons. To evaluate the performance of the proposed method, we compare it with other mainstream pose estimation methods, including Mask-RCNN[15], CPN[13], CornerNet[20] and Hourglass[21]. The data processing of these methods follow the proposed one, and these methods fine-tuned the pretrained network for keypoint estimation on the collected dataset. The results are listed in Table.1, among which our method achieves 72.2 AP, and average of all E_d is 0.58 . It can be seen clearly that our proposed model outperforms the compared counterparts in two metrics. For the compared methods, they suffer from overfitting problem due to the limited size of training data, and there is no obvious difference among their results. Moreover, there is no obvious difference among these methods, and the Hourglass performs little better which probably because of the multi-scale feature extraction. Visualized results of all experiments are shown in Fig.5, in which the red points of the spine are the ground truth, and the green ones are the inference location. Then, two adjacent points are connected with a straight line in order(see Fig.5.).

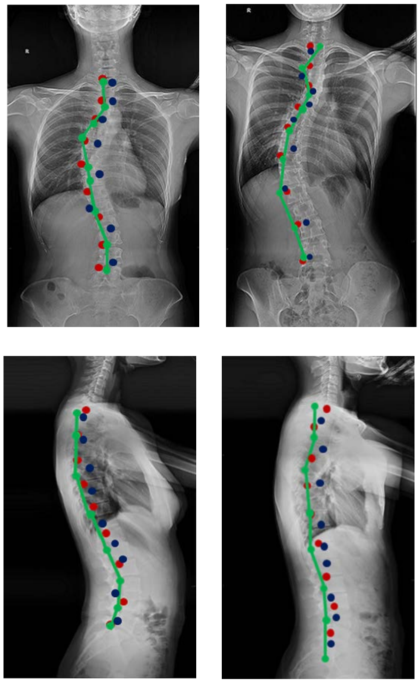


Figure 5: Visualized results of our method on the collected dataset. The green points denote the predicted keypoints, and the green lines are the predicted spinal curves. The blue points denote the average results of comparison methods.

Table 2: Ablation study on collected dataset without a component of method.

Methods	backone	E_d	AP	AP^{50}	AP^{75}
sSE	HRNet-w32	1.32	65.2	74.3	69.2
cSE	HRNet-w32	1.09	70.0	80.3	75.9
SE	HRNet-w32	1.42	60.3	72.1	66.2
One-scale	HRNet-w32	1.01	68.1	76.1	70.7
Our method	HRNet-w48	0.98	69.8.	83.2	77.5

Ablation Study. In this section, we conduct an ablation study to demonstrate the efficacy of the major components through the same processing and training strategy. In particular, we respectively perform the experiments with such four variants as the model trained without, SE modules, cSE modules, sSE modules and multi-scale feature extraction. The ablation results are listed in Table. 2. It shows that all components in our models can bring improvement of the keypoint estimation, especially when adding the sSE modules to extract the spatial information with different feature scales. Besides, the ablation study also illustrates multi-scale feature extraction is meaningful for keypoint estimation.

4 CONCLUSION

In this paper, we propose a keypoint estimation approach to assess the spine curve for IS-diagnosis from the small dataset. It can alleviate the overfiting problem of the training network with limited data. Furthermore, we present a dot-annotation strategy to label the images, which is more simple and effective than the pixel-wise annotations. The numerical experiments show that our approach has achieved better performance compared with mainstream pose estimation methods. In this way, the proposed approach can provide a explainable result to better assist doctor in preliminary diagnosis of idiopathic scoliosis. For further work, we will employ prior medical knowledge to guide the feature extraction for improving the accuracy of models and facilitate computer-aided diagnosis in clinical practice.

REFERENCES

[1] Philipp Ernst, Georg Hille, Christian Hansen, Klaus D Tonnies, and Marko Rak. A cnn-based framework for statistical assessment of spinal shape and curvature in whole-body mri images of large populations. pages 3–11, 2019.

[2] Tianyu Liu, Wenhui Fan, and Cheng Wu. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine*, 101:101723, 2019.

[3] Asim Smailagic, Pedro Costa, Hae Young Noh, Devesh Walawalkar, Kartik Khandelwal, Adrian Galdran, Mostafa Mirshekari, Jonathon Fagert, Susu Xu, Pei Zhang, et al. Medal: Accurate and robust deep active learning for medical image analysis. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 481–488. IEEE, 2018.

[4] Jingru Yi, Pengxiang Wu, Qiaoying Huang, Hui Qu, and Dimitris N Metaxas. Vertebra-focused landmark detection for scoliosis assessment. *arXiv: Image and Video Processing*, 2020.

[5] Hongbo Wu, Christopher S Bailey, Parham Rasoulinejad, and Shuo Li. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet. pages 127–135, 2017.

[6] Zhongyi Han, Benzheng Wei, Ashley Mercado, Stephanie Leung, and Shuo Li. Spine-gan: Semantic segmentation of multiple spinal structures. *Medical Image Analysis*, 50:23–35, 2018.

[7] Hyunkwang Lee, Sehyo Yune, Mohammad Mansouri, Myeongchan Kim, Shahein Tajmir, Claude Emmanuel Guerrier, Sarah A Ebert, Stuart R Pomerantz, Javier Romero, Shahmir Kamalian, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering*, 3(3):173–182, 2019.

[8] Xiyu Lu, Xinlei Chen, Guang Sun, Depeng Jin, Ning Ge, and Lieguang Zeng. Uwb-based wireless body area networks channel modeling and performance evaluation. In *2011 7th International Wireless Communications and Mobile Computing Conference*, pages 1929–1934. IEEE, 2011.

[9] Joseph Bullock, Carolina Cuesta-Lázaro, and Arnau Quera-Bofarull. Xnet: A convolutional neural network (cnn) implementation for medical x-ray image segmentation suitable for small datasets. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10953, page 109531Z. International Society for Optics and Photonics, 2019.

[10] Xinlei Chen, Yu Wang, Jiayou He, Shijia Pan, Yong Li, and Pei Zhang. Cap: Context-aware app usage prediction with heterogeneous graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3:1–25, 03 2019.

[11] Huan Qi, Sally Collins, and J Alison Noble. Automatic lacunae localization in placental ultrasound images via layer aggregation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 921–929. Springer, 2018.

[12] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE transactions on medical imaging*, 36(2):674–683, 2016.

[13] Yilun Chen, Zhicheng Wang, Yuxiang Peng, Zhiqiang Zhang, Gang Yu, and Jian Sun. Cascaded pyramid network for multi-person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7103–7112, 2018.

[14] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*, 2018.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[17] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5693–5703, 2019.

[18] Samuel Kadoury, Farida Cheriet, and Hubert Labelle. Personalized x-ray 3-d reconstruction of the scoliotic spine from hybrid statistical and image-based models. *IEEE Transactions on medical imaging*, 28(9):1422–1435, 2009.

[19] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Concurrent spatial and channel squeeze & excitationin fully convolutional networks. In *International conference on medical image computing and computer-assisted intervention*, pages 421–429. Springer, 2018.

[20] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 734–750, 2018.

[21] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016.