Automation of Spine Curve Assessment in Frontal Radiographs Using Deep Learning of Vertebral-Tilt Vector

KANG CHEOL KIM[ID][1], **HYE SUN YUN**[ID][1], **SUNGJUN KIM**[ID][2], **AND JIN KEUN SEO**[ID][1]

[1]Department of Computational Science and Engineering, Yonsei University, Seoul 03722, South Korea
[2] Center for Clinical Imaging Data Science, Department of Radiology and Research Institute of Radiological Science, Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul 06273, South Korea

Corresponding author: Sungjun Kim (agn70@yuhs.ac)

ABSTRACT In this paper, an automated and visually explainable system is proposed for a scoliosis assessment from spinal radiographs, which deals with the drawback of manual measurements, which are known to be time-consuming, cumbersome, and operator dependent. Deep learning techniques have been successfully applied in the accurate extraction of Cobb angle measurements, which is the gold standard for a scoliosis assessment. Such deep learning methods directly estimate the Cobb angle without providing structural information of the spine which can be used for diagnosis. Although conventional segmentation-based methods can provide the spine structure, they still have limitations in the accurate measurement of the Cobb angle. It would be desirable to build a clinician-friendly diagnostic system for scoliosis that provides not only an automated Cobb angle assessment but also local and global structural information of the spine. This paper addresses this need through the development of a hierarchical method which consisting of three major parts. (1) A confidence map is used to selectively localize and identify all vertebrae in an accurate and robust manner, (2) vertebral-tilt field is used to estimate the slope of an individual vertebra, and (3) the Cobb angle is determined by combining the vertebral centroids with the previously obtained vertebral-tilt field. The performance of the proposed method was validated, resulting in circular mean absolute error of 3.51° and symmetric mean absolute percentage error of 7.84% for the Cobb angle.

INDEX TERMS Frontal radiographs, spine, scoliosis, Cobb angle, deep learning.

I. INTRODUCTION

Adolescent idiopathic scoliosis is a structural spinal deformity mainly in the coronal plane [7]. Because radiography is fast, inexpensive, and simple compared with computed tomography and magnetic resonance imaging, frontal radiography is commonly used in diagnosis of scoliosis and for monitoring the progression. Radiography takes advantage of the ability to generate an entire spine image of a standing patient while reflecting the 3D rotatory nature of the scoliotic deformity [18].

The Cobb angle is commonly used to measure the lateral curvature of the spine in the coronal plane from a frontal radiograph, and is defined by the angle between two lines parallel to the upper plate of the superior end vertebra and

the lower plate of the inferior end vertebra [12] as shown in Fig. 1. If the endplates are not well visualized, the boundary of pedicles are used to compute Cobb angle [12], [18], [19]. A manual measurement of the Cobb angle is time-consuming, cumbersome, and operator-dependent, resulting in high inter- and intra-observer variations. Even the intra-observer variability of the Cobb angle, which is known to be less than the interobserver variability, was reported to range as much as 5° to 10° [20], [24].

The human measurement of Cobb angle is known to be variable; a spine curve is practically considered progressed on radiographs when the Cobb angle increases by 5° or more, per consecutive clinic visit [29]. This could be an arbitrary criterion that can potentially mislead the patient care. Hence, there is a significant need to improve the reproducibility of the Cobb angle measurement through automatic estimation.

The associate editor coordinating the review of this manuscript and approving it for publication was Qichun Zhang[ID].

Numerous computer-aided methods for an automatic estimation of the Cobb angle have been developed by many investigators. These can be divided roughly into two categories: segmentation-based methods and direct estimation methods. Segmentation-based methods use an active contour model [1], customized filter [2], and charged-particle models [26] for vertebral segmentation to calculate the Cobb angle. Unfortunately, these methods are not robust because an accurate segmentation of the vertebra is extremely difficult owing to an unclear vertebral boundary in the radiographs. Direct estimation methods [30], [33] attempt to extract the correlation between spine features (e.g., landmarks) from radiographs and the Cobb angle estimation without segmentation. However, these landmark-based methods suffer from an accurate and robust estimation because small errors in the landmarks can cause serious errors in the Cobb angle.

Recently, Wu *et al.* [34] proposed an Multi-View Correction Network to achieve a fully automated comprehensive scoliosis assessment by leveraging the correlation between the frontal and lateral radiograph. Wang *et al.* [35] proposed an Multi-View Extrapolation Network for an accurate Cobb angle measurement in both frontal and lateral view by taking advantage of multiple views [34] and high-precision calculation [6]. Chen *et al.* [6] developed Adaptive Error Correction Net combined with a high-precision calculation for directly calculating the Cobb angle in a single frontal radiograph.

Despite their highly accurate results, these methods still suffer from lack of visual interpretability because they provide only the Cobb angle as output, without providing the most tilted vertebra. The designation of the most tilted vertebra is important for the decision of the curve progression and surgery planning. Therefore, it would be desired to build a clinician-friendly diagnosis system that provides highly accurate and reproducible Cobb angle measurement in a visually interpretable manner.

This paper proposes a hierarchical deep learning method designed to build a clinician-friendly system that provides intermediate decision process. The advantage of the proposed method is that it can directly visualize the end vertebrae by calculating a tilted angle of an individual vertebra, which could be hardly provided by the previously proposed methods [6], [34], [35]. The proposed method consists of three main steps, considering both the local and global information of vertebrae, which mimic the decision-making process of a clinician:

1) Localization and identification of the individual thoracic and lumbar vertebrae using a confidence map.
2) Estimation of the slope of the vertebrae using the vertebral-tilt field.
3) Cobb angle measurement using a confidence map and vertebral-tilt field.

To develop a visually interpretable and highly accurate system, we combined the segmentation-based method and direct estimation method. Similar to the segmentation method, which can visualize the results, the vertebral-tilt field provides a prediction at each pixel inside the vertebral region.
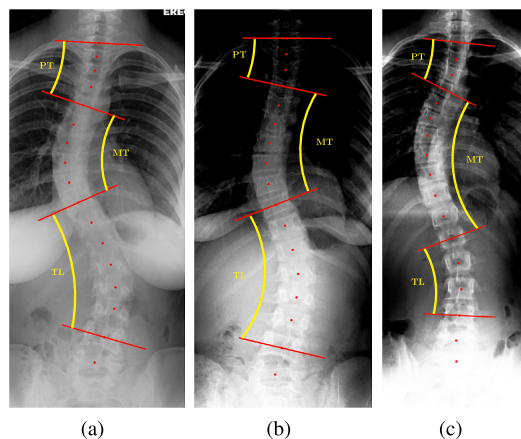

(a)   (b)   (c)

FIGURE 1. The Cobb angle is commonly used to measure the lateral curvature of the spine in the coronal plane. It is defined by the angle between two lines parallel to the upper plate of the superior end vertebra and the lower plate of the inferior end vertebra. Three Cobb angles, namely, the proximal-thoracic (PT), main thoracic (MT), and thoracic-lumbar (TL) anlges, are needed for scoliosis assessment.

Here, it predicts a vector that provides the slope of a vertebra in direct manner. Thereafter, this vertebral-tilt field, combined with the localization and identification of the vertebrae from a confidence map, provides an accurate slope on an individual vertebra in a visually interpretable manner.

The proposed method has three major contributions. (1) A highly accurate and robust Cobb angle measurement is achieved by the confidence map and vertebral-tilt field. (2) A visually explainable system is developed to improve the clinician's workflow. (3) A vertebral-tilt field is proposed for accurate estimation of slope of vertebrae.

The performance of our method is evaluated on 128 anterior-posterior (AP) radiographs with 481 labeled training data confirmed by radiologists. The experiment results show that the proposed method provides an accurate and robust performance for an identification of the vertebrae and Cobb angle measurement.

We achieved a 3.51° circular mean absolute error (CMAE) and 7.84% symmetric mean absolute percentage error (SMAPE) for the Cobb angle.

II. METHODS

Let $I(\mathbf{x})$ represent the intensity of the grayscale AP X-ray image at pixel position $\mathbf{x} = (x_1, x_2) \in \Omega$ where $\Omega = \{(x_1, x_2) : x_1 = 1, \cdots, h, \ x_2 = 1, \cdots, w\}$ represents a pixel grid in an image. Then, image I can be viewed as a matrix $I \in \mathbb{R}^{h \times w}$. The goal is to develop a fully automated method for a Cobb angle measurement from AP radiographs I. An automated measurement of the Cobb angle from radiographs requires dealing with the overlapping shadows of other thoracoabdominal bone and soft tissue structures. In addition, it is necessary to distinguish between the cervical and thoracic vertebrae which are adjacent and have a similar shape in frontal radiographs.
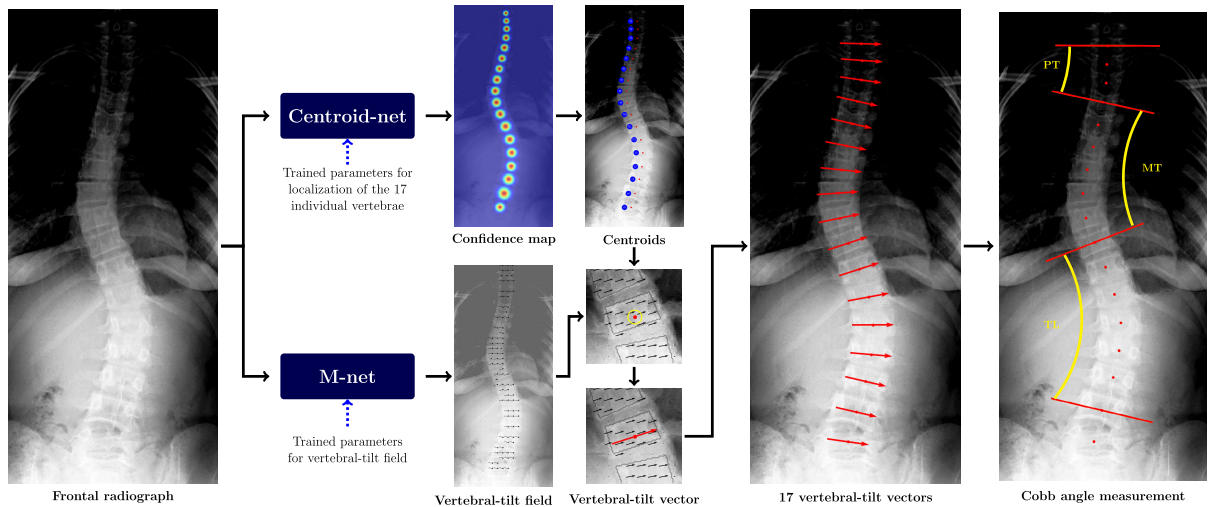
FIGURE 2. An overview of the proposed method. During the training phase, we train the different two neural networks to find proper parameters of the Centroid-net and the M-net. In the test phase, the Centroid-net is used to localize and identify all thoracic and lumbar vertebrae. The vertebral-tilt field is given by the M-net. Then, the 17 vertebral-tilt vectors which represent slope of 17 vertebrae are computed by combining centroids and vertebral-tilt field. Finally, we calculate the three Cobb angles using 17 vertebral-tilt vectors.

Our method consists of the three parts: localization and identification of the thoracic and lumbar vertebrae, an estimation of the slope of the vertebrae, and a Cobb angle measurement.

The schematic overview is shown in Fig. 2. A confidence map is used to localize and identify the 17 vertebrae. For an accurate and robust estimation of the Cobb angle, we take advantage of the vertebral-tilt field to describe the slope of individual vertebrae. The Cobb angle can be accurately determined by 17 vertebral-tilt vectors, which are given by the vertebral centroids and vertebral-tilt field.

A. LOCALIZATION AND IDENTIFICATION OF THE VERTEBRAE

In our method, we first predict the centroids of the 12 thoracic and 5 lumbar vertebrae in image I, where the output are expressed by the vector $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_{17}) \in \mathbb{R}^{2 \times 17}$, where $\mathbf{p}_j = (p_{j,1}, p_{j,2})$ represents the centroid of the j-th vertebra.

For an estimation of the 17 centroids, we employ a confidence map [5], [23], [31], [32] representing the belief of the centroids at each pixel position $\mathbf{x} = (x_1, x_2)$ in image I (see Fig. 2). To obtain the confidence map, we generate an individual confidence map $\psi_j : \mathbf{x} \mapsto \mathbb{R}$ for $j = 1, \cdots, 17$ that is defined by

$$\psi_j(\mathbf{x}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{p}_j||_2^2}{2\sigma_j^2}\right), \qquad (1)$$

where σ_j^2 is given by 1/8 of the height of j-th vertebra in image I. Next, these 17 confidence maps are integrated into the confidence map $\Psi : \mathbf{x} \mapsto \mathbb{R}$ obtained by the following:

$$\Psi(\mathbf{x}) = \max\{\psi_1(\mathbf{x}), \cdots, \psi_{17}(\mathbf{x})\}. \qquad (2)$$

Here, $\Psi(\mathbf{x})$ represents the maximum among all values of $\psi_j(\mathbf{x})$ with $j = 1, \cdots, 17$ at pixel position $\mathbf{x}$.

A confidence map regression function: $f_c : I \mapsto \Psi_{\text{rfn}}$ will be learned using a deep learning technique with a labeled training dataset $\mathcal{D}_c := \{I^{(n)}, \mathbf{P}^{(n)}, \Psi^{(n)}\}_{n=1}^N$.

The proposed network, called a Centroid-net, consists of three neural network functions: (i) a feature extraction network $f_{\text{ext}} : I \mapsto I_*$ shown in Fig. 3(a), (ii) an initial prediction network $f_{\text{init}} : I_* \mapsto \tilde{\Psi}_{\text{init}}$ in Fig. 3(b), and (iii) a refinement network $f_{\text{rfn}} : (I_*, \tilde{\Psi}_{\text{init}}) \mapsto \Psi_{\text{rfn}}$ shown in Fig. 3(c).

Here, f_{ext} initially produces a set of feature maps $I_* = f_{\text{ext}}(I) \in \mathbb{R}^{\frac{h}{4} \times \frac{w}{4} \times 512}$ which is an input of the network f_{init}. The next two networks sequentially predict the coarse initial confidence map $\tilde{\Psi}_{\text{init}} = f_{\text{init}}(I_*) \in \mathbb{R}^{\frac{h}{8} \times \frac{w}{8}}$ and the final confidence map $\Psi_{\text{rfn}} = f_{\text{rfn}}(\tilde{\Psi}_{\text{init}}, I_*) \in \mathbb{R}^{h \times w}$ by taking advantage of the refinement of the initial prediction $\tilde{\Psi}_{\text{init}}$ where the initial confidence map $f_{\text{init}}(I_*)$ is concatenated with the intermediate feature map I_* as an input of f_{rfn}.

The Centroid-net is designed to achieve a large receptive field size at a pixel of the output layer with a sequential prediction structure to capture the long range dependency between the 17 vertebrae. We adopt a convolutional neural network to learn three functions $f_{\text{ext}}, f_{\text{init}}$, and f_{rfn}. Fig. 3 shows the architecture of the Centroid-net. These networks $f_{\text{ext}}, f_{\text{init}}$, and f_{rfn} are learned simultaneously, using the training data $\mathcal{D}_c := \{I^{(n)}, \mathbf{P}^{(n)}, \Psi^{(n)}\}_{n=1}^N$.

In the Centroid-net, we use the weighted loss function to improve the prediction accuracy for T1 (the first thoracic vertebrae), because it is difficult to distinguish between T1 and C7 (the last cervical vertebra), as shown in Fig. 4. The proposed weighted loss function is:

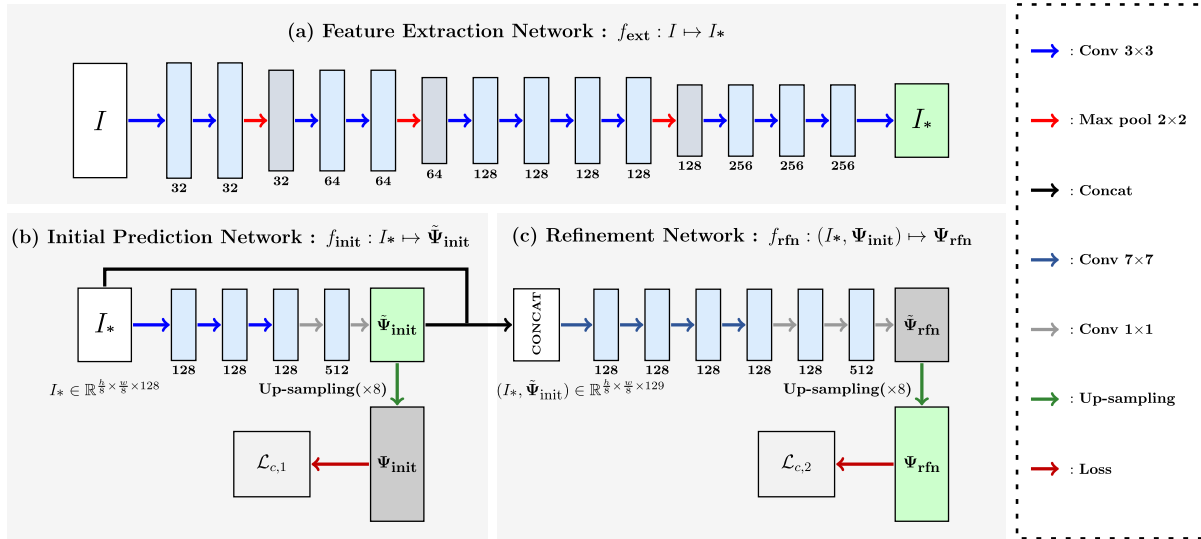$$\mathcal{L}_c = \frac{1}{N} \sum_{n=1}^N (\mathcal{L}_{c,1}^{(n)} + \mathcal{L}_{c,2}^{(n)}), \qquad (3)$$

FIGURE 3. Architecture of the proposed Centroid-net for localizing the 17 thoracic and lumbar vertebrae. The function f_{init} returns the initial confidence map providing the belief of the centroids of the 17 vertebrae. The output of f_{rfn} is a confidence map providing the belief of all centroids of 17 the vertebrae. Here, white blocks and green blocks denote the input and output of each network, respectively.
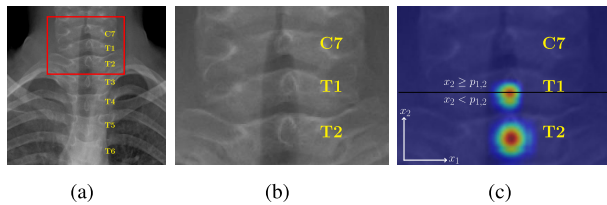

FIGURE 4. (a) shows the region of cervical and thoracic vertebra. (b) shows the region in the red box in (a). As shown in (b), the C7 and T1 vertebrae have a similar shape. (c) shows the ground-truth confidence map. During the training, the loss is computed in the region below black line ($x_2 < p_{1,2}$) while ignoring the loss occurred in the region of cervical vertebrae ($x_2 \geq p_{1,2}$).

FIGURE 5. Localization of the 17 vertebrae. (a) Test image I. (b) The confidence map $\Psi_{\text{rfn}} = f_{\text{rfn}}(I_*, f_{\text{init}}(I))$. (c) Local maxima denoted by red dots. (d) Determination of the centroids of the 12 thoracic and 5 lumbar vertebrae. In (d), the wrong candidates denoted by red box (whose scores are less than half of the mean score) are removed. Finally, the 17 candidates starting from the bottom candidate are selected as shown in (d).

where $\mathcal{L}_{\text{c},1}^{(n)}$ represents the intermediate loss given by

$$
\begin{cases}
\mathcal{L}_{\text{c},1}^{(n)} = \sum_{\mathbf{x} \in \Omega} \omega^{(n)}(\mathbf{x}) \left| g_1^{(n)}(\mathbf{x}) - \Psi^{(n)}(\mathbf{x}) \right|^2 \\
\text{with } g_1^{(n)} = \mathcal{U}\left(f_{\text{init}}(f_{\text{ext}}(I^{(n)})) \right),
\end{cases}
\tag{4}
$$

and $\mathcal{L}_{\text{c},2}^{(n)}$ represents the final loss given by

$$
\begin{cases}
\mathcal{L}_{\text{c},2}^{(n)} = \sum_{\mathbf{x} \in \Omega} \omega^{(n)}(\mathbf{x}) \left| g_2^{(n)}(\mathbf{x}) - \Psi^{(n)}(\mathbf{x}) \right|^2 \\
\text{with } g_2^{(n)} = f_{\text{rfn}}(f_{\text{ext}}(I^{(n)}), f_{\text{init}}(f_{\text{ext}}(I^{(n)}))).
\end{cases}
\tag{5}
$$

Here, $\mathcal{U}$ is an $8\times$ upsampling operator using bicubic interpolation and $\omega^{(n)}$ is the weight given by

$$
\omega^{(n)}(\mathbf{x}) = \begin{cases} 0 & \text{if } x_2 > p_{1,2}^{(n)} \\ 1 & \text{otherwise.} \end{cases}
\tag{6}
$$

The above weight $\omega^{(n)}$ is designed to calculate the loss only in the region containing the thoracic and lumbar vertebrae,
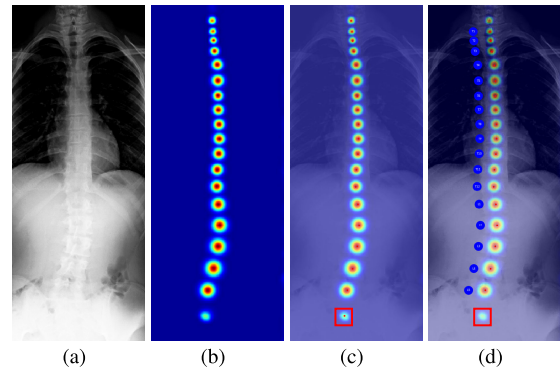
whereas $\omega^{(n)} = 0$ is within the region containing the cervical vertebra. This weight $\omega^{(n)}$ is used to focus on predicting 12 thoracic vertebrae (T1-12) and 5 lumbar vertebrae (L1-5) locations, while ignoring the prediction in the region containing cervical vertebrae. This weighted loss approach can accurately and robustly predict T1, while avoiding the difficulty of distinguishing between T1 and C7 (see Section III-F for further details).

The proposed Centroid-net involving three networks f_{ext}, f_{init}, and f_{rfn} is determined by minimizing the loss function in (3) using the training data $\mathcal{D}_{\text{c}} = \{I^{(n)}, \mathbf{P}^{(n)}, \Psi^{(n)}\}_{n=1}^{N}$.

The Centroid-net maps from I (frontal radiograph) to Ψ_{rfn} (confidence map), as shown in Fig. 5(a) and (b). From the confidence map Ψ_{rfn}, it is easy to determine the centroids $P = (p_1, \cdots, p_{17}) \in \mathbb{R}^{2 \times 17}$ of the 17 vertebrae.
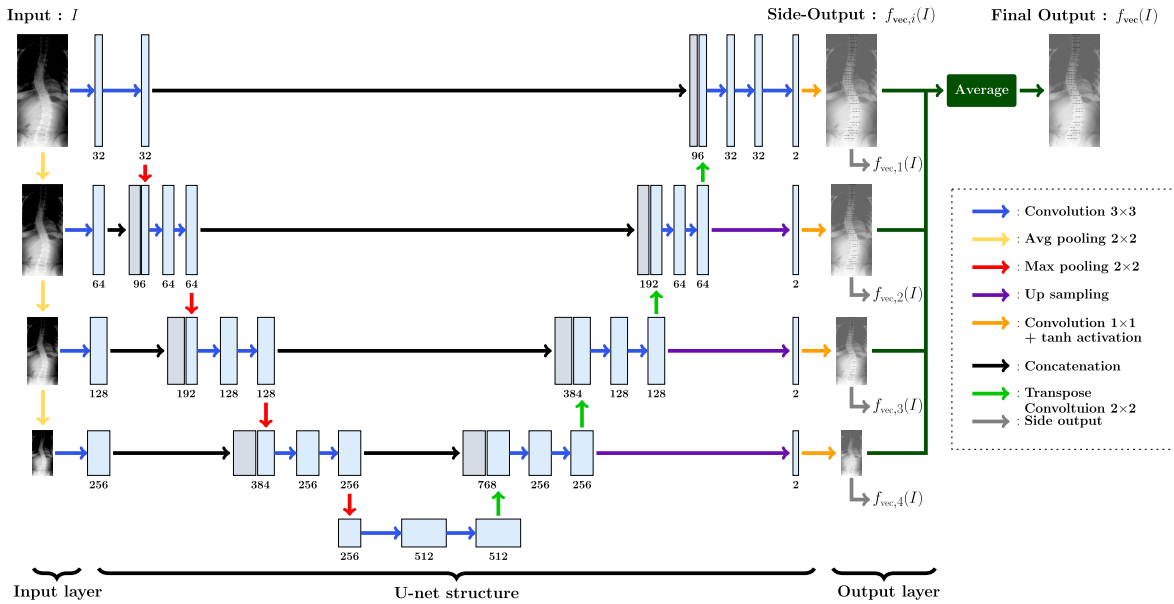
FIGURE 6. M-net architecture for learning of vertebral-tilt field. This M-net is designed to learn local and global structure at the same time.

First, the Otsu's thresholding [21] is applied to the confidence map Ψ_{rfn} in Fig. 5(b) to eliminate small local perturbations which are local maxima distant from vertebrae. The local maxima after thresholding are shown in Fig. 5(c). These local maxima are the candidates of the centroids. Next, we need to select 17 centroids $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_{17}) \in \mathbb{R}^{2 \times 17}$ from the several candidates. To do so, we set the score as the value of Ψ_{rfn} at each local maximum point. We exclude those candidates whose scores are less than half of the mean score. See red box in Fig. 5(c) and (d). Finally, we select the candidates 17 starting from the bottom candidate, as shown in Fig. 5(d).
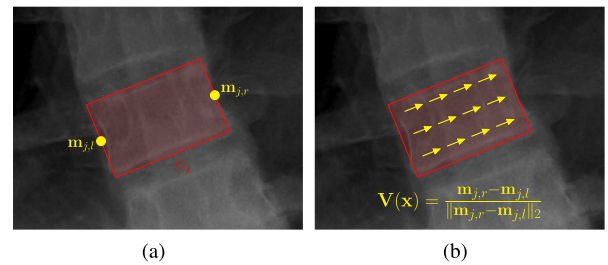

FIGURE 7. (a) The j-th vertebral region Ω_j (red-colored box) which is defined by the inside region of polygon from four corner points. The two yellow points in (a) denote the left and right middle points. (b) The unit vector at each pixel in Ω_j.

B. LEARNING VERTEBRAL-TILT FIELD

This section describes a method for providing a vertebral-tilt field which will be used for determining the 17 vertebral-tilt vectors. The vertebral-tilt field, denoted by $\mathbf{V}$, aims to describe the slope of each vertebra in image I, as shown in Fig. 7. To learn a neural network $f_{\mathrm{vec}} : I \mapsto \mathbf{V}$, we use the M-net [8] with labeled training data $\mathcal{D}_{\mathrm{vec}} = \{I^{(n)}, \mathbf{V}^{(n)}\}_{n=1}^{N}$. The architecture of the M-net is based on U-net [25], and two major parts are added in the input and output layers. Three parts are explained as following.

The U-Net is a convolutional neural network architecture developed for biomedical image segmentation. The architecture of the U-net consists of two parts. (1) The encoding path performs 3×3 convolutions followed by a rectified linear unit (ReLU), and max pooling. (2) The decoding path applies upsampling using 2×2 transpose convolutions and 3×3 convolutions, followed by ReLU in which the up-sampled output is concatenated with a high-resolution feature in the encoding path as shown in Fig. 6.

In the input layer, an image pyramid constructed by multi-scale images is used to integrate a multi-level receptive field. Here, the image is down-sampled by the average pooling and convolution with ReLU applied to the down-sampled image.

In the output layer, a side-output layer is used to learn local and global information at the same time. A multi-label loss function with a side output is used to deal with the vanishing gradient problem by replenishing the back-propagated gradients [8], [32]. At the output layer, a 1×1 convolution and an element-wise tangent hyperbolic activation function are applied.

Here, the f_{vec} is expressed as follows:

$$f_{\mathrm{vec}}(I) = \frac{1}{4} \sum_{i=1}^{4} f_{\mathrm{vec},i}(I), \qquad (7)$$

where $f_{\mathrm{vec},i}$ is the function producing the i-th side output, and the f_{vec} is learned by minimizing the following
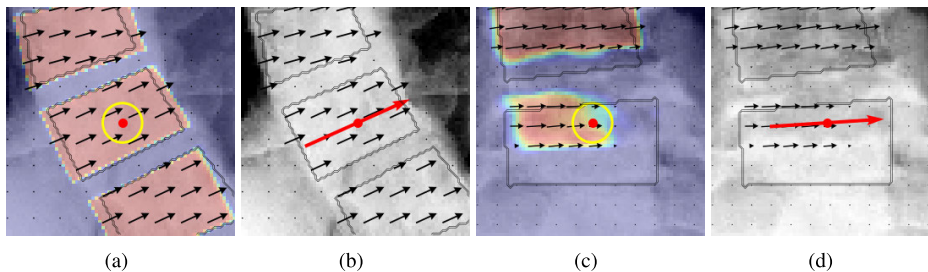
FIGURE 8. Visualization of learned vector fields and weights for two cases. In (a)-(d), the black arrows represent visualization of the vector in the vector field and the black contours represent j-th vertebral region Ω_j. In (a) and (c), the vector field $\mathbf{V}$ in j-th vertebral region Ω_j is overlaid with absolute value of vector field. The red dot and the yellow circle in (a) and (c) denote the j-th centroid $\mathbf{p}_j$ and contour of disk D_j, respectively. In (b) and (d), the estimated j-th vertebral-tilt vector $\mathbf{v}_j$ is denoted by red arrows. In (a), the vector field is learned well throughout the Ω_j while providing accurate the j-th vertebral-tilt vector. As shown in (d), the proposed method provide reasonable j-th vertebral-tilt vector even if the vector field is not learned well over Ω_j, but only in the corner region.

multi-label loss:

$$\mathcal{L}_{\text{vec}} = \frac{1}{4N} \sum_{n=1}^{N} \sum_{i=1}^{4} \sum_{\mathbf{x} \in \Omega} \omega^{(n)}(\mathbf{x}) \left\| f_{\text{vec},i}(I^{(n)})(\mathbf{x}) - \mathbf{V}^{(n)}(\mathbf{x}) \right\|^2. \quad (8)$$

Here, $\omega^{(n)}$ is the same weight as in (6).

Now, we will describe how to generate the ground-truth $\mathbf{V}^{(n)}$. Given image I, we first take a rectangular domain Ω_j occupying the j-th vertebral region, as shown in Fig. 7. The vector field $\mathbf{V}$ is zero outside $\cup_{j=1}^{17} \Omega_j$. In the j-th vertebral region, $\mathbf{V}$ is determined by

$$\mathbf{V}(\mathbf{x}) = \frac{\mathbf{m}_{j,r} - \mathbf{m}_{j,l}}{\|\mathbf{m}_{j,r} - \mathbf{m}_{j,l}\|}, \quad (9)$$

where $\mathbf{m}_{j,r}$ and $\mathbf{m}_{j,l}$ are the right and left midpoints, respectively, as shown in Fig. 7.

C. COBB ANGLE MEASUREMENT

From the neural networks described in the previous sections, we obtained a map from I to $(\mathbf{P}, \mathbf{V})$. Now, it remains to determine three Cobb angles, $\Theta = (\Theta_1, \Theta_2, \Theta_3)$. We chose a disk D_j centered at $\mathbf{p}_j$ with radius 5 that is contained in the region of the j-th vertebra Ω_j (see Fig. 8). We computed the weighted average of $\mathbf{V}$ over the disk D_j:

$$\mathbf{v}_j = \sum_{\mathbf{x} \in D_j} \frac{\|\mathbf{V}(\mathbf{x})\|}{\sum_{\mathbf{x} \in D_j} \|\mathbf{V}(\mathbf{x})\|} \mathbf{V}(\mathbf{x}), \quad j = 1, \cdots, 17. \quad (10)$$

This $\mathbf{v}_j$ is called j-th vertebral-tilt vector and it provides a slope of the j-th vertebra, denoted by θ_j as follow:

$$\theta_j = \arctan\left(\frac{\mathbf{v}_j \cdot (0,1)}{\mathbf{v}_j \cdot (1,0)} \right). \quad (11)$$

Using these 17 vertebral slopes $(\theta_1, \cdots, \theta_{17})$, we first determined the end vertebrae in three regions: the proximal thoracic (apex between T1 and T3), the main thoracic (apex between T3 and T12), and the thoracolumbar/lumbar (apex between T12 and L4). Then, the three Cobb angles $\Theta = (\Theta_1, \Theta_2, \Theta_3)$ are given by angle between the end vertebrae in
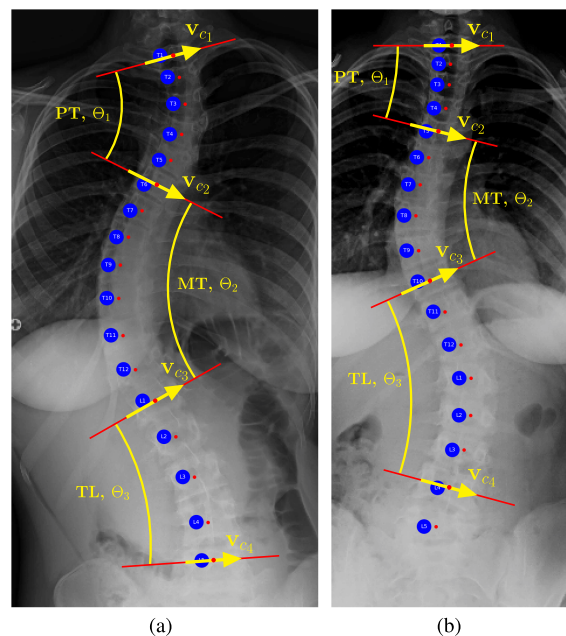

FIGURE 9. Cobb angle measurement. Two cases are shown in (a) and (b). In (a), the MT is the major curve. Then, the vertebral-tilt vectors on the four end vertebrae are determined in the order of $\mathbf{v}_{c_2}$, $\mathbf{v}_{c_3}$, $\mathbf{v}_{c_1}$, and $\mathbf{v}_{c_4}$. In (b), the TL is the major curve. In this case, we determine the vertebral-tilt vectors in the order of $\mathbf{v}_{c_3}$, $\mathbf{v}_{c_4}$, $\mathbf{v}_{c_2}$, and $\mathbf{v}_{c_1}$. The j-th Cobb angle Θ_j is given by the angle between $\mathbf{v}_{c_j}$ and $\mathbf{v}_{c_{j+1}}$.

the three regions, respectively. Here, apex is the vertebra or disk which is the most distant from the center of the vertebral column [12].

Here, we provide a more detailed explanation on how to evaluate the three Cobb angles $\Theta = (\Theta_1, \Theta_2, \Theta_3)$. Evaluation of Θ is determined by finding four end vertebrae and using the corresponding vertebral-tilt vectors, denoted by $\mathbf{v}_{c_1}$, $\mathbf{v}_{c_2}$, $\mathbf{v}_{c_3}$, and $\mathbf{v}_{c_4}$ (see Fig. 9). Now, we explain how to determine the four end vertebrae. Let $\boldsymbol{\theta}_{j,k}$ denote the angle between $\mathbf{v}_j$ and $\mathbf{v}_k$. Then, $\boldsymbol{\theta}_{j,k}$ satisfies

$$\cos(\boldsymbol{\theta}_{j,k}) = \frac{\mathbf{v}_j \cdot \mathbf{v}_k}{\|\mathbf{v}_j\| \|\mathbf{v}_k\|}. \quad (12)$$

Let $c_{\min}$ and $c_{\max}$ be numbers in $\{1, 2, \cdots, 17\}$ given by

$$c_{\min} = \min \left\{ \operatorname*{argmax}_{(j,k) \in \{1,\cdots,17\} \times \{1,\cdots,17\}} \boldsymbol{\theta}_{j,k} \right\}, \quad \text{and}$$

$$c_{\max} = \max \left\{ \operatorname*{argmax}_{(j,k) \in \{1,\cdots,17\} \times \{1,\cdots,17\}} \boldsymbol{\theta}_{j,k} \right\}. \qquad (13)$$

Here, $c_{\min}$- and $c_{\max}$-th vertebrae can be viewed as the upper and lower end vertebrae, respectively, of the major curve that has the largest Cobb angle, where MT or TL can be the major curve [12]. For example, in Fig. 9(a), MT is the major curve and therefore, $c_2 = c_{\min}$ and $c_3 = c_{\max}$. On the other hand, in Fig. 9(b), TL is the major curve and therefore $c_3 = c_{\min}$ and $c_4 = c_{\max}$.

Now it remains to determine remaining two end vertebrae. In the case when MT is the major curve, the remaining two vertebral-tilt vectors $\mathbf{v}_{c_1}$ and $\mathbf{v}_{c_4}$ are determined by

$$c_1 = \operatorname*{argmax}_{k \in \{1,\cdots,c_2-1\}} \boldsymbol{\theta}_{k,c_2}, \quad \text{and}$$

$$c_4 = \operatorname*{argmax}_{k \in \{c_3+1,\cdots,17\}} \boldsymbol{\theta}_{c_3,k}. \qquad (14)$$

On the other hand, when TL is the major curve, the remaining two vertebral-tilt vectors $\mathbf{v}_{c_1}$ and $\mathbf{v}_{c_2}$ are determined by

$$c_1 = \operatorname*{argmax}_{k \in \{1,\cdots,c_2-1\}} \boldsymbol{\theta}_{k,c_2}, \quad \text{and}$$

$$c_2 = \operatorname*{argmax}_{k \in \{1,\cdots,c_3-1\}} \boldsymbol{\theta}_{k,c_3}. \qquad (15)$$

Then, the three Cobb angles $\Theta = (\Theta_1, \Theta_2, \Theta_3)$ are determined by

$$\Theta_j = \boldsymbol{\theta}_{c_j, c_{j+1}} \quad \text{for } j = 1, 2, 3. \qquad (16)$$

III. EXPERIMENTS AND RESULTS

In this experiments, Python 3.6 and PyTorch 1.1 [22] were used to implement the proposed method. All training and evaluation were conducted on a workstation equipped with the two Intel(R) Xeon(R) E5-2630 v4 @ 2.20GHz CPUs, 128GB of DDR4 RAM, and four NVIDIA GeForce GTX 1080ti 11GB GPUs.

A. DATA

For the training and evaluation, spinal AP X-ray images and their label were provided by the Digital Imaging Group, London, ON, Canada [33]. All X-ray images were collected from individual patients. The provided images include 481 AP X-ray images for training and 128 AP X-ray images for testing. For training, we split the training data into 431 and 50 for training and validation, respectively. The labeled data include 3 Cobb angles and 68 landmarks representing the four corner points of the 12 thoracic and 5 lumbar vertebrae. The labeled data were manually annotated by two experts in London Health Sciences Center, London, ON, Canada [6], [33].

We call this dataset an internal dataset to distinguish it from external dataset described below. For an external validation,

we also used an external dataset provided by a different hospital. The external dataset include 20 AP X-ray images that were collected from individual patients.

The methods for generating the training data $\mathcal{D}_c := \{I^{(n)}, \mathbf{P}^{(n)}, \Psi^{(n)}\}_{n=1}^N$ and $\mathcal{D}_{\text{vec}} = \{I^{(n)}, \mathbf{V}^{(n)}\}_{n=1}^N$ are as follows:

1) For the given original X-ray images with a size of $h_o^{(n)} \times w_o^{(n)}$, we resized the all images to $512 \times w_o^{(n)}(512/h_o^{(n)})$ to generate $I^{(n)}$.
2) The intensity of resized image $I^{(n)}$ was scaled to the range of 0 to 1.
3) The centroids of each vertebra were given by the intersection of the middle of the width and the middle of the height in each vertebral body.
4) For the training of Centroid-net, we generated a ground-truth confidence map $\Psi^{(n)}$ using (1) and (2) with the ground-truth centroids $\mathbf{P}^{(n)}$.
5) For the training of M-net, we generated the ground-truth vertebral direction field $\mathbf{V}^{(n)}$ using (9).
6) For the data augmentation, we applied the random brightness, random contrast adjustment, and random rotation within an angle of $-10°$ to $10°$.

B. TRAINING OF THE PROPOSED NETWORK

We trained the proposed neural network by minimizing the loss functions in (3) and (8) using the Adam method [11]. Here, we choose a batch size of 4 by considering our computational capability. Batch normalization [9] was also applied. The learning rate was set to 10^{-4}. We trained the Centroid-net and the M-net for 1000 and 1500 epochs, respectively. The training was finished when the validation loss stopped decreasing.

C. QUANTITATIVE ANALYSIS OF COMPUTATIONAL EFFORT

We quantitatively analyze the computational effort, including computation time and memory requirements of the proposed neural networks.

1) COMPUTATION TIME

We provide the computation time of Centroid-net and M-net for training and test processes. In the training process, we measured the average time per epoch, which includes data loading, data augmentation, forward computation, and backward computation with optimization process. In the test process, the average time per batch was recorded with a batch size of 1. The test time included data loading and forward computation. The computation time is summarized in Table 1. Here, we used a single GeForce GTX 1080ti 11GB GPU to measure the training and test time.

2) MEMORY REQUIREMENTS

To estimate the total memory requirements, it requires to compute the number of all network parameters and intermediate activations [28]. To obtain the total memory in units of byte from this number, we multiply by 4 because
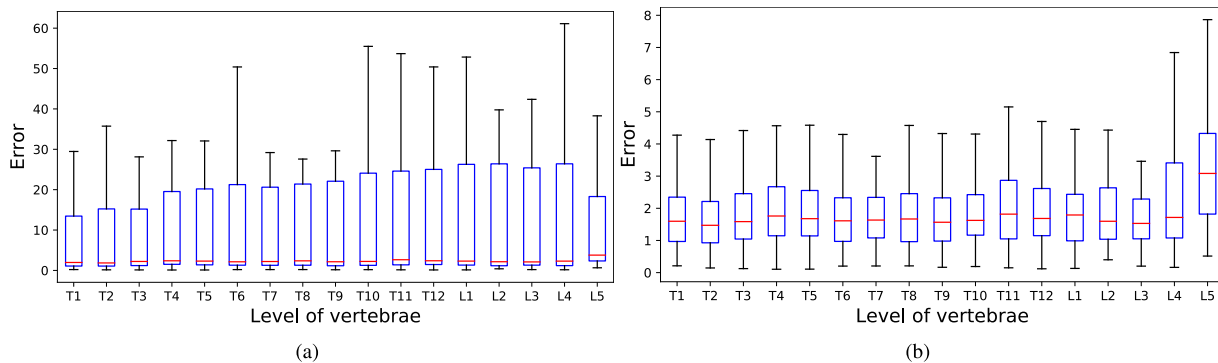
FIGURE 10. The centroids detection error of the proposed method is given by box plots. Median error is represented by red lines. (a) shows error between the output of the proposed method and ground-truth centroids and (b) shows the error between the output and closet ground-truth centroids.

TABLE 1. Quantitative analysis of computation time. For each training and test process, average computation time per epoch and batch are represented.

	Centroid-net	M-net
Training time(second/epoch)	29.1583	148.2518
Test time(second/batch)	0.0062	0.0105

every floating-point occupies the 4 bytes in a single-precision system.

We computed the amount of memory occupied by the network parameters. The Centroid-net has 1130080 trainable parameters, which occupy 4.31MB in single-precision floating-point format. The M-net has 10014888 trainable parameters occupying 38.20MB.

To estimate the memory requirements in the training process, we computed the total number of intermediate activations in forward pass and gradients in backward pass. Therefore, the required memory can be estimated as

$$\sum_{\text{layer}} N_{\text{batch}} \times N_{\text{activation}} \times 2 \times 4\text{byte}, \quad (17)$$

where N_{batch} is the batch size and $N_{\text{activation}}$ is the number of activations in each layer. Here, we multiply by 2 in the consideration of the backward pass, which occupies same memory as the forward pass. The Centroid-net requires 1.46GB with a batch size of 4 and the M-net requires 3.28GB with a batch size of 4.

D. QUANTITATIVE EVALUATION AND COMPARISON OF THE RESULTS

In this section, we provide quantitative evaluation of the proposed method on the internal testing dataset that includes 128 AP X-ray images.

1) IDENTIFICATION AND DETECTION OF VERTEBRAE

For a quantitative evaluation of the Centroid-net, we used the distance error between the output of the proposed method and ground-truth centroids in the pixel space. Fig. 10(a) shows the boxplot of the center position detection error.

We achieved a median error of 1.11 for the 17 vertebrae. A higher error occurred when the Centroid-net failed to predict the L5 vertebra. The identification of 17 vertebrae was deemed correct when the model predicted the 17 vertebrae with a distance error of less than 20 pixels. We achieved an identification rate of 90.6%. We also computed the distance error between the output and closest ground-truth centroids, as shown in Fig. 10(b). This error shows how close the predicted centroids are to the centroids of the vertebrae, regardless of the vertebral level.

2) COBB ANGLE MEASUREMENT

For an evaluation of the three Cobb angles given by the proposed method, we used circular mean absolute error (CMAE) [4] and symmetric mean absolute percentage error (SMAPE) [13], [17].

The CMAE between $\Theta^{(n)}$ and $\Theta_{\text{GT}}^{(n)}$ is defined using the mean of circular mean (CMEAN) as follows:

$$\text{CMAE} = \frac{1}{N} \sum_{n=1}^{N} \text{CMEAN}\left(\left| \Theta^{(n)} - \Theta_{\text{GT}}^{(n)} \right| \right), \quad (18)$$

where

$$\text{CMEAN}\left(\left| \Theta^{(n)} - \Theta_{\text{GT}}^{(n)} \right| \right)$$

$$= \arctan\left(\frac{\sum_{i=1}^{3} \sin\left(\left| \Theta_i^{(n)} - \Theta_{\text{GT},i}^{(n)} \right| \right)}{\sum_{i=1}^{3} \cos\left(\left| \Theta_i^{(n)} - \Theta_{\text{GT},i}^{(n)} \right| \right)} \right). \quad (19)$$

Here, $\Theta = (\Theta_1, \Theta_2, \Theta_3)$ indicates the three Cobb angles given by the proposed method and $\Theta_{\text{GT}} = (\Theta_{\text{GT},1}, \Theta_{\text{GT},2}, \Theta_{\text{GT},3})$ is the ground-truth of the three Cobb angles labeled by experts. The CMEAN was used to evaluate the angular quantity correctly. For example, the absolute error between the two angles 358° and 2° is 356°, whereas the difference in the actual angle is only 4°. To be precise, we first convert the three angles Θ from degrees into

TABLE 2. Comparison of the results for several different methods are reported. Evaluations of the Cobb angle measurement using multiple metrics are reported. The CMAE is represented with mean and standard deviation and the value of SMAPE is represented.

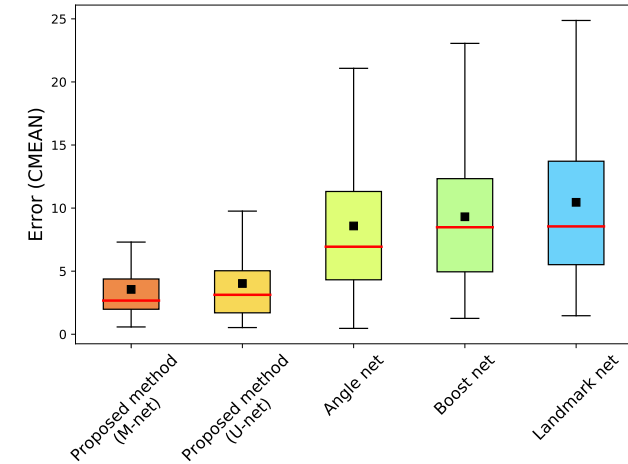| Method	Metric	
	CMAE(°)	SMAPE(%)
Proposed Method(M-net)	**3.51 ± 3.89**	**7.84**
Proposed Method(U-net)	3.90 ± 4.79	8.79
Angle Net	8.58 ± 7.60	20.35
Boost Net	9.31 ± 7.77	23.44
Landmark Net	10.46 ± 8.86	26.94

FIGURE 11. Box plots of circular mean error for several methods. The box plots represent interquartile ranges of circular mean error. Red lines denote the median value and black squares denote the mean value reported in Table 2.

radians to compute the sine and cosine in the CMEAN. Next, we convert the value of CMEAN from radians to degrees.

The SMAPE is defined as follows:

$$\text{SMAPE} = \frac{1}{N} \sum_{n=1}^{N} \frac{\sum_{i=1}^{3} \left| \Theta_i^{(n)} - \Theta_{\text{GT},i}^{(n)} \right|}{\sum_{i=1}^{3} \left(\Theta_i^{(n)} + \Theta_{\text{GT},i}^{(n)} \right)} \times 100\%. \quad (20)$$

Here, the SMAPE is a prediction accuracy represented by relative error and it has advantage of scale-independency and robustness to outliers.

We compare the proposed method with other existing methods: Angle-net in [6], Boost Net in [33], and Landmark Net in [6]. Additionally, we also used the U-net [25] based proposed method instead of using M-net. The quantitative evaluation results for several methods are reported in Table 2. As shown in Table 2, the M-net based proposed method achieves better performance in terms of CMAE and SMAPE. Box plots are represented to show the distribution of circular mean error (19) for several methods, as shown in Fig. 11.

To show the robustness of the proposed method against noise in a radiograph, we provide the performance evaluation of CMAE and SMAPE by adding different levels of Gaussian noise to the radiograph. The value of a noisy radiograph at pixel position $\mathbf{x}$ is defined by $(1 + r\epsilon)I(\mathbf{x})$ with $\epsilon \sim \mathcal{N}(0, 1)$, where r is a noise level (e.g., $r = 0.05$ for 5% noise).

TABLE 3. Comparison of the results for different levels of Gaussian noise. The CMAE and SMAPE are represented with different noise levels.

Noise level	Metric	
	CMAE(°)	SMAPE(%)
$r = 0.00$	**3.51 ± 3.89**	**7.84**
$r = 0.05$	3.66 ± 4.20	8.19
$r = 0.10$	4.33 ± 5.91	9.57
$r = 0.15$	4.68 ± 6.17	10.28
$r = 0.20$	5.67 ± 7.61	12.69
$r = 0.25$	6.82 ± 8.48	15.36
$r = 0.30$	7.88 ± 9.60	18.00

TABLE 4. Quantitative evaluation of Cobb angle measurements on the external testing dataset. The CMAE is represented with mean and standard deviation and the value of SMAPE is represented.

Method	Metric	
	CMAE(°)	SMAPE(%)
Proposed Method(M-net)	**4.11 ± 3.98**	**6.44**
Proposed Method(U-net)	4.38 ± 5.89	6.87

As shown in Table. 3, the proposed method still provides better performance than the existing methods in [6], [33] even though we added the Gaussian noise to the input radiograph. The experimental result also shows that the proposed method has robustness against noise in radiographs. Note that when we trained our model, we did not use the addition of random Gaussian noise as data augmentation.

E. QUANTITATIVE EVALUATION ON THE EXTERNAL DATASET

In the previous section, we demonstrated that the proposed method provides accurate and robust Cobb angle estimation on the internal testing dataset. Furthermore, to test generalization ability and robustness of the proposed method, we additionally assessed the Cobb angle measurement performance using 20 frontal radiographs from a different hospital. The end vertebrae designation and Cobb angle measurements of these radiographs were performed by two experienced radiologists in consensus. The quantitative evaluation of Cobb angle measurements on external dataset is reported in Table. 4. The proposed method achieved small error for this external dataset, showing that the proposed method has robustness and generalization ability.

F. COMPARISON BETWEEN CONFIDENCE MAP RESULTS WITH AND WITHOUT USING WEIGHTED LOSS

We next compare the localization performance with and without using a weighted loss function. Fig. 12 shows that a Centroid-net using a weighted loss function outperforms the result without a weighted loss function. As shown in Fig. 12(b) and (e), the Centroid-net without a weighted loss function fails to predict the T1 vertebra. We analyzed the results as follows: The model trained using a conventional loss function has to predict the T1 vertebra but not the C7 vertebra, which is adjacent to T1 and has similar a pattern as T1. This sometimes causes the model to fail in predicting the
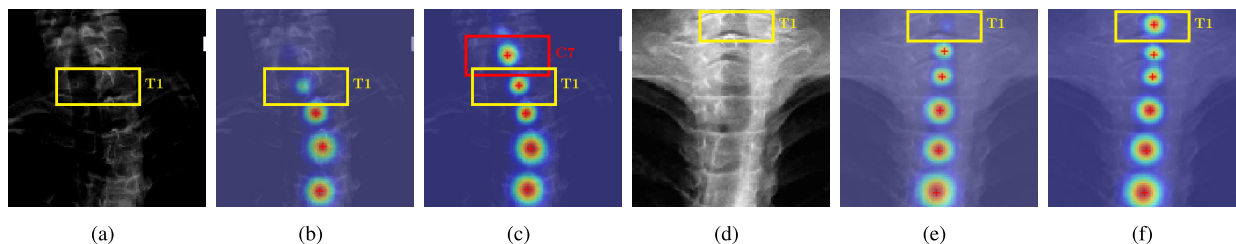
FIGURE 12. Comparison between confidence map results without ((b) and (e)) and with ((c) and (f)) using weighted loss function. (a) and (d) show the X-ray images. The yellow boxes show the thoracic vertebra region. In (b) and (e), the Centroid-net fails to predict the thoracic vertebra. In (c) and (f), the Centroid-net predicts the thoracic vertebra correctly. As shown in the red box in (c), the model predicts the cervical vertebra region owing to the weighted loss function.
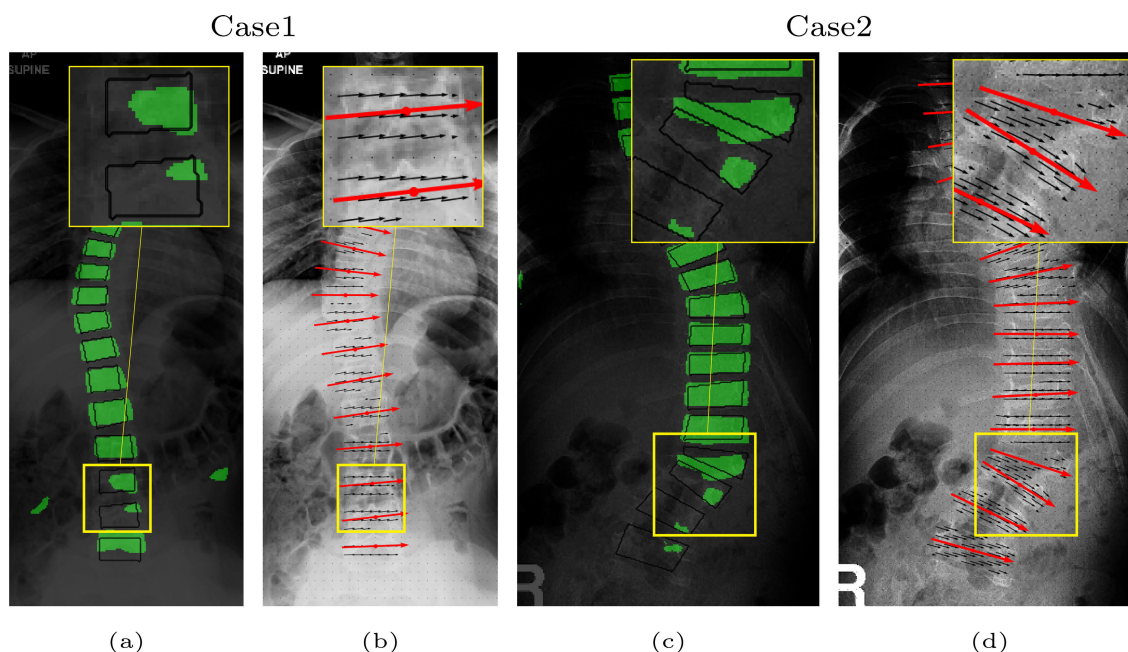

FIGURE 13. Comparison between the proposed method and segmentation-based method. (a) and (c) show the segmentation results. The black line in (a) and (c) denote the boundary of ground-truth segmentation. (b) and (d) show the vertebral-tilt field (black arrows) and predicted vertebral-tilt vector (red arrows). As shown in yellow box in (d), the proposed method provides slope of vertebra accurately even though vertebral-tilt field is not learned well over corresponding vertebral region.

T1 vertebra (Fig. 12(e)), or to predict T1 with low confidence (Fig. 12(b)). This problem arises from the fact that one of the two vertebrae with a similar shape has to be predicted and the other does not. On the other hand, when using the weighted loss function, we do not calculate the loss in the cervical vertebral region, resulting in predicting the thoracic and lumbar vertebrae accurately. In this case, the Centroid-net predicts the cervical vertebra with high probability owing to its similar shape.

G. COMPARISON BETWEEN THE PROPOSED METHOD AND SEGMENTATION-BASED METHOD

We qualitatively compared the results between the proposed method and the segmentation-based method to show the advantage of the proposed method through a vector field approach. The M-net was used to segment the 17 vertebrae from frontal radiographs. As shown in Fig. 13(a) and (c),

the segmentation method fail to describe a vertebral region accurately. In this case, the segmentation-based method cannot provide the slope of the vertebra since it requires highly accurate boundary segmentation. The segmentation-based methods require additional edge detection method such as hough transform [2], [3] to identify the vertebral endplates. On the other hand, the proposed method shown in Fig. 13(b) and (d) provides the accurate estimation slope of vertebra denoted by the red arrows even though the vertebral-tilt field did not encode the vector at each pixel in the region of vertebrae as shown in yellow box in Fig. 13(d).

H. QUALITATIVE EVALUATION OF THE PROPOSED METHOD

For a qualitative evaluation of the proposed method, we visualized the results on internal testing dataset with six selected subjects, as shown in Fig. 14.
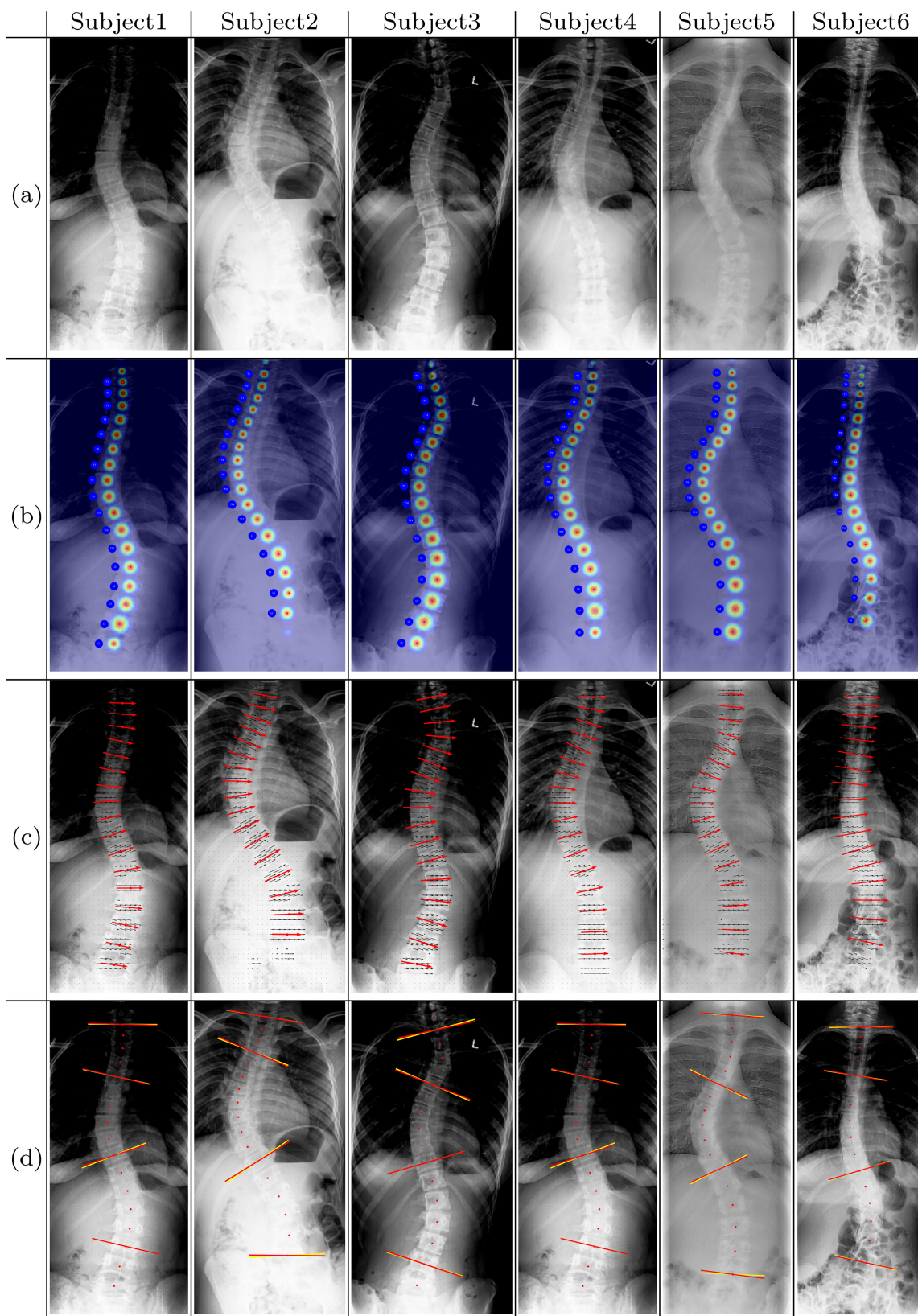
FIGURE 14. Qualitative evaluation of the results for six subjects. (a) AP X-ray images, (b) confidence map with estimated centroids, (c) vertebral-tilt vectors with vertebral-tilt field, and (d) the results of Cobb angle measurements. In (D), the yellow lines denote the ground-truth, and the red lines denote the results of the proposed method.

The results show that the proposed method properly provides 17 centroids (Fig. 14(b)) and 17 vertebral-tilt vectors (Fig. 14(c)), which are used for Cobb angle measurement. In Fig. 14(d), four end vertebrae given by the proposed method are presented with ground truth.

IV. DISCUSSION AND CONCLUSION

In this paper, we proposed a visually explainable Cobb angle measurement method using deep learning by considering clinician's decision process. The accurate and visually explainable Cobb angle measurement is important because of the following reasons: (1) Inaccurate measurement may lead clinicians to misinterpret scoliosis curve progression. (2) A visually explainable scoliosis assessment algorithm that not only calculates Cobb angle, and also identifies the most tilted vertebrae of the curve can improve the clinicians' workflow in the real clinical practice.

However, no existing method has achieved both accurate and visually explainable measurement of the Cobb angle in terms of clinical performance. The direct estimation methods suffers from lack of interpretability, even though it achieved highly accurate results. The indirect estimation methods face inherently inaccurate measurements owing to the dependency on the quality of landmark estimation or boundary segmentation of vertebrae, even though they visualize the intermediate decision process using the anatomical structure of spine.

To overcome the above-mentioned difficulties, we integrated the advantages of direct estimation method into the proposed indirect method. First, we used the confidence map regression method to localize and identify all vertebrae by taking a fully convolutional structure [27], while conventional coordinate regression methods require deep layers with a large number of network parameters and only take a fixed size of image as an input [23]. Next, the vertebral-tilt field was used to describe the slope of vertebrae by assigning a vector at each pixel inside region of vertebrae. This vector provides the slope of vertebra in a direct manner. An advantage of the vertebral-tilt field is that it can estimate the slope of the vertebrae even if the vectors were not well learned over the region of vertebra, as shown in Fig. 13(d). Finally, the Cobb angle was provided by combining the confidence map and vertebral-tilt field results. In this study, the vertebral-tilt field was implemented by the M-net which has shown improved performance in medical image segmentation because vertebral-tilt field provide pixel-wise dense prediction like image segmentation [27].

We demonstrated that the proposed method achieved a highly accurate Cobb angle estimation through a visually explainable system based on the confidence map and vertebral-tilt field. The performance evaluation on both the internal and the external testing dataset shows that the proposed method has robustness over frontal radiographs from different hospitals.

The proposed method has room for improvement. We believe that uncertainty quantification of the proposed Cobb angle measurement will be an important in our future study, where Bayesian deep learning method in [10], [14] or Gaussian process regression in [15], [16] could be adopted.

REFERENCES

[1] A. H and G. K. Prabhu, "Automatic quantification of spinal curvature in scoliotic radiograph using image processing," *J. Med. Syst.*, vol. 36, no. 3, pp. 1943–1951, Jun. 2012.

[2] H. Anitha, A. K. Karunakar, and K. V. N. Dinesh, "Automatic extraction of vertebral endplates from scoliotic radiographs using customized filter," *Biomed. Eng. Lett.*, vol. 4, no. 2, pp. 158–165, Jun. 2014.

[3] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, Jan. 1981.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York, NY, USA: Springer, 2002.

[5] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1302–1310.

[6] B. Chen, Q. Xu, L. Wang, S. Leung, J. Chung, and S. Li, "An automated and accurate spine curve analysis system," *IEEE Access*, vol. 7, pp. 124596–124605, 2019.

[7] R. M. Castelein, W. C. Chu, A. J. Danielsson, M. Dobbs, T. B. Grivas, C. A. Gurnett, K. D. Luk, A. Moreau, P. O. Newton, I. A. Stokes, S. L. Weinstein, and R. G. Burwell, "Adolescent idiopathic scoliosis," *Nature Rev. Disease Primers*, vol. 1, pp. 1–21, Jun. 2015.

[8] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.

[9] S. Batch normalization and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 1–5.

[10] K. Endall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5574–5584.

[11] D. P. Adam Kingma and J. Ba, "Adam: A method for Stochastic Optimization," *Proc. 3rd Int. Conf. Learn. Represent.*, 2014, pp. 1–10.

[12] H. Kim, H. S. Kim, E. S. Moon, C.-S. Yoon, T.-S. Chung, H.-T. Song, J.-S. Suh, Y. H. Lee, and S. Kim, "Scoliosis imaging: What radiologists should know," *RadioGraphics*, vol. 30, no. 7, pp. 1823–1842, Nov. 2010.

[13] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016.

[14] L. Akshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6402–6413.

[15] K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, "Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries," *IEEE Trans. Transport. Electrific.*, vol. 5, no. 4, pp. 1225–1236, Dec. 2019.

[16] K. Liu, Y. Li, X. Hu, M. Lucu, and W. D. Widanage, "Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3767–3777, Jun. 2020.

[17] S. Makridakis, "Accuracy measures: Theoretical and practical concerns," *Int. J. Forecasting*, vol. 9, no. 4, pp. 527–529, 1993.

[18] D. Malfair, A. K. Flemming, M. F. Dvorak, P. L. Munk, A. T. Vertinsky, M. K. Heran, and D. A. Graeb, "Radiographic evaluation of scoliosis: Review," *Amer. J. Roentgenol.*, vol. 94, no. 3, pp. 8–22, 2010.

[19] S. S. Mehta, H. N. Modi, S. Srinivasalu, T. Chen, S. W. Suh, J.-H. Yang, and H. R. Song, "Interobserver and intraobserver reliability of cobb angle measurement: Endplate versus pedicle as bony landmarks for measurement: A statistical analysis," *J. Pediatric Orthopaedics*, vol. 29, no. 7, pp. 749–754, Oct. 2009.

[20] R. T. Morrissy, G. S. Goldsmith, E. C. Hall, D. Kehl, and G. H. Cowie, "Measurement of the cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error," *J. Bone Joint Surgery*, vol. 72, no. 3, pp. 320–327, Mar. 1990.

[21] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[22] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037. [Online]. Available: https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library

[23] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Integrating spatial configuration into heatmap regression based CNNs for landmark localization," *Med. Image Anal.*, vol. 54, pp. 207–219, May 2019.

[24] J. E. H. Pruijs, M. A. P. E. Hageman, W. Keessen, R. van der Meer, and J. C. van Wieringen, "Variation in cobb angle measurements in scoliosis," *Skeletal Radiol.*, vol. 23, no. 7, pp. 517–520, Oct. 1994.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, vol. 9351, 2015, pp. 234–241.

[26] T. A. Sardjono, M. H. F. Wilkinson, A. G. Veldhuizen, P. M. A. van Ooijen, K. E. Purnama, and G. J. Verkerke, "Automatic cobb angle determination from radiographic images," *Spine*, vol. 38, no. 20, pp. E1256–E1262, Sep. 2013.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[28] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[29] F. E. Silva and L. G. Lenke, "Adolescent idiopathic scoliosis," *Surgical Manage. Spinal Deformities*, vol. 107, pp. 97–118, Oct. 2009.

[30] H. Sun, X. Zhen, C. Bailey, P. Rasoulinejad, Y. Yin, and S. Li, "Direct estimation of spinal cobb angles by structured multi-output regression," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 529–540.

[31] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1799–1807.

[32] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4724–4732.

[33] H. Wu, C. Bailey, P. Rasoulinejad, and S. Li, "Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostNet," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 10433, 2017, pp. 127–135.

[34] H. Wu, C. Bailey, P. Rasoulinejad, and S. Li, "Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-net," *Med. Image Anal.*, vol. 48, pp. 1–11, Aug. 2018.

[35] L. Wang, Q. Xu, S. Leung, J. Chung, B. Chen, and S. Li, "Accurate automated cobb angles estimation using multi-view extrapolation net," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101542.

HYE SUN YUN received the B.S. degree in mathematics from Ewha Womans University, Seoul, South Korea, in 2018. She is currently pursuing the Ph.D. degree in computational science and engineering with Yonsei University, South Korea. Her research interests include cephalometry, computerized tomography, and deep learning for medical image analysis.
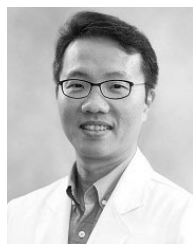
SUNGJUN KIM received the M.D. degree from the Yonsei University College of Medicine, Seoul, South Korea, in 1996, and the Ph.D. degree from Gangwon National University, Gangwon, South Korea, in 2010. Since 2011, he has been a Professor with the Department of Radiology, Yonsei University College of Medicine, South Korea. He participated as a Primary Investigator in an investigation published in *Scientific Reports* (Performance of the Deep Convolutional Neural Network Based Magnetic Resonance Image Scoring Algorithm For Differentiating Between Tuberculous and Pyogenic Spondylitis). He has published many articles-based on clinical researches using MRI. His research interests include musculoskeletal radiology, analysis of medical imaging, including radiography, ultrasonography, CT, and MRI, and machine learning application for medical image analysis and reconstruction.

KANG CHEOL KIM received the B.S. degree in physics from Chung-Ang University, Seoul, South Korea, in 2016. He is currently pursuing the Ph.D. degree in computational science and engineering with Yonsei University, South Korea. His research interests include nonlinear ill-posed inverse problem, image processing, computer vision, and deep learning for medical image analysis.

JIN KEUN SEO received the Ph.D. degree from the University of Minnesota, in 1991. Since 1995, he has been a Professor with Yonsei University, South Korea. He is currently the Director of the BK21plus of Computational Science and Engineering. He wrote books entitled *Nonlinear Inverse Problems in Imaging* (Wiley Press) and *Electro-Magnetic tissue properties MRI* (Imperial College Press). His research interests include inverse problems, mathematical modeling, image processing, partial differential equations, harmonic analysis, and deep learning for medical image analysis.

• • •