

Accepted Manuscript

Automated Comprehensive Adolescent Idiopathic Scoliosis
Assessment using MVC-Net

Hongbo Wu, Chris Bailey, Parham Rasoulinejad, Shuo Li

PII: S1361-8415(18)30287-1
DOI: [10.1016/j.media.2018.05.005](https://doi.org/10.1016/j.media.2018.05.005)
Reference: MEDIMA 1372



To appear in: *Medical Image Analysis*

Received date: 2 August 2017
Revised date: 24 April 2018
Accepted date: 11 May 2018

Please cite this article as: Hongbo Wu, Chris Bailey, Parham Rasoulinejad, Shuo Li, Automated Comprehensive Adolescent Idiopathic Scoliosis Assessment using MVC-Net, *Medical Image Analysis* (2018), doi: [10.1016/j.media.2018.05.005](https://doi.org/10.1016/j.media.2018.05.005)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- An effective end-to-end method of automated quantitative spinal curvature estimation for comprehensive AIS assessment
- A brand-new convolutional layer for effective multi-view feature learning
- A novel objective function and training algorithm for efficient multi-task learning

ACCEPTED MANUSCRIPT

Automated Comprehensive Adolescent Idiopathic Scoliosis Assessment using MVC-Net

Hongbo Wu^{a,b}, Chris Bailey^c, Parham Rasoulinejad^c, and Shuo Li^{a,b,*}

^a*Dept. of Medical Imaging, Western Univeristy, London, ON, Canada*

^b*Digital Imaging Group (DIG), London, ON, Canada*

^c*London Health Sciences Center, London, ON, Canada*

Abstract

Automated quantitative estimation of spinal curvature is an important task for the ongoing evaluation and treatment planning of Adolescent Idiopathic Scoliosis (AIS). It solves the widely accepted disadvantage of manual Cobb angle measurement (time-consuming and unreliable) which is currently the gold standard for AIS assessment. Attempts have been made to improve the reliability of automated Cobb angle estimation. However, it is very challenging to achieve accurate and robust estimation of Cobb angles due to the need for correctly identifying all the required vertebrae in both Anterior-posterior (AP) and Lateral (LAT) view x-rays. The challenge is especially evident in LAT x-ray where occlusion of vertebrae by the ribcage occurs. We therefore propose a novel Multi-View Correlation Network (MVC-Net) architecture that can provide a fully automated end-to-end framework for spinal curvature estimation in multi-view (both AP and LAT) x-rays. The proposed MVC-Net uses our newly designed multi-view convolution layers to incorporate joint features of multi-view x-rays, which allows the network to mitigate the occlusion problem by utilizing the structural dependencies of the two views. The MVC-Net consists of three closely-linked components: (1) a series of X-modules for joint representation of spinal structure (2) a Spinal Landmark Estimator network for robust spinal landmark estimation, and (3) a Cobb Angle Estimator network for

*Corresponding author

accurate Cobb Angles estimation. By utilizing an iterative multi-task training algorithm to train the Spinal Landmark Estimator and Cobb Angle Estimator in tandem, the MVC-Net leverages the multi-task relationship between landmark and angle estimation to reliably detect all the required vertebrae for accurate Cobb angles estimation. Experimental results on 526 x-ray images from 154 patients show an impressive 4.04° Circular Mean Absolute Error (CMAE) in AP Cobb angle and 4.07° CMAE in LAT Cobb angle estimation, which demonstrates the MVC-Net's capability of robust and accurate estimation of Cobb angles in multi-view x-rays. Our method therefore provides clinicians with a framework for efficient, accurate, and reliable estimation of spinal curvature for comprehensive AIS assessment.

Keywords: CNN, ConvNet, AIS, Scoliosis, Spinal Curvature, Multi-View, X-ray, Multi-Task Learning

1. Introduction

Adolescent Idiopathic Scoliosis (AIS) is the most common type of spinal deformity that occurs in children at the onset of puberty (Weinstein et al., 2008). Large cross-continental studies have shown that the prevalence of AIS can be as high as 5.2% and progression of large spinal deformities leads to poor quality of life and complications from injury to the heart and lungs (Asher and Burton, 2006). Clinicians currently make treatment decisions by assessing the degree of spinal deformity. It is therefore essential to have a reliable way of measuring spinal deformations.

However, spinal curvature assessment in clinical practice is both time-consuming and unreliable. The current gold standard for assessing spinal deformities is the manual measurement of Cobb angles (Cobb, 1948) based on anterior-posterior (AP) and lateral (LAT) x-ray images, which relates to the extent of lateral and sagittal curvature of the spine. This procedure shown in Fig. 1 relies on clinicians to identify the most tilted vertebrae endplates on the x-ray images (Vrtovec et al., 2009) and then measuring the Cobb angles between those vertebrae. The

low image contrast of x-ray along with occlusion of vertebrae makes it difficult to reliably measure Cobb angles. Therefore, the current clinical workflow will benefit with a more robust automated method of assessing spinal curvature. Existing efforts in automatically evaluating spinal curvature have been limited to the AP view, which measures the lateral curvature of the spine and is the easiest angles to measure. Sagittal spinal curvature (also known as kyphosis and lordosis) in LAT x-rays has been neglected by the literature due to the extreme difficulty in locating the thoracic vertebra occluded by the ribcage (Fig. 1b).

We therefore propose a Multi-view Correlation Network (MVC-Net) for Cobb angle estimation in both AP and LAT view x-rays, which provides a fully automated and reliable solution for comprehensive spinal curvature assessment in the clinical workflow. The proposed MVC-Net is designed to accomplish two tasks: (1) learn robust multi-view convolutional features which alleviates the problem of vertebral occlusion, and (2) exploit the correlation between vertebral landmarks and spinal curvature for robust automated end-to-end assessment of AIS. Fig. 2 shows our proposed MVC-Net architecture for fully automated comprehensive AIS assessment.

1.1. Previous Work

Existing work for computer aided quantitative assessment of spinal curvature fall in two categories: (1) **Segmentation** and (2) **Direct Estimation**.

Segmentation. Segmentation-based methods aim to quantify spinal curvature by first segmenting the required anatomical structures and then computing the relevant measurements based on the segmentation. Previous attempts at automated spinal curvature assessment used active contouring (Anitha and Prabhu, 2012), filtering (Anitha et al., 2014), and physics models (Sardjono et al., 2013) in order to locate the required vertebrae for Cobb angle measurement. While these methods are successfully applied to spinal curvature quantification (mainly Cobb angle measurements), the main disadvantages with these methods are the reliance on dedicated feature engineering and user bias. Feature engineering includes the application of filtering techniques such as hough trans-

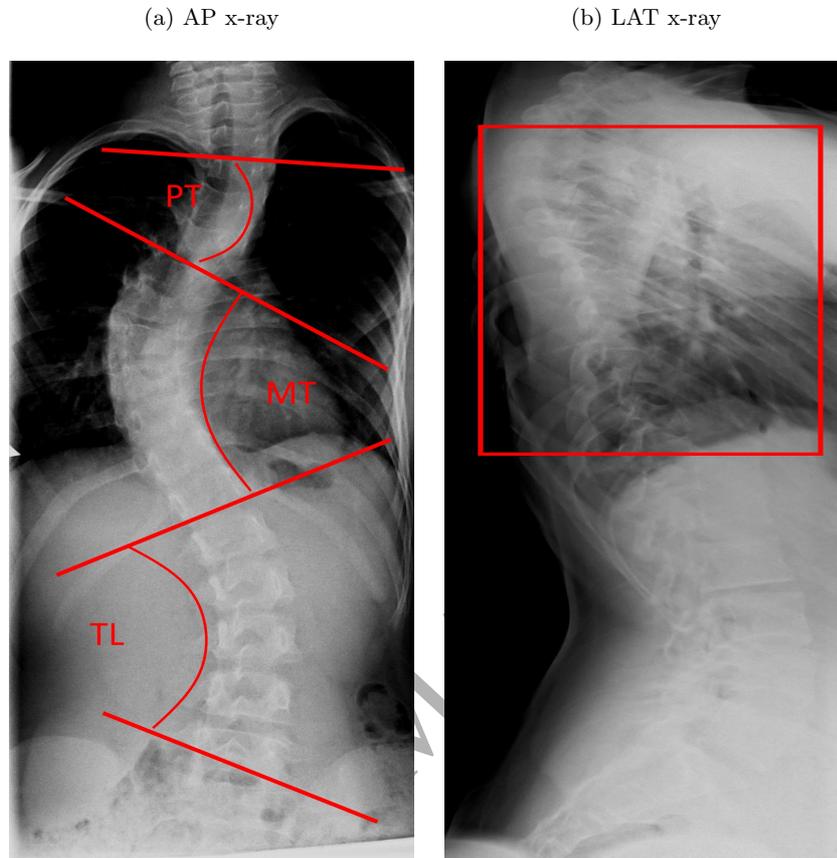


Figure 1: Assessment of AIS using Cobb Angles. (a) Proximal-Thoracic (PT), Main Thoracic (MT), Thoracic-Lumbar (TL) angles are measured to assess the extent of spinal curvature. (b) Angle measurement in LAT view is unreliable due to vertebral occlusion by the ribcage as indicated by the red box.

formations (Anitha and Prabhu, 2012; Anitha et al., 2014; Zhang et al., 2009) and morphological constraints such as active contours (Anitha and Prabhu, 2012). User bias occurs while selecting the region of interest (Zhang et al., 2010) when applying filtering and determining the relevant vertebra based on endplate angle (Zhang et al., 2009). Suboptimally engineered features and strong user bias may prevent segmentation-based methods from robust and efficient spinal curvature assessment.

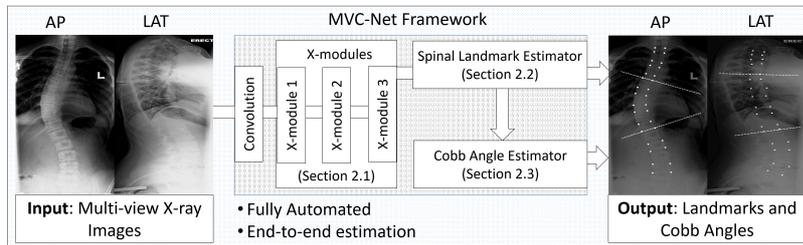


Figure 2: End-to-end framework for comprehensive AIS spinal curvature assessment using our MVC-Net. The MVC-Net is the first method to automatically estimate Cobb angles in both AP and LAT X-rays.

Direct Estimation. Direct estimation methods aim to capture the relationship between medical images and clinical measurements without the need for segmentation. These methods have recently gained huge success in quantitative estimation such as cardiac volume and spinal curvature in single view data (Xue et al., 2017b,a; Sun et al., 2017; Wu et al., 2017; Zhen et al., 2015), and have historically benefited from multi-task and machine learning for capturing the image feature-clinical measurement relationship. Xue et al. (2017b) used a Recurrent Neural Network in conjunction with a Convolutional Neural Network (ConvNet) to encode both temporal and spatial information for cardiac volume estimation. Sun et al. (2017) aimed to improve robustness of spinal curvature assessment by consolidating the tasks of vertebral landmark detection with Cobb angle estimation by exploiting the dependency between the two tasks. Wu et al. (2017) achieves robust spinal landmark estimation by automatically removing deleterious outlier features.

Despite their effectiveness in single view estimations, these methods are inadequate for the task of spinal curvature assessment in multi-view x-rays since the features learned do not explicitly capture the 3D spatial correlation between the AP and LAT views. Since these methods do not take into account the underlying relationship between AP and LAT x-rays, they are incapable of alleviating the problem of vertebrae occlusion. Therefore, accurate automated quantitative estimation in multi-view (e.g. AP and LAT for 2D, axial and sagit-

tal for 3D) data is still an outstanding clinical challenge that warrants further investigation.

1.2. Overview of MVC-Net Architecture

Our proposed MVC-Net architecture achieves fully automated comprehensive scoliosis assessment through joint multi-view feature learning and explicit reinforcement of output correlations through multi-task learning. The MVC-Net architecture is designed to creatively leverage the correlation between AP and LAT x-rays in order to capture latent multi-view representation of the spinal x-ray while utilizing the task-related features of spinal landmarks for accurate spinal curvature estimation. The MVC-Net architecture consists of three closely-linked components: (1) a series of cross-linked convolution layers for explicitly modeling joint multi-view representation of the spinal structure, (2) a Spinal Landmark Estimator (SLE) network that leverages the joint features for robust spinal landmark detection, and (3) a Cobb Angle Estimator (CAE) network that leverages the correlation between AP/LAT x-ray features and spinal landmarks for accurate Cobb angle estimation.

The inclusion of spinal landmarks in spinal curvature estimation provides three clear advantages: (1) Allows the network to utilize the dependency between Cobb angles and spinal landmarks (angles can be computed from landmarks) for positively reinforcing their reciprocal relationship during training, (2) allows clinicians to visually assess the accuracy of the estimation, and (3) opens up the option for clinicians to measure other spinal quantities such as height of spinal disc for osteoporosis assessment.

1.3. Contributions

We summarize our contributions as follows:

- For the first time, our proposed automated end-to-end framework provides an efficient and reliable method of quantitative spinal curvature estimation for AIS assessment in multiple views.

- We proposed a novel multi-view convolution layer that takes advantage of multi-view inputs in order to learn a more comprehensive representation of the latent spinal structure.
- We proposed a novel objective function and multi-task training algorithm that effectively optimizes the estimation of Cobb angles and spinal landmarks by exploiting the reciprocal relationship between the two tasks.

2. Proposed MVC-Net Architecture

The MVC-Net architecture achieves fully automated comprehensive AIS assessment through joint multi-view input feature learning and explicit reinforcement of reciprocal relationship between spinal landmark and Cobb angle. The MVC-Net architecture is designed to creatively leverage the correlation between AP and LAT views in order to acquire robust spinal landmarks for comprehensive scoliosis assessment. The MVC-Net architecture, illustrated in Fig. 3, captures the full extent of spinal structure for robust Cobb angle estimation using the X-modules, Joint Regression loss, and Iterative Landmark-Angle Training. The implementation of this model was done in Keras (Chollet, 2015). **For the sake of brevity in this section, we will designate W as the weight matrix and b as the bias vector of a particular layer.**

2.1. Joint Feature Learning using X-modules

Since both AP and LAT x-rays contain parts of the same physical structure, learning a joint representation of the two views helps alleviate the impact of information corruption such as occlusion. We therefore propose a series of X-module connections specifically designed to capture the underlying physical correlation of a structure using multi-view images. As shown in Fig. 4, the X-module greatly improves the robustness of feature space by utilizing the joint representation of AP and LAT image features.

Each X-module consists of three operations:

1. Weighted spatial summation of feature maps from each view.

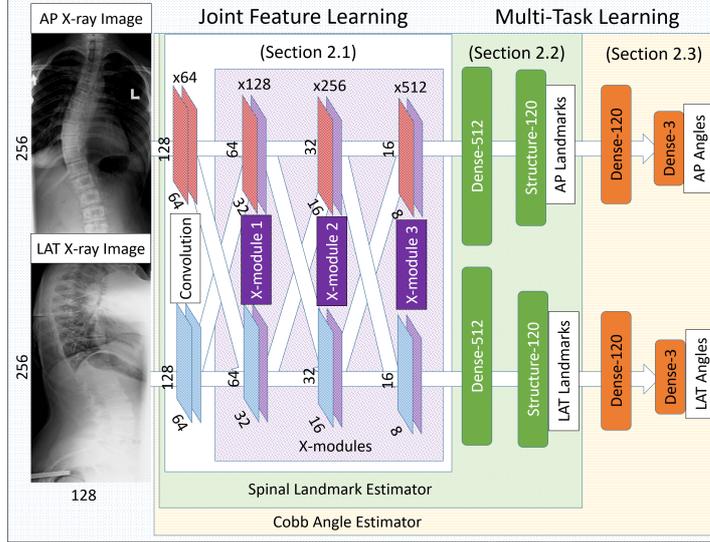


Figure 3: The MVC-Net provides accurate spinal landmarks and Cobb angles by optimizing the Spinal Landmark Estimator and Cobb Angle Estimator networks in tandem. The MVC-Net consists of (1) a series of X-modules shown in purple, (2) the Spinal Landmark Estimator shown in green, and (3) the Cobb Angle Estimator shown in orange.

2. Concatenation of the resulting summation with the original feature maps for each view (AP and LAT).
3. Spatial convolution of the concatenated feature maps for each view.

We define a spatial summation of feature maps as

$$F_c^\Sigma = Q_c * F_c^{AP} + (1 - Q_c) * F_c^{LAT}, \quad (1)$$

where F_c is the feature map of channel c , Q_c is the coefficient matrix (the same size as F_c) of the weighted spatial summation for each channel c . The $*$ denotes element-wise multiplication. This spatial summation procedure enables the network to explicitly model the 3D spatial correlation of the two views as a joint representation. The merged feature maps are subsequently channel-wise concatenated with the original feature maps from each respective view $\chi \in \{AP, LAT\}$ to allow the network to choose which features to use during

training:

$$G_c^X = [F_c^X, F_c^\Sigma]. \quad (2)$$

A convolution layer is then applied to the concatenated feature maps to process the features for the next layer.

$$F_c^X = \text{conv}(G_c^X). \quad (3)$$

We can thus formulate the joint feature learning as the composite function $J^X = \mathcal{F}(X^X)$ where \mathcal{F} is the series of convolution (*conv*) layer and X-modules (F^X), and X^X is the multi-view x-ray image.

All of the spatial convolutions in our model use 4×4 filters downsampled with a stride of 2 (without pooling), followed by batch normalization, Parametric Relu (He et al., 2015) activation function, and 25% additive Gaussian dropout regularizer (Hinton et al., 2012). We doubled the number of filter channels after each downsample.

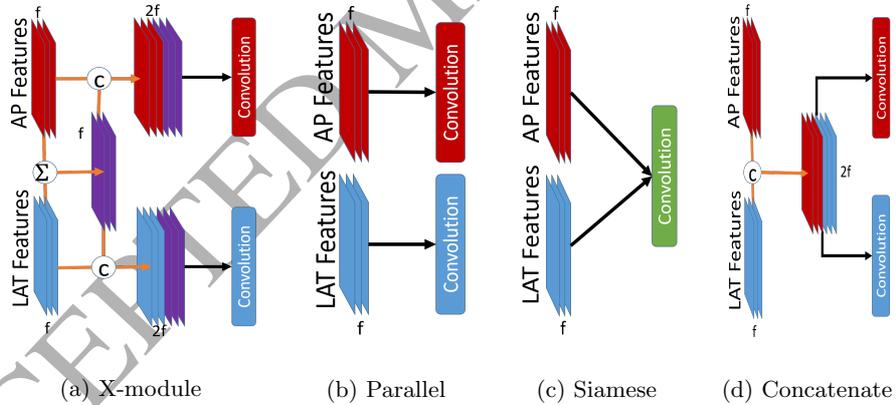


Figure 4: The X-module (a) is the only type of connection that utilizes joint features for learning discriminative representations of multi-view (AP and LAT) x-rays. Parallel (b) and Concatenate (d) modules do not explicitly capture joint multi-view features while Siamese (c) module does not have dedicated convolution layer for each view and is therefore susceptible to catastrophic interference (McCloskey and Cohen, 1989). The Σ corresponds to **weighted spatial summation of the AP/LAT features**, C represents concatenation across channels, and f denotes the number of filters.

2.2. Landmark Regression using Spinal Landmark Estimator

The Spinal Landmark Estimator (SLE) network maps the learned joined image features to spinal landmarks for more accurate Cobb angle estimation. The SLE first projects the joint features of each view (AP and LAT) to a \mathbb{R}^{512} hidden layer and then produces the landmark locations using a spinal structured output layer.

We define the landmark feature projection as:

$$P_{lm}^x = \tanh(J^x W + b), \quad (4)$$

where \tanh is the hyperbolic tangent function.

The spinal structured output layer contains a structural dependency matrix that captures the dependencies of the landmark coordinates. The structural dependency matrix S is used at the output in order to force a prior on the detected landmarks. S is a symmetric square matrix of binary values in which adjacent spinal landmarks are represented by 1 while distant landmarks are represented by 0 (Eqn. 5). This structural matrix was hand-crafted based on the ordering of physical location of the landmarks during annotation of the ground truth and remains constant during training. Fig. 5 shows two representative examples of landmark dependency: adjacent landmarks on the same vertebra or vertebral disc are considered to be dependent (represented by 1 in the matrix S).

$$\mathbf{S} \in \mathbb{R}^{120 \times 120} = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & 119 & 120 \end{matrix} \\ \begin{pmatrix} 1 & 1 & \cdot & 0 & 0 \end{pmatrix} & 1 \\ \begin{pmatrix} 1 & 1 & \cdot & 0 & 0 \end{pmatrix} & 2 \\ \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} & \dots \\ \begin{pmatrix} 0 & 0 & \cdot & 1 & 1 \end{pmatrix} & 119 \\ \begin{pmatrix} 0 & 0 & \cdot & 1 & 1 \end{pmatrix} & 120 \end{matrix} \quad (5)$$

This constraint limits the degree of freedom affecting each predicted spinal landmark to only neighbouring landmarks in order to improve regression accu-

racy. The output of the SLE (Y_{lm}^x) is thus defined as:

$$Y_{lm}^x = \sigma[(P_{lm}^x \cdot W + b) \cdot S], \quad (6)$$

where σ is the sigmoid logistic function and S is the landmark dependency matrix as defined in Eqn. 5.

The Spinal Landmark Estimator can thus be formalized as the mapping function $M_{lm} : X^x \rightarrow L^x$ where $X^x \in \mathbb{R}^{256 \times 128}$ are the AP/LAT x-ray images; and $L^x = [x^{60}, y^{60}]$ are the scaled x/y-coordinates of the AP/LAT spinal landmarks. In order to accommodate the differences in resolution, we scaled the

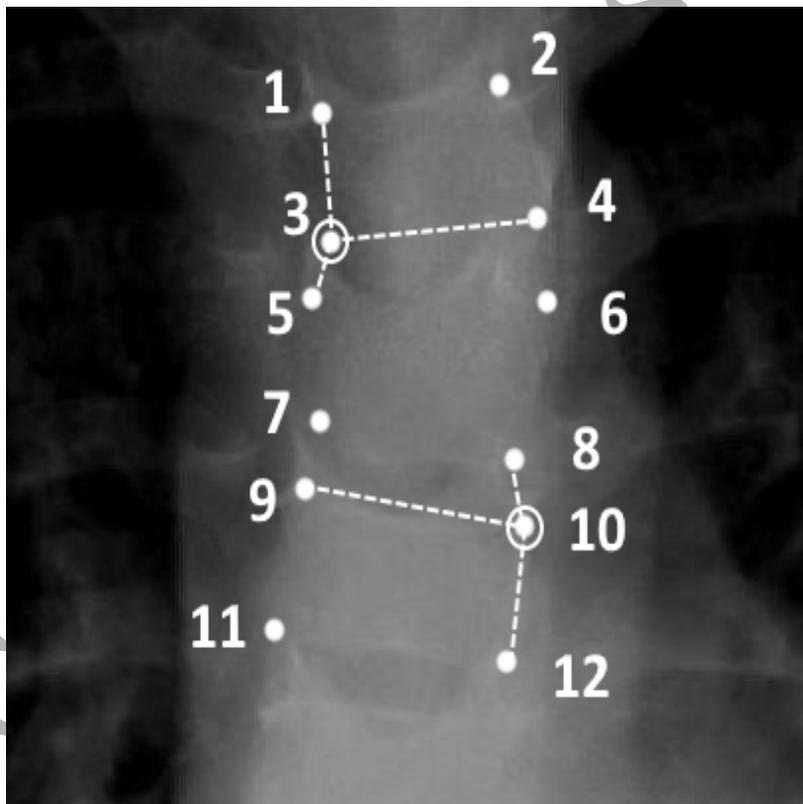


Figure 5: The neighbourhood of landmark 3 includes 1/4 (adjacent landmarks on same vertebra) and 4/5 (adjacent landmarks on same disc). The neighbourhood of landmark 10 includes 8/9 (adjacent landmarks on same disc) and 9/12 (adjacent landmarks on same vertebra). The dashed lines represent the connectivity of landmark dependency of the circled landmarks.

landmark coordinates to between 0 and 1 by dividing the original x/y positions with the width/height of the original image respectively.

2.3. Angle Regression using Cobb Angle Estimator

The Cobb Angle Estimator (CAE) is a dedicated network for mapping joint spinal features to Cobb angles. The CAE first projects the landmark feature projection to a common feature space of \mathbb{R}^{120} (same dimension as landmark coordinates) and add the result to the landmark output to adaptively mitigate the error associated with landmark prediction. The adjusted landmarks are then used to predict the Cobb angles via a densely connected output layer. The CAE is thus encouraged to explicitly map relevant multi-view x-ray features (with respect to important spinal landmarks) to the corresponding Cobb angles by utilizing the landmarks learned by the SLE.

We define the angle feature projection as:

$$P_{an}^x = \tanh(P_{lm}^x W + b), \quad (7)$$

where P_{lm}^x is the AP/LAT landmark feature projection as defined in Eqn. 4. We then add the projection to the landmarks to compute the adjusted landmark coordinates \hat{Y}_{lm}^x

$$\hat{Y}_{lm}^x = Y_{lm}^x + P_{an}^x. \quad (8)$$

Finally, the output of the CAE (Y_{an}^x) is computed as:

$$Y_{an}^x = \hat{Y}_{lm}^x \cdot W + b. \quad (9)$$

We can formulate this network as the mapping function $M_{an} : X^x \rightarrow A^x$ where $\{\alpha_i\}_{i=1}^3 \in A^x$ are the PT, MT, and TL Cobb angles of respective AP/LAT x-ray images.

Since spinal landmarks are essential for computing Cobb angles, we have explicitly tied a task-related relationship between the landmark and angle estimation by directly connecting the CAE to SLE in an end-to-end fashion. This allows the CAE to more effectively utilize multi-task learning for more accurate Cobb angle estimation.

3. Joint Regression Loss

We propose a novel objective function named Joint Regression Loss (JR-loss) that improves spinal landmark and Cobb angle estimation accuracy by combining a *correlation loss* with a *robust regression loss*. The JR-loss not only minimizes estimation error but also improves the correlation between the prediction and ground truth.

3.1. Landmark Loss

Robust Regression Loss The traditional loss function for regression problems is the Mean Squared Error (MSE) loss:

$$\mathcal{L}_{mse}(X^x, L^x, \theta) = \mathbf{MEAN}(M_{lm}(X^x; \theta) - L^x)^2. \quad (10)$$

where X^x is an X-ray image, L^x is its corresponding ground truth landmark, θ is the parameters of the SLE network M_{lm} as described in section 2.2, and **MEAN** is the element-wise arithmetic mean.

However, this loss is not robust since large errors become squared and will therefore skew the overall mean of the loss. In order to minimize the effect of large errors during training, we use the log of hyperbolic cosine as the objective function instead:

$$\mathcal{L}_{reg}(X^x, L^x, \theta) = \mathbf{MEAN}[\log \frac{1}{2}(e^{M_{lm}(X^x; \theta) - L^x} + e^{L^x - M_{lm}(X^x; \theta)})]. \quad (11)$$

Correlation Loss To make sure that our predictions not only minimize the error but also conform to the spinal curvature of our ground truth, we use the Pearson loss defined as:

$$\rho = \frac{\mathbf{MEAN}[M_{lm}(X^x; \theta)L^x] - \mathbf{MEAN}[M_{lm}(X^x; \theta)]\mathbf{MEAN}[L^x]}{\mathbf{STD}[M_{lm}(X^x; \theta)]\mathbf{STD}[L^x]}, \quad (12)$$

where **STD** is the standard deviation.

Since the output of Pearson coefficient ρ ranges from -1 (negative correlation) to 1 (perfect correlation), we rearranged it as follows in order to force the output to be between 0 (perfect correlation) and 2 (negative correlation):

$$\mathcal{L}_{cor}(X^x, L^x, \theta) = 1 - \rho. \quad (13)$$

The joint regression loss function for optimizing SLE is thus defined as:

$$\mathcal{L}_{lm}(X^X, L^X, \theta) = \mathcal{L}_{cor}(X^X, L^X, \theta) + \phi \mathcal{L}_{reg}(X^X, L^X, \theta), \quad (14)$$

where ϕ is a scaling factor controlling the relative importance of the regression loss term.

Since we want to ensure equal importance for both accuracy and correlation, a grid search of ϕ between 1 and 10 (at interval of 1) was performed to select the optimal value of ϕ that balances the magnitudes of the robust regression and correlation loss within the first epoch. The empirically determined value of $\phi = 5$ was used in all of our experiments.

3.2. Angle Loss

While the joint regression loss works well for euclidean quantities like distance, it is not robust in dealing with circular quantities due to the phase-wrapping property of angles. For instance, the euclidean distance between 5° and 355° is 350° while in reality, the difference is only 10° . We will therefore have to modify the loss function for circular quantities to alleviate this. To do so, we can consider performing arithmetics in polar coordinate space and then converting the result back to angular space using trigonometry.

The angle α can be decomposed as polar coordinates on the unit circle:

$$\alpha \rightarrow \{x = \cos(\alpha), y = \sin(\alpha)\}. \quad (15)$$

Given a list of N angles $[\alpha_0, \dots, \alpha_N]$, we can define their circular mean as:

$$\bar{x} = \frac{1}{N} \sum_{i=0}^N \cos(\alpha_i) \quad (16)$$

$$\bar{y} = \frac{1}{N} \sum_{i=0}^N \sin(\alpha_i) \quad (17)$$

$$\mathbf{CMEAN}([\alpha_0, \dots, \alpha_N]) = \arctan\left(\frac{\bar{y}}{\bar{x}}\right). \quad (18)$$

When computing the loss for Cobb angles, we replace the arithmetic mean (**MEAN**) with the circular mean (**CMEAN**) as defined by Eqn. 18. The robust

regression loss (Eqn. 11) for angles thus becomes:

$$\mathcal{L}^{\alpha}_{reg}(X^x, A^x, \Theta) = \mathbf{CMEAN}[\log \frac{1}{2}(e^{M_{an}(X^x; \Theta) - A^x} + e^{A^x - M_{an}(X^x; \Theta)}), \quad (19)$$

where A^x is the 6 ground truth Cobb angles and Θ is the parameters of the CAE network M_{an} as described in section 2.3.

The correlation loss (Eqn. 13) is also modified to:

$$\mathcal{L}^{\alpha}_{cor}(X^x, A^x, \Theta) = 1 - \frac{\mathbf{CMEAN}[M_{an}(X^x; \Theta)A^x] - \mathbf{CMEAN}[M_{an}(X^x; \Theta)]\mathbf{CMEAN}[A^x]}{\mathbf{STD}[M_{an}(X^x; \Theta)]\mathbf{STD}[A^x]}. \quad (20)$$

The objective function for optimizing CAE thus becomes:

$$\mathcal{L}_{an}(X^x, A^x, \Theta) = \mathcal{L}^{\alpha}_{cor}(X^x, A^x, \Theta) + \phi \mathcal{L}^{\alpha}_{reg}(X^x, A^x, \Theta). \quad (21)$$

4. Iterative Landmark-Angle Training (ILAT) Algorithm

The ILAT algorithm allows the MVC-Net to leverage the similarities between tasks of the SLE and CAE for more robust spinal curvature estimation. To accomplish this, the AP/LAT x-ray images and their corresponding landmark coordinates and Cobb angles are used to train the MVC-Net through Mini-batch Stochastic Gradient Descent optimization with momentum of 0.95 and a starting learning rate of 0.01. A two-stage alternating optimization scheme was used to train each component network of the MVC-Net iteratively to ensure a synergistic effect when optimizing the two related tasks (Fig. 6).

For each batch during training, we optimized the Spinal Landmark Estimator and Cobb Angle Estimator in tandem (one after the other). The joint optimization scheme allows the MVC-Net to leverage the reciprocal relationship between spinal landmarks and Cobb angles for more accurate Cobb angle estimation (since Cobb angles can be computed from spinal landmarks). We trained the model for 100 epochs while halving the learning rate after every 10 epochs. The general outline of the iterative 2-stage training scheme is summarized in Alg. 1.

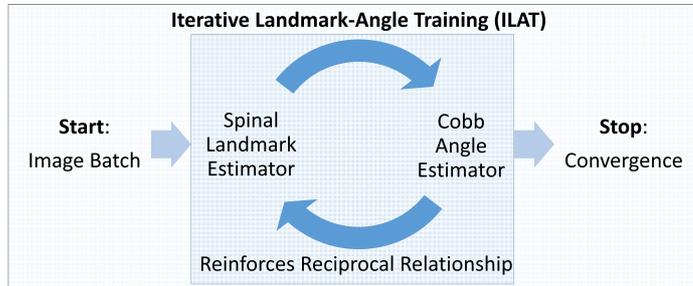


Figure 6: The ILAT algorithm optimizes the Landmark Estimator and Cobb Angle Estimator in tandem for each training batch during training such that the reciprocal relationship between spinal landmarks and Cobb angles is reinforced.

Algorithm 1 Iterative Landmark-Angle Training

- 1: Initialization: randomly initialize Spinal Landmark Estimator (SLE) parameters θ and Cobb Angle Estimator (CAE) parameters Θ .
 - 2: Set $\phi = 5$
 - 3: **repeat**
 - 4: **for** $b \in \{1, \dots, nbatches\}$ **do**
 - 5: Compute loss of SLE using 14
 - 6: Update θ using backpropagation
 - 7: Compute loss of CAE using 21
 - 8: Update Θ using backpropagation
 - 9: **end for**
 - 10: **until** Convergence
-

5. Experimental Setup

5.1. Dataset

Our dataset consists of 526 spinal x-ray images from 154 patients equally divided between AP and LAT views. The images were provided by local clinicians and all of them show signs of scoliosis to varying extent. The average resolution of the images is 0.26mm/pixel. In order to reduce potential problems caused by different image resolutions, we scaled all the images to 128 pixels in

width and 256 pixels in height. We also normalized the labeled landmark coordinates as a percentage of the scaled image (i.e. the midpoint $\{64, 128\}$ of the image becomes $\{0.5, 0.5\}$). Since the cervical vertebrae (vertebrae of the neck) are seldom involved in spinal deformity (Group, 2008), 15 vertebrae composed of the thoracic and lumbar vertebrae was selected by an expert as part of the spinal curvature evaluation. Each vertebra is marked by four landmarks with respect to its four corners resulting in 60 points per spinal image. These landmarks comprises the ground truth (GT) of our data and are used to train the SLE. The Cobb angle GT data is then computed from the GT landmarks and used for training the CAE. At testing time, the images are divided into training and validation set for each fold such that no patient is included in both sets. The results from our experiments were attained by averaging performance over 10-fold patient-wise crossvalidation.

5.2. Data Augmentation

In order to allow our model to generalize better to unseen data, we introduce the necessary invariance through dynamic data augmentation during training. At training time, the images are randomly transformed to account for such invariances. In terms of spinal x-ray images we want to introduce shift, intensity, and rotation invariance.

We therefore augmented the data with:

- Randomly introducing Poisson distributed noise to the images to simulate intensity variance.
- Randomly rotating the images up to 5° to allow for flexibility in rotation.
- Randomly shifting the images by 1% to encourage shift invariance.

5.3. Performance Metric

Spinal Landmark. We evaluate the accuracy of our Spinal Landmark estimation using the following two metrics: (1) The Mean Absolute Error (MAE) and (2) Pearson correlation coefficient (ρ) between the predicted values and annotated ground truth.

The MAE is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N \mathbf{MEAN}[|M_{lm}(X_i^\chi; \theta) - L_i^\chi|] \quad (22)$$

where N is the number of $\chi \in \{AP, LAT\}$ x-ray images. The ρ -correlation is defined as:

$$\rho = \frac{1}{N} \sum_{i=1}^N \mathbf{MEAN}\left[\frac{\mathbf{MEAN}[M_{lm}(X_i^\chi; \theta)L_i^\chi] - \mathbf{MEAN}[M_{lm}(X_i^\chi; \theta)]\mathbf{MEAN}[L_i^\chi]}{\mathbf{STD}[M_{lm}(X_i^\chi; \theta)]\mathbf{STD}[L_i^\chi]}\right] \quad (23)$$

Cobb Angle. For Cobb angle estimation, we used circular MAE and Symmetric Mean Absolute Error (SMAPE) instead of ρ -correlation to represent the relative error.

The circular MAE is defined as:

$$CMAE = \frac{1}{N} \sum_{i=1}^N \mathbf{CMEAN}[|M_{an}(X_i^\chi; \Theta) - A_i^\chi|] \quad (24)$$

The SMAPE metric is defined as:

$$SMAPE = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{SUM}[|M_{an}(X_i^\chi; \Theta) - A_i^\chi|]}{\mathbf{SUM}[M_{an}(X_i^\chi; \Theta) + A_i^\chi]} \times 100\% \quad (25)$$

where **SUM** is the element-wise summation of a vector.

6. Results and Analysis

6.1. Overview of Results

The effectiveness of our MVC-Net for comprehensive spinal curvature assessment has been thoroughly validated through two steps. (1) We first demonstrate that our method has the best capability of accurately and robustly assessing spinal curvature compared to current state of the art methods. (2) We then rationalize the effectiveness of our three innovations (X-module, JR-loss, ILAT) by examining in isolation the effectiveness of each component. For our experiments, we showed results for both directly estimated angles as well as angles measured from the directly estimated landmarks. In almost all cases, directly estimated angles using multi-task learning outperformed angles measured from the landmarks, which validates the advantages of multi-task learning.

As shown in table 1, the MVC-Net contains obvious advantage over existing methods that only use a single-view. The accurate estimation of spinal landmarks and Cobb angles can be attributed to the robust multi-view feature representation capabilities of X-modules and multi-task training of the SLE and CAE. Specifically, our method achieved the lowest estimation error of 0.0398 ± 0.0153 (scaled), 0.0459 ± 0.0198 (scaled), 4.04 ± 0.99 (deg), 4.07 ± 1.19 (deg) with respect to AP/LAT landmarks (scaled pixels), and AP/LAT angles (degrees).

Table 1: The **MVC-Net** achieved the lowest error in landmark and Cobb angle estimation on 10-fold crossvalidation compared to other methods. The number after \pm represents the standard deviation.

Directly Estimated Landmarks				
Method	AP		LAT	
	MAE (scaled)	ρ	MAE (scaled)	ρ
Sun et al. (2017)	0.0419 ± 0.0181	0.945 ± 0.0508	0.0517 ± 0.0217	0.921 ± 0.0734
Criminisi et al. (2011)	0.0442 ± 0.0148	0.954 ± 0.0332	0.0505 ± 0.0183	0.943 ± 0.0523
Wu et al. (2017)	0.0400 ± 0.0160	0.951 ± 0.0389	0.0509 ± 0.0218	0.923 ± 0.0762
MVC-Net	0.0398 ± 0.0153	0.956 ± 0.0172	0.0459 ± 0.0198	0.945 ± 0.0214
Directly Estimated Angles				
Method	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
Sun et al. (2017)	6.26 ± 2.81	37.08 ± 16.91	6.07 ± 2.08	24.23 ± 9.10
Criminisi et al. (2011)	5.67 ± 3.16	35.62 ± 15.41	6.20 ± 2.39	20.55 ± 10.53
Wu et al. (2017)	5.37 ± 2.98	41.35 ± 13.33	5.54 ± 2.95	25.14 ± 12.68
MVC-Net	4.04 ± 0.99	35.85 ± 15.46	4.07 ± 1.19	23.72 ± 9.87
Angles Measured from Predicted Landmarks				
Method	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
Sun et al. (2017)	6.37 ± 3.52	41.34 ± 20.42	6.85 ± 3.42	24.75 ± 12.33
Criminisi et al. (2011)	6.27 ± 3.38	42.62 ± 21.49	6.34 ± 2.58	22.12 ± 12.47
Wu et al. (2017)	6.44 ± 4.67	43.60 ± 20.73	6.67 ± 4.54	26.00 ± 12.62
MVC-Net	6.08 ± 3.07	45.48 ± 19.96	6.40 ± 2.96	32.40 ± 14.69

6.2. Benefits of X-Modules

The results in Table 2 indicate that the X-module is the best method for incorporating joint feature learning. The X-module was compared against three

other methods of incorporating joint features, namely Parallel, Siamese, and Concatenate (Fig. 4b-d). The Parallel connection uses a separate convolution layer for each of the two views, the Concatenate connection concatenates the feature maps of the two x-ray views while Siamese connection shares the same layer for both views. The network trained using the X-module achieved the lowest MAE and highest correlation coefficient compared to networks using Parallel, Siamese, or Concatenate connections. The visual validation in Fig. 7 shows that the X-module produces more discriminative feature maps as indicated by the presence of noticeable spinal structures in both AP and LAT features. The inclusion of these spinal structures in the feature map helps mitigate the vertebrae occlusion problem, resulting in higher overall accuracy as is validated by our results. Fig. 8 provides visual confirmation that the X-module connection achieves superior performance compared to other types of connections.

6.3. Benefits of Joint Regression Loss Function

The novel JR-loss function not only improves robustness of spinal landmark detection but also directly reinforces the accuracy of Cobb angle estimation. The JR-loss optimizes both regression accuracy as well as the output correlation, which enables the network to utilize the general trend of spinal landmarks for more accurate Cobb angle estimation. Table 3 compares our method against networks with the same architecture trained using the following loss functions: (1) the mean squared error loss ‘mse-loss’ (Eqn. 10), (2) the cosine-proximity loss ‘cos-loss’ (Scipy, 2014), (3) the pearson correlation loss ‘cor-loss’ (Eqn. 13), and (4) the robust regression loss ‘reg-loss’ (Eqn. 11). The JR-loss demonstrates obvious advantage by achieving lower MAE error and standard deviation compared to networks trained with other loss functions.

6.4. Benefits of ILAT Algorithm

Table 4 shows that the network trained using the ILAT algorithm has the best performance compared to the sequential, angles-only, and landmark-only training methods. The sequential training scheme involves first optimizing

Table 2: The X-module represents the best connectivity for multi-view feature learning. The number after \pm represents the standard deviation.

Directly Estimated Landmarks				
Method	AP		LAT	
	MAE (scaled)	ρ	MAE (scaled)	ρ
Siamese	0.0501 \pm 0.0242	0.926 \pm 0.0231	0.0528 \pm 0.0394	0.895 \pm 0.0458
Parallel	0.0430 \pm 0.0164	0.944 \pm 0.0425	0.0567 \pm 0.0209	0.908 \pm 0.0432
Concat	0.0494 \pm 0.0180	0.932 \pm 0.0476	0.0505 \pm 0.0252	0.942 \pm 0.0374
X-Module	0.0398 \pm 0.0153	0.956 \pm 0.0172	0.0459 \pm 0.0198	0.945 \pm 0.0214
Directly Estimated Angles				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
Siamese	5.45 \pm 2.62	44.16 \pm 17.18	5.44 \pm 2.03	45.62 \pm 18.28
Parallel	5.18 \pm 2.73	44.80 \pm 20.45	5.56 \pm 3.26	43.07 \pm 18.42
Concat	4.94 \pm 2.76	46.2 \pm 11.49	5.59 \pm 3.64	41.78 \pm 19.68
X-Module	4.04 \pm 0.99	35.85 \pm 15.46	4.07 \pm 1.19	23.72 \pm 9.87
Angles Measured from Predicted Landmarks				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
Siamese	6.82 \pm 4.00	47.56 \pm 21.43	9.14 \pm 5.76	39.40 \pm 19.85
Parallel	6.09 \pm 3.26	44.80 \pm 20.45	5.75 \pm 3.10	29.39 \pm 12.89
Concat	6.94 \pm 3.94	44.61 \pm 11.50	7.14 \pm 5.41	35.31 \pm 15.85
X-Module	6.08 \pm 3.07	45.48 \pm 19.96	6.40 \pm 2.96	32.40 \pm 14.69

the SLE until convergence followed by optimizing the CAE. Angles-only and landmark-only refers to solely training the CAE or SLE respectively. The ILAT algorithm demonstrates up to 3 times lower MAE error compared to the sequentially trained network commonly used in traditional pipelines. This is attributed to the ILAT algorithm bringing the network parameters closer to a global minimum during each stage of the gradient descent by taking advantage of the interdependency between spinal landmarks and Cobb angles. Further comparison with the angles-only and landmark-only method proves that the ILAT uses positive reinforcement between spinal landmark and Cobb angle prediction in order to reach a lower error.

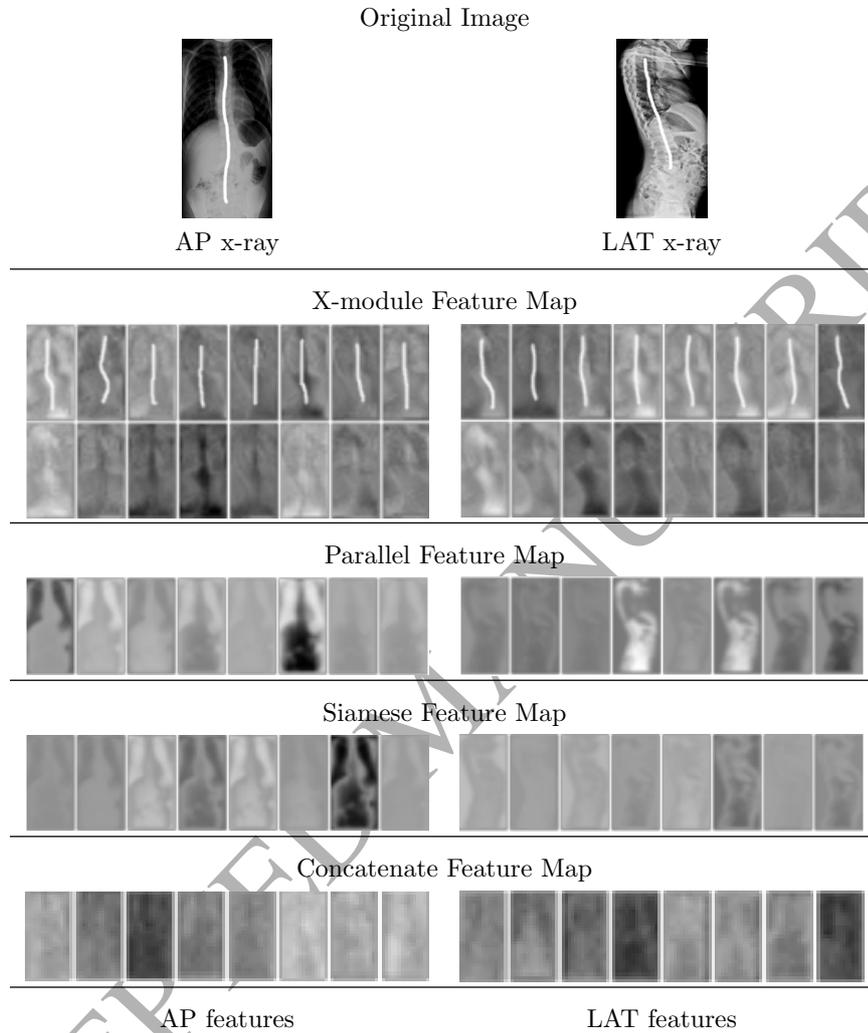
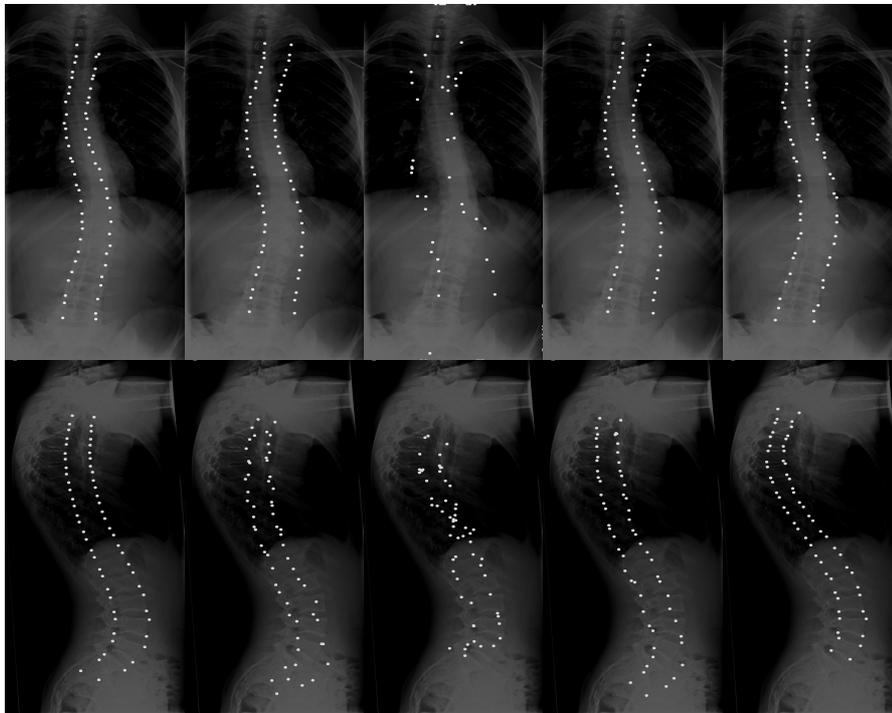


Figure 7: The X-module is able to learn more discriminative features compared to Parallel, Siamese, and Concatenate connections as shown by the white outline in the first row of the feature maps (drawn by the authors for visualization purpose). The X-module feature map exhibits more defined spinal structure similar to the original x-ray image compared to the feature maps produced by other connection types.

6.5. Analysis

The proposed MVC-Net architecture achieved the lowest average MAE of 0.0398 (AP), 0.0459 (LAT) scaled pixel values and the highest correlation co-



(a) MVC-Net (b) Parallel (c) Siamese (d) Concat (e) Label

Figure 8: The model trained using X-modules (a) shows the best landmark quality. The spinal landmarks detected by our method more closely resembles the labeled (e) landmarks compared to other connection types (b – d).

efficient of 0.956 (AP), 0.945 (LAT) for predicting spinal landmarks and 4.04 (AP), 4.07 (LAT) degrees for Cobb angle estimation. This is due to the contributions of 1) the X-Module, which successfully learned joint AP and LAT features, 2) the novel JR-loss objective function, which ensures both regression accuracy and consistency of the landmark coordinates, and 3) the strategically designed ILAT algorithm, which leverages the relationship between spinal landmarks and Cobb angles for more accurate estimations. The directly estimated angles using multi-task learning unequivocally outperformed angles measured from landmarks, leading to a lower error and standard deviation. Furthermore, our results demonstrate that the high performance can only be achieved by

Table 3: The network trained using the JR-loss achieved the lowest error in 10-fold cross-validation compared to other loss functions. The number after \pm represents the standard deviation.

Directly Estimated Landmarks				
Method	AP		LAT	
	MAE (scaled)	ρ	MAE (scaled)	ρ
mse-loss	0.0608 ± 0.0220	0.927 ± 0.0512	0.0628 ± 0.0252	0.906 ± 0.0522
cos-loss	0.0507 ± 0.0291	0.942 ± 0.0474	0.0567 ± 0.0338	0.933 ± 0.0394
cor-loss	0.0560 ± 0.0211	0.935 ± 0.0341	0.0557 ± 0.0303	0.926 ± 0.0427
reg-loss	0.0523 ± 0.0225	0.923 ± 0.0369	0.0563 ± 0.0265	0.919 ± 0.0412
JR-loss	0.0398 ± 0.0153	0.956 ± 0.0172	0.0459 ± 0.0198	0.945 ± 0.0214
Directly Estimated Angles				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
mse-loss	5.87 ± 4.30	48.33 ± 18.44	5.84 ± 3.31	49.05 ± 18.53
cos-loss	4.83 ± 2.46	36.60 ± 12.53	5.07 ± 2.58	27.72 ± 15.80
cor-loss	5.14 ± 2.57	40.58 ± 12.08	5.87 ± 2.36	32.21 ± 14.08
reg-loss	5.34 ± 2.27	41.61 ± 13.59	5.61 ± 2.04	25.76 ± 12.31
JR-loss	4.04 ± 0.99	35.85 ± 15.46	4.07 ± 1.19	23.72 ± 9.87
Angles Measured from Predicted Landmarks				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
mse-loss	7.06 ± 5.02	48.40 ± 20.08	8.13 ± 5.37	36.34 ± 15.71
cos-loss	7.48 ± 6.04	46.83 ± 20.59	7.80 ± 4.31	44.72 ± 19.72
cor-loss	6.53 ± 4.18	45.05 ± 21.48	7.81 ± 4.75	35.13 ± 15.35
reg-loss	7.32 ± 5.77	47.39 ± 19.05	7.97 ± 6.64	48.43 ± 19.05
JR-loss	6.08 ± 3.07	45.48 ± 19.96	6.40 ± 2.96	32.40 ± 14.69

training the SLE and CAE iteratively, which indicates that reciprocal multi-task learning of both landmarks and angles is essential for achieving higher accuracy and precision.

It is widely known by the spinal surgeon community that there is a large inter- and intra-observer variability in manual Cobb angle measurements. Pruijs et al. (1994) determined the variability of Cobb angle measurements by physicians to be 3.2 degrees standard deviation. Meanwhile, the standard deviation

Table 4: The Iterative-training algorithm achieved the lowest error in landmark estimation on 10-fold crossvalidation compared to sequential and angles-only method. The number after \pm represents the standard deviation.

Directly Estimated Landmarks				
Method	AP		LAT	
	MAE (scaled)	ρ	MAE (scaled)	ρ
sequential	0.091 ± 0.0537	0.728 ± 0.242	0.0951 ± 0.0392	0.792 ± 0.153
angles-only	--	--	--	--
landmark-only	0.0417 ± 0.0165	0.951 ± 0.0380	0.0492 ± 0.0219	0.936 ± 0.0621
ILAT	0.0398 ± 0.0153	0.956 ± 0.0172	0.0459 ± 0.0198	0.945 ± 0.0214
Directly Estimated Angles				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
sequential	8.27 ± 5.25	47.99 ± 21.28	6.59 ± 2.84	33.31 ± 15.28
angles-only	8.24 ± 3.54	54.04 ± 18.50	6.49 ± 2.86	30.48 ± 13.29
landmark-only	--	--	--	--
ILAT	4.04 ± 0.99	35.85 ± 15.46	4.07 ± 1.19	23.72 ± 9.87
Angles Measured from Predicted Landmarks				
	CMAE (deg)	SMAPE (%)	CMAE (deg)	SMAPE (%)
sequential	7.98 ± 4.84	44.84 ± 18.50	6.99 ± 3.50	42.78 ± 18.00
angles-only	--	--	--	--
landmark-only	6.51 ± 3.86	43.98 ± 21.98	7.01 ± 4.42	33.32 ± 16.63
ILAT	6.08 ± 3.07	45.48 ± 19.96	6.40 ± 2.96	32.40 ± 14.69

of our approach was 0.99 for AP and 1.19 for LAT x-ray images, which validates our approach as a more robust method of measurement. Table 5 shows the similarity in data distribution (mean and standard deviation) of Cobb angles between the ground truth and our prediction, which indicates that our predictions indeed reflect the distribution of our dataset. However, since our dataset does not contain x-ray images with metal artifacts, potential degradation in landmark accuracy might occur when processing images with metal artifacts caused by spinal bracing. While our approach was thoroughly validated on x-ray images acquired at our clinic, further investigation using x-ray images acquired at different clinics or inclusion of images with various imaging artifacts is warranted.

Table 5: The distribution of Cobb angles in the prediction resembles the distribution in our dataset. The number after \pm is the standard deviation (STD) of the angles.

Distribution of Cobb Angles in Our Dataset				
	AP		LAT	
Angle	Min \leftrightarrow Max	Mean \pm STD	Min \leftrightarrow Max	Mean \pm STD
PT	0.0119 \leftrightarrow 34.7	6.74 \pm 7.69	0.0446 \leftrightarrow 47.4	8.53 \pm 6.67
MT	0.0766 \leftrightarrow 66.0	11.2 \pm 10.6	0.00631 \leftrightarrow 58.8	6.80 \pm 7.25
TL	0.0176 \leftrightarrow 32.3	7.67 \pm 6.48	1.05 \leftrightarrow 49.6	18.2 \pm 7.78
Distribution of Cobb Angles in Our Prediction				
	AP		LAT	
Angle	Min \leftrightarrow Max	Mean \pm STD	Min \leftrightarrow Max	Mean \pm STD
PT	1.66 \leftrightarrow 35.8	8.05 \pm 5.80	1.88 \leftrightarrow 25.0	8.06 \pm 4.05
MT	1.75 \leftrightarrow 41.9	10.1 \pm 6.65	2.10 \leftrightarrow 23.1	7.43 \pm 3.44
TL	1.74 \leftrightarrow 28.8	7.89 \pm 4.60	7.79 \leftrightarrow 35.5	16.8 \pm 4.29

7. Conclusion

We proposed a fully automated spinal curvature estimation framework for comprehensive assessment of AIS using MVC-Net. The MVC-Net creatively utilizes the joint feature learning capabilities of the X-module, our novel JR-loss function, and specially designed ILAT algorithm to provide accurate and robust spinal curvature assessments in multi-view x-rays. It is the first time that multi-view x-rays have been used to solve the problem of vertebrae occlusion. When validated on our large dataset of 526 x-ray images, the MVC-Net was able to achieve automated estimation of Cobb angles in both AP and LAT x-rays. The highly accurate spinal landmarks and Cobb angles produced by our framework can not only be used by clinicians for comprehensive AIS assessment, but also be further extended to other clinical applications such as assessment of osteoporosis.

8. Acknowledgments

Computations were performed using the data analytics Cloud at SHARC-NET (<http://www.sharcnet.ca>) provided through the Southern Ontario Smart Computing Innovation Platform (SOSCIP); the SOSCIP consortium is funded by the Ontario Government and the Federal Economic Development Agency for Southern Ontario. The authors also wish to thank Dr. Jinhui Qin for assistance with the computing environment.

References

References

- Anitha, H., Karunakar, A., Dinesh, K., 2014. Automatic extraction of vertebral endplates from scoliotic radiographs using customized filter. *Biomedical Engineering Letters* 4, 158–165.
- Anitha, H., Prabhu, G., 2012. Automatic quantification of spinal curvature in scoliotic radiograph using image processing. *Journal of medical systems* 36, 1943–1951.
- Asher, M.A., Burton, D.C., 2006. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis* 1, 2.
- Chollet, F., 2015. Keras. <https://github.com/fchollet/keras>.
- Cobb, J., 1948. Outline for the study of scoliosis. *Instr Course Lect* 5, 261–275.
- Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2011. *Regression Forests for Efficient Anatomy Detection and Localization in CT Studies*. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 106–117.
- Group, S.D.S., 2008. *Radiographic Measurement Manual*. Medtronic Sofamor Danek USA.

- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. CoRR abs/1502.01852. URL: <http://arxiv.org/abs/1502.01852>.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. CoRR abs/1207.0580. URL: <http://arxiv.org/abs/1207.0580>.
- McCloskey, M., Cohen, N.J., 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation* 24, 109–165.
- Prujjs, J.E.H., Hageman, M.A.P.E., Keessen, W., van der Meer, R., van Wieringen, J.C., 1994. Variation in cobb angle measurements in scoliosis. *Skeletal Radiology* 23, 517–520. URL: <https://doi.org/10.1007/BF00223081>, doi:10.1007/BF00223081.
- Sardjono, T.A., Wilkinson, M.H., Veldhuizen, A.G., van Ooijen, P.M., Purnama, K.E., Verkerke, G.J., 2013. Automatic cobb angle determination from radiographic images. *Spine* 38, 1256–1262.
- Scipy, C., 2014. Cosine distance. <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.spatial.distance.cosine.html>.
- Sun, H., Zhen, X., Bailey, C., Rasoulinejad, P., Yin, Y., Li, S., 2017. Direct estimation of spinal cobb angles by structured multi-output regression, in: *International Conference on Information Processing in Medical Imaging*, pp. 529–540.
- Vrtovec, T., Pernuš, F., Likar, B., 2009. A review of methods for quantitative evaluation of spinal curvature. *European Spine Journal* 18, 593–607.
- Weinstein, S.L., Dolan, L.A., Cheng, J.C., Danielsson, A., Morcuende, J.A., 2008. Adolescent idiopathic scoliosis. *The Lancet* 371, 1527–1537.

- Wu, H., Bailey, C., Rasoulinejad, P., Li, S., 2017. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet, in: *Medical Image Computing and Computer Assisted Interventions*, pp. 127–135.
- Xue, W., Islam, A., Bhaduri, M., Li, S., 2017a. Direct multitype cardiac indices estimation via joint representation and regression learning. *IEEE Transactions on Medical Imaging* .
- Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S., 2017b. Direct estimation of regional wall thicknesses via residual recurrent neural network, in: *The 25th biennial international conference on Information Processing in Medical Imaging (IPMI 2017)*.
- Zhang, J., Lou, E., Hill, D.L., Raso, J.V., Wang, Y., Le, L.H., Shi, X., 2010. Computer-aided assessment of scoliosis on posteroanterior radiographs. *Medical & biological engineering & computing* 48, 185–195.
- Zhang, J., Lou, E., Le, L.H., Hill, D.L., Raso, J.V., Wang, Y., 2009. Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. *Journal of Digital Imaging* 22, 463472.
- Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S., 2015. Direct and simultaneous four-chamber volume estimation by multi-output regression, in: *MICCAI*, Springer. pp. 669–676.