applied sciences

MDPI

Article

Light-Convolution Dense Selection U-Net (LDS U-Net) for Ultrasound Lateral Bony Feature Segmentation

Sunetra Banerjee [1], Juan Lyu [2], Zixun Huang [3], Hung Fat Frank Leung [3], Timothy Tin-Yan Lee [4], De Yang [4], Steven Su [1], Yongping Zheng [4] and Sai-Ho Ling [1,*]

[1] School of Biomedical Engineering, University of Technology, Sydney, NSW 2007, Australia; sunetra.banerjee@student.uts.edu.au (S.B.); steven.su@uts.edu.au (S.S.)
[2] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; lvjuan91@sina.com
[3] Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China; zixun.huang@connect.polyu.hk (Z.H.); frank-h-f.leung@polyu.edu.hk (H.F.F.L.)
[4] Department of Biomedical Engineering, Hong Kong Polytechnic University, Hong Kong, China; timothy.lee@connect.polyu.hk (T.T.-Y.L.); de.de.yang@connect.polyu.hk (D.Y.); yongping.zheng@polyu.edu.hk (Y.Z.)
* Correspondence: steve.ling@uts.edu.au

Abstract: Scoliosis is a widespread medical condition where the spine becomes severely deformed and bends over time. It mostly affects young adults and may have a permanent impact on them. A periodic assessment, using a suitable modality, is necessary for its early detection. Conventionally, the usually employed modalities include X-ray and MRI, which employ ionising radiation and are expensive. Hence, a non-radiating 3D ultrasound imaging technique has been developed as a safe and economic alternative. However, ultrasound produces low-contrast images that are full of speckle noise, and skilled intervention is necessary for their processing. Given the prevalent occurrence of scoliosis and the limitations of scalability of human expert interventions, an automatic, fast, and low-computation assessment technique is being developed for mass scoliosis diagnosis. In this paper, a novel hybridized light-weight convolutional neural network architecture is presented for automatic lateral bony feature identification, which can help to develop a fully-fledged automatic scoliosis detection system. The proposed architecture, Light-convolution Dense Selection U-Net (LDS U-Net), can accurately segment ultrasound spine lateral bony features, from noisy images, thanks to its capabilities of smartly selecting only the useful information and extracting rich deep layer features from the input image. The proposed model is tested using a dataset of 109 spine ultrasound images. The segmentation result of the proposed network is compared with basic U-Net, Attention U-Net, and MultiResUNet using various popular segmentation indices. The results show that LDS U-Net provides a better segmentation performance compared to the other models. Additionally, LDS U-Net requires a smaller number of parameters and less memory, making it suitable for a large-batch screening process of scoliosis without a high computational requirement.

Keywords: lateral bony feature; depthwise separable convolution; segmentation; scoliosis; ultrasound; U-Net

Citation: Banerjee, S.; Lyu, J.; Huang, Z.; Leung, H.F.F.; Lee, T.T.-Y.; Yang, D.; Su, S.; Zheng, Y.; Ling, S.-H. Light-Convolution Dense Selection U-Net (LDS U-Net) for Ultrasound Lateral Bony Feature Segmentation. *Appl. Sci.* **2021**, *11*, 10180. https://doi.org/10.3390/app112110180

Academic Editor: Hartmut Schneider

Received: 4 October 2021
Accepted: 27 October 2021
Published: 30 October 2021

1. Introduction

Scoliosis is a deformation of the spinal cord, usually in an S or C shape with a curvature generally greater than 10°, that occurs in the coronal plane between the dorsal and ventral parts. In some cases, the degree of curvature is stable but in other cases, it increases over time. This ailment primarily starts from adolescence [1]. Hence, it is necessary to screen the children. The current study shows that the overall occurrence of adolescent idiopathic scoliosis (AIS) is 0.47–5.2% of the total population [2], and is quite prevalent in many regions such as China (5%), Hong Kong (3–4%), and the USA (2%). There are no

early signs of idiopathic scoliosis. As the spinal deformation aggravates, patients develop some physical irregularities such as uneven shoulders, inflated curvature of the spine, disproportional alignment of hips, or back pain and discomfort. If not treated properly, progressive scoliosis may cause constant back pain, breathing problems, or even physical disability for life [3–5]. The early detection of scoliosis is critical to prevent the worsening of the spine over time.

The conventional method of diagnosis of scoliosis is using Cobb's measurement technique to measure the curvature angle of the spine on a standing radiograph [6]. If two lines are drawn so that one is perpendicular to the upper endplate of the most titled upper vertebra and another is perpendicular to the lower endplate of the most tiled lower vertebra, then the angle between them is known as the Cobb angle [6]. However, the traditional radiography method is not preferred for periodic diagnosis of scoliosis as ionising radiation may cause cancer. According to research, excessive ionising radiation exposure can cause breast cancer in girls and may result in leukaemia and prostate cancer in general [7]. Levy et al. established that an AIS patient who had undergone frequent X-ray scans had a 2–4% higher chance of becoming affected by cancer vis-à-vis normal young people [7]. In improved X-ray modalities such as EOS, where the radiation dosage is 8–10 times lower than traditional X-rays, there is still an accumulation of an ionising radiation dose that could lead to adverse effects in the long run [8–10].

Alternatively, magnetic resonance imaging (MRI), a radiation-free technology, can also be used for the 3D assessment of spine curvature. However, for scoliosis detection, the patient has to be in a standing posture, something that is a challenge in MRI as such scanning would require specialized MRI installation and a long operation time [11].

Ultrasound imaging is a non-radiating imaging modality that is used extensively in the medical field due to its low cost and high portability. This modality works on the principle of capturing the reflected ultrasound from the cortical surface of the internal organs and mapping their topographical information [12]. A recent invention in this field is freehand 3D ultrasound, where the conventional 2D ultrasound is combined with position sensors [13,14].

Zheng et al. developed a 3D ultrasound system called the Scolioscan system [15]. The Scolioscan system is made of an ultrasound scanner with a built-in linear probe, a frame structure, an electromagnetic spatial sensing device, and propriety software. The procedure involves freehand scanning with the probe from the bottom to the top of the patient's back, covering the whole spine area, while the electromagnetic spatial sensing device continuously detects the probe's position and orientation. B-mode images, collected along with their corresponding position and orientation information, are used to reconstruct into a 3D image thereby forming the coronal view of the spine using the volume projection image (VPI) method [16]. Through this research, Zheng et al. demonstrated the reliability of the new Scolioscan system for scoliosis diagnosis [15] and efficacy vis-à-vis the conventional radiographic Cobb method [2]. Medium to strong correlation between the Scolioscan angles and X-ray Cobb angles was evidenced, along with very good intra and inter-observer reliability and an intra-class correlation >0.87. Scolioscan, therefore, can be taken as a promising 3D ultrasound imaging system for scoliosis screening and monitoring curve progression [2].

However, the current technique requires manual annotation for the measurement of the spine curvature angle, which in turn is dependent on the examiner's judgment, expertise, and measurement speed. As a result, the manual intervention would limit the number of cases that can be handled at a given time, depending on the availability of the examiner.

As an improvement, a semi-automatic measurement of spine curvature angle was developed. It used a 6th order polynomial curve fitting method to estimate the spine curve equation [17]. However, in this method, tangent lines were inputted manually with human judgment. Zhou et al. then worked on an automatic measurement of spine curvature using 3D ultrasound image pre-processing with phase congruency and a newly developed

two-fold threshold strategy [18]. This method overcame the limitations of manual measurement of spine curvature angles (e.g., variations in measurements). However, Zhou et al. concluded that the computational time of this method was uneconomic as the program spent a lot of time computing the phase congruencies of the images [18]. This method was also not helpful for fast processing and detection.

There are two serious challenges associated with any ultrasound image that degrade the quality of output images. These are (i) speckle noise, which is caused by reasons such as air gap between the device and patient's body (Figure 1b), inconsistencies of ultrasound gel layer during scanning, improper scanning speed (Figure 1c), or other system losses; and (ii) low image contrast produced because the speed of sound varies for various tissues, and it is often difficult to separate fat and water-based tissues. As speckle noise appears as information, it makes the clinical data hard to differentiate [19], and the low contrast of the image causes unnecessary distractions for medical practitioners.

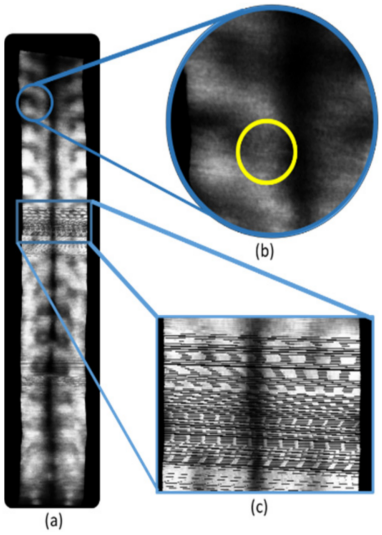

Figure 1. Various noise in ultrasound spine image (**a**), raw image (**b**), and (**c**) types of speckle noises.

1.1. Need for Segmentation of Bony Features

An alternative way of measuring a new parameter, called ultrasound curve angle (UCA), which is equivalent to radiographic Cobb angle [20], is illustrated in Figure 2. A 3D volume projection image is sliced into nine 2D coronal plane images of different depths for scoliosis detection. One such 2D ultrasound spine image is shown in Figure 2a. Figure 2b illustrates the various segments of the spine—the thoracic bony features, rib, T12 level, and lumbar bony features. Figure 2c,d outlines the current methodology used by a medical expert to calculate the UCA. The most tilted thoracic bony features (TBFs) and lumbar bony features (LBFs) are identified by expert judgment. Further steps involve drawing lines through the centre of the identified TBF and LBF pairs (Figure 2c) and allocating lines to measure the UCA (Figure 2d) [20]. Hence, the clear segmentation of bony features is an important step for the UCA measurement process.

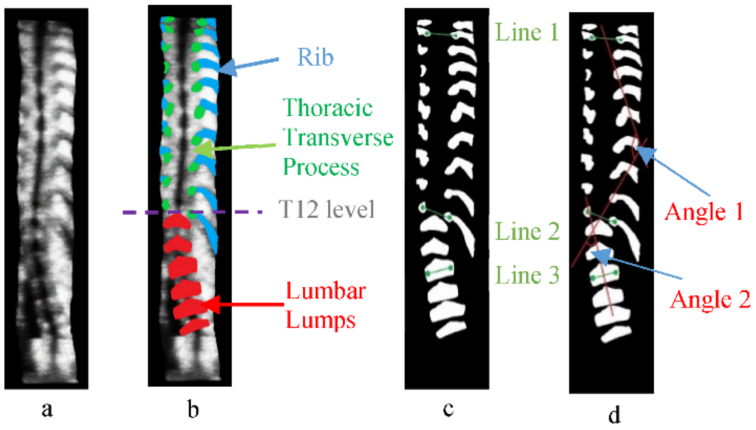

Figure 2. Ultrasound curve angle (UCA) measurement of (**a**) original ultrasound spine image and (**b**) marked lateral bony features with three types of anatomical features. (**c**,**d**) Illustration of methodology.

1.2. Variabilities in Lateral Bony Features

The locations, shapes, and sizes of the bony features were different (Figure 3a–c) as were the UCA (Figure 3d–f). The slopes between two endpoints of two adjacent bony features were also different (Figure 3d–f). The variability of LBFs and TBFs presents a unique challenge.

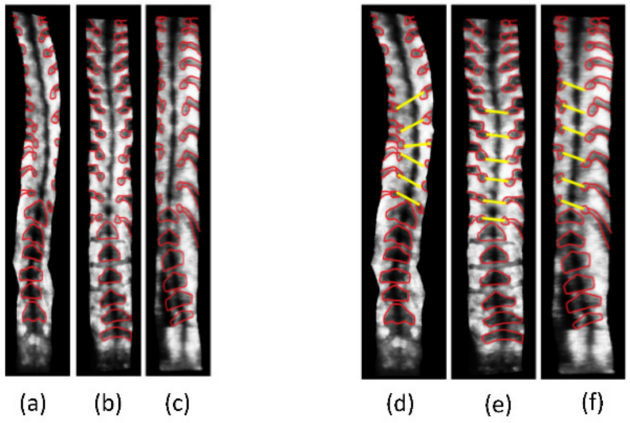

Figure 3. (**a**–**c**): Lateral bony features of different shapes, sizes, and locations. (**d**–**f**) Slopes between two adjacent bony features (thoracic) are different.

1.3. Medical Image Segmentation Using CNN

Machine learning techniques have been successfully applied in the field of medical diagnosis using ultrasound. Vedula et al. proposed the transformation of speckled, blurry ultrasound images into better quality images using a convolutional neural network (CNN) [21]. In the past few years, CNNs have been successfully used in various biomedical image processing tasks including image classification [22–24], feature extraction [25], and segmentation [26–29]. Semantic segmentation has become a prior interest area in medical imaging [27,28,30–32]. However, automatic semantic segmentation of biomedical images could be difficult when there is a variability of shapes and sizes of the anatomy between patients as well as low contrast of surrounding tissues [33].

Among the many machine learning techniques, the U-Net architecture is particularly successful in biomedical image segmentation [26]. U-Net can efficiently segment images with a very limited number of the annotated training dataset. A U-Net consists of a multi-layer deep encoder network that extracts spatial features from the image, and a corresponding multi-layer deep decoder network that up-samples the feature maps to

predict the segmentation masks. It uses the self-learning property of the convolution kernel to input the original image and obtain the classification result. By increasing the number of layers, the U-Net can extract considerably complex and detailed image features. Generally, the shallower layers of U-Net are found to be capable of extracting some general features of images, whereas the deeper layers could extract more specific features. U-Net and U-Net like models have been used successfully in segmenting 2D or 3D ultrasound images of breast lesion [34], human placenta [35], liver [36], spine anatomy [37–39], kidney [40], etc.

MultiResUNet [41], as an improved version of U-Net, was introduced to segment very challenging images that cannot possibly be undertaken by basic U-Net such as images having irregular shapes with multi-scaled features. In its architecture, Inception blocks [23] were introduced in place of the conventional convolution layers of U-Net to help with multilayer feature extraction. Additionally, to reduce the semantic gap between encoder–decoder features, the normal convolution layers were replaced with residual connections that made training easier [42].

Attention U-Net [43] improves the conventional U-Net with an attention gate for medical imaging that automatically learns to focus on target structures of varying shapes and sizes. Because of the small number of training data, dense connectivity is needed in biomedical image processing. Dense connectivity can be successfully incorporated within the encoder and decoder path [44,45]. Another semantic segmentation using Atrous convolutions was introduced using the DeepLab family for better segmentation [46–48].

Recently, there is an increasing need to implement deep learning techniques for medical problems on mobile phones, embedded systems, or any PC with high diagnostic accuracy and a low computational requirement. Most of the CNN models are over-parameterized and need high computing power and memory for training and inferencing [49]. Depthwise separable convolution layers are the solution to this problem [50]. Depthwise separable convolution layers are successful in forming image classification models in two ways: (a) they provide better models (e.g., Xception model [51]) than the conventional convolution layers, with a considerably smaller number of parameters, and (b), they reduce the memory space requirement when reducing the number of parameters (e.g., the MoblieNets family of architectures [52]). Additionally, it has been found that the regular convolution operations can be considered to be equivalent to the depthwise separable convolution operations [53]. Hence, compared to conventional convolutional layers, using depthwise convolution layers will have similar performance in terms of accuracy, but require a smaller number of training parameters.

Inspired by the above work, we aimed to develop a low-computation novel hybridized deep learning architecture to suitably segment the bony features in ultrasound spine images. The proposed architecture has three main aspects: (a) the basic U-Net structure is adopted as the network architecture, but the conventional convolutional layers are replaced with dense depthwise separable convolution layers to increase the computational efficiency; (b) selection gates [43] are deployed for a smarter identification of the target bony features; and (c) the encoders–decoders are connected using multi-scale skip-pathways [41] to enhance feature fusion. The segmentation result of the proposed architecture was compared quantitatively and qualitatively with the basic U-Net [26], Attention U-Net [43], and MultiResUNet [41] models.

2. Methodology

Our objective was to automatically segment, using a suitable CNN architecture, the lateral bony features (LBFs and TBFs) in an ultrasound spine image, which is plagued with speckle noise and low contrast. This paper has two main contributions:

(a) Development of a suitable architecture that can produce better segmentation results vis-à-vis contemporary models while handling the inherent drawbacks of ultrasound images; and
(b) Enhancement of the computational efficacy for a lightweight architecture.

In a nutshell, a lightweight version of U-Net that contains densely connected depthwise separable convolution followed by pointwise convolution, multiscale skip connection, and selection gates is proposed. It is inspired by several salient features used in other models such as the U-Net, MultiResUNet, and Attention U-Net. Figure 4 illustrates the architecture of the proposed network of Light-Convolution Dense Selection U-Net (LDS U-Net). This network model was built on the concept of depthwise separable convolution and has three main features: (a) novel light dense blocks, (b) improvised multi-scaled skip connections, and (c) selection gates. The details of depthwise separable convolution and the three features are given as follows.

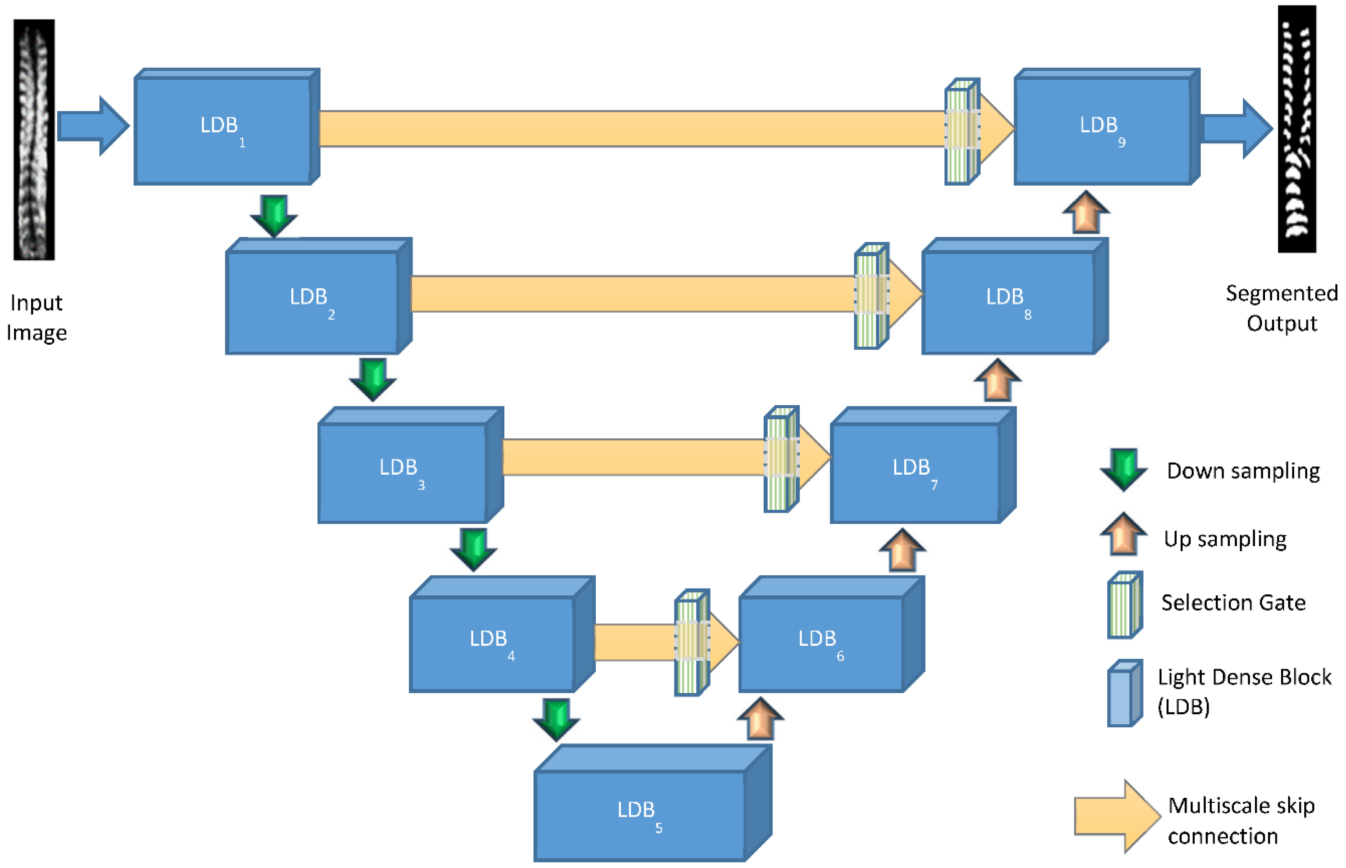

Figure 4. Proposed architecture of LDS U-Net.

2.1. Depthwise Separable Convolution

Depthwise separable convolutions lessen the number of parameters and computation used in convolutional operations while increasing representational efficacy [54]. This kind of convolution can be applied to information such as spatial, depth dimensions, and the number of channels. While normal convolution deals with a single convolution operation, a depthwise separable convolution separates a kernel into two different kernels that carry out two convolution operations, namely, the depthwise convolution and the pointwise convolution. In the depthwise convolution, a spatial convolution is conducted independently over each channel of the input. It is followed by a pointwise convolution, where a 1×1 convolution is conducted to map the depthwise channel output into a new channel space [51].

A standard convolution layer works by applying a convolution kernel to all channels of the input image and takes a weighted sum of the input pixels covered by the kernel sliding across all input channels of the image. This means that for a standard convolution, no matter how many input channels are available, the number of output channels is one.

However, in depthwise separable convolutions, features are only learned from the input channels. Therefore, the output layer has the same number of channels as the input.

Suppose in a convolution operation, the input is of size $P_f \times P_f \times K$ with the feature map f and generates an output of size $P_g \times P_g \times L$ with the feature map g, where P_f is the spatial width and height of the input feature map; K is the number of input channels; P_g is the spatial width and height of the output feature map; and L is the number of output channels. Then, for a conventional convolution operation (Figure 5a) with the convolution kernel N of size $P_n \times P_n \times K \times L$, where P_n is the spatial dimension of the kernel, the computation cost is given by the equation:

$$G_c = P_n \,.\, P_n \,.\, K \,.\, L \,.\, P_f \,.\, P_f \tag{1}$$

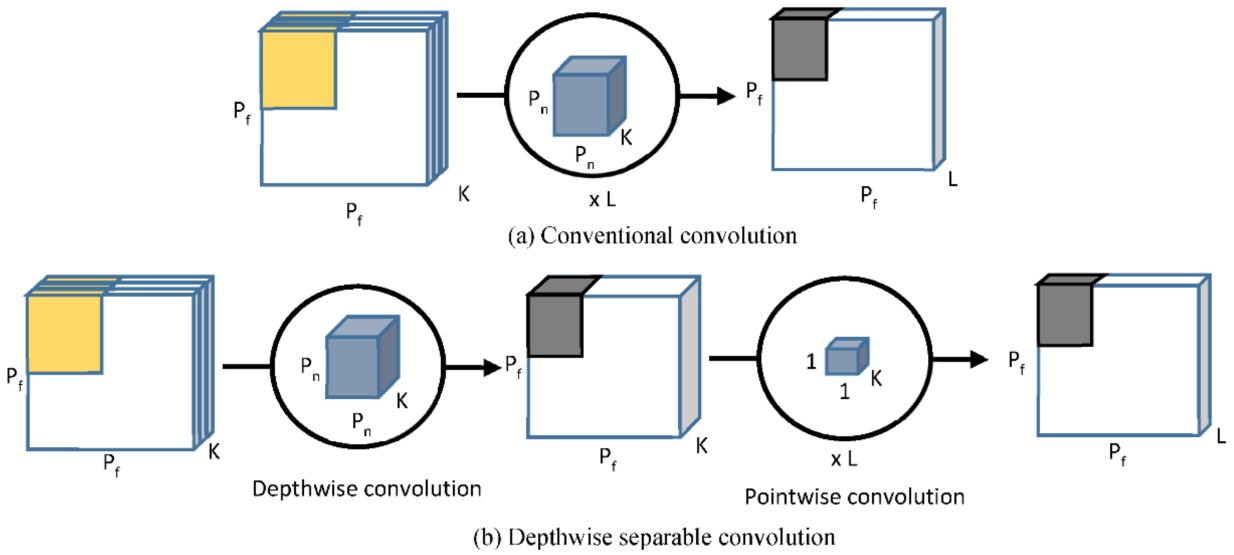

Figure 5. Schematic representation of (**a**) conventional vs. (**b**) depthwise separable convolution.

On using the depthwise separable convolution (Figure 5b), the computation cost will be the aggregation of depthwise and pointwise convolutions and is given by the equation:

$$G_d = P_n \,.\, P_n \,.\, K \,.\, P_f \,.\, P_f + K \,.\, L \,.\, P_f \,.\, P_f \tag{2}$$

Combining Equations (1) and (2), the reduction in computation can be represented by the equation:

$$\begin{aligned} G &= \frac{G_d}{G_c} = \frac{P_n \,.\, P_n \,.\, K \,.\, P_f \,.\, P_f + K \,.\, L \,.\, P_f . P_f}{P_n \,.\, P_n \,.\, K \,.\, L \,.\, P_f \,.\, P_f} \\ &= \frac{1}{L} + \frac{1}{P_n^2} \end{aligned} \tag{3}$$

Equation (3) represents the reduction in computation requirements of the depthwise separable convolution compared to the conventional convolution, resulting in considerably lower computing and parameter cost of the network.

As conventional U-Net, with its few layers, is not deep enough to perform this particular segmentation task, adding more layers directly and making it deeper, may solve the segmentation problem. However, a deeper neural network tends to develop gradient vanishing and redundant computation in network training [55]. To overcome these associated problems, few modifications are required to enhance the learning process of the network.

2.2. Light Dense Block

The first modification is related to the conventional dense network, which uses regular convolutional layers and has the advantage of parameter simplicity, vanishing-gradient

minimization, and feature reuse [56]. In this paper, a proposed light dense block was designed as the main building block of Light-Convolution Dense Selection U-Net (LDS U-Net). The basic structure of the light dense block is shown in Figure 6 and here, unlike a conventional dense network, all the convolutional layers are depthwise separable convolution layers.

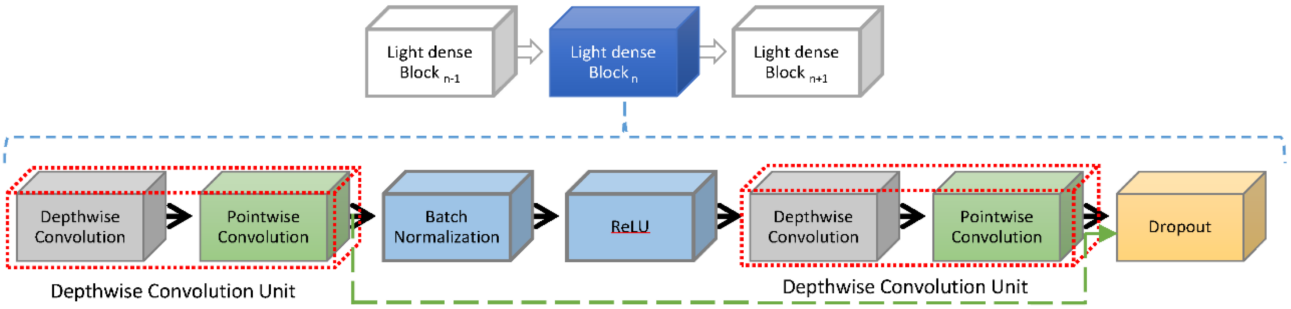

Figure 6. Proposed light dense block.

The first layer of the light dense block is a depthwise convolution unit that consists of a depthwise convolution block followed by a pointwise convolution block. The next layers are the batch normalization, rectified linear unit (ReLU) activation function, another depthwise convolution unit, and a dropout layer. The first depthwise convolution unit is also connected densely to the dropout layer, as shown by the green dotted arrow in Figure 6. Through this new design, the light dense block delivers the same advantages as a conventional dense block, but with a smaller number of parameters.

2.3. Multi-Scale Path

In the basic U-Net architecture, there are skip pathways between the respective layers of the encoding and decoding side, and shortcut paths before the max-pooling layers in the encoder side and after the deconvolution layers in the decoder side. Often, spatial information gets lost during the max-pooling operation, and the skip connections help the network to propagate information from the encoder side to the decoder side. However, the skip pathways often come up with a problem of a semantic gap during the feature fusion because the first layer of the encoder, which extracts the low-level features, is connected to the terminal layer of the decoder, which deals with more high-level features. Additionally, because of the added complexities of variabilities in sizes, shapes, and positions of bony features, both the low and high level features would have to be retained for detailed segmentation.

To reduce the discrepancy between the encoder–decoder features and enhance the feature fusion, a multi-scale skip path was proposed in this paper. A multi-scale inception [41] module was incorporated between the encoder and decoder layers to enhance the low-level features extracted in the encoder side. Through this connection, the low-level features will undergo further processing before merging with the high-level features on the decoder side. Moreover, instead of the usual convolutional layers, multi-scale inception layers are used. They improve the utilization rate of computing resources by increasing the depth and width of the network while keeping the computational budget constant [57].

Inception modules have been proven to be very promising in enlarging receptive fields and capturing more context information [58]. It enhances the depiction capability of low-level features. The inception module adopts multiple branches with different kernel sizes to capture multi-scale information. This methodology is key to dealing with the problem of handling the high variability of shapes, sizes, and positions of bony features in ultrasound images of the spine. However, as the inception module is very computationally demanding, the normal convolution operation in the traditional inception module is replaced by the depthwise convolution in this proposed model.

In the proposed light dense block, a sequence of two 3×3 depthwise convolutional layers is used to carry out feature extraction. In the skip path, a sequence of 3×3 convolution blocks is used, as shown in Figure 7, instead of bigger 5×5 and 7×7 blocks. Therefore, the outputs from the three 3×3 convolution blocks are concatenated to enhance the receptive field and reduce the semantic gap between the encoder and the decoder. A residual connection of a 1×1 convolution block is also presented in the skip path to make the learning procedure stable.

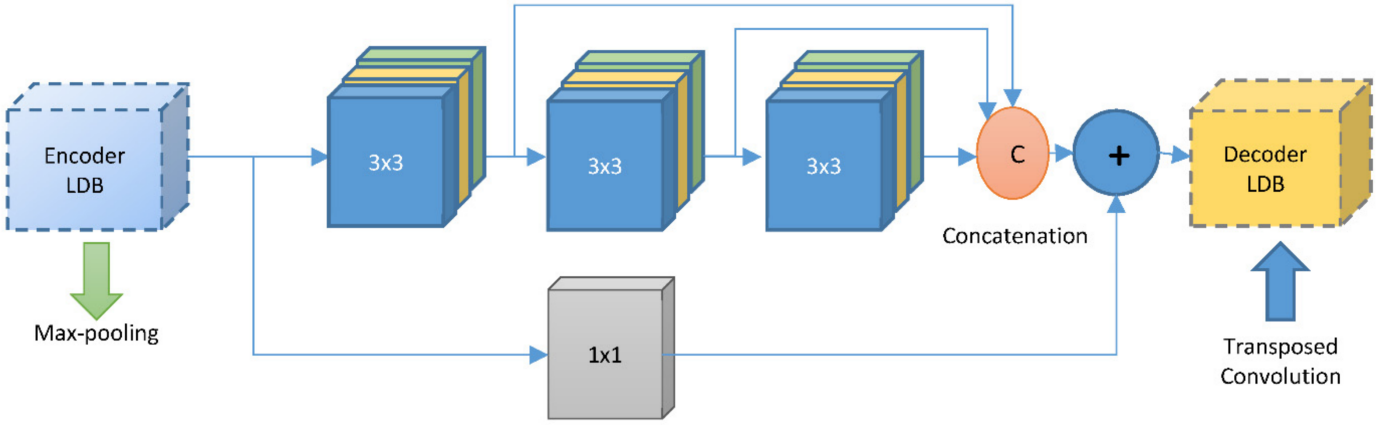

Figure 7. Multi-scale inception module.

In the proposed inception module, the technique to control the number of filters of the convolution layers inside the block was adapted [41]. A parameter P was assigned to control the number of filters as given by Equation (4),

$$P = \lambda \times F \tag{4}$$

where F is the number of filters in the corresponding layers like the basic U-Net, and λ is a scaler coefficient. The number of filters was set to F = [32, 64, 128, 256, 512], along with the five layers of the LDS U-Net architecture, respectively, and λ was chosen as 1.67 to ensure that the model structure was similar to the basic U-Net. The numbers of filters were set to $\frac{P}{6}$, $\frac{P}{3}$, and $\frac{P}{2}$ in the three corresponding convolutional layers for extracting multiscale features.

2.4. Selection Gate

In an ultrasound spine image, the noise in the extraneous regions can appear as information and can hinder the segmentation process. To tackle this problem smartly and efficiently, selection gates were applied in the proposed network. The selection gate [43] discards extraneous regions in the input spine ultrasound image to ignore the associated noise and allows for preferential attention to the target bony features to select the relevant features of importance. A conventional selection gate, improvised with depthwise separable convolution layers, was integrated into the proposed architecture, as shown in Figure 8. This aimed to reduce the computational overhead while increasing the segmentation accuracy.

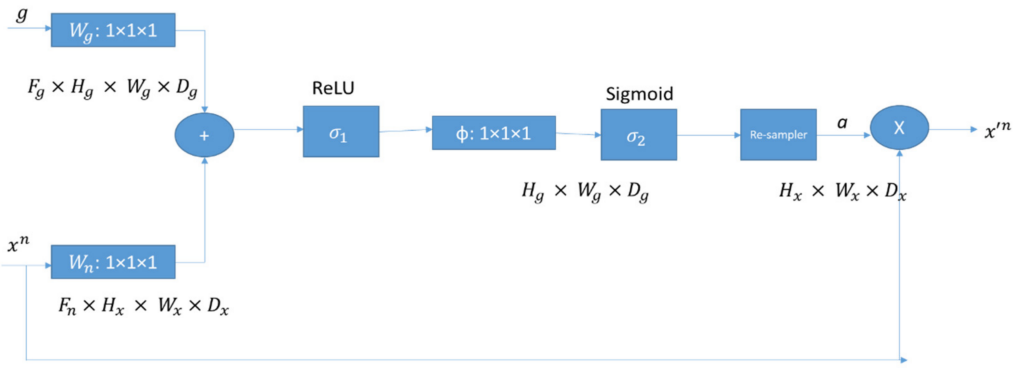

Figure 8. Selection gate.

The light dense convolutional layers extract deeper features when it processes deeper layers gradually. Suppose that after processing the n-th layer, the generated feature map is x^n. As shown in Figure 8, a is the attention coefficient that identifies and extracts only the desired parts of the image for better segmentation; F_n is the number of feature maps in layer n; H_x, W_x, D_x are the height, width, and dimension of the n-th layer, respectively; g is a gating vector booked from the preceding layer of the network, which is a coarser layer; F_g is the number of feature maps in the layer g; and H_g, W_g, D_g are the height, width, and dimension of g, respectively.

The selection gate works with two input vectors: x^n and g. g has a lower dimension and better feature representation as it comes from a deeper layer compared to x^n. x^n, after processing by a stride convolution, and g, after processing through a $1 \times 1 \times 1$ convolution, merge elementwise. After that, a ReLU activation layer σ_1 and a $1 \times 1 \times 1$ convolution operation take place, which reduces the dimension of the resultant vector. Additionally, $\sigma_1\left(x^n_{i,c}\right) = \max\left(0, x^n_{i,c}\right)$, where i and c stand for the spatial and channel dimensions, respectively. Then, a sigmoid layer σ_2 scales the vector to the range [0, 1] to generate the attention coefficient a, where $\sigma_2(x_{i,c}) = \frac{1}{1+\exp\left(-x_{i,c}\right)}$ is the sigmoid activation function. Grid resampling of attention coefficients is conducted using trilinear interpolation. A value of close to 1 indicates more significant features. The attention coefficient is multiplied element-wise to x^n after resampling. The output of the selection gate is ${x'}^n_{i,c} = x^n_{i,c} \cdot a^n_i$.

2.5. Ablation Study

Ablation experiments were carried out to gauge the contribution of each feature of the LDS U-Net. Three models were successively developed before arriving at the final segmentation architecture. These intermittent models were independently evaluated using the available dataset.

The first intermitted model was a modified U-Net with light dense block and named as the light dense (LD) model. In this model, the basic convolution layers of U-Net were replaced by the newly developed light-dense blocks. In the second intermitted model, trainable selection gates were introduced, which was trained to isolate the relevant areas of interest, amidst noisy areas and facilitated the flow of only the relevant information within the network. This is called the light convolution selection (LCS) model. The third intermitted model was developed from the LD model by introducing multiscale skip-paths, which would allow for the seamless propagation of richer information through the network. This model is named the light-dense inception (LDI) model. In the final version, (i.e., the LDS model), both the selection gate and multi-scale path are included for the segmentation of bony features with various shapes and sizes.

3. Experimental Setup

3.1. Dataset

The input images used in this research are collected using the Scolioscan system (Figure 9) (Model SCN801, Telefield Medical Imaging Ltd., developed in Hong Kong), which generates 3D volume projection images (VPI) using a 3D ultrasound imaging technique. The experimental procedures involving human subjects were approved by the Institutional Review Board. The subjects gave informed consent for their inclusion in this study as required, and the work adheres to the Declaration of Helsinki.

Figure 9. Scolioscan system [2].

A total of 109 images, collected from 109 patients (82 females and 27 males) with an age range of 15.6 ± 2.7 years and having different degrees of spine deformity, were used retrospectively.

Nine 2D coronal images of different depths were extracted from one 3D ultrasound VPI image [16]. Since the quality of individual images varies greatly, human experts were employed to manually assess the clarity of the lateral bony features represented in an image and select the best image for a given patient. Subsequently, 109 2D vertebral coronal images formed the input dataset for this work, and each 2D image was then resized uniformly to 2574 × 640-pixel and converted to the '.png' format.

The truth mask for each 2D coronal image was carried out by experts from Hong Kong Polytechnic University. Some sample input 2D coronal images, along with their expert

generated truth masks, are shown in Figure 10a–d. The labelling of the truth mask was conducted based on some key features: (1) There should be six lumbar bony features (LBF); (2) the thoracic bony features (TBFs) in the lumbar region are not labelled as they are not generally visible in ultrasound images; (3) if the last pair of TBFs (i.e., the T12 level) is not visible, it is labelled according to the judgment of the experts. The input raw datasets of 109 spine ultrasound images were randomly split into one training set of 79 images and one testing set of 30 images.

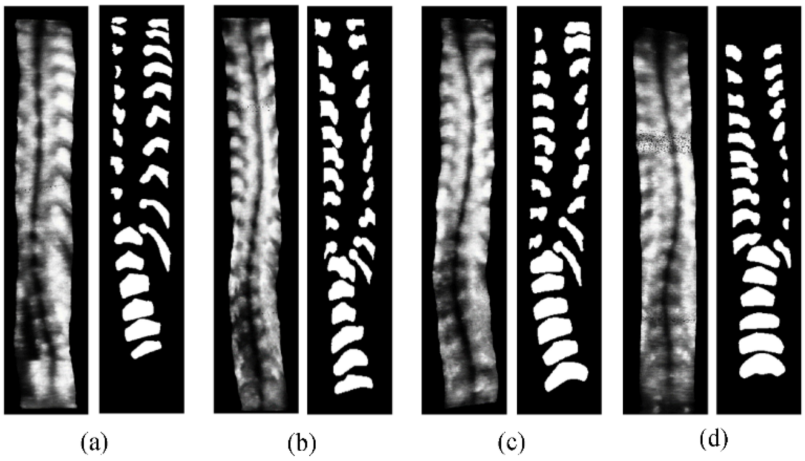

Figure 10. (**a**–**d**) Four image sets each having a raw image plus a truth mask.

3.2. Pre- and Post-Processing and Data Augmentation

The size of each raw ultrasound image was 2574 × 640 pixels and was resized to 256 × 64 pixels through image pre-processing, maintaining the aspect ratio of the original image. The resized ultrasound images were randomly flipped and rotated for data augmentation. The objective of the experiment was to develop a segmentation architecture and to benchmark its performance against the original U-Net, Attention U-Net, and MultiResUNet. This requires no specific pre-processing except that the input images were resized to fit into the GPU memory, and the pixel values were divided by 255 to bring them into the [0, 1] range. During post-processing, the image was resized back to the original size of the raw images to ensure that the image size did not impact the ultrasound curve angle (UCA) calculation for scoliosis.

3.3. Implementation Details

Anaconda, or more specifically, Spyder software, was used to conduct the experiments. The network models were implemented using Keras [59] with Tensorflow backend [60]. The experiments were conducted on a GPU laptop with NVIDIA GeForce RTX 2060.

The general working principle of any semantic segmentation algorithm is to investigate each pixel and anticipate whether it represents a point of interest or merely a part of the background. Alternatively, this principle can also be treated as a pixel-wise binary classification problem with the objective of the segmentation algorithm to minimize the binary cross-entropy loss function.

For image A, let the corresponding truth mask (TM) be B, and the predicted segmentation output be B'. For a pixel ma, the TM value is b_{ma} and the network predicted output is b'_{ma}. The binary cross-entropy loss for that image is defined as:

$$Cross\ Entropy\ (A, B, B') = \sum_{ma \epsilon A} \left(-(b_{ma}) \log\left(b'_{ma}\right) + (1 - b_{ma}) \log\left(1 - b'_{ma}\right)\right) \quad (5)$$

For a batch containing p images, the loss function L becomes,

$$L = \frac{1}{p}\sum_{i=1}^{p} Cross\ Entropy\ \left(A_i, B_i, B_i'\right) \tag{6}$$

The goal of the model is to minimize the binary cross-entropy loss and the model is trained using the Adam optimizer [61]. The Adam optimizer adaptively computes different learning rates for different parameters from estimates of the first and second moments of the gradients. All the models, used in this research, were trained up to 120 epochs since 120 epochs were found to be the saturation point for model accuracy, and no further progress was observed beyond this point. Finally, 3-fold cross-validation of the dataset was used to validate the consistency of the model.

3.4. Evaluation Metrics

For quantitative performance evaluation, four very popular evaluation indices were employed: Jaccard similarity (JS) [62], DICE coefficient (DC) [63], F1 score [64], and pixel accuracy:

$$Jaccard\ Similarity = \frac{J \cap \hat{J}}{J \cup \hat{J}} \tag{7}$$

$$Dice\ coefficient = 2\left(\frac{J \cap \hat{J}}{J + \hat{J}}\right) \tag{8}$$

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FP} \tag{10}$$

where $\hat{J}$ is the predicted segmentation output from the method to be evaluated, and J is the expert suggested truth mask. The J contours are references for further segmentation analysis [65]. TP, TN, FP and FN are the truly positive, true negative, false positive, and false negative, respectively. True positive denotes the pixels present in both the truth mask and predicted segmented region. True negative denotes the pixels present in neither segmented truth mask nor predicted segmented region. False positive signifies the pixels present only in the predicted segmented region and false negative signifies the pixels present in the suggested truth mask only.

4. Results

This section is presented in two parts. In the first part, quantitative and qualitative comparisons of the LD model (light dense model), LDI model (light-dense inception model), and LCS model (light convolution selection model) are shown together with the Light-Convolution Dense Selection U-Net (LDS U-Net) model to evaluate the importance of key features used in the segmentation of spine ultrasound images with variable shapes and sizes of bony features and to assess the overall effectiveness of the proposed network. In the second part, an extensive analysis for the newly proposed model and comparisons with basic U-Net, Attention U-Net, and MultiResUNet are made to evaluate the performance of LDS U-Net to other contemporary models.

4.1. Evaluation of Performance of Key Features

Three key features are used in the proposed LDS U-Net model–light dense blocks, multi-scale paths, and selection gates. At the onset, it is important to assess whether these features play an important role in achieving the objectives. Hence, during the ablation study, models were made by isolating the desired key features and comparing the semantic segmentation performance. At first, both the multi-scale paths and selection gates were removed from the main model. The model is named as the light dense (LD) model and

it successfully segments the thoracic and lumbar bony features. Next, the selection gates are added with the LD model, and named as the light convolution selection (LCS) model. As an alternate improvement to the LD model, the conventional shortcut connection was replaced with the inception skip path to form the light dense inception (LDI) model. Finally, all of the above models were implemented together as the light dense selection U-Net (LDS U-Net) model.

4.1.1. Quantitative Evaluation (Ablation Study)

Table 1 outlines the segmentation performances for the ablation study. The LCS model generated better segmentation than the LD model (1.07% better Dice score). It implies that the model with a selection gate can detect the target lateral bony features better than the LD model. The LDI model further improves the performance by introducing the multi-scale path with a 1.53% higher Dice score than the LCS model. Finally, both the selection gate and multi-scale paths were included in the LDS U-Net model, and it generates a Dice score of 3.26% higher than that of the LDI model. In terms of Jaccard index, F1 score, and accuracy, the proposed LDS U-Net performed the best. Hence, the LDS U-Net model was chosen as the proposed model. This model also obtained the smallest standard deviations in the four evaluation indices.

Table 1. Quantitative evaluation of ablation study.

Method	Avg. Jaccard Index (Std Dev)	Avg. Dice Score (Std Dev)	Avg. F1 Score (Std Dev)	Avg. Accuracy (Std Dev)
LD	0.7123 (±0.039)	0.8204 (±0.036)	0.8412 (±0.030)	0.9108 (±0.025)
LCS	0.7280 (±0.037)	0.8292 (±0.033)	0.8532 (±0.031)	0.9203 (±0.021)
LDI	0.7339 (±0.034)	0.8419 (±0.032)	0.8704 (±0.029)	0.9367 (±0.020)
Proposed LDS U-Net	**0.7415 (±0.03)**	**0.8694 (±0.028)**	**0.8885 (±0.025)**	**0.9592 (±0.020)**

LD: light dense; LCS: light convolution selection; LDI: light dense inception; LDS: light dense selection.

4.1.2. Qualitative Evaluation (Ablation Study)

The aim was to segment and identify the bony features from the ultrasound spine image for automatic scoliosis detection. The LD model successfully segmented the thoracic and lumbar bony features, as shown in Figure 11c. Figure 11d shows the LCS model, which provided better results in the segmentation of the left thoracic bony features (Nos. 5, 6, 7, 8, and 9) compared to Figure 11c. It implies that the model with a selection gate detects the target bony features better than the LD model. To further improve the LCS model, the conventional shortcut connection is replaced with the inception skip path to form the LDI model. From Figure 11e, it is clear that the LDI model produces better segmentation of the left thoracic bony features compared to the LD and LDC models. Finally, the proposed LDS U-Net model, as shown in Figure 11f, generated more accurate boundary segmentation of the left thoracic bony features than the LD, LCS, or LDI models. Figure 12 shows several more qualitative visual comparisons of bony feature segmentation using the LD, LCS, LDI, and proposed LDS U-Net models.

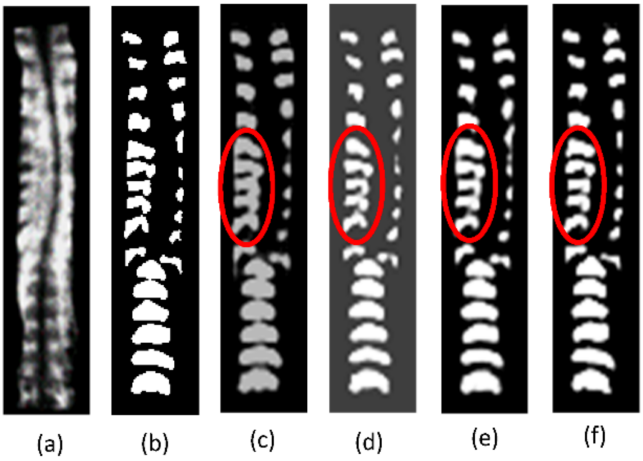

Figure 11. Individual segmentation result of the spine ultrasound image: (**a**) Raw image, (**b**) truth mask, (**c**) LD model, (**d**) LCS model, (**e**) LDI model, and (**f**) proposed LDS U-Net model.

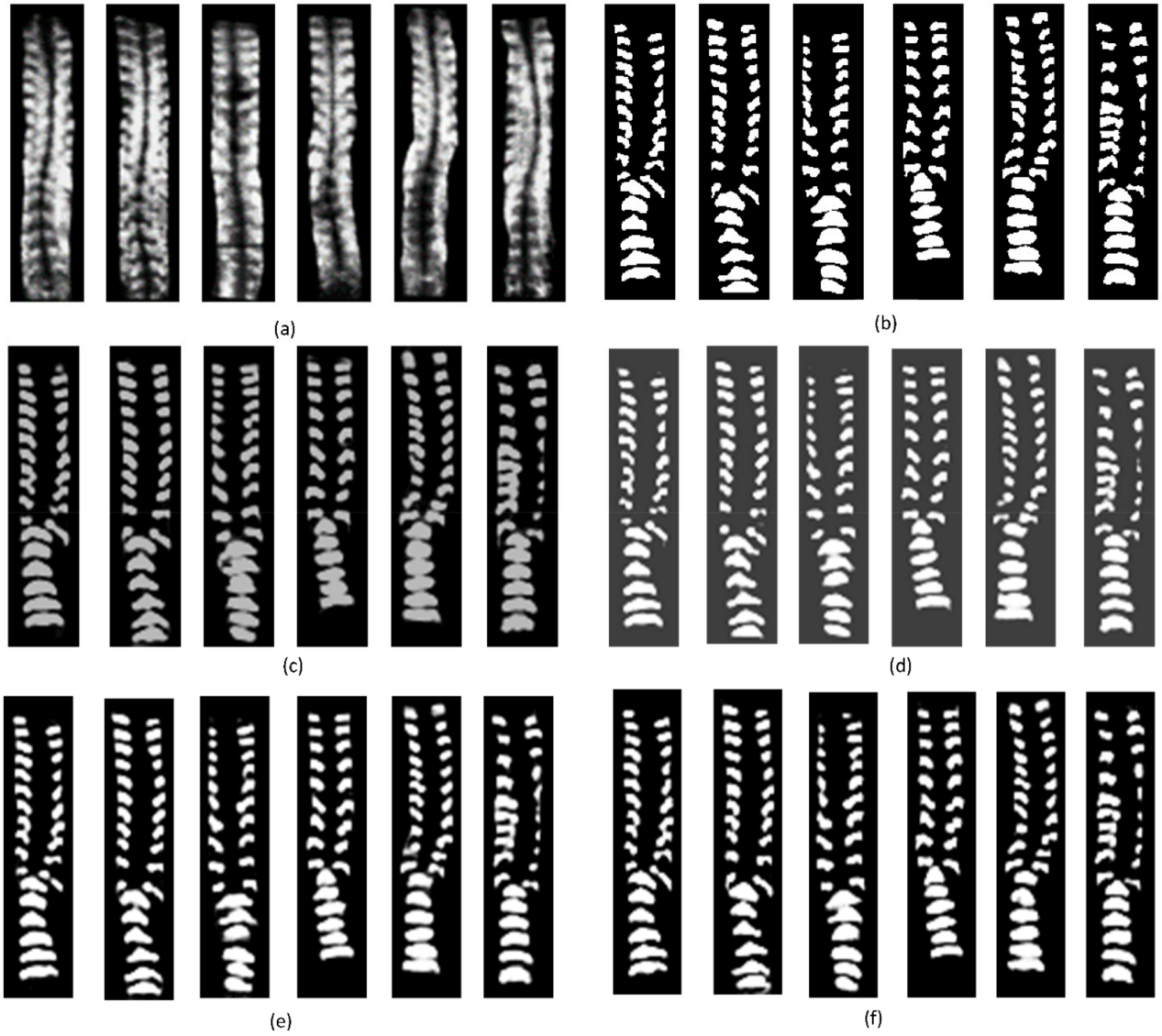

Figure 12. Qualitative comparison of 6 different case models: (**a**) Raw image (**b**) truth mask, (**c**) LD model, (**d**) LCS model, (**e**) LDI model, and (**f**) proposed LDS U-Net model.

4.2. Comparison of LDS U-Net with Other Contemporary Models

The LDS U-Net was compared with three architectures, namely U-Net, Attention U-Net, and MultiResUNet. Each of these models has its unique advantages. However, when it comes to the segmentation of a noisy ultrasound image, these architectures might not support deep-level data extraction and high image variability.

4.2.1. LDS U-Net Outperforms U-Net, Attention U-Net, and MultiResUNet in the Segmentation of Ultrasound Spine Image Dataset

In Table 2, the segmentation output of the proposed LDS U-Net is compared with other models using four segmentation evaluation indices: Jaccard index, Dice score, F1 score, and pixel accuracy. Dice score is the most direct evaluation index. The average Dice score of LDS U-Net was 0.8964, which was 2.79%, 4.57% and 6.89%, higher than MultiResUNet, Attention U-Net, and U-Net, respectively. In terms of the Jaccard index, F1-score, and pixel accuracy, the results of LDS U-Net were much better than the other mentioned networks. The standard deviations of the results of the proposed LDS U-Net also showed a smaller spread. The proposed LDS U-Net is more capable of learning almost all the features from the training dataset.

Table 2. Quantitative performance evaluation of various architectures.

Method	Avg. Jaccard Index (Std Dev)	Avg. Dice Score (Std Dev)	Avg. F1 Score (Std Dev)	Avg. Accuracy (Std Dev)
U-Net	0.7015 (±0.035)	0.8133 (±0.039)	0.8327 (±0.0296)	0.8919 (±0.0271)
Attention U-Net	0.7189 (±0.033)	0.8297 (±0.037)	0.8401 (±0.0296)	0.9195 (±0.0269)
MultiResUNet	0.7264 (±0.032)	0.8458 (±0.033)	0.8658 (±0.0289)	0.9398 (±0.0258)
Proposed LDS U-Net	**0.7415 (±0.03)**	**0.8694 (±0.0285)**	**0.8885 (±0.0256)**	**0.9592 (±0.020)**

4.2.2. LDS U-Net Gives the Best Identification of Bony Features

The detection of exact locations of lateral bony features (thoracic and lumbar) is very crucial for UCA calculation. However, the associated speckle noise makes the identification process more challenging because it suppresses many important features. Four sets of images are shown in Figure 13. Each set consists of an input image, a truth mask, and the segmentation results using U-Net, Attention U-Net, and MultiResUNet and the proposed LDS U-Net model.

From the set shown in Figure 13a, it is clear that the right side upper TBFs of the spine image are missing in the U-Net segmentation. For Attention U-Net, these are just visible, and for MultiResUNet, and LDS U-Net, they are more prominent. In the case of LBFs, identification is not as straightforward. The T12 region and six LBFs are visible in the truth mask. U-Net can identify the T12 region and five LBFs. The Attention U-Net and MultiResUNet could identify only four of the LBFs. The proposed LDS U-Net could identify all LBFs and the T12 region, which indicates that this model is capable of segmenting LBFs correctly.

From Figure 13b, it is clear that U-Net is not capable of segmenting all the bony features in the six different segments of LBFs. Attention U-Net and MultiResUNet were somehow able to distinguish the six features, but some features were unclear and provided misleading information. On the other hand, LDS U-Net provided a clearer identification of the T12 region and six LBFs, although the last two features were still conjoined.

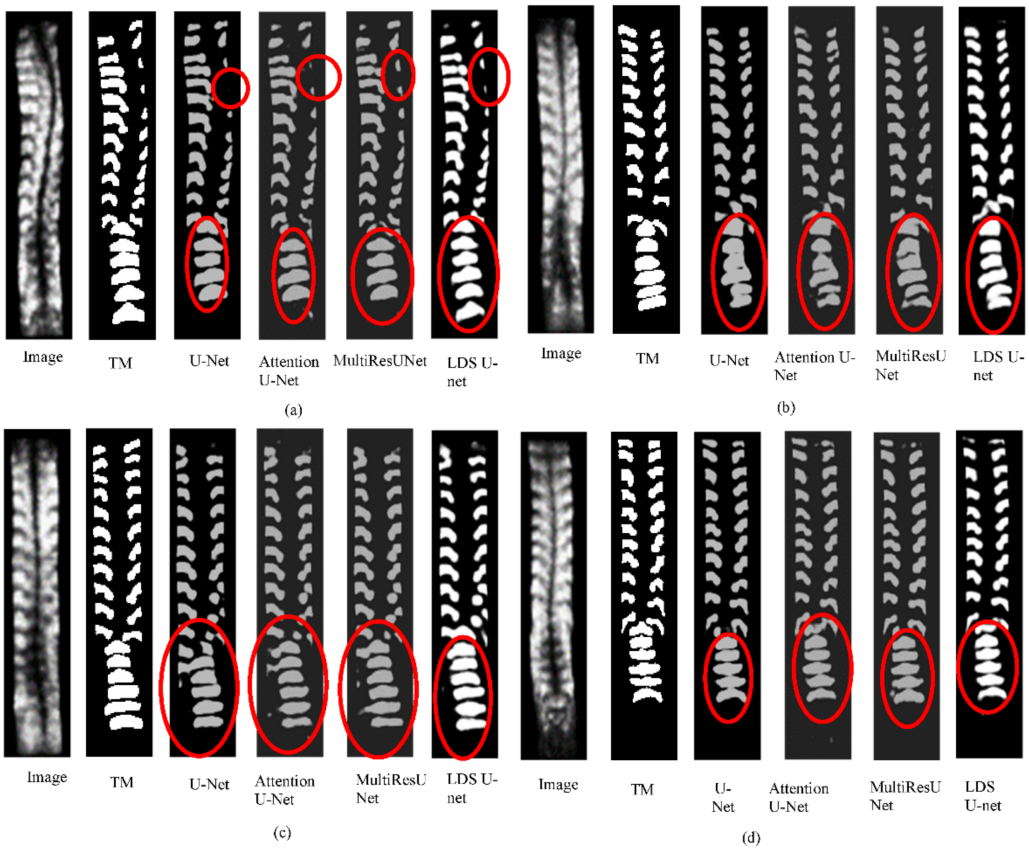

Figure 13. Four spine image sets (**a**–**d**) consisting of the input image, truth mask, and segmentation results using U-Net, Attention U-Net, MultiResUNet, and the proposed LDS U-Net model.

Similar observations about TBFs and LBFs can be made in Figure 13c. In U-Net, the two uppermost LBFs were not segmented correctly, but appeared to be fragmented, while in Attention U-Net and MultiResUNet, they appeared to be elongated. LDS U-Net clearly distinguished all the thoracic and lumbar bony features.

Figure 13d shows that U-Net and MultiResUNet were unable to segment the T12 region and all LBFs from the ultrasound spine image. On the other hand, Attention U-Net could identify all six LBFs. However, in both cases, the T12 region was conjoined with the uppermost lumber bony feature. The proposed LDS U-Net could segment the T12 region and all the lumbar bony features consistently.

4.2.3. Evaluation of the Number of Bony Features Identified

Identification of distinct bony features is a key criterion for evaluating the performance of various segmentation models. In a truth mask of a spine ultrasound, depending on the scanning coverage area, there are nine pairs of TBFs, a T12 region (one pair of bones), and six LBFs, making a total of 26 bony features. The total number of distinguishable bony features in the outputs of each model is manually recorded. In many cases, the bony features are either conjoined or fragmented and are not recorded as meaningful features. Table 3 shows that LDS U-Net provides a larger number of meaningful bony features when compared to other models. The proposed LDS U-Net could provide the highest percentage of segmented images with all 26 features fully detected (76.59%).

Table 3. Comparison between number of bony features identified (avg).

Truth Mask	U-Net	Attention U-Net	MultiResUNet	LDS U-Net
9 pairs of TBF, T12 region, and 6 LBFS—Total: 26 bony features	25.11	25.23	25.30	**25.45**
% of images where all 26 features were fully detected	69.72%	71.56%	73.23%	**76.59%**

4.2.4. Evaluation of Computational Requirement

In Table 4, the computational requirement of the contemporary models and the proposed model are compared in terms of the total number of parameters and program storage requirements. LDS U-Net architecture offers advantages such as (i) low computational requirements due to the usage of depthwise separable convolution layers; (ii) feature reuse enabled through the application of novel light dense block; and (iii) enhancement of the depth and width of the network, with no incremental computational budget, through the redesigned multi-scale inception layer. As shown in Table 4, these advantages of the proposed model translate to a lower parameter requirement and smaller memory footprint when compared to U-Net and Attention U-Net. Although MultiResUNet requires a fewer number of parameters than LDS U-Net, the overall segmentation performance of MultiResUNet was lower, as discussed in the previous sections.

Table 4. Evaluation of computational requirement.

Method	Total Number of Parameters	Size in Megabytes (MB)
U-Net	31 Mil	355 MB
Attention U-Net	37.1 Mil	433.49 MB
MultiResUNet	7.2 Mil	146.01 MB
Proposed LDS U-Net	**9.1 Mil**	**127.5 MB**

5. Discussion

Being a radiation-free and economic imaging modality, the 3D ultrasound imaging system has the potential to become a very popular diagnosis technique to detect scoliosis [15]. However, this necessitated the establishment of new measurement indices as, in the X-ray and Cobb angle method, scanning is conducted to assess anterior spinal deformity. A new index called spinous process angle (SPA) was developed exclusively for ultrasound scanning to measure the posterior spinal deformity. In a nutshell, if lines are drawn through the most tilted parts of the spinous column profile of coronal ultrasound images, then the angle that is formed is known as the SPA. The SPA measurement focuses on the middle dark line of the spine ultrasound image as the main region of interest. Subsequently, several scoliosis assessment techniques, both manual and automatic, have been researched to measure SPA and to demonstrate its equivalence to the gold-standard Cobb angle [16,18]. However, there is an overall drawback in the SPA measurement approach as there was a high inherent tendency of the technique to underestimate the severity of the curvature of scoliosis compared to the Cobb angle [1,15]. Subsequently, an alternate index was developed called the ultrasound curvature angle (UCA). Unlike SPA, in the UCA technique, the lateral bony features are the main regions of interest and further research was undertaken to demonstrate that it is equivalent to the X-ray Cobb angle [20]. Though a manual method of UCA measurement is established [20], there is a need to find a suitable method to automate the UCA measurement so that the technique can be made fast and scalable.

Clear demarcation of lateral bony features is a vital step in UCA measurement. In traditional X-ray, the images are relatively noise-free and clear features can be identified

easily. However, in the case of ultrasound modality, scanned images have low contrast, and are often plagued by speckle noise. This makes the differentiation of the bony features challenging. In addition, the fact that the lateral bony features are numerous and can significantly vary in shape, size, and location, adds to the overall complexity of the problem. The first step of automatic UCA measurement is clear segmentation of the lateral bony features and this paper attempted to accomplish this first step of image segmentation using deep learning.

U-Net is one of the most successful architectures for biomedical image processing. As a starting point, the ultrasound spine images were segmented using basic U-Net [26]. However, the segmentation output was not satisfactory (Avg. Dice score: 0.8133) as the output images had many missed or conjoined bony features and the architecture required a large number of parameters, which directly increased the computation cost and memory size [66]. As the first modification, the conventional convolution layers of U-Net were replaced with proposed light dense block and depthwise convolutional operations that replaced the conventional convolution operations. The new architecture was called the light dense (LD) model. In LD model, each depthwise convolution unit is densely connected with another depthwise convolution unit within the same light dense block (shown in Figure 6). This architecture incorporates feature propagation and boosts feature reuse while effectively reducing the number of parameters. Though the LD model obtained better segmentation accuracy (Avg. Dice score: 0.8204) with a fewer number of parameters, it was still unable to adequately distinguish the TBFs and LBFs and generated conjoined bony features in the output as the input images were noisy. The two issues, high noise in images and variability of bony features, were tackled separately as follows:

(a) To improve the segmentation clarity and tackle the 'noisy' information, trainable selection gates were employed in the LD model as the next modification and the new architecture was called the light convolution dense selection (LCS) model. Through this modification, it was anticipated that the gating mechanism [43] would smartly suppress irrelevant information such as noise from the feature maps and be able to identify a greater number of bony features from the useful information in the image. Past research shows that the gating mechanism can be effectively used to extract the selective features from the noisy ultrasound images as it suppresses noise and enables the network to make segmentation predictions based on class-specific features [67]. In effect, the selection gate blocked extraneous information from information-rich feature maps and thereby improved the segmentation output for noisy ultrasound images (Avg. Dice score: 0.8292).

(b) To handle the problem of large variabilities in shape, size, and locations of the TBFs and LBFs, the conventional skip-pathways were replaced by multi-scale [41] skip-paths and the architecture was named as the light-dense inception (LDI) model. Feature fusion was enhanced by using this skip-path as the inconsistencies between encoder–decoder features were bridged. This path solves the problem of the semantic gap between the encoder and decoder side and enhances the feature fusion and improves the segmentation output [41]. Through this modification, the low-level features underwent additional processing and the multi-scale inception layers, adapted in this skip-path, enabled the network to extract more features from different scales. This provided a significant boost to the segmentation performance (Avg. Dice score: 0.8419) compared to the initial LD model.

As a final step, the two sub-models were combined into the final architecture and called the Light-Convolution Dense Selection U-Net (LDS U-Net) model. As anticipated, the combination of selection gates and multi-scale skip pathway gave a much improved segmentation accuracy (Avg. Dice score: 0.8694) compared to both the LD model and basic U-Net model.

To understand the performance of LDS U-Net vis-a-vie other contemporary models, the proposed model was compared with the basic U-Net, MultiResUNet, and Attention U-Net using the same 109 volume projection ultrasound spine image datasets. First, LDS U-Net quantitatively outperformed the other models with 0.7415 Jaccard index, 0.8694 Dice coefficient, 0.8885 F1 score, and 0.9592 pixel accuracy. Second, it is able to better identify

the thoracic and lumbar bony features most consistently when qualitatively compared to other models. The segmentation outcomes of LDS U-Net, through manual observation and measurement of distinguishable bony features, are also very promising. Third, as it is constructed using light dense block, LDS U-Net was also found to be computationally efficient with the least number of parameters and smallest program memory size. Finally, in handling noisy ultrasound images with high variability, the LDS U-Net was successful in correctly segmenting 76.59% of total images (shown in Table 3) with complete identification of all 26 bony features, which is more consistent than any of the other models.

During the research, it was also observed that there were few cases where the LDS U-Net did not perform well. For instance, in Figure 14, the demarcation between the T12 level and 1st LBF is obscure. In these scenarios, even the other architectures failed to give proper segmentation of this section. Upon closer scrutiny, it was found that the quality of the input image was not adequate for automatic segmentation and experts were able to generate the truth mask only due to their prior knowledge on six lumbar bony features. Hence, in such cases, the image segmentation becomes quite challenging both by the human eye and by machine learning algorithms. As an improvement and for a clearer demarcation of the T12 level and 1st LBF region, the authors postulate that other images, from different depths, can be included in the segmentation process. Second, as an overall improvement in segmentation performance, future work should be aimed at better removal of the excessive noise by making the architecture denser. Additionally, as the research is restricted to the availability of 109 image datasets, the consistency of the performance can be improved by including more image datasets from diverse patient groups.

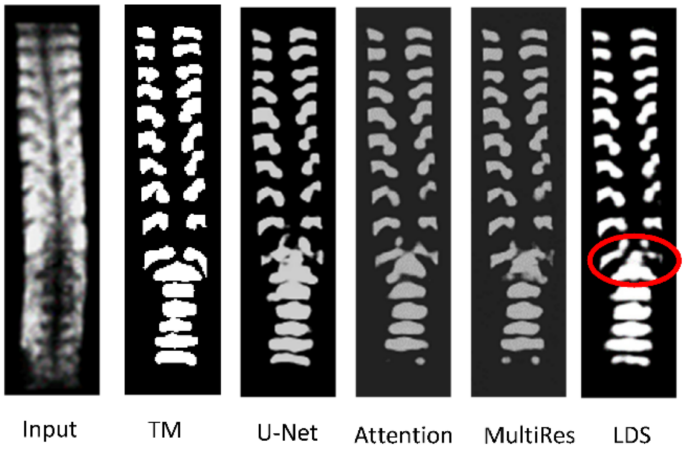

Figure 14. LDS U-Net is unable to distinguish the T12 level and 1st LBF.

6. Conclusions

In this work, a novel CNN architecture, LDS U-Net, was developed to identify and segment lateral bony features from spine ultrasound images. As ultrasound images are contaminated with noise and are of low contrast, segmentation work is more challenging in this case when compared to other imaging modalities. The research establishes that LDS U-Net was successful in segmenting lateral bony features and can be included as a step prior to the actual UCA calculation. In a nutshell, the proposed LDS U-Net has three main elements: (a) light dense blocks that reduce the number of parameters used in the architecture and thereby reduce the computation time; (b) selection gates that smartly discard the extraneous regions and ensure explicit flow of only the relevant information within the network; and (c) multi-scale paths that help with clearer identification of lateral bony features through improved feature propagation. The next step, post automatic segmentation, would be to work on automating the process of the ultrasound curvature angle (UCA) measurement using segmented bony regions as input and validating the results with extensive clinical trials and evaluations.

Author Contributions: Conceptualization, S.B., S.-H.L. and Y.Z.; methodology, S.B. and S.-H.L.; software, S.B. and S.-H.L.; validation, S.-H.L., H.F.F.L., T.T.-Y.L., D.Y., S.S. and Y.Z.; formal analysis, J.L. and Z.H.; investigation, T.T.-Y.L., D.Y. and Y.Z.; resources, S.-H.L. and Y.Z.; data curation, S.B., J.L. and Z.H.; writing—original draft preparation, S.B. and S.-H.L.; writing—review and editing, S.B., S.-H.L., H.F.F.L., D.Y. and T.T.-Y.L.; visualization, S.-H.L. and Y.Z.; supervision, S.-H.L., H.F.F.L., S.S. and Y.Z.; project administration, S.-H.L. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The project was partially supported by the Hong Kong Research Grant Council Research Impact Fund (R5017-18).

Institutional Review Board Statement: The experimental procedures involving human subjects were approved by the Institutional Review Board of The Hong Kong Polytechnic University (HSEARS20180906005, 7 September 2018). The subjects gave informed consent for their inclusion in this study as required, and the work adheres to the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank the Australian Government for the Research Training Program Stipends (RTPS) awarded to the first author.

Conflicts of Interest: No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

References

1. Kim, H.; Kim, H.S.; Moon, E.S.; Yoon, C.S.; Chung, T.S.; Song, H.T.; Suh, J.S.; Lee, Y.H.; Kim, S. Scoliosis imaging: What radiologists should know. *Radiographics* **2010**, *30*, 1823–1842. [CrossRef] [PubMed]
2. Zheng, Y.-P.; Lee, T.T.Y.; Lai, K.K.L.; Yip, B.H.K.; Zhou, G.Q.; Jiang, W.W.; Cheung, J.C.W.; Wong, M.S.; Ng, B.K.W.; Cheng, J.C.Y.; et al. A reliability and validity study for Scolioscan: A radiation-free scoliosis assessment system using 3D ultrasound imaging. *Scoliosis Spinal Disord.* **2016**, *11*, 13. [CrossRef] [PubMed]
3. Ran, H.; Zhi-hong, W.; Jiang-na, H. Scoliosis on pulmonary function. *Acta Acad. Med. Sin.* **2011**, *33*, 102–106.
4. Li, S.; Yang, J.; Zhu, L.; Li, Y.; Peng, H.; Lin, Y.; Li, X.; Huang, Z.; Wang, H. Left ventricular mechanics assessed by 2-dimensional speckle tracking echocardiography in children and adolescents with idiopathic scoliosis. *Clin. Spine Surg.* **2017**, *30*, E381–E389. [CrossRef]
5. Liu, D.; Yang, Y.; Yu, X.; Yang, J.; Xuan, X.; Yang, J.; Huang, Z. Effects of specific exercise therapy on adolescent patients with idiopathic scoliosis: A prospective controlled cohort study. *Spine* **2020**, *45*, 1039. [CrossRef]
6. Cobb, J. Outline for the study of scoliosis. In *American Academy of Orthopaedic Surgeons Instructional Course Lectures*; J.W. Edwards: Wuhan, China, 1948; Volume 5, pp. 261–275.
7. Levy, A.R.; Goldberg, M.S.; Mayo, N.E.; Hanley, J.A.; Poitras, B. Reducing the lifetime risk of cancer from spinal radiographs among people with adolescent idiopathic scoliosis. *Spine* **1996**, *21*, 1540–1547. [CrossRef]
8. McKenna, C.; Wade, R.; Faria, R.; Yang, H.; Stirk, L.; Gummerson, N.; Sculpher, M.; Woolacott, N. EOS 2D/3D X-ray imaging system: A systematic review and economic evaluation. *Health Technol. Assess. Winch. Engl.* **2012**, *16*, 1. [CrossRef]
9. Rehm, J.; Germann, T.; Akbar, M.; Pepke, W.; Kauczor, H.U.; Weber, M.A.; Spira, D. 3D-modeling of the spine using EOS imaging system: Inter-reader reproducibility and reliability. *PLoS ONE* **2017**, *12*, e0171258. [CrossRef]
10. Vergari, C.; Gajny, L.; Courtois, I.; Ebermeyer, E.; Abelin-Genevois, K.; Kim, Y.; Langlais, T.; Vialle, R.; Assi, A.; Ghanem, I.; et al. Quasi-automatic early detection of progressive idiopathic scoliosis from biplanar radiography: A preliminary validation. *Eur. Spine J.* **2019**, *28*, 1970–1976. [CrossRef]
11. Ungi, T.; King, F.; Kempston, M.; Keri, Z.; Lasso, A.; Mousavi, P.; Rudan, J.; Borschneck, D.P.; Fichtinger, G. Spinal curvature measurement by tracked ultrasound snapshots. *Ultrasound Med. Biol.* **2014**, *40*, 447–454. [CrossRef]
12. Suzuki, S.; Yamamuro, T.; Shikata, J.; Shimizu, K.; Iida, H. Ultrasound measurement of vertebral rotation in idiopathic scoliosis. *J. Bone Jt. Surg. Br. Vol.* **1989**, *71*, 252–255. [CrossRef]
13. Huang, Q.-H.; Zheng, Y.-P.; Lu, M.-H.; Chi, Z. Development of a portable 3D ultrasound imaging system for musculoskeletal tissues. *Ultrasonics* **2005**, *43*, 153–163. [CrossRef]
14. Huang, Q.-H.; Zheng, Y.-P.; Li, R.; Lu, M.-H. 3-D measurement of body tissues based on ultrasound images with 3-D spatial information. *Ultrasound Med. Biol.* **2005**, *31*, 1607–1615. [CrossRef]
15. Zhou, G.-Q.; Zheng, Y.-P. Assessment of scoliosis using 3-D ultrasound volume projection imaging with automatic spine curvature detection. In Proceedings of the 2015 IEEE International Ultrasonics Symposium (IUS), Taipei, Taiwan, 21–24 October 2015; pp. 1–4.
16. Cheung, C.-W.J.; Zhou, G.-Q.; Law, S.-Y.; Mak, T.-M.; Lai, K.-L.; Zheng, Y.-P. Ultrasound volume projection imaging for assessment of scoliosis. *IEEE Trans. Med Imaging* **2015**, *34*, 1760–1768. [CrossRef]

17. Zhou, G.-Q.; Jiang, W.-W.; Lai, K.-L.; Lam, T.-P.; Cheng, J.C.-Y.; Zheng, Y.-P. Semi-automatic Measurement of Scoliotic Angle Using a Freehand 3-D Ultrasound System Scolioscan. In Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing, Paphos, Cyprus, 31 March–2 April 2016; Springer: New York, NY, USA, 2016; pp. 341–346.
18. Zhou, G.-Q.; Jiang, W.-W.; Lai, K.-L.; Zheng, Y.-P. Automatic measurement of spine curvature on 3-D ultrasound volume projection image with phase features. *IEEE Trans. Med Imaging* **2017**, *36*, 1250–1262. [CrossRef]
19. Loupas, T.; McDicken, W.; Allan, P.L. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE Trans. Circuits Syst.* **1989**, *36*, 129–135. [CrossRef]
20. Lee, T.T.-Y.; Lai, K.K.-L.; Cheng, J.C.-Y.; Castelein, R.M.; Lam, T.-P.; Zheng, Y.-P. 3D ultrasound imaging provides reliable angle measurement with validity comparable to X-ray in patients with adolescent idiopathic scoliosis. *J. Orthop. Transl.* **2021**, *29*, 51–59.
21. Vedula, S.; Senouf, O.; Bronstein, A.M.; Michailovich, O.V.; Zibulevsky, M. Towards CT-quality ultrasound imaging using deep learning. *arXiv* **2017**, arXiv:1710.06304.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems, Proceedings of the 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012*; Curran Associates, Inc.: Red Hook, NY, USA, 2013; pp. 1097–1105.
23. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
24. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
25. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Computer Vision—ECCV 2014, Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014*; Springer: New York, NY, USA, 2014; pp. 818–833.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: New York, NY, USA, 2015; pp. 234–241.
27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2015; pp. 3431–3440.
28. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
29. Hu, R.; Dollár, P.; He, K.; Darrell, T. Girshick RLearning to Segment Every Thing. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 22–27.
30. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
31. Li, J.; Wu, Y.; Zhao, J.; Guan, L.; Ye, C.; Yang, T. Pedestrian detection with dilated convolution, region proposal network and boosted decision trees. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AL, USA, 14–19 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 4052–4057.
32. Sun, W.; Wang, R. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 474–478. [CrossRef]
33. Roth, H.R.; Shen, C.; Oda, H.; Oda, M.; Hayashi, Y.; Misawa, K.; Mori, K. Deep learning and its application to medical image segmentation. *Med Imaging Technol.* **2018**, *36*, 63–71.
34. Amiri, M.; Brooks, R.; Behboodi, B.; Rivaz, H. Two-stage ultrasound image segmentation using U-Net and test time augmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 981–988. [CrossRef]
35. Han, M.; Bao, Y.; Sun, Z.; Wen, S.; Xia, L.; Zhao, J.; Du, J.; Yan, Z. Automatic segmentation of human placenta images with U-Net. *IEEE Access* **2019**, *7*, 180083–180092. [CrossRef]
36. Thomson, B.R.; Nijkamp, J.; Ivashchenko, O.; van der Heijden, F.; Smit, J.N.; Kok, N.F.; Kuhlmann, K.F.; Ruers, T.J.; Fusaglia, M. Hepatic vessel segmentation using a reduced filter 3D U-Net in ultrasound imaging. *arXiv* **2019**, arXiv:1907.12109.
37. Lyu, J.; Bi, X.; Banerjee, S.; Huang, Z.; Leung, F.H.; Lee, T.T.Y.; Yang, D.D.; Zheng, Y.P.; Ling, S.H. Dual-task ultrasound spine transverse vertebrae segmentation network with contour regularization. *Comput. Med Imaging Graph.* **2021**, *89*, 101896. [CrossRef]
38. Banerjee, S.; Ling, S.H.; Lyu, J.; Su, S.; Zheng, Y.-P. Automatic Segmentation of 3D Ultrasound Spine Curvature Using Convolutional Neural Network. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2039–2042.
39. Huang, Z.; Wang, L.W.; Leung, F.H.; Banerjee, S.; Yang, D.; Lee, T.; Lyu, J.; Ling, S.H.; Zheng, Y.P. Bone Feature Segmentation in Ultrasound Spine Image with Robustness to Speckle and Regular Occlusion Noise. *arXiv* **2020**, arXiv:2010.03740.
40. Ravishankar, H.; Venkataramani, R.; Thiruvenkadam, S.; Sudhakar, P.; Vaidya, V. Learning and incorporating shape models for semantic segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017*; Springer: New York, NY, USA, 2017; pp. 203–211.

41. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87. [CrossRef]
42. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv* **2016**, arXiv:1602.07261.
43. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
44. Guan, S.; Khan, A.A.; Sikdar, S.; Chitnis, P.V. Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 568–576. [CrossRef] [PubMed]
45. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **2017**, *26*, 4509–4522. [CrossRef] [PubMed]
46. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
47. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
48. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
49. Denil, M.; Shakibi, B.; Dinh, L.; Ranzato, M.A.; de Freitas, N. Predicting parameters in deep learning. In *Advances in Neural Information Processing Systems, Proceedings of the 27th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013*; Curran Associates, Inc.: Red Hook, NY, USA, 2013; Volume 26, pp. 2148–2156.
50. Sifre, L.; Mallat, P.S. Rigid-Motion Scattering For Image Classification Author. Ph.D. Thesis, Ecole Polytechnique, Palaiseau, France, 2014.
51. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
52. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
53. Guo, J.; Li, Y.; Lin, W.; Chen, Y.; Li, J. Network decoupling: From regular to depthwise separable convolutions. *arXiv* **2018**, arXiv:1808.05517.
54. Kaiser, L.; Gomez, A.N.; Chollet, F. Depthwise separable convolutions for neural machine translation. *arXiv* **2017**, arXiv:1706.03059.
55. Bi, L.; Kim, J.; Kumar, A.; Fulham, M.; Feng, D. Stacked fully convolutional networks with multi-channel learning: Application to medical image segmentation. *Vis. Comput.* **2017**, *33*, 1061–1071. [CrossRef]
56. Pleiss, G.; Chen, D.; Huang, G.; Li, T.; van der Maaten, L.; Weinberger, K.Q. Memory-efficient implementation of densenets. *arXiv* **2017**, arXiv:1707.06990.
57. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. DRINet for medical image segmentation. *IEEE Trans. Med Imaging* **2018**, *37*, 2453–2462. [CrossRef]
58. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
59. Chollet, F. Keras. 2015. Available online: https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1887532 (accessed on 26 October 2021).
60. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
61. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
62. Jaccard, P. The distribution of the flora in the alpine zone. *New Phytol.* **1912**, *11*, 37–50. [CrossRef]
63. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
64. Huang, H.; Xu, H.; Wang, X.; Silamu, W. Maximum F1-score discriminative training criterion for automatic mispronunciation detection. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 787–797. [CrossRef]
65. Abbasian Ardakani, A.; Bitarafan-Rajabi, A.; Mohammadzadeh, A.; Mohammadi, A.; Riazi, R.; Abolghasemi, J.; Homayoun Jafari, A.; Bagher Shiran, M. A hybrid multilayer filtering approach for thyroid nodule segmentation on ultrasound images. *J. Ultrasound Med.* **2019**, *38*, 629–640. [CrossRef]
66. Gadosey, P.K.; Li, Y.; Agyekum, E.A.; Zhang, T.; Liu, Z.; Yamak, P.T.; Essaf, F. Sd-unet: Stripping down u-net for segmentation of biomedical images on platforms with low computational budgets. *Diagnostics* **2020**, *10*, 110. [CrossRef]
67. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef]