SPGNet: A Shape-prior Guided Network for Medical Image Segmentation

Zhengxuan Song[1] , **Xun Liu**[2] , **Wenhao Zhang**[3] , **Yongyi Gong**[4] , **Tianyong Hao**[5] and **Kun Zeng**[1,*]
[1]Sun Yat-sen University
[2]The Third Affiliated Hospital of Sun Yat-sen University
[3]University of the West of England
[4]Guangdong University of Foreign Studies
[5]South China Normal University
songzhx6@mail2.sysu.edu.cn, naturestyle@163.com, Wenhao.Zhang@uwe.ac.uk,
gongyongyi@gdufs.edu.cn, haoty@m.scnu.edu.cn, zengkun2@mail.sysu.edu.cn

Abstract

Given the intricacy and variability of anatomical structures in medical images, some methods employ shape priors to constrain segmentation. However, limited by the representational capability of these priors, existing approaches often struggle to capture diverse target structure morphologies. To address this, we propose SPGNet to guide segmentation by fully exploiting category-specific shape knowledge. The key idea is to enable the network to perceive data shape distributions by learning from statistical shape models. We uncover shape relationships via clustering and obtain statistical prior knowledge using principal component analysis. Our dual-path network comprises a segmentation path and a shape-prior path that collaboratively discern and harness shape prior distribution to improve segmentation robustness. The shape-prior path further serves to refine shapes iteratively by cropping features from the segmentation path, guiding the segmentation path and directing attention specifically to the edges of shapes which could be most significantly susceptible to segmentation error. We demonstrate superior performance on chest X-ray and breast ultrasound benchmarks.

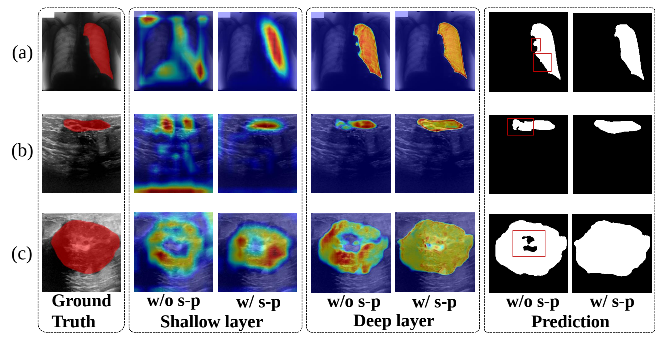

Figure 1: The first column shows the ground truth masks for a set of right lung X-rays, benign breast tumor ultrasound, and malignant tumor ultrasound images. The second to fifth columns present segmentation path visual attention maps in shallow encoder and deep decoder layers, without or with the collaborative shape-prior path. The sixth and seventh columns depict segmentation outputs with or without the shape-prior path. The network guided by shape priors demonstrates improved shape integrity and smoother edges.

1 Introduction

Medical image segmentation has always been critical to medical image processing. Currently, most mainstream methods focus primarily on high-precision pixel-level supervision. Despite significant achievements in a variety of segmentation tasks in differing domains, limitations persist when dealing with medical images. Primarily, anatomical structures in medical images often exhibit shape patterns and geometric information that generic pixel-level supervision could fails to leverage fully, especially when shape regularities are prominent and perceptibly advantageous.

Past studies indicate integrating shape prior knowledge could benefit traditional segmentation algorithms [Nosrati *et al.*, 2016]. For instance, introducing shape priors to level set contour evolution techniques proved to enhance the accuracy of segmentation [Chen *et al.*, 2002]. Integrating elastic shape priors into frameworks was also shown to be able to align the outcomes closer to the actual shape variations [Schoenemann *et al.*, 2007]. However, these conventional segmentation methods often struggle with complex medical image scenarios due to constraints in data distribution assumptions and susceptibility to noise and contextual information.

Deep learning methods, in comparison, exhibit greater flexibility in adapting to diverse shapes and backgrounds [Bohlender *et al.*, 2021]. It has been shown that this paradigm of segmentation methods could also benefit from shape priors to better cater for specific task requirements. For example, a model based on sparse representation and local repulsive deformation was proposed for normalizing the former deep convolutional neural network segmentation and constraining the segmentation results within an effective shape domain [Xing *et al.*, 2015]. Although post-processing has often demonstrated effective for further improving segmentation [Li *et al.*, 2017; Medley *et al.*, 2019], its usefulness still hinges on the dependability of the segmentation model. Another model [Lee *et al.*, 2019] based on template deformation implemented seg-

[*]Corresponding author

mentation by deforming the shape prior template. However, this approach may be constrained by the expressiveness of the priors and would likely be undermined by structural variations in medical images, such as when handling images of various organs with distinct shapes and morphological characteristics.

While these methods aim to constrain segmentation using shape priors, they have not entirely endowed deep neural networks with the capacity to perceive shape patterns in the data. We propose to enable deep learning methods to fully incorporate learned priors from training data as a form of regularization. This enables segmentation networks to be influenced by prior shape knowledge when classifying pixels. As shown in Figure 1, with shape prior guidance, attention in shallow layers concentrates completely on the target area. In deep layers, the excellent alignment between the attention map and the target area is maintained, while the most activated attention areas shift towards the edges which serves to improve the accuracy and smoothness of segmentation. Conversely, a network without guidance exhibits fragmented and misaligned attention highly susceptible to noise and variability.

As shown in Figure 2, our end-to-end dual-path collaborative network integrates a multi-class statistical shape model for incorporating a wealth of shape priors. The segmentation path focuses on dense pixel-level classification, while the shape-prior path regresses prior flows. We introduce a collaboration module establishing interactions between encoders to enhance robustness by exchanging features. Simultaneously, by collaboratively learning multi-class statistical shape model deformation, the segmentation encoder is enabled to perceive shape distribution and utilise this explicit shape prior knowledge to guide segmentation. We also use cropped local features from the segmentation path to refine shapes across different scales, guiding attention to focus on boundaries.

Our contributions are summarized as follows:

- We propose a novel dual-path collaborative segmentation network, SPGNet, which embeds explicit and diverse shape priors. The dual-path structure enhances the representation capability of segmentation path encoders. The segmentation path, guided by the explicit shape priors, reinforces shape understanding and enhances attention at target edges, addressing single-path deficiencies in capturing shape features. We also design a cluster strategy to learn shape regularities from the training set.
- We explore the collaborative effects of SPGNet and demonstrate the effectiveness of the shape-prior path in improving segmentation accuracy. Specifically, we validate the efficacy of each component within the shape-prior path.
- We evaluate SPGNet on a chest X-ray dataset with prominent shape regularities and a breast ultrasound dataset with potential regularities. Results demonstrate superior accuracy over baselines and existing state-of-the-art methods, particularly in edge smoothness. We also validate its adaptability to medical images with low signal-to-noise ratios, blurry boundaries, and significant shape and positional variations of lesions. The superiority of our method has been demonstrated.

2 Relate Works

Shape Clustering and Statistical Modeling. Shape clustering groups shapes by extracting descriptors and clustering based on similarity distances. For instance, a skeleton-based approach captures intrinsic structural information for same-class shapes, clustering using a node-matching matrix [Shen *et al.*, 2013]. However, skeleton-based methods often express relatively coarse shape features. In contrast, a hierarchical clustering method is used in a different study for contour-based shapes to learn probabilistic models from shape clusters [Srivastava *et al.*, 2005]. In another approach, to address unlabeled longitudinal shape data, a flexible nonlinear mixture model is established by learning average shape trajectories and variances for each cluster [Debavelaere *et al.*, 2020]. For instance segmentation, k-means clustering of training masks are used to obtain centres of shape clusters, establishing a linear prior model e.g., [Kuo *et al.*, 2019]. A pipeline combining segmentation, clustering, and modeling has also been proposed [Bruse *et al.*, 2017]. With the rise of deep neural networks, exploring clustered shape information with statistical models still has significant potential.

Shape-prior Guided Segmentation. Numerous methods have attempted to leverage shape priors for segmentation. For example, to address blurred overlapping regions in the cell cytoplasm, a generator utilizing a prior template to generate masked was proposed [Song *et al.*, 2020]. Similarly, a cyclic registration network was also designed to integrate anatomical context specificity with priors [Jiang and Veeraraghavan, 2022]. For a different medical application, a deep neural network was designed to predict PCA layers for improving the segmentation of the left ventricle in ultrasound images [Milletari *et al.*, 2017]. Generating threshold-based priors and optimizing outputs via a spatial transform network was also proposed as another strategy for harnessing prior shape knowledge [Zhao *et al.*, 2021]. More recently, it was also found that introducing an additional branch transforms visible regions into complete areas through supervision, and therefore facilitates holistic shape understanding [Gao *et al.*, 2023]. Similarly, another study proposed a generative invariant shape prior network that introduced a branch to learn invariant priors, mimicking human perceptual learning of basic shapes [Li *et al.*, 2023]. Other than considering two-dimensional shapes, a different method used three-dimensional reconstructed shapes as priors and reconstructed occluded objects before projecting them to predict complete mask [Li *et al.*, 2022]. However, these methods may struggle to handle complex and heterogeneous targets due to insufficient shape diversity [Zhao *et al.*, 2021; Jiang and Veeraraghavan, 2022]. Implicit priors may fail to generalize [Gao *et al.*, 2023; Li *et al.*, 2023]. In contrast, our approach establishes an explicit multi-class shape statistical model to guide segmentation.

3 Methodology

3.1 Overview

In Figure 2, we introduce SPGNet, a novel image segmentation algorithm with embedded shape priors. Section 3.2 discusses offline multi-class shape statistical modeling, Sec-

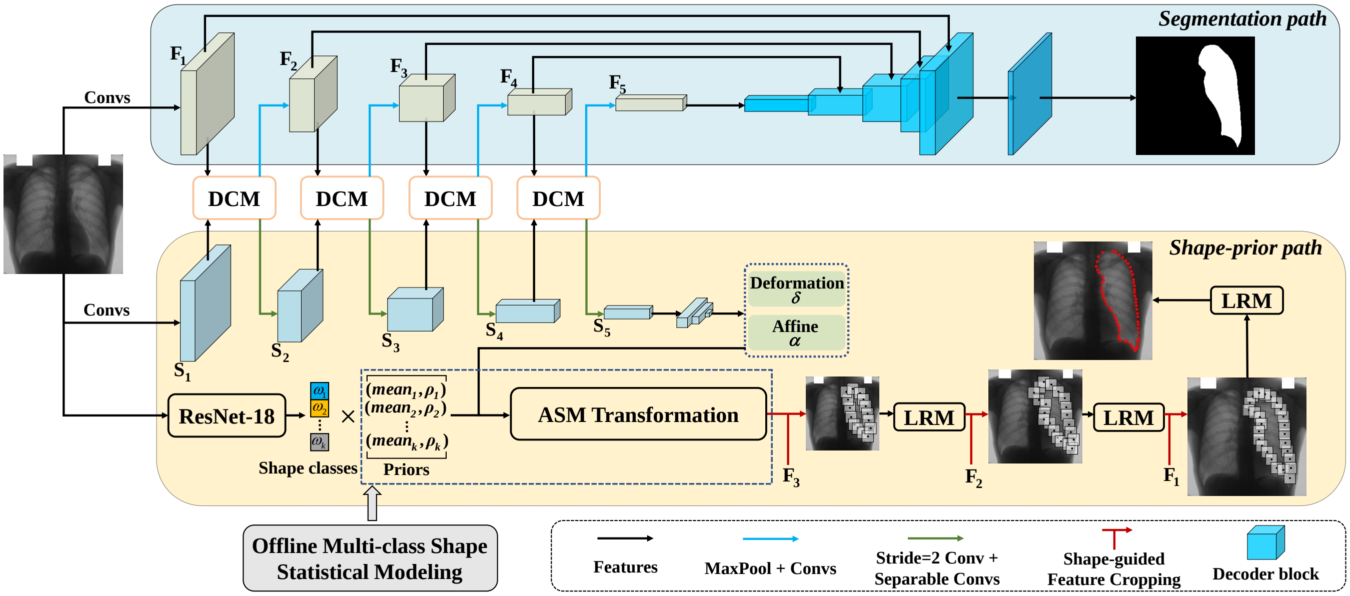

Figure 2: SPGNet is a dual-path collaborative network consisting of three main components: offline multi-class shape statistical modeling, segmentation path, and shape prior path.

tion 3.3 details SPGNet's internal modules and Section 3.4 introduces the hybrid loss function for network training.

3.2 Offline Multi-class Shape Statistical Modeling

Shape Preparation. As shown in Figure 3, we derive shape data from mask annotations, focusing on regions with heightened curvature to build a statistical model with prominent shape features. The shape generation process involves: (i) Employing Bézier curves to smoothly fit mask contours, yielding a set of smoothed contour points. (ii) Determining the number of sampling points as p and calculating the absolute curvature for each point, followed by normalization. A base value is introduced to prevent neglecting points with extremely small curvature. The ratio of the sum of normalized curvatures to the number of sampling points serves as the sampling distance. (iii) Iterating through contour points, saving the current sum of curvatures and the number of sampled points, denoted as k. When the sum exceeds k times the sampling distance, the point is saved as a sample point and k increments by 1. The traversal ends when k reaches p. The sampled p points constitute the shape contour, represented as $S^i = ((x_1^i, y_1^i), ..., (x_p^i, y_p^i))^{\mathrm{T}} \in \mathbb{R}^{p\times 2}$.

Procrustes Shapes Agglomerative Clustering. We employed the agglomerative clustering method, utilizing the Procrustes shape distance as the shape similarity metric, which requires aligning shapes before computing distances. The Procrustes shape distance calculation between two aligned shapes S^1 and S^2 is calculated by:

$$P_d = \sqrt{\sum_{j=1}^{p}[(x_j^1 - x_j^2)^2 + (y_j^1 - y_j^2)^2]} \tag{1}$$

In Figure 3, the agglomerative clustering of shapes involves these steps: (i) Standardizing and aligning all shapes using

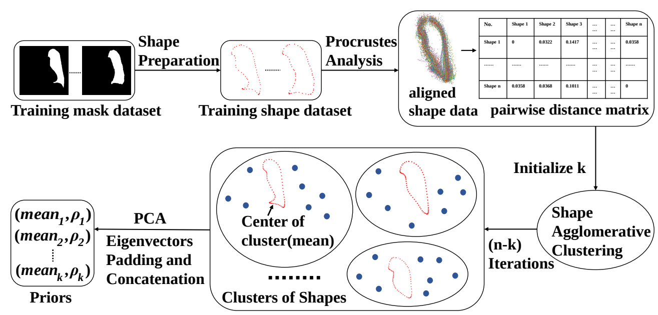

Figure 3: This is the flowchart depicting the process of computing multi-class shape priors.

the Procrustes analysis [Cootes *et al.*, 1995] in the training set. (ii) Calculating the pairwise distance matrix using Procrustes shape distance. (iii) Determining the number of clusters k. Starting with $k = 2$ and iteratively incrementing k, we select the k value that maximizes the variance of distances between k cluster centers as the initialization. We limit k to a maximum of 100 to prevent excessive redundancy in categories. (iv) Initiating agglomerative clustering with complete linkage for calculating inter-cluster distances, minimizing dissimilar shape aggregation between clusters. After all iterations, we obtain k shape clusters.

Multi-class Shape Priors and Modeling. We compute the mean shape within each cluster and conduct principal component analysis (PCA) on cluster shapes to derive eigenvectors representing major variations. The number of eigenvectors per cluster aligns with the maximum across clusters, denoted as t. Concatenating mean shapes and eigenvectors yields shape priors, denoted as $Priors = \{(mean_i, \rho_i)_{i=1}^k\} \in \mathbb{R}^{k\times(1+t)\times p\times 2}$. Simultaneously, shape category labels C are obtained from the shape clusters. The Active Shape Model (ASM) [Cootes *et al.*, 1995], a statistical model for shapes, necessitates a set of shape training samples $Set = \{(S^i)_{i=1}^n\}$ for model construction. Initially, we compute the average shape vector for all shapes:

$$\overline{S} = \frac{1}{n}\sum_{i=1}^{n} S^i \tag{2}$$

We employ PCA for dimensionality reduction on the n shape training samples, creating an approximate model that encapsulates the entire training dataset:

$$S \approx \overline{S} + b\Phi \tag{3}$$

where $\Phi = (\rho_1, \rho_2, ..., \rho_j) \in \mathbb{R}^{j\times p\times 2}$ represents the first j major eigenvectors corresponding to the eigenvalues of the covariance matrix. $b \in \mathbb{R}^{1\times j}$ denotes the shape deformation parameters. Subsequently, we can utilize a neural network to predict the deformation parameters b, facilitating the fitting of the shape statistical model to the shapes in the training set. Additionally, we introduce an affine transformation function a with parameters $\gamma = (s_\gamma, \theta_\gamma, (t_1)_\gamma, (t_2)_\gamma) \in \mathbb{R}^{1\times 4}$. The four parameters include the scale parameter s_γ, rotation parameter θ_γ, and translation parameters $(t_1)_\gamma$ and $(t_2)_\gamma$. For a set of points $S = (S_x, S_y) \in \mathbb{R}^{p\times 2}$, affine transformations can be described as:

$$a(S,\gamma) = \begin{bmatrix} s_\gamma\, cos(\theta_\gamma) & -s_\gamma \sin(\theta_\gamma) & (t_1)_\gamma \\ s_\gamma\, sin(\theta_\gamma) & s_\gamma \cos(\theta_\gamma) & (t_2)_\gamma \end{bmatrix} \begin{bmatrix} S_x \\ S_y \\ 1 \end{bmatrix} \tag{4}$$

In summary, an ASM Transformation model is denoted as:

$$T_{asm}(\overline{S}, \Phi, \gamma, b) = a(\overline{S} + b\Phi, \gamma) \tag{5}$$

The ASM Transformation function is embedded in the network computational process. $Priors$ are precomputed offline before training and inference, entering the network as constant parameters.

3.3 SPGNet

Dual-path Collaboration Module(DCM). In the encoder sections of both paths, we aim for mutual attention during training, allowing the segmentation path to attend to learned features from the shape-prior path and vice versa. To achieve this, we introduce the dual-path collaboration module (DCM), incorporating spatial and channel attention for enhanced feature interaction. As shown in Figure 4, the DCM takes features $F_i \in \mathbb{R}^{C_i\times H_{F_i}\times W_{F_i}}$ and $S_i \in \mathbb{R}^{C_i\times H_{S_i}\times W_{S_i}}$ from the two paths as input, with S_i resized to match F_i dimensions. Spatial attention is computed by averaging along the channel dimension and applying the sigmoid activation function $\sigma(\cdot)$. For channel attention, global average pooling (GAP) is applied along the spatial dimension, and linear layers, along with the sigmoid activation function $\sigma(\cdot)$, calculate channel attention. The concatenated features, after passing through consecutive convolutional layers, are multiplied separately by spatial attention and channel attention, before the results are summed. This process introduces spatial and channel attention, fostering interactive features. A residual structure combines these interactive attention features with features from each path. The resulting features F_{i+1} and S_{i+1} are obtained after passing through the encoder blocks of each path.

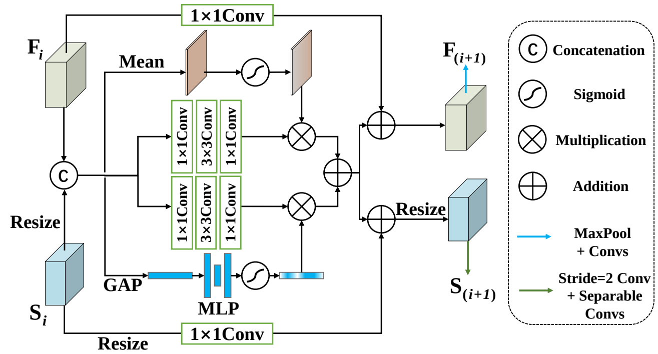

Figure 4: The structure of the Dual-path Collaboration Module (DCM).

Segmentation Path. In Figure 2, the segmentation path employs an encoder-decoder architecture with skip connections. All subsequent encoder blocks, except the initial one, receive interactive attention features from the DCM. Each encoder block is composed of two sets of sub-blocks and a max-pooling layer. The sub-blocks include a convolutional layer, ReLU activation and batch normalization. In the decoder section, features from the preceding layer are initially upsampled using bilinear interpolation. Following concatenation with the skip connection input from the encoder, the combined features enter the decoder. The ultimate layer of the decoder employs a convolutional layer to derive the probability distribution map for dense pixel-wise classification.

ASM Transformation. We acquired prior shape knowledge, denoted as $Priors$, which is loaded into the network before training. We utilised the ResNet18 [He *et al.*, 2016] as the shape classifier backbone in the classification network. During training, the classifier learns the shape category weights for input images, and the appropriate shape prior is obtained by multiplying the classifier output $Classes \in \mathbb{R}^k$ with $Priors$. Specifically, the classifier output is used to calculate the shape mean $mean \in \mathbb{R}^{1\times p\times 2}$ and t eigenvectors $\rho \in \mathbb{R}^{t\times p\times 2}$. In the shape-prior path, the shape encoder block employs a combination of spatially separable convolution and convolutional layers with a stride of 2, focusing on capturing edge and shape features. The output of the last encoder block is fed through continuous convolutional layers to obtain deformable parameters $\delta \in \mathbb{R}^{1\times t}$ and affine parameters $\alpha \in \mathbb{R}^{1\times 4}$. The shape L^0 is obtained through ASM Transformation:

$$L^0 = T_{asm}(mean, \rho, \alpha, \delta) \tag{6}$$

Shape-guided Feature Cropping. As illustrated in Figure 5, given the input shape points $L^{(t-1)} \in \mathbb{R}^{p\times 2}$ and the features $F_{(4-t)}$ from the segmentation path, where $t = 1, 2, 3$, and setting the relative length of the clipping patch l, we first convert $F_{(4-t)}$ into $f_{(4-t)} \in \mathbb{R}^{c\times w_{f_{(4-t)}}\times h_{f_{(4-t)}}}$ through convolution. For a shape point in the spatial direction of $f_{(4-t)}$, denoted as (L_x^{t-1}, L_y^{t-1}), we clip out c patches, each containing $n \times n$ sampled feature points. The feature at each sampled point is calculated on $f_{(4-t)}$ using bilinear interpolation, and we retain the bottom-left relative position coordinates of the patch, denoted as $(L_x^{t-1}, L_y^{t-1})_{lt}$. By using p shape points, we

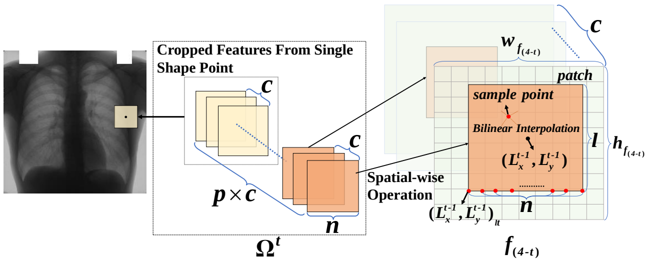

Figure 5: Shape-guided Feature Cropping

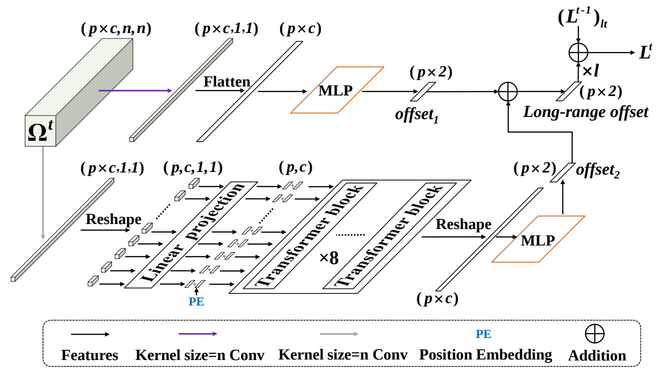

Figure 6: The structure of the Long-range Refinement Module (LRM)

obtain the clipped feature sequence $\Omega^t \in \mathbb{R}^{(p\times c)\times n\times n}$ and the bottom-left position sequence $L_{lt}^{t-1} \in \mathbb{R}^{p\times 2}$.

Long-range Refinement Module(LRM). The global correlation of shape points is crucial for accuracy. In Figure 6, we introduce the Long-range Refinement Module (LRM), which takes Ω^t, L_{lt}^{t-1}, and the patch side length l as inputs. This facilitates the adjustment of shape points within the patch. LRM consists of a parallel structure employing convolution and a multilayer perceptron. Additionally, it utilizes a transformer encoder [Zheng *et al.*, 2021] to handle the feature sequence. In the first direction, a $n \times n$ convolutional layer extracts global features from Ω^t, followed by flattening and passing through a multilayer perceptron with two linear layers and ReLU activation, resulting in $offset_1 \in \mathbb{R}^{p\times 2}$. In the second direction, another $n \times n$ convolutional layer extracts Ω^t and projects it onto token sets with positional embedding. Afterwards, an 8-layer transformer encoder with multi-head self-attention is employed to establish global feature correlations. The output, reshaped into linear features, passes through a multilayer perceptron, yielding $offset_2 \in \mathbb{R}^{p\times 2}$. The refined shape points are computed as follows:

$$L^t = L_{lt}^{(t-1)} + (offset_1 + offset_2) \times l \tag{7}$$

In our model, the coarse output of the ASM Transformation is denoted as L^0, and the outputs of the three stages of LRM are (L^1, L^2, L^3).

3.4 Hybrid Loss Function

We formulated a hybrid loss function for training SPGNet. The total loss $\mathcal{L}$ is a weighted sum of the segmentation loss $\mathcal{L}_{seg}$ derived from the segmentation path and the shape loss $\mathcal{L}_{shape}$, along with the classification loss $\mathcal{L}_c$. The segmentation loss is calculated as follows:

$$\mathcal{L}_{seg} = \lambda_1 \mathcal{L}_{ce}(\hat{Y}, Y) + \lambda_2 \mathcal{L}_{dice}(\hat{Y}, Y) \tag{8}$$

where $\hat{Y}$ is the output probability segmentation map from our segmentation path, and Y represents the ground truth for the probability segmentation mask. The losses $\mathcal{L}_{ce}$ and $\mathcal{L}_{dice}$ correspond to the cross-entropy and dice loss [Li *et al.*, 2019], respectively. The classification loss is calculated as:

$$\mathcal{L}_c = \lambda_3 \mathcal{L}_{bce}(\hat{C}, C) \tag{9}$$

where $\hat{C}$ represents the predicted shape category probabilities from classifier, and C refers to the ground truth for shape categories. $\mathcal{L}_{bce}$ corresponds to the binary cross-entropy loss [Chen *et al.*, 2023]. The shape loss calculation is as follows:

$$\mathcal{L}_{shape} = \sum_{i=4}^{7} \lambda_i \mathcal{L}_{l1}(L^{i-4}, L) \tag{10}$$

where L represents the ground truth for the shape, and $\mathcal{L}_{l1}$ refers to the L1 loss [Feng *et al.*, 2018]. The total loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = \mathcal{L}_{seg} + \mathcal{L}_c + \mathcal{L}_{shape} \tag{11}$$

4 Experiment

4.1 Dataset and Pre-Processing

In our comprehensive evaluation, we utilized two publicly available datasets: chest X-rays and breast ultrasound image data. Below, we outline the databases employed for these experiments.

JSRT Dataset. The JSRT database [Shiraishi *et al.*, 2000] consists of 247 high-resolution X-ray images, comprising 154 conventional chest X-rays with lung nodules selected from 14 medical centers and 93 chest X-rays without lung nodules. The dataset provides manually annotated masks for three anatomical structures of interest: the left lung, right lung, and heart. We separated the masks for these three anatomical structures for individual evaluations.

Breast Ultrasound Dataset (BUS). The Breast Ultrasound Dataset (BUS) [Al-Dhabyani *et al.*, 2020] includes 780 breast ultrasound images categorized into three classes: 133 normal, 437 benign, and 210 malignant images. For our experiments, we combined samples from both benign and malignant classes. The dataset offers manually annotated masks for tumor regions.

4.2 Implementation Details and Metrics

We implemented all evaluation methods on a server equipped with an NVIDIA GeForce RTX 4090 GPU. For a fair quantitative comparison, all methods for dense pixel-level classification underwent evaluation using the same 5-fold cross-validation scheme, maintaining a standardized input resolution of 256×256. To prevent overfitting, consistent data augmentation was applied in experiments, including random rotation,

Method	Left Lung Dice↑	Left Lung Jaccard↑	Left Lung HD95↓	Right Lung Dice↑	Right Lung Jaccard↑	Right Lung HD95↓	Heart Dice↑	Heart Jaccard↑	Heart HD95↓
UNet [Ronneberger *et al.*, 2015]	97.87	95.87	5.06	97.26	94.72	5.92	95.63	91.70	9.48
AttUnet [Oktay *et al.*, 2018]	97.88	95.87	4.57	97.32	94.83	5.71	95.72	91.88	8.34
UNet++ [Zhou *et al.*, 2020]	97.91	95.94	4.35	97.27	94.74	**5.58**	95.63	91.71	8.88
TransUnet [Chen *et al.*, 2021]	97.53	95.20	5.76	96.87	93.99	7.34	94.85	90.31	10.77
SAUNET [Sun *et al.*, 2020]	97.64	95.90	4.51	96.89	94.03	7.00	95.01	90.57	8.71
SPGNet$_{seg}$(ours)	**98.01**	**96.10**	**3.91**	**97.40**	**95.01**	5.64	**95.96**	**92.27**	**8.09**
HybridGNet+2IGSC [Gaggion *et al.*, 2022]	93.51	87.9	8.64	91.72	84.82	11.45	90.05	82.11	12.12
Joint+HDC [Bransby *et al.*, 2023]	95.83	92.05	6.78	94.84	90.29	8.70	93.14	87.28	**9.83**
SPGNet$_{shape}$(ours)	**97.39**	**94.93**	**4.57**	**97.28**	**94.73**	**5.62**	**94.97**	**90.33**	9.95

Table 1: Comparison with state-of-the-art methods on JSRT. Above the central horizontal line, a comparison is made for methods based on dense pixel-level classification, while below, a comparison is conducted for methods based on points regression.

Method	Breast Tumor Dice↑	Breast Tumor Jaccard↑	Breast Tumor HD95↓
UNet [Ronneberger *et al.*, 2015]	73.15	63.89	43.2
AttUnet [Oktay *et al.*, 2018]	74.60	66.05	30.25
UNet++ [Zhou *et al.*, 2020]	72.94	64.52	31.76
TransUnet [Chen *et al.*, 2021]	71.84	62.66	38.27
SAUNET [Sun *et al.*, 2020]	73.52	65.00	31.28
UNext-L [Valanarasu *et al.*, 2022]	67.03	56.73	46.11
AAUNet [Chen *et al.*, 2023]	77.68	68.94	29.10
SPGNet$_{seg}$(ours)	**78.40**	**69.70**	**26.46**

Table 2: Comparison with state-of-the-art methods based on dense pixel-level classification on BUS.

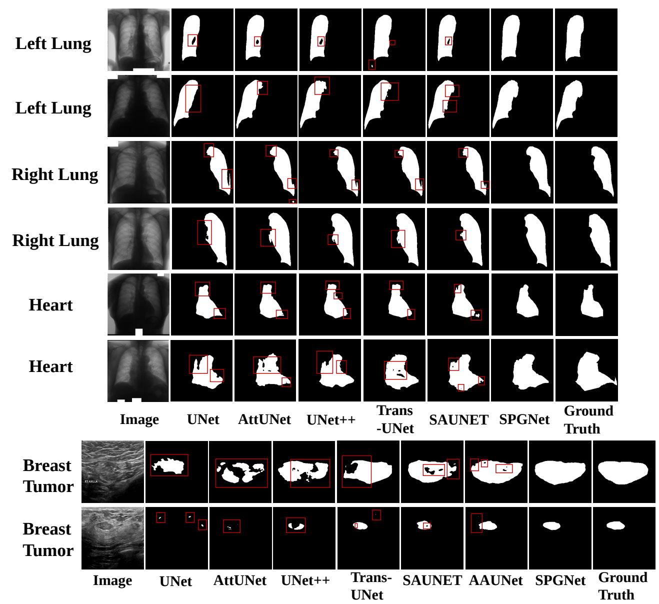

Figure 7: Qualitative experimental examples were conducted on the BUS and JSRT datasets, where errors in the comparative method are highlighted with red rectangular boxes.

random vertical flipping, and random changes in brightness and contrast. We performed 150 epochs of training on the JSRT dataset and 300 epochs on the Breast Ultrasound Dataset (BUS) while keeping the remaining training hyperparameters consistent. The batch size was set to 16, utilizing the Adam optimizer. The initial learning rate was 0.0001, with weight decay at 0.0005, and a learning rate decay of 90% every 15 epochs. During the validation phase, we assessed segmentation performance using the Dice coefficient(%) (Dice), Jaccard index(%) (Jaccard), and 95% Hausdorff Distance(mm) (HD95).

4.3 Comparison with State-of-the-art

Results on JSRT. Table 1 summarizes the experimental outcomes for the left lung, right lung, and heart components in JSRT. Our method, employing a shape point sampling of 128, surpasses other dense pixel-level classification approaches in the average Dice score. Although the improvement in average Dice may not be as pronounced compared to other methods, as depicted in Figure 7, our approach distinctly excels in edge smoothness and accuracy. Moreover, in methods based on points regression, our approach (with a shape point sampling of 64) achieves significantly higher average Dice scores in the shape path than other state-of-the-art points regression methods. Training was conducted in both scenarios, using our parameters and the optimal parameters specified in their respective papers. Compared to the state-of-the-art method [Bransby *et al.*, 2023], our method demonstrates improvements in Dice scores of 1.56%, 2.44%, and 1.83% for the left lung, right lung, and heart, respectively. These results highlight the superior performance of SPGNet on datasets exhibiting prominent shape patterns.

Results on BUS. Table 2 presents the experimental results on the BUS dataset. The segmentation path of our method(with a shape point sampling of 128) achieved an average Dice score superior to other state-of-the-art methods based on dense pixel-level classification. For SPGNet, the Dice and Jaccard scores reached 78.4% and 69.7%, respectively, This represents an improvement of 0.72% and 0.76% compared to the previous state-of-the-art method [Chen *et al.*, 2023]. As illustrated in Figure 7, our method excels in accurately identifying the location of breast tumors and maintaining overall shape and edge smoothness. The experimental results demonstrate that SPGNet can leverage shape priors to enhance performance on datasets with underlying shape patterns.

Method	$\mathcal{L}_{seg}$	$\mathcal{L}_c$	$\mathcal{L}_{shape}$	Dice of BUS	Dice of Right Lung
Baseline	-	-	-	74.04	97.49
DCM	$\lambda_1\lambda_2$	-	-	75.69	97.52
DCM+T_{asm}	$\lambda_1\lambda_2$	λ_3	λ_4	78.56	97.64
DCM+T_{asm} + (Crop+LRM)×1	$\lambda_1\lambda_2$	λ_3	$\lambda_4\lambda_5$	78.86	97.66
DCM+T_{asm} + (Crop+LRM)×2	$\lambda_1\lambda_2$	λ_3	$\lambda_4\lambda_5\lambda_6$	79.37	97.69
DCM+T_{asm} + (Crop+LRM)×3	$\lambda_1\lambda_2$	λ_3	$\lambda_4\lambda_5\lambda_6\lambda_7$	**79.63**	**97.71**

Table 3: Ablation study on the effectiveness of components in the shape-prior path, with fold 1 of the BUS dataset and fold 4 of JSRT(right lung) as the validation set.

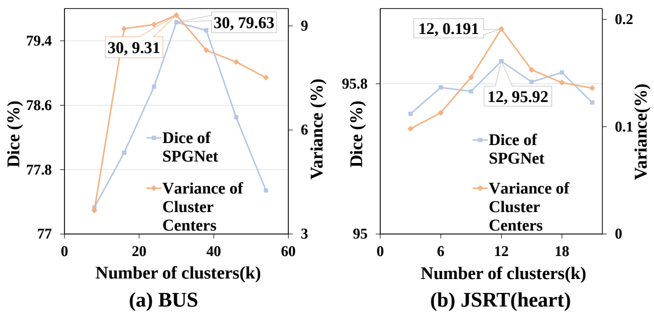

Figure 8: Ablation on the number of clusters in BUS and JSRT (Heart), with fold 1 of the BUS and JSRT as the validation set.

4.4 Ablation Study

Ablation of Each Component in Shape-prior Path. Table 3 illustrates the performance enhancement trends of the segmentation path with the introduction of each component in our dual-path collaborative network compared to the baseline. To validate each module, various coefficients in the mixed loss function were adjusted to reflect the combined effects. It was observed that three factors primarily contributed to the improvement: (i) DCM (Dual-path Collaboration Module): Coefficients λ_1 and λ_2 for $\mathcal{L}_{seg}$ were set during training to highlight the impact of DCM. Utilizing DCM, which enables collaborative learning of input image features by dual-path encoders, yielded enhancements of 1.65% for BUS and 0.03% for the right lung, emphasizing the effectiveness of our dual-path collaborative network structure. (ii) Supervision on Shape: Building upon the first factor (DCM), coefficients λ_3 and λ_4 for $\mathcal{L}_c$ and $\mathcal{L}_{shape}$ were introduced. This supervises the shape weights and guides the ASM Transformation output shape, leading to a significant 2.87% improvement for BUS and 0.12% for the right lung. This underscores the effectiveness of our offline-modeled, multi-class shape priors in guiding the segmentation network, demonstrating that deep learning segmentation networks, guided by diverse shape prior information, can significantly compensate for deficiencies in shape perception. (iii) Multi-stage Shape Refinement: Building upon the second factor (DCM+T_{asm}), coefficients λ_5, λ_6, and λ_7 were sequentially introduced for each stage of shape refinement. With the introduction of each refinement stage, there has been a relative improvement of 0.3%, 0.81% and 1.07% for BUS, and 0.02%, 0.05% and 0.07% for the right

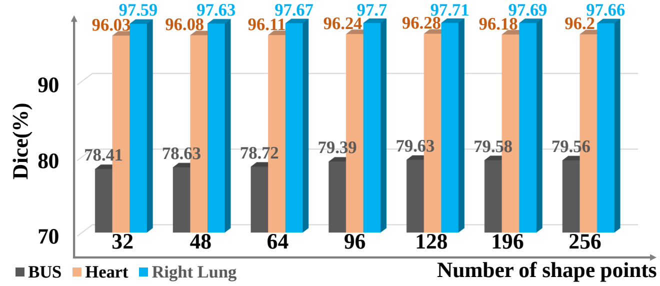

Figure 9: Ablation on the number of shape points in BUS and JSRT (Heart) was performed, with fold 1 of BUS, fold 3 of JSRT (Heart) and fold 4 of JSRT (Right Lung) as the validation sets.

lung, respectively, compared to the result of (DCM+T_{asm}). This demonstrates that our combination of multi-stage shape refinement modules (Crop+LRM) is effective. The experimental results indicate that the shape-prior path could lead to improved performance in the segmentation path.

Ablation on the Number of Clustering. We validated the impact of the number of clusters in Procrustes shapes agglomerative clustering on segmentation performance. Our criterion is to select the value that maximizes the variance of cluster distances as the offline clustering quantity for SPGNet. As shown in [Figure 8 (a)], when we chose the clustering quantity k value according to our rule (our rule selected 30), the variance of cluster centers reached 9.31%, and SPGNet achieved a dice value of 79.63%. The segmentation performance is higher compared to situations when we chose the values of the cluster of numbers (k) around the one selected according to our criterion. A similar trend can be observed in [Figure 8(b)]. This suggests that selecting the value that maximizes the variance of cluster distances can approximate an optimal quantity of shape types, avoiding situations with insufficient categories or redundancy. This serves as a suitable rule for selecting the clustering quantity to enhance segmentation performance.

Ablation on the Number of Shape Points. We assessed the impact of varying quantities of shape points on segmentation performance, as depicted in Figure 9. Seven scenarios, ranging from 32 to 256 shape points, were selected to observe their influence on segmentation performance. As the number of shape points increased, our segmentation performance gradually improved, reaching optimal performance at 128 points. It can be observed that the number of points (e.g., 96, 128, 192, 256) that effectively represent shapes leads to a significant improvement in model performance, with little variation among them. Therefore, selecting sufficient and reasonable numbers of points can better promote model performance improvement.

5 Conclusion

In this paper, we introduce a novel shape-prior guided segmentation network. The core concept is to enable the network to grasp the shape distribution within the data by learning from statistical shape model enriched with shape prior knowledge, thereby enhancing segmentation accuracy. We validate the effectiveness of the proposed network through extensive experiments on two public datasets.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (Grant No. 81873631, 81370866, 81070612), Guangzhou Science and Technology Planning Project (Grant No. 202002020047), and Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515011078), and the Royal Society (Grant Reference: IEC\NSFC\201041).

References

[Al-Dhabyani *et al.*, 2020] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

[Bohlender *et al.*, 2021] Simon Bohlender, Ilkay Oksuz, and Anirban Mukhopadhyay. A survey on shape-constraint deep learning for medical image segmentation. *IEEE Reviews in Biomedical Engineering*, 2021.

[Bransby *et al.*, 2023] Kit Mills Bransby, Greg Slabaugh, Christos Bourantas, and Qianni Zhang. Joint dense-point representation for contour-aware graph segmentation. *arXiv preprint arXiv:2306.12155*, 2023.

[Bruse *et al.*, 2017] Jan L Bruse, Maria A Zuluaga, Abbas Khushnood, Kristin McLeod, Hopewell N Ntsinjana, Tain-Yen Hsia, Maxime Sermesant, Xavier Pennec, Andrew M Taylor, and Silvia Schievano. Detecting clinically meaningful shape clusters in medical image data: metrics analysis for hierarchical clustering applied to healthy and pathological aortic arches. *IEEE Transactions on Biomedical Engineering*, 64(10):2373–2383, 2017.

[Chen *et al.*, 2002] Yunmei Chen, Hemant D Tagare, Sheshadri Thiruvenkadam, Feng Huang, David Wilson, Kaundinya S Gopinath, Richard W Briggs, and Edward A Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50:315–328, 2002.

[Chen *et al.*, 2021] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[Chen *et al.*, 2023] Gongping Chen, Lei Li, Yu Dai, Jianxun Zhang, and Moi Hoon Yap. Aau-net: An adaptive attention u-net for breast lesions segmentation in ultrasound images. *IEEE Transactions on Medical Imaging*, 42(5):1289–1300, 2023.

[Cootes *et al.*, 1995] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[Debavelaere *et al.*, 2020] Vianney Debavelaere, Stanley Durrleman, Stéphanie Allassonnière, and Alzheimer's Disease Neuroimaging Initiative. Learning the clustering of longitudinal shape data sets into a mixture of independent or branching trajectories. *International Journal of Computer Vision*, 128:2794–2809, 2020.

[Feng *et al.*, 2018] Zhen-Hua Feng, Josef Kittler, Muhammad Awais, Patrik Huber, and Xiao-Jun Wu. Wing loss for robust facial landmark localisation with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2235–2245, 2018.

[Gaggion *et al.*, 2022] Nicolás Gaggion, Lucas Mansilla, Candelaria Mosquera, Diego H Milone, and Enzo Ferrante. Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis. *IEEE Transactions on Medical Imaging*, 42(2):546–556, 2022.

[Gao *et al.*, 2023] Jianxiong Gao, Xuelin Qian, Yikai Wang, Tianjun Xiao, Tong He, Zheng Zhang, and Yanwei Fu. Coarse-to-fine amodal segmentation with shape prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1262–1271, 2023.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Jiang and Veeraraghavan, 2022] Jue Jiang and Harini Veeraraghavan. One shot pacs: Patient specific anatomic context and shape prior aware recurrent registration-segmentation of longitudinal thoracic cone beam cts. *IEEE Transactions on Medical Imaging*, 41(8):2021–2032, 2022.

[Kuo *et al.*, 2019] Weicheng Kuo, Anelia Angelova, Jitendra Malik, and Tsung-Yi Lin. Shapemask: Learning to segment novel objects by refining shape priors. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9207–9216, 2019.

[Lee *et al.*, 2019] Matthew Chung Hai Lee, Kersten Petersen, Nick Pawlowski, Ben Glocker, and Michiel Schaap. Tetris: Template transformer networks for image segmentation with shape priors. *IEEE transactions on medical imaging*, 38(11):2596–2606, 2019.

[Li *et al.*, 2017] Yuanwei Li, Chin Pang Ho, Matthieu Toulemonde, Navtej Chahal, Roxy Senior, and Meng-Xing Tang. Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model. *IEEE transactions on medical imaging*, 37(5):1081–1091, 2017.

[Li *et al.*, 2019] Cheng Li, Hui Sun, Zaiyi Liu, Meiyun Wang, Hairong Zheng, and Shanshan Wang. Learning cross-modal deep representations for multi-modal mr image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 57–65. Springer, 2019.

[Li *et al.*, 2022] Zhixuan Li, Weining Ye, Tingting Jiang, and Tiejun Huang. 2d amodal instance segmentation guided by 3d shape prior. In *European Conference on Computer Vision*, pages 165–181. Springer, 2022.

[Li *et al.*, 2023] Zhixuan Li, Weining Ye, Tingting Jiang, and Tiejun Huang. Gin: Generative invariant shape prior for amodal instance segmentation. *IEEE Transactions on Multimedia*, 2023.

[Medley *et al.*, 2019] Daniela O Medley, Carlos Santiago, and Jacinto C Nascimento. Deep active shape model for robust object fitting. *IEEE Transactions on Image Processing*, 29:2380–2394, 2019.

[Milletari *et al.*, 2017] Fausto Milletari, Alex Rothberg, Jimmy Jia, and Michal Sofka. Integrating statistical prior knowledge into convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 161–168. Springer, 2017.

[Nosrati *et al.*, 2016] Nosrati, Masoud S, Hamarneh, and Ghassan. Incorporating prior knowledge in medical image segmentation: a survey. *arXiv preprint arXiv:1607.01092*, 2016.

[Oktay *et al.*, 2018] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[Schoenemann *et al.*, 2007] Schoenemann, Thomas, Cremers, and Daniel. Globally optimal image segmentation with an elastic shape prior. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–6. IEEE, 2007.

[Shen *et al.*, 2013] Wei Shen, Yan Wang, Xiang Bai, Hongyuan Wang, and Longin Jan Latecki. Shape clustering: Common structure discovery. *Pattern Recognition*, 46(2):539–550, 2013.

[Shiraishi *et al.*, 2000] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000.

[Song *et al.*, 2020] Youyi Song, Lei Zhu, Baiying Lei, Bin Sheng, Qi Dou, Jing Qin, and Kup-Sze Choi. Shape mask generator: Learning to refine shape priors for segmenting overlapping cervical cytoplasms. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pages 639–649. Springer, 2020.

[Srivastava *et al.*, 2005] Anuj Srivastava, Shantanu H Joshi, Washington Mio, and Xiuwen Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on pattern analysis and machine intelligence*, 27(4):590–602, 2005.

[Sun *et al.*, 2020] Jesse Sun, Fatemeh Darbehani, Mark Zaidi, and Bo Wang. Saunet: Shape attentive u-net for interpretable medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pages 797–806. Springer, 2020.

[Valanarasu *et al.*, 2022] Valanarasu, Jeya Maria Jose, Patel, and Vishal M. Unext: Mlp-based rapid medical image segmentation network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 23–33. Springer, 2022.

[Xing *et al.*, 2015] Fuyong Xing, Yuanpu Xie, and Lin Yang. An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging*, 35(2):550–566, 2015.

[Zhao *et al.*, 2021] Chen Zhao, Yan Xu, Zhuo He, Jinshan Tang, Yijun Zhang, Jungang Han, Yuxin Shi, and Weihua Zhou. Lung segmentation and automatic detection of covid-19 using radiomic features from chest ct images. *Pattern Recognition*, 119:108071, 2021.

[Zheng *et al.*, 2021] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890, 2021.

[Zhou *et al.*, 2020] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2020.