ModelMix: A New Model-Mixup Strategy to Minimize Vicinal Risk across Tasks for Few-scribble based Cardiac Segmentation

Ke Zhang and Vishal M. Patel

Department of Electrical and Computer Engineering, The Johns Hopkins University
kzhang99@jhu.edu

Abstract. Pixel-level dense labeling is both resource-intensive and time-consuming, whereas weak labels such as scribble present a more feasible alternative to full annotations. However, training segmentation networks with weak supervision from scribbles remains challenging. Inspired by the fact that different segmentation tasks can be correlated with each other, we introduce a new approach to few-scribble supervised segmentation based on model parameter interpolation, termed as *ModelMix*. Leveraging the prior knowledge that linearly interpolating convolution kernels and bias terms should result in linear interpolations of the corresponding feature vectors, ModelMix constructs virtual models using convex combinations of convolutional parameters from separate encoders. We then regularize the model set to minimize vicinal risk across tasks in both unsupervised and scribble-supervised way. Validated on three open datasets, *i.e.*, ACDC, MSCMRseg, and MyoPS, our few-scribble guided ModelMix significantly surpasses the performance of the state-of-the-art scribble supervised methods. Our code is available at https://github.com/BWGZK/ModelMix.

Keywords: Weakly supervised learning · Scribble annotation · Segmentation · Mixup

1 Introduction

Creating large-scale fully annotated medical image datasets is both time-consuming and burdensome. To address this bottleneck, researchers have investigated weak annotations [17], such as image-level labels, sparse labels, and noisy labels. Among these alternatives, scribble is a particularly attractive choice due to its advantages in annotating complex structures [3]. We propose to explore few-shot scribble supervised segmentation, which further reduces the annotation effort by exploiting several scribble annotated images and a large amount of unlabeled images.

Few-shot scribble supervised segmentation is particularly challenging due to the scarcity of annotations. Existing methods mainly exploit the labeled pixels [1,3,9,14,15] and regularize the model training with priors [19,23,24]. These

methods are susceptible to the supervision amount and might easily over-fit to annotated samples.

Several studies have been conducted to explore augmentation techniques based on mixup, including Mixup [21], Cutout [5], Cutmix [20], PuzzleMix [10] and Comixup [11]. These works interpolate images to generate unseen samples. However, it is difficult to directly interpolate labels for different datasets due to disparity in label categories. Meanwhile, considering the domain gap, directly blending images from distinct segmentation tasks may lead to unrealistic results.

To address the above challenges, we propose to mix model parameters of complementary tasks. Our method is based on two assumptions: (1) we assume that different segmentation tasks are intrinsically related, such as myocardial pathology and cardiac structure segmentation. (2) we suppose the individual models for each task are distributed in the vicinity of the general model. One direct solution is to train a general model for correlated tasks. However, the model is faced with the trade-off between learning domain-invariant knowledge and domain-specific information, especially when the model size is not large enough. To tackle this, we propose to train individual models for each task, while regularize them with vicinal principle. Specifically, we construct virtual models from the individual models, and then apply vicinal regularization to the virtual models.

We propose to mix model parameters learned from correlated tasks, termed as ModelMix, for few-shot scribble-guided cardiac segmentation. Firstly, we train individual models for each task and employ techniques such as data augmentation and dropout to enhance the network's robustness. Secondly, we conduct linear interpolation between randomly selected convolution parameters from the encoders of separate models to construct virtual networks. Thirdly, we apply vicinal constraints between the hybrid and non-hybrid versions of the model to encourage the mixed networks to generalize across all individual tasks.

The contributions of this work are two-fold. (1) We propose a holistic ModelMix strategy to learn from complementary segmentation tasks. Specifically, we construct mixed models and regularize them to have consistent performance with each individual model. (2) Evaluated on three public datasets of ACDC, MSCMRseg, and MyPS, our ModelMix demonstrated its advantage over existing scribble supervised segmentation approaches.

2 Method

As shown in Figure 1, our ModelMix linearly interpolates the parameters of encoder convolutional layer for different tasks, and requires the segmentation of mixed model and individual model to be consistent. The proposed ModelMix is composed of three stages: (1) The individual models are trained end-to-end for different tasks. (2) For the encoder trained for a separate task, we randomly select the convolutional layers and mix the parameters with linear interpolation while keeping the parameters of other layers unchanged. (3) Both supervised and

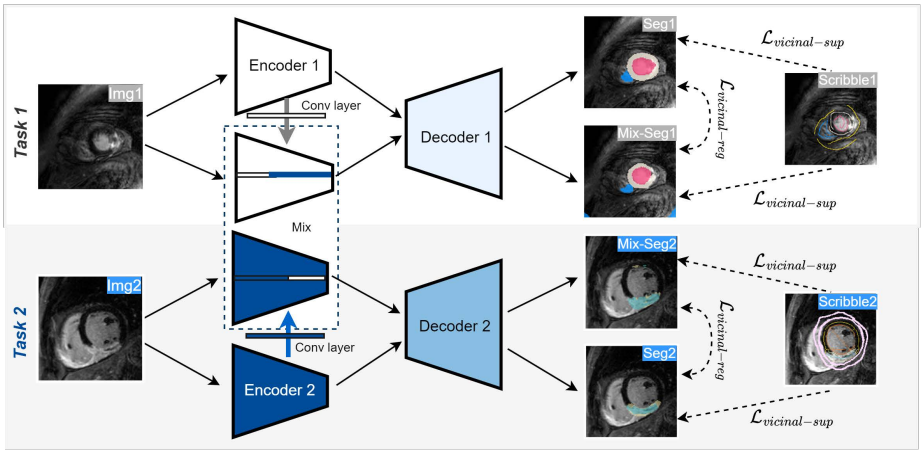

Fig. 1: Overview of the proposed ModelMix framework for cardiac segmentation from scribble supervision. ModelMix performs linear interpolation of the encoder convolutional parameters across different tasks, ensuring that the segmentation of the mixed model aligns with that of the individual models.

unsupervised vicinal regularization are applied to require consistent segmentation results between hybrid and non-hybrid models.

2.1 Train for separate tasks

For n segmentation tasks $\{t_i\}_{i=1}^{n}$, we adopt the same encoder architecture and separate segmentation decoder, denoted as $\{e_1, e_2, \cdots, e_n\}$ and $\{h_1, h_2, \cdots, h_n\}$, respectively. The encoders have the same architecture but do not share model weights. To improve the robustness of neural network, we introduce both image- and model-level perturbation to the training process.

For perturbations at the image level, we conduct enhancement operations both within individual images and between different images. For each individual image, we randomly cutout a square area of pixels. Let x denote the input image and $\mathbb{1}_c$ be the binary cutout mask, the obtained result is represented as $x^c = \mathbb{1}_c \odot x$. For perturbation across images, we perform linear interpolation between images from the same dataset. Let α be the mix ratio sampled from beta distribution, we define the perturbed image as $x' = \alpha x_1^c + (1-\alpha) x_2^c$. Inspired by [22], we require the mix-equivalence of image segmentation with invariant loss $\mathcal{L}_{\text{inv}}$ by maximizing the cosine similarity (cos):

$$\mathcal{L}_{\text{inv}} = -\cos\left(f(x'), \alpha f(\mathbb{1}_{c_1} \odot x_1) + (1-\alpha) f(\mathbb{1}_{c_2} \odot x_2)\right). \tag{1}$$

For model-level perturbation, we simply apply dropout to the encoder of each separate model. The design serves to mitigate over-fitting and improve the robustness of neural networks.

2.2 Minimize vicinal risk across tasks

For segmentation tasks, our objective is to minimize the mean of the loss function l across the distribution of models P, referred to as the expected risk. Similarly, the empirical vicinal risk indicates the average loss function across constructed vicinal virtual models. Empirically, we approximate the model distribution using the model set $\{f_i\}_{i=1}^n$:

$$R_\delta(f) = \int l(f_i(x), y) dP_\delta(f) = \frac{1}{n}\sum_{i=1}^{n} l(f_i(x), y), \tag{2}$$

where $\delta(f_i)$ is the Dirac mass centered at f_i. According to the vicinal risk minimization principle [4], the distribution of function f is extended to:

$$P_v(\tilde{f}) = \frac{1}{n}\sum_{i,j} v(\tilde{f}_{ij}|f_i, f_j). \tag{3}$$

We propose a generic vicinal distribution v by *ModelMix*, which assesses the likelihood of finding the virtual model $\tilde{f}$ near the existing models f_i and f_j. For complementary tasks pair set Ω, we sample the task pair $(t_i, t_j) \in \Omega$. We assume different tasks share the general encoder and separate segmentation decoder, *i.e.*, $f_i(x) = h_i(\bar{e}(x))$ Then, we construct the general vicinal distribution as follows:

$$\mu(\tilde{f}|f_i, f_j) = \frac{1}{n}\sum_{i,j} E[\delta(\tilde{f}_{ij} = h(\tilde{e})|f_i, f_j)], \tag{4}$$

where $\tilde{e}$ denotes the constructed virtual encoder. We assume e_i is distributed in the vicinity of the general backbone $\bar{e}$ for each separate task t_i. Then, we simply extend the vicinal distribution of $\tilde{e}$ by incorporating the prior knowledge that linear interpolation of the convolution parameters results in linear interpolation of the relevant features. Taking the convolution operation as $g(\boldsymbol{x}, \boldsymbol{k}, b)$ with kernel $\boldsymbol{k}$ and bias b, the constructed convolution $\tilde{g}(\boldsymbol{x}, \boldsymbol{k}, b)$ is defined as:

$$\begin{aligned} \tilde{g}_{ij}(\boldsymbol{x}, \boldsymbol{k}, b) &= \lambda g_i(\boldsymbol{x}, \boldsymbol{k}, b) + (1-\lambda) g_j(\boldsymbol{x}, \boldsymbol{k}, b) \\ &= \boldsymbol{x} * [\lambda \boldsymbol{k}_i + (1-\lambda)\boldsymbol{k}_j] + \lambda\,[b_i + (1-\lambda) b_j]\,, \end{aligned} \tag{5}$$

where λ is the mixed ratio sampled from beta distribution. For each encoder, we randomly select a convolutional layer from the encoder to perform interpolation between tasks and keep the parameters of other layers unchanged.

2.3 Vicinal regularization

We apply vicinal regularization in both unsupervised and scribble-supervised manner to enable the constructed virtual model have consistent performance with individual models on each single task. Firstly, we regularize the output of constructed model $\tilde{f}_{ij}$ to be consistent with f_i:

$$\mathcal{L}_{\text{vicinal-reg}} = \sum_{i,j} \mathcal{L}_{\cos}(\tilde{f}_{ij}(x), f_i(x)) = -\frac{\tilde{f}_{ij}(x) \cdot f_i(x)}{\|\tilde{f}_{ij}(x)\|_2 \cdot \|f_i(x)\|_2}, \tag{6}$$

which minimizes the cosine similarity between $\tilde{f}_{ij}(x)$ and $f_i(x)$. Secondly, we leverage the supervision of scribble annotations, and calculate the supervised loss for $\tilde{f}_{ij}$, f_i, respectively:

$$\mathcal{L}_{\text{vicinal-sup}} = \sum_{i,j} \left[\mathcal{L}_{\text{sup}}(y, \tilde{f}_{ij}(x)) + \mathcal{L}_{\text{sup}}(y, f_i(x)) \right] \tag{7}$$

$$\mathcal{L}_{\text{sup}}(y, f(x)) = -\left[y \log(f(x)) + \frac{2yf(x)}{y + f(x)} \right], \tag{8}$$

where $\mathcal{L}_{\text{sup}}$ is the combination of cross entropy and Dice loss calculated for annotated pixels, and y denotes the scribble annotations. Then, our overall training objective is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{inv}} + \mathcal{L}_{\text{vicinal-reg}} + \mathcal{L}_{\text{vicinal-sup}}. \tag{9}$$

3 Experiment

3.1 Experiment Setup

Datasets: ACDC[1] includes cardiac MR images obtained from 100 patients, with the annotations of the right ventricle (RV), left ventricle (LV), and myocardium (MYO). Following [23], we partitioned the 100 subjects into three groups: a training set comprising 70 subjects, a validation set with 15 subjects, and a test set of 15 subjects. For weak-supervision studies, we utilized expert manually annotated scribble annotations released by [19]. **MSCMRseg**[2] consists of late gadolinium enhancement (LGE) cardiac MR images from 45 patients diagnosed with cardiomyopathy. We adopt scribble annotations from [22]. Following [6], we divide the dataset randomly into 25 training images, 5 validation images, and 15 test images for evaluation. **MyoPS**[3] comprises 45 paired multi-sequence CMR images encompassing Balanced Steady-State Free Precession (BSSFP), Late Gadolinium Enhancement (LGE), and T2-weighted CMR sequences. Compared to the structural segmentation task, MyoPS is more challenging due to the diverse representation of pathology across different patients, posing difficulties in portraying scar and edema characteristics. We leverage the scribbles released by [24] and divided the dataset into 20 pairs for training, 5 for validation, and 20 for testing following [13]. *For all datasets, we train the models using five randomly selected scribble-annotated volumes in conjunction with other unlabeled images from the training set.*

Implementation: We utilized the UNet [16] architecture as the foundational structure for our segmentation network. To introduce perturbations and enhance training, a dropout layer (with a ratio of 0.5) was incorporated before each convolutional block. We apply random rotations, flips, and crop to augment the

[1] https://www.creatis.insa-lyon.fr/Challenge/acdc/databasesTraining.html
[2] http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mscmrseg19/data.html
[3] https://zmiclab.github.io/zxh/0/myops20/

Table 1: Component ablations: 5-scribble based segmentation results in Dice scores of the ablation study. **Bold** denotes the best result, underline indicates the best but one.

Methods	$\mathcal{L}_{\text{pce}}$	$\mathcal{L}_{\text{inv}}$	$\mathcal{L}_{\text{vicinal}}$ sup	$\mathcal{L}_{\text{vicinal}}$ reg	Share Enc	MyoPS(5 scribble) Scar	MyoPS(5 scribble) Edema	MyoPS(5 scribble) Avg	MSCMRseg(5 scribble) LV	MSCMRseg(5 scribble) MYO	MSCMRseg(5 scribble) RV	MSCMRseg(5 scribble) Avg
#1	✓	×	×	×	×	.310±.257	.153±.090	.231±.199	.373±.346	.337±.254	.139±.143	.283±.264
#2	✓	✓	×	×	×	.348±.189	.531±.106	.440±.177	.700±.234	.416±.190	.501±.460	.539±.319
#3	✓	✓	✓	×	×	.480±.074	.625±.197	.552±.160	.823±.104	.717±.117	.508±.419	.683±.275
#4	✓	✓	✓	✓	×	**.518±.081**	**.630±.128**	**.574±.117**	**.922±.050**	**.799±.095**	**.696±.267**	**.805±.181**
#5	✓	✓	-	-	✓	.242±.166	.386±.150	.314±.168	.581±.094	.462±.101	.346±.265	.463±.188
						MyoPS(5 scribble)			ACDC(5 scribble)			
#1	✓	×	×	×	×	.310±.257	.153±.090	.231±.199	.585±.249	.466±.147	.279±.176	.443±.231
#2	✓	✓	×	×	×	.348±.189	.531±.106	.440±.177	.580±.240	.544±.183	.486±.294	.537±.244
#3	✓	✓	✓	×	×	.449±.059	.644±.113	.546±.134	.803±.161	.686±.124	.462±.303	.650±.253
#4	✓	✓	✓	✓	×	**.543±.170**	**.588±.194**	**.566±.173**	**.829±.145**	**.743±.127**	**.728±.131**	**.767±.140**
#5	✓	✓	-	-	✓	.179±.173	.258±.227	.219±.195	.783±.142	.473±.129	.143±.133	.466±.295

training dataset, with resulting images resized to 256 × 256 for network input. The intensity of each slice is scaled to a range of 0-1. All models are trained with batch size of 16 and a learning rate of $1e^{-3}$. We use evaluation metrics of Dice scores and Haussdorff Distance (HD) to assess segmentation accuracy. Implementation was conducted over 1000 epochs on eight NVIDIA 3090Ti 24GB GPUs.

3.2 Ablation study

We verified the effectiveness of ModelMix components on two task combinations, *i.e*, MyoPS & MSCMRseg and MyoPS & ACDC. All models are trained using five scribbles and then evaluated on the validation set. *Five* ablated models are implemented, comprising the partial cross-entropy loss ($\mathcal{L}_{\text{pce}}$) computed from annotated pixels, Mix invariant loss ($\mathcal{L}_{\text{inv}}$), and the vicinal loss ($\mathcal{L}_{\text{vicinal-sup}}$, $\mathcal{L}_{\text{vicinal-reg}}$). We also compared the results to the baseline trained with shared encoder (Share Enc). Details are summarized in Table 1.

Components of ModelMix: When integrating Mix-based augmentations, there is a significant improvement in the performance of model #2 compared to model #1, with an average Dice increase of 20.9%, 25.6%, and 9.4% on MyoPS, MSCMRseg, and ACDC, respectively. When combined with Supervised vicinal loss ($\mathcal{L}_{\text{vicinal-sup}}$), Model #3 obtained remarkable performance gain on all datasets, demonstrating that learning from complementary tasks can promote the performance of each individual task. When leveraging unlabeled pixels with unsupervised vicinal loss ($\mathcal{L}_{\text{vicinal-reg}}$) , the average Dice Scores of model #4 on MyoPS, MSCMR, and ACDC are further boosted to 57.4%, 80.5%, and 76.7% , respectively.

Combination of Tasks: We combine MyoPS with MSCMRseg and ACDC

Table 2: The 5-scribble supervised segmentation results on MSCMRseg dataset.

Methods	Dice				HD(mm)			
	LV	MYO	RV	Avg	LV	MYO	RV	Avg
PCE	.454±.271	.356±.182	.102±.072	.304±.241	111.87±14.40	109.72±20.30	123.62±19.33	115.07±16.02
Mixup	.440±.102	.310±.127	.021±.013	.257±.200	259.42±14.18	210.00±12.37	251.98±15.6	240.47±25.96
Cutout	.315±.103	.307±.153	.166±.110	.263±.139	259.42±14.18	240.06±16.38	252.18±15.42	250.56±17.04
CycleMix	.517±.086	.421±.108	.007±.007	.315±.237	213.20±35.65	151.36±55.12	260.56±12.66	208.37±58.88
ShapePU	.758±.191	.567±.168	.059±.026	.461±.331	209.04±16.09	234.08±18.15	237.86±14.13	226.99±20.45
WSL4	.809±.079	.653±.109	.599±.261	.687±.191	140.95±69.06	147.74±59.93	**95.07±60.53**	127.92±67.49
w/ MyoPS	**.875±.077**	**.754±.079**	**.722±.201**	**.784±.145**	**78.05±16.11**	**69.85±30.45**	99.20±46.81	**82.36±35.09**
FullSup-UNet	.775±.158	.604±.147	.572±.207	.651±.191	23.50±21.79	34.03±19.25	81.29±11.29	46.27±30.91
FullSup-nnUNet	.885±.085	.757±.147	.757±.201	.799±.160	21.48±29.68	13.50±12.99	18.27±12.51	17.75±19.87

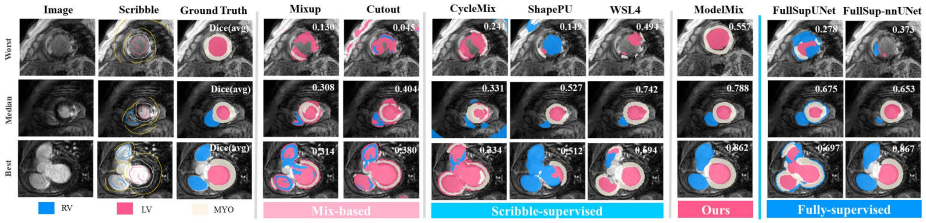

Fig. 2: The qualitative comparison on MSCMRseg dataset. The three images are the worst, median and best cases selected by the average Dice Score.

respectively and compared their segmentation performance. Note that images from MSCMRseg and MyoPS both contain enhanced pathological information, while ACDC only includes structural information. Therefore, the complementarity between MyoPS and MSCMR is greater than the combination of MyoPS and ACDC. As expected, the performance improvement of MyoPS with MSCMR is slightly better than that of MyoPS with ACDC, improving by 13.4% and 12.6% over separate trained models, respectively.

Comparison with shared encoder: In Table 1, we further compare our method with the shared encoder baseline of Model #5. Employing a shared encoder (Share Enc) for tasks, Model #5 exhibits inferior performance compared to Model #2, which is trained using separate encoders. This disparity arises because when training supervision is limited, employing a single encoder to generalize across various tasks poses greater challenges to model training. Leveraging separate encoders and constructed virtual models in the vicinity between individual models, our approach (Model #4) notably surpasses both Model #2 and Model #5 by large margins.

3.3 Performance and Comparisons

We compare our method against three sets of *10* baselines. The first set includes six scribble-guided models, *i.e.*, partial cross-entropy (PCE) [18], Mixup [21],

Table 3: The 5-scribble guided segmentation results on MyoPS dataset.

Methods	Ratio	ACDC	MSCMR	Dice			HD(mm)		
				Scar	Edema	Avg	Scar	Edema	Avg
PCE	×	×	×	.242±.170	.122±.077	.182±.144	76.22±37.24	124.89±21.27	100.55±38.77
CVIR [7]	✓	×	×	.288±.191	.085±.034	.186±.170	45.01±18.44	125.27±20.83	85.14±45.04
nnPU [12]	✓	×	×	.290±.166	.236±.078	.263±.131	126.51±35.27	125.05±20.69	125.78±28.55
ours	×	✓	×	.455±.251	.518±.140	.487±.203	77.52±34.68	81.85±28.37	79.69±31.35
	×	×	✓	**.488±.263**	**.575±.147**	**.532±.215**	70.18±34.19	75.57±28.13	72.87±31.02
	×	✓	✓	.474±.269	.545±.158	.509±.221	**41.26±20.38**	**46.63±18.71**	**43.95±19.50**
FullSup-UNet	-	×	×	.423±.253	.445±.149	.434±.205	117.61±35.08	119.13±22.7	118.37±29.17
FullSup-nnUNet	-	×	×	.496±.252	.563±.141	.529±.204	43.86±37.27	45.14±33.86	44.50±35.15

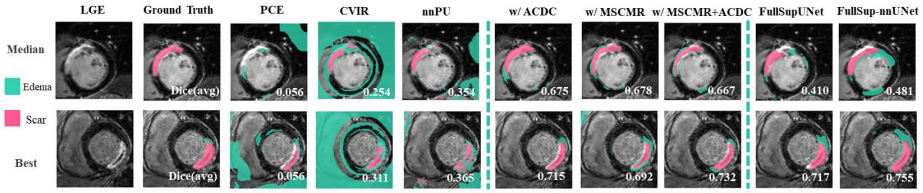

Fig. 3: The qualitative comparison on MyoPS dataset. The two images are the median and best cases selected by the average Dice Score.

Cutout [5], CycleMix [22], ShapePU [23], and WSL4 [15]. The second set comprises two semi-supervised methods of positive and unlabeled learning, including CVIR [7] and nnPU [12]. These methods are provided with the additional information of label class ratio, and then adapted for scribble supervised segmentation. Finally, we include results from fully supervised UNet [2] (FullSup-UNet) and fully supervised nnUNet [8] (FullSup-nnUNet) for reference.

Comparison with scribble-supervised methods: Table 3 presents the quantitative comparisons on the MSCMR dataset. When mixing with the MyoPS model, our ModelMix(w/ MyoPS) significantly outperforms all other scribble-supervised methods by an average of 9.7% in Dice Score. This is affirmed by the qualitative results of Figure 2, which visualizes the results of the worst, median, and best cases by the average Dice Scores of all compared methods.

Comparison with semi-supervised methods: Table 3 summarizes the results on the MyoPS dataset. Considering that traditional pseudo-label based methods (*i.e.*, WSL4) are used to segment regular structures and fail to converge on this challenging task, we compare our methods against the adapted semi-supervised methodologies such as CVIR [7] and nnPU [12], with additional input of category mix ratios. Mixing with model parameters of ACDC and MSCMRseg separately, our ModelMix obtains remarkable performance gain, exceeding the second-best comparison methods by 22.4% and 26.9% in Dice, respectively. When combined with both ACDC and MSCMRseg, ModelMix learns robust anatomical priors and demonstrates the competitive performance on both Dice and HD. Figure 3 visualizes the median and best cases selected by average

Dice of all compared methods. One can observe that our ModelMix achieves more realistic segmentation results than other methods when mixed with complementary tasks (w/ ACDC, MSCMR, ACDC+MSCMR).

4 Conclusion

We introduce ModelMix, a simple and model-agnostic approach to blend model parameters of complementary tasks. Our method minimize the vicinal risk of virtual models, which are constructed through linear interpolation of encoder convolutional parameters and their corresponding features. Extensive evaluations on three public datasets demonstrate that ModelMix achieves state-of-the-art performance in the challenging task of few-shot scribble-supervised segmentation.

Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P.M., Rueckert, D.: Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 586–594. Springer (2018)
2. Baumgartner, C.F., Koch, L.M., Pollefeys, M., Konukoglu, E.: An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 111–119. Springer (2017)
3. Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner, C.F.: Learning to segment medical images with scribble-supervision alone. In: DLMIA/ML-CDS@MICCAI (2018)
4. Chapelle, O., Weston, J., Bottou, L., Vapnik, V.: Vicinal risk minimization. Advances in neural information processing systems **13** (2000)
5. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
6. Gao, S., Zhou, H., Gao, Y., Zhuang, X.: Bayeseg: Bayesian modeling for medical image segmentation with interpretable generalizability. arXiv preprint arXiv:2303.01710 (2023)
7. Garg, S., Wu, Y., Smola, A.J., Balakrishnan, S., Lipton, Z.: Mixture proportion estimation and pu learning: A modern approach. Advances in Neural Information Processing Systems **34** (2021)
8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
9. Ji, Z., Shen, Y., Ma, C., Gao, M.: Scribble-based hierarchical weakly supervised learning for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 175–183. Springer (2019)

10. Kim, J.H., Choo, W., Song, H.O.: Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In: International Conference on Machine Learning (ICML) (2020)
11. Kim, J., Choo, W., Jeong, H., Song, H.O.: Co-mixup: Saliency guided joint mixup with supermodular diversity. In: International Conference on Learning Representations (2021)
12. Kiryo, R., Niu, G., du Plessis, M.C., Sugiyama, M.: Positive-unlabeled learning with non-negative risk estimator. vol. 30 (2017)
13. Li, L., Wu, F., Wang, S., Luo, X., Martín-Isla, C., Zhai, S., Zhang, J., Liu, Y., Zhang, Z., Ankenbrand, M.J., et al.: Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. Medical Image Analysis **87**, 102808 (2023)
14. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3159–3167 (2016)
15. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528–538. Springer (2022)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
17. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis **63**, 101693 (2020)
18. Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C.: Normalized cut loss for weakly-supervised cnn segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1818–1827 (2018)
19. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Transactions on Medical Imaging **40**(8), 1990–2001 (2021)
20. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: International Conference on Computer Vision (ICCV) (2019)
21. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. International Conference on Learning Representations (2018), https://openreview.net/forum?id=r1Ddp1-Rb
22. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11656–11665 (2022)
23. Zhang, K., Zhuang, X.: Shapepu: A new pu learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 162–172. Springer (2022)
24. Zhang, K., Zhuang, X.: Zscribbleseg: Zen and the art of scribble supervised medical image segmentation. arXiv preprint arXiv:2301.04882 (2023)