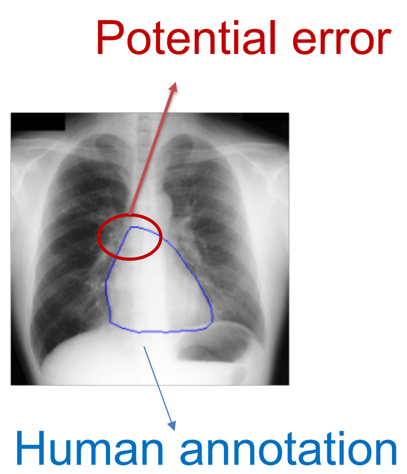
Superpixel-guided Iterative Learning from Noisy Labels for Medical Image Segmentation

Shuailin Li*, Zhitong Gao*, and Xuming He
{lishl, gaozht, hexm}@shanghaitech.edu.cn

Problem Setup

Learning Segmentation from Noisy Labels

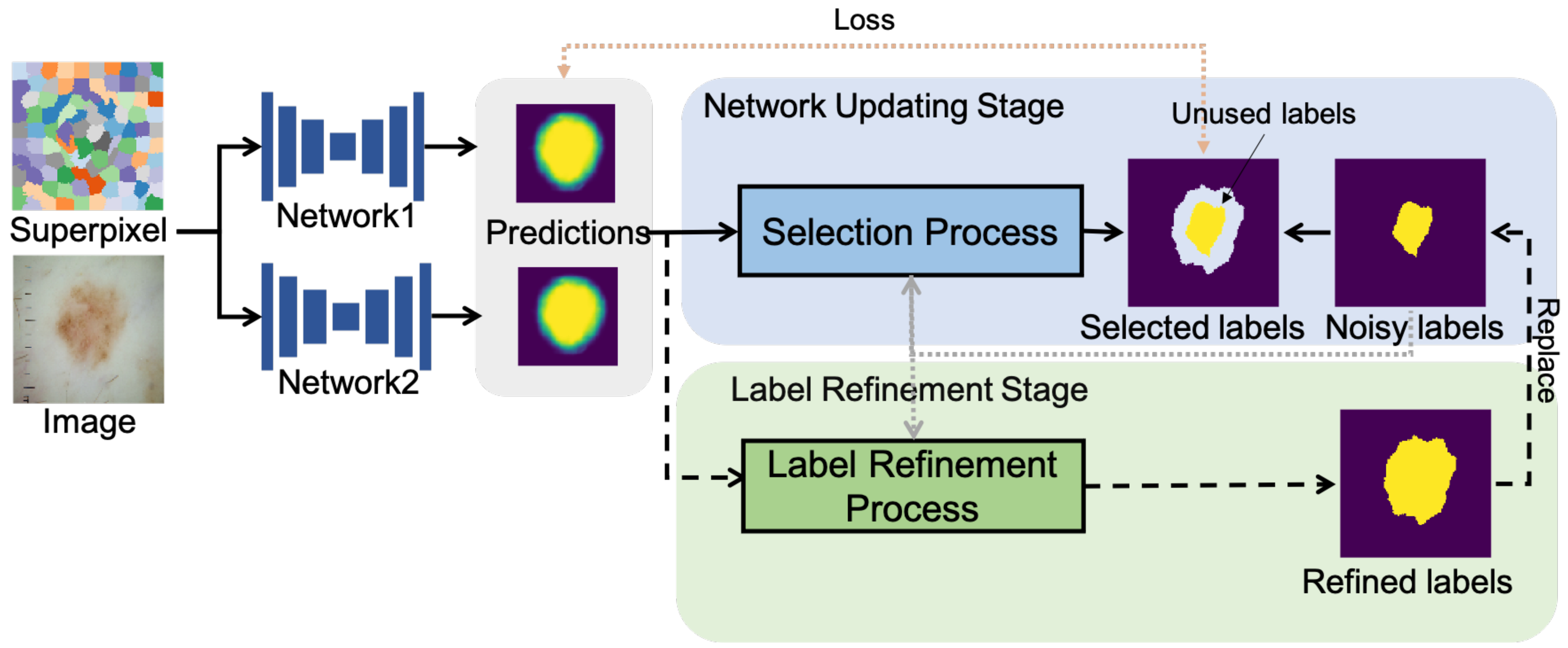
- Segmentation annotations often include a varying amount of noise.
- We aim to robustly learn a semantic segmentation network.



Main Ideas

- Exploit structural constraint in segmentation labels
  - adopt a superpixel representation.
- Develop an iterative learning scheme that combines
  - noise-aware training of segmentation network
  - noisy label refinement

Methodology



Overview: we use **superpixels** as our guidance in an **iterative learning** process which jointly updates network parameters and refines noisy labels.

Superpixel Representation

- Assume the pixels share similar labels in each superpixel.
- Treat each superpixel as a data sample during training.

Iterative Model Learning

- Network updating: select a portion of small-loss superpixels to jointly update networks, with the loss

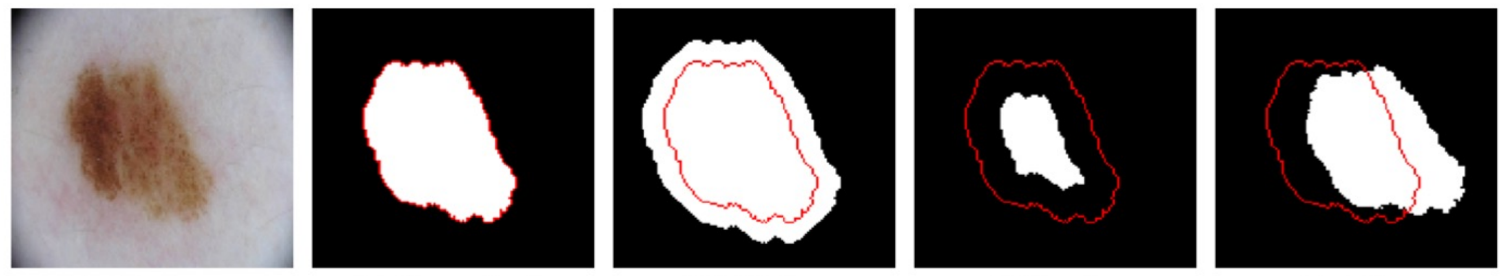
$$\ell^{sp} = (1-\lambda) * (\ell_{ce}(\mathbf{P}_s^1, \mathbf{Y}_s) + \ell_{ce}(\mathbf{P}_s^2, \mathbf{Y}_s)) + \lambda * \ell_{kl}(\mathbf{P}_s^1, \mathbf{P}_s^2)$$

$$\mathbf{P}_s^i(c,k) = \frac{1}{N(k)} \sum_{j:S_j=k} \mathbf{P}^i(c,j), \qquad \mathbf{Y}_s(c,k) = \frac{1}{N(k)} \sum_{j:S_j=k} \mathbb{1}(Y_j = c)$$

- Label refinement: relabel a subset of unreliable superpixel labels according to model predictions.

Experiment

Simulated Noise Patterns



(a) Skin image (b) Groundtruth (c) Dilation (d) Erosion (e) Affine

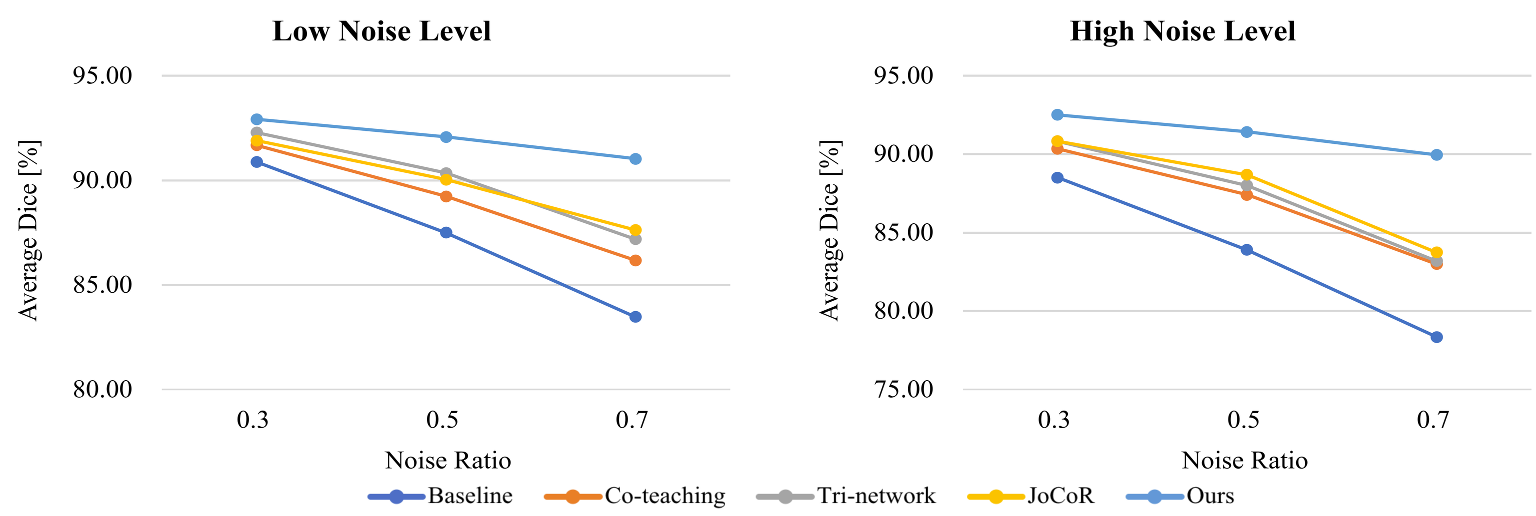
Experimental Results

* We report average Dice over the last 10 epochs.

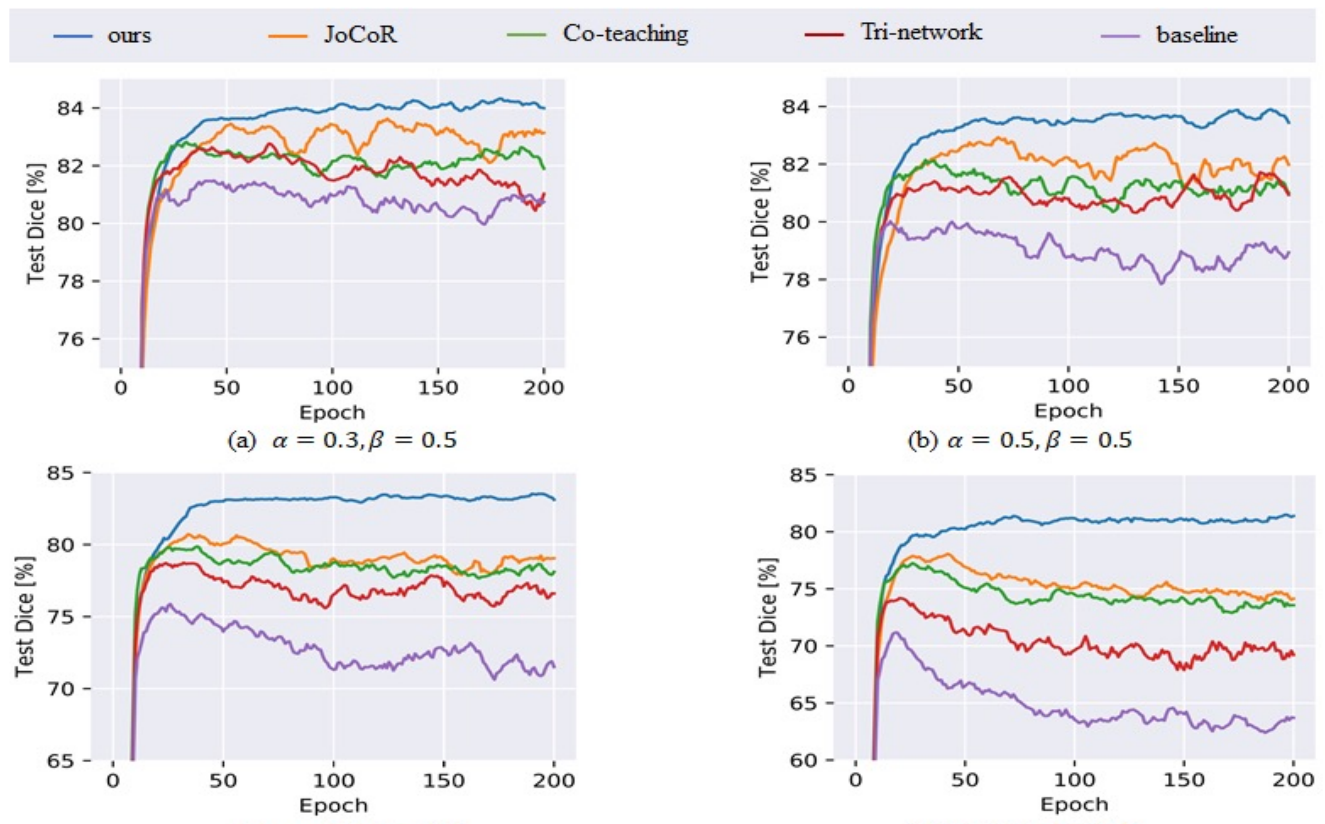
- Quantitative results on ISIC dataset. Test dice. (α, β control noise ratio and noise level;)

	Baseline	Co-teaching	Tri-network	JoCoR	Ours
Original data	82.49	82.72	82.96	83.64	**84.26**
$\alpha=0.3, \beta=0.5$	80.75	81.44	81.50	82.65	**84.00**
$\alpha=0.3, \beta=0.7$	79.46	81.47	80.73	81.58	**83.34**
$\alpha=0.5, \beta=0.5$	78.95	81.22	80.94	82.41	**83.90**
$\alpha=0.5, \beta=0.7$	75.44	80.06	80.24	81.06	**83.19**
$\alpha=0.7, \beta=0.5$	76.61	79.61	79.55	80.55	**83.83**
$\alpha=0.7, \beta=0.7$	71.51	78.50	76.61	79.05	**83.12**
$\alpha=1.0, \beta=0.5$	71.13	76.69	75.61	78.43	**82.23**
$\alpha=1.0, \beta=0.7$	63.71	73.68	70.01	74.30	**81.39**

- Quantitative results on JSRT dataset. Test dice.

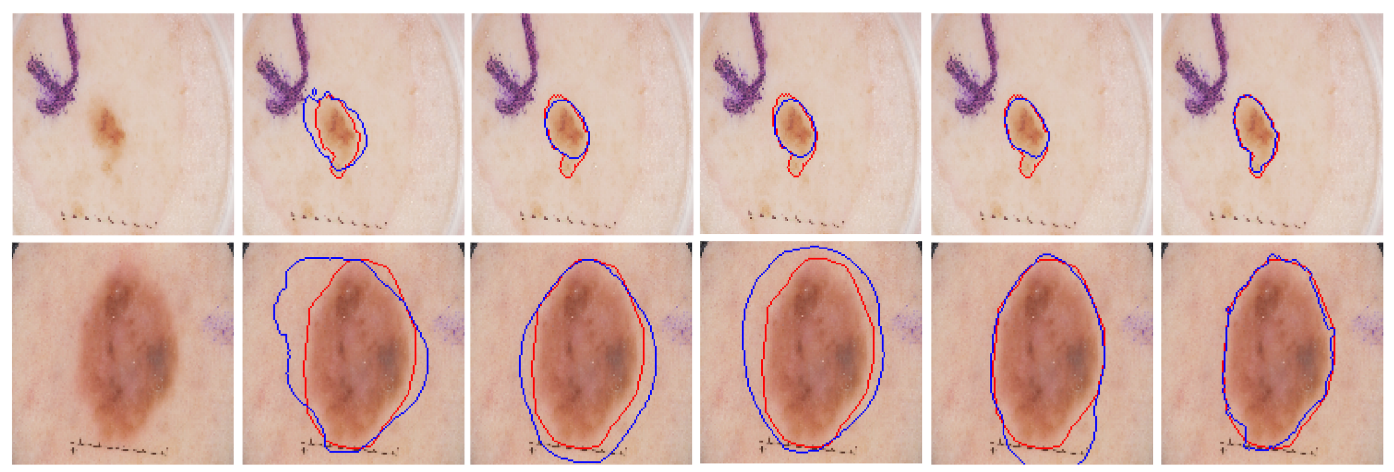


- Curves of test dice vs. epoch on four different noise settings. ISIC.



(a) $\alpha=0.3, \beta=0.5$ (b) $\alpha=0.5, \beta=0.5$ (c) $\alpha=0.7, \beta=0.7$ (d) $\alpha=1.0, \beta=0.7$

- Qualitative comparisons on ISIC dataset ($\alpha=0.7, \beta=0.7$). (red: groundtruth, blue: predicted mask)



(a) Skin image (b) Baseline (c) Co-teaching (d) Tri-network (e) JoCoR (f) Ours

Ablation Study

- Ablation study on our model components (ISIC, $\alpha=0.7, \beta=0.7$).

Method	Superpixel	Selection	Label Refinement	Dice[%]
Ours	✓	✓	✓	**83.12**
	×	✓	✓	81.15
	✓	×	✓	79.32
	✓	✓	×	80.56