A Spatially Constrained Deep Convolutional Neural Network for Nerve Fiber Segmentation in Corneal Confocal Microscopic Images using Inaccurate Annotations

Ning Zhang[1], Susan Francis[2], Rayaz A. Malik[3], Xin Chen[4]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Canada

[2]Sir Peter Mansfield Imaging Centre, University of Nottingham, UK

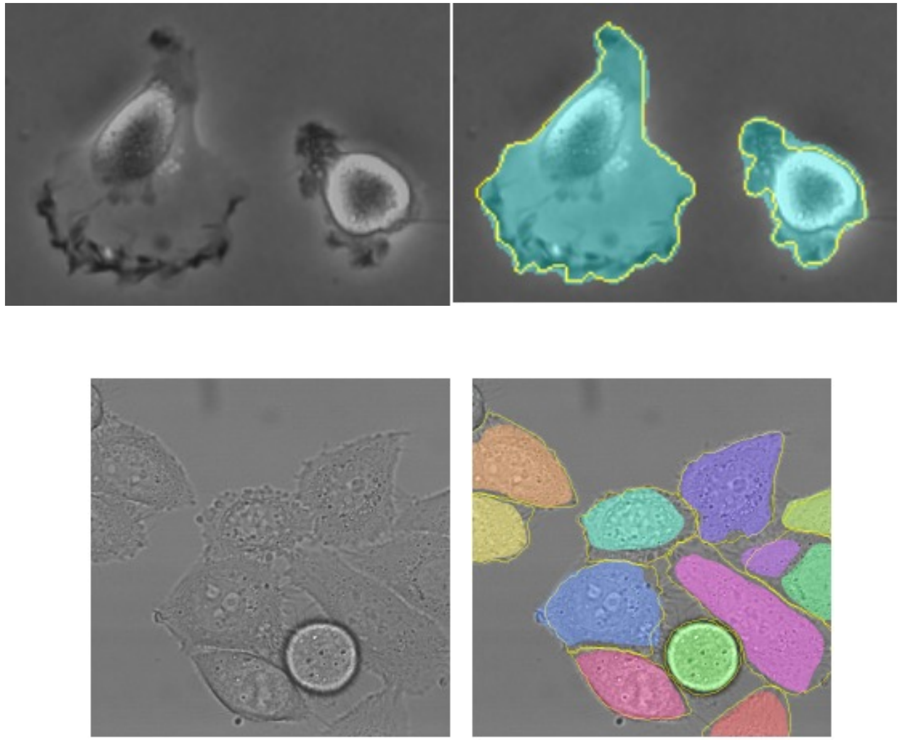
[3] Weill Cornell Medicine-Qatar, Qatar

[4] IMA Group, School of Computer Science, University of Nottingham, UK

Motivation

- Deep convolutional neural networks (DCNN) achieve superior performance in image segmentation when accurate labels are available for fully supervised learning.

- Obtain accurate labelling is challenging in medical images.



Fully supervised segmentation result of U-net [1]

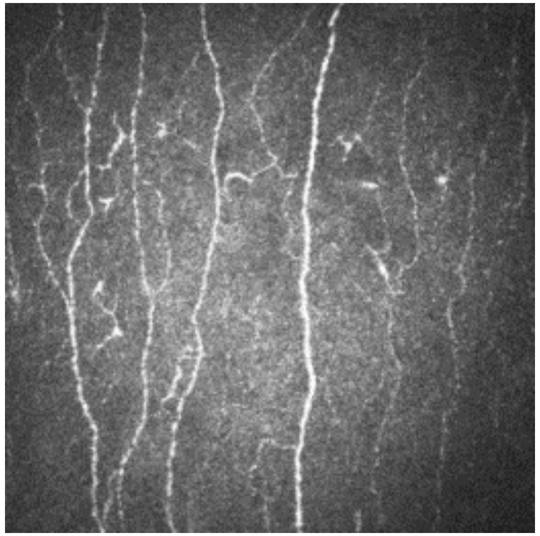
[1] *Olaf Ronneberger, et al,MICCAI, 2015, pp. 234–241.*

Aim

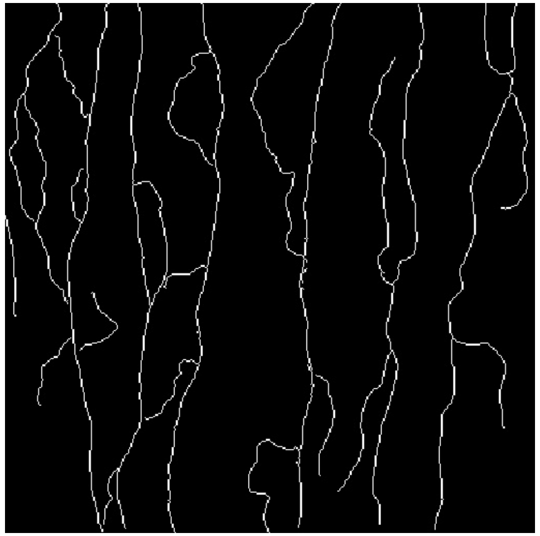
➢Train DCNN to achieve good performance using inaccurate labels

➢Challenges

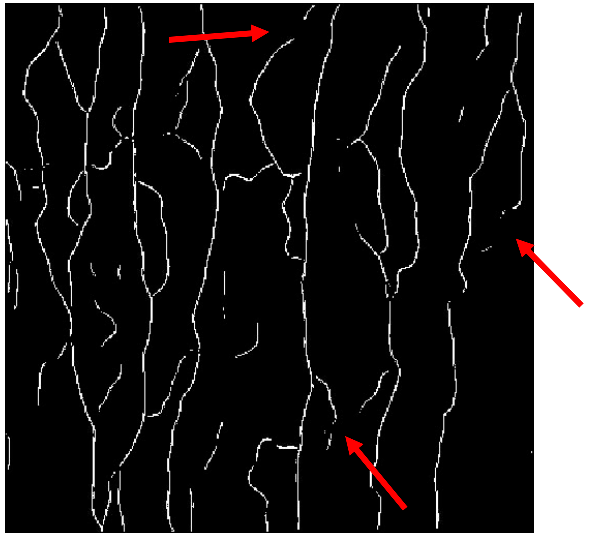
- Fully supervised learning will be misled by these inaccurate labels
- Ignore spatial consistency of labels



(a) CCM image



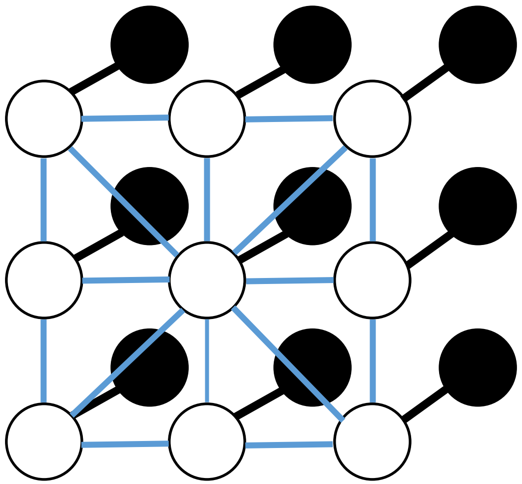
(b) Inaccurate label for training



(c) Output from U-net

Method

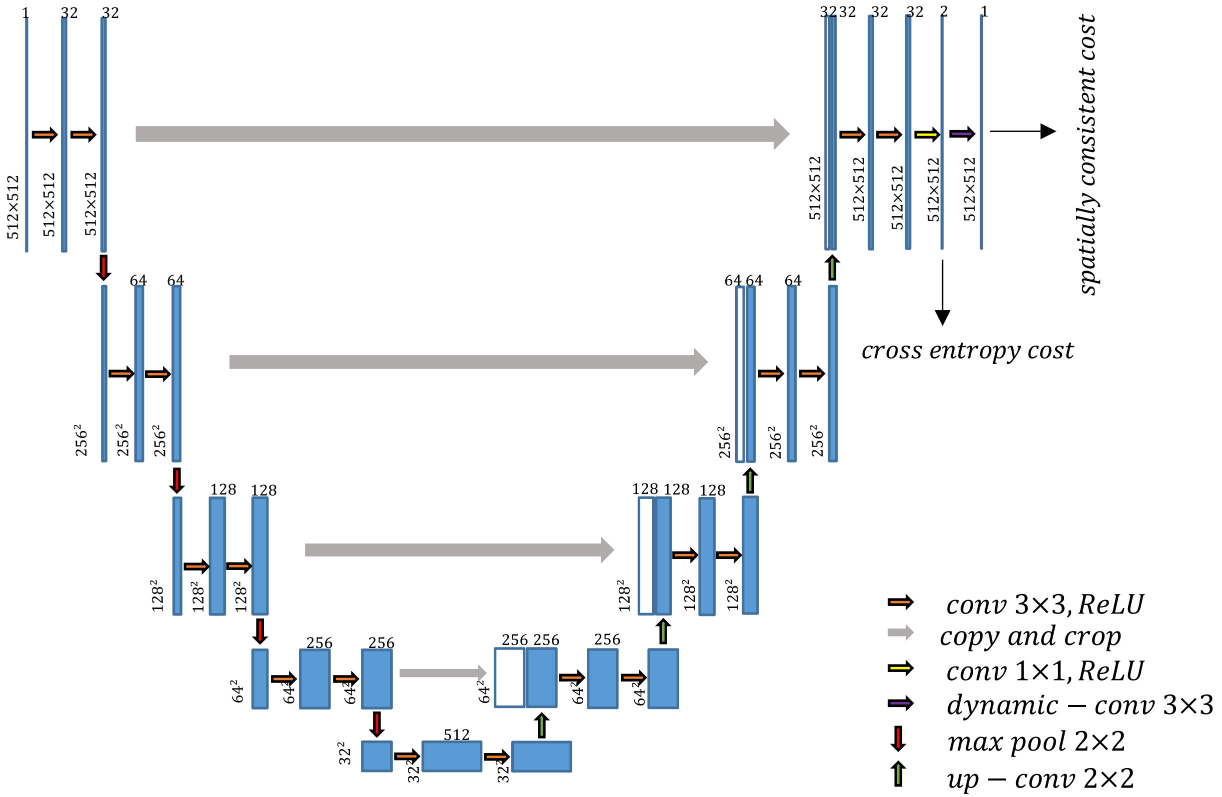
- Integrate conditional random field (CRF) into DCNNs to incorporate a spatial consistent constraint.
- Cost function: $L = L_{cross\ entropy} + L_{pairwise}$
  - Learning driven by both local image information and data labels



● Observed image pixel intensity

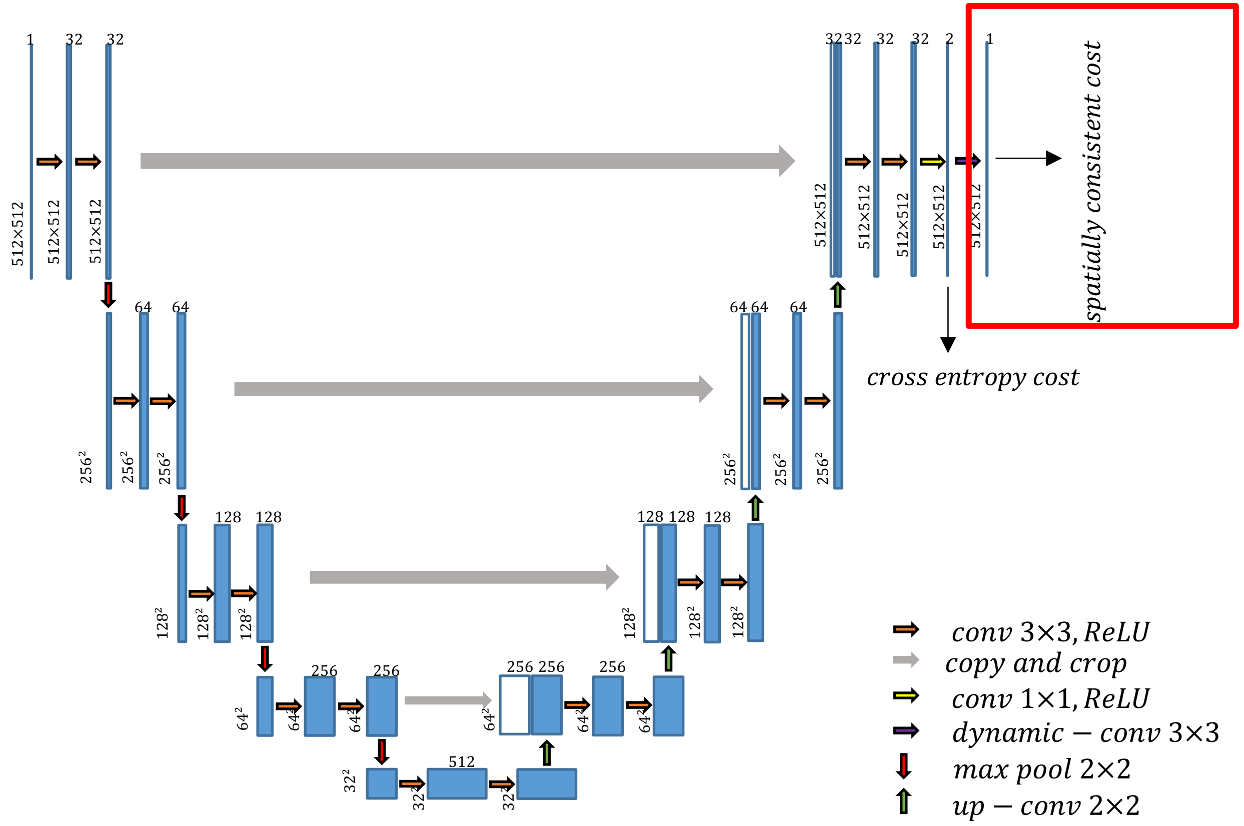
○ Label of pixel

Network structure



- **Encoding path:**
  - 4 layers of 3×3 convolutional operation followed by 2×2 max pooling
- **Decoding path:**
  - 4 layers of 2×2 up-convolutional operation followed by 3×3 convolutional operation
- **Skip connection**
- **Dynamic convolutional layer**

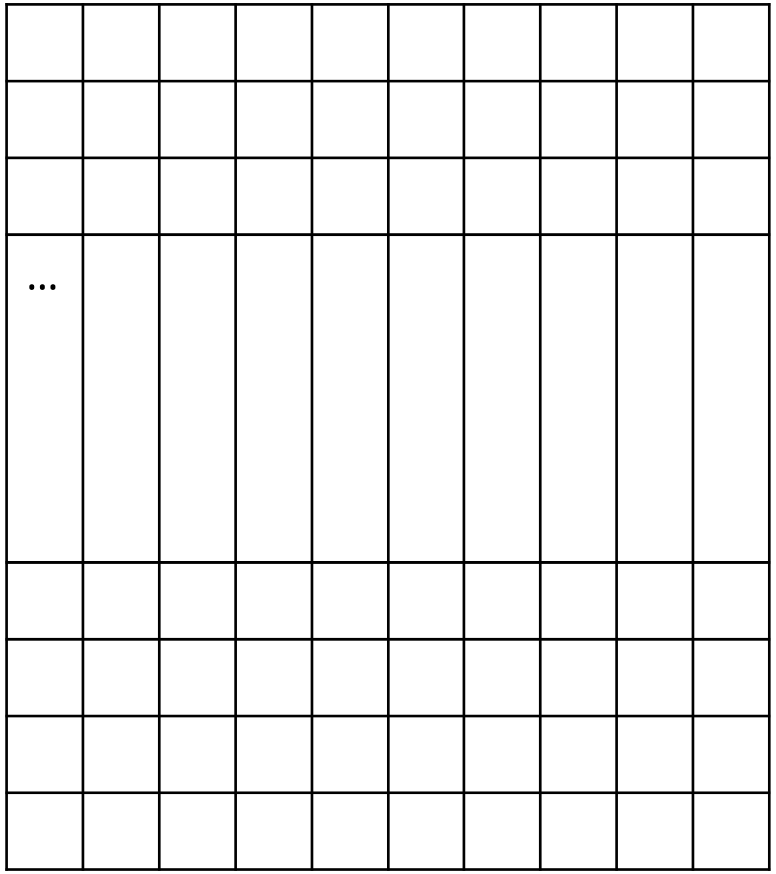
Network structure



Dynamic convolutional layer:

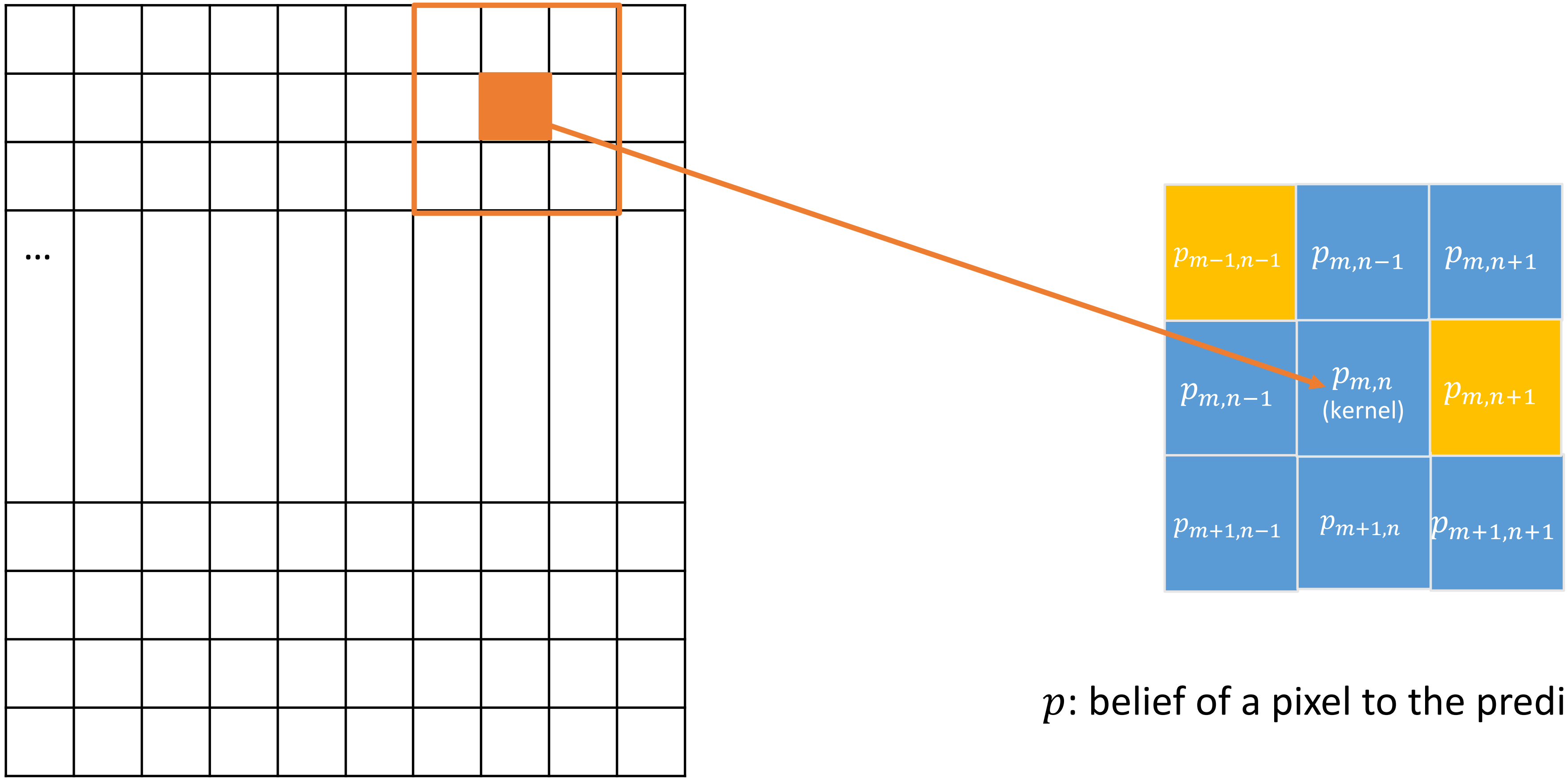
$$L_{pairwise} = \sum_i \psi_i = \sum_i \frac{\sum_j \psi_{i,j}}{Z_i}$$

CRF as 3-by-3 dynamic convolutional layer



Output probability map from DCNN

CRF as 3-by-3 dynamic convolutional layer

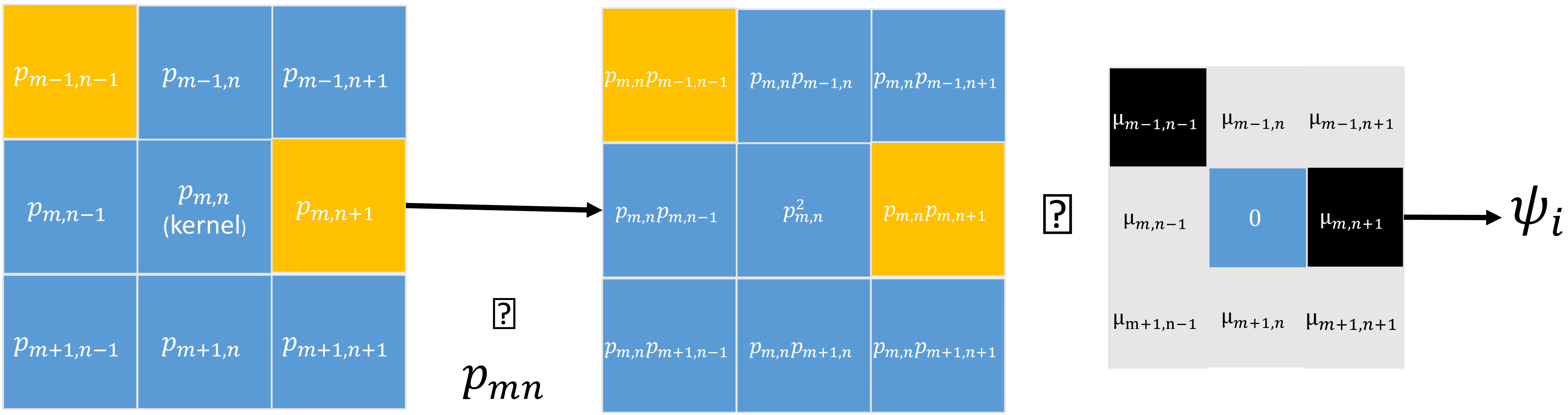


p: belief of a pixel to the predicted label

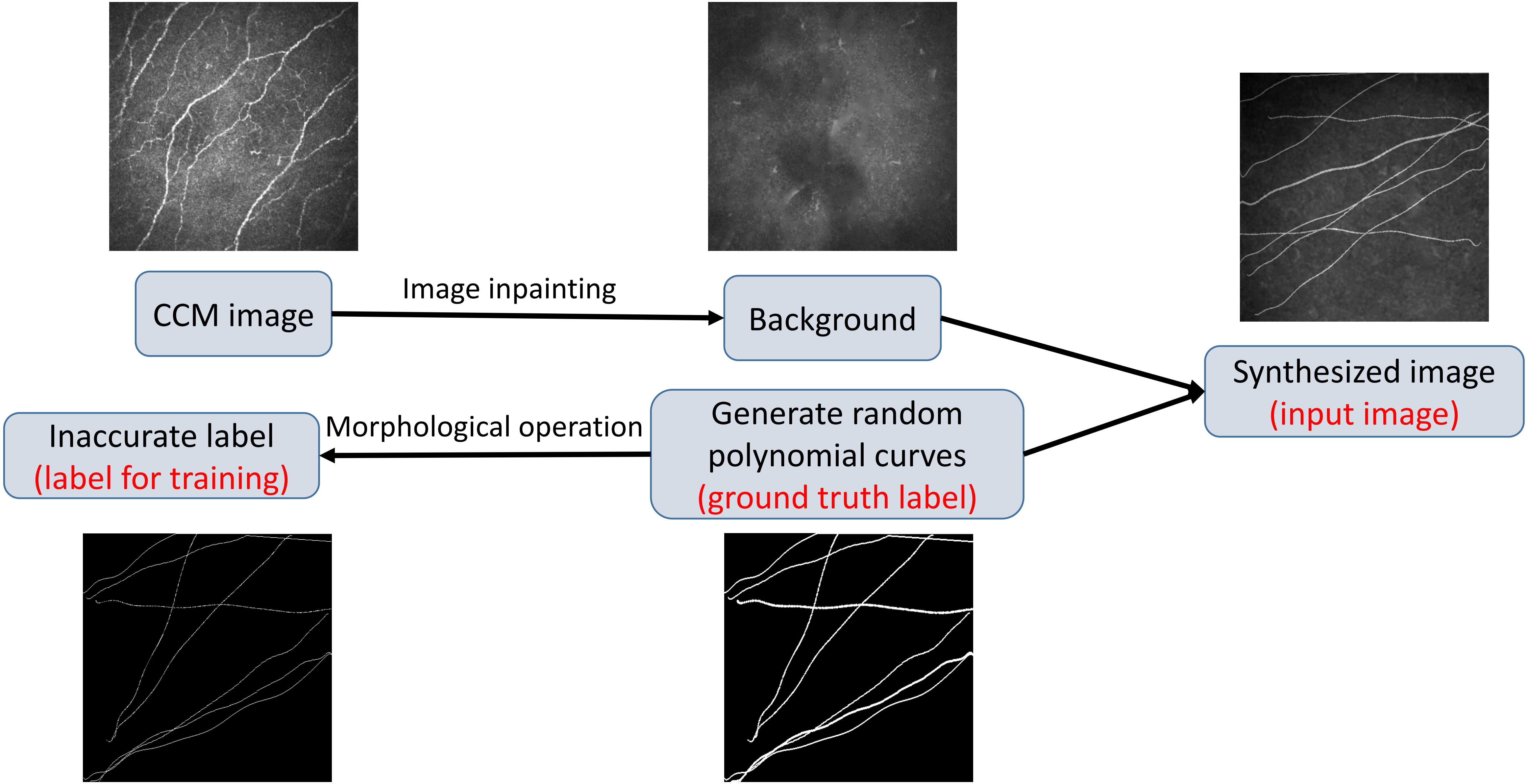
Output probability map from DCNN

CRF as 3-by-3 dynamic convolutional layer

$$\psi_{i,j} = \mu_{i,j} p_i p_j \qquad \psi_i = \sum_j \psi_{i,j} \qquad \mu_{i,j} = \begin{cases} -\exp\left(-\frac{(\mathrm{I_i}-\mathrm{I_j})^2}{2\sigma^2}\right) & (j \in R, \hat{y}_i = \hat{y}_j) \\ +\exp\left(-\frac{(\mathrm{I_i}-\mathrm{I_j})^2}{2\sigma^2}\right) & (j \in R, \hat{y}_i \neq \hat{y}_j) \\ 0 \ (j \notin R) \end{cases}$$



Synthesized dataset



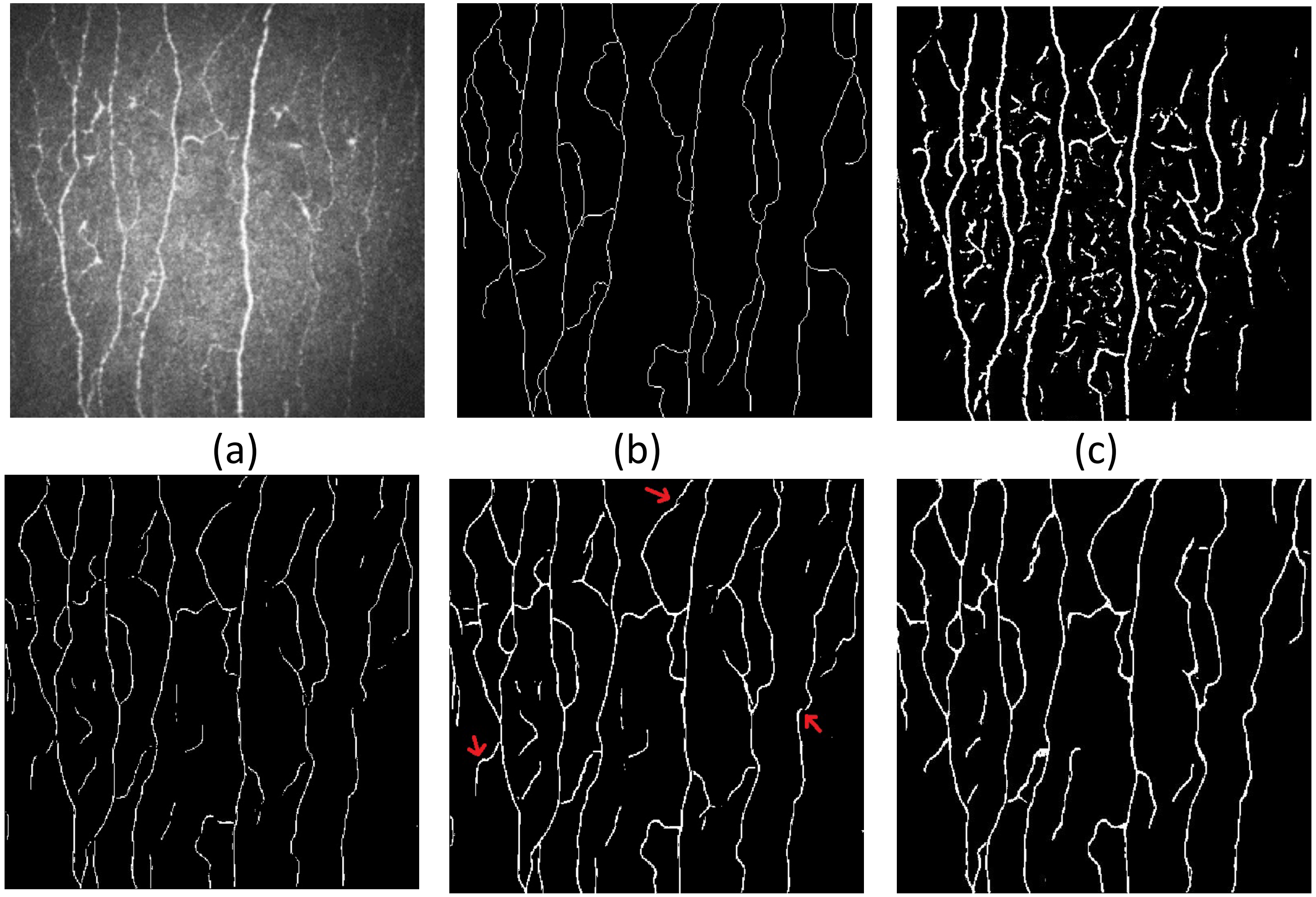
Quantitative results of synthesized dataset

- 500 training, 500 testing, image size: 512×512
- 300 epochs, 25% dropout, Adam optimizer with learning rate 10^{-4}
- Running time of dynamic convolution (per iteration): 0.036s

Method	Dice	Precision	Recall
Baseline	0.38 ± 0.04	0.94 ± 0.06	0.24 ± 0.03
Chen [2]	0.67 ± 0.12	0.58 ± 0.15	0.84 ± 0.07
U-net [1]	0.60 ± 0.10	0.98 ± 0.02	0.45 ± 0.03
U-net+CRF [3]	0.64 ± 0.13	0.97 ± 0.03	0.50 ± 0.04
Proposed	0.80 ± 0.12	0.90 ± 0.04	0.75 ± 0.06

[1]Olaf Ronneberger, et al, MICCAI, 2015, pp. 234–241.
[2] Chen X, et al. IEEE Transactions on Biomedical Engineering, 2016, 64(4): 786-794.
[3] Chen L C, et al. arXiv preprint arXiv:1412.7062, 2014.

Qualitative results of real CCM Images


(a) (b) (c) (d) (e) (f)

(a) CCM image

(b) Inaccurate manual annotation

(c) Result of Chen's method

(d) Result of U-net

(e) Result of U-net+CRF

(f) Result of our method

Summary

- Our method integrates CRF with DCNN to achieve image segmentation with inaccurate labels.
  - Implemented the CRF as a dynamic convolutional layer
- Advantages of our method
  - More accurate segmentation results using inaccurate labels.
  - Avoid overfitting to the data as local image information is also considered simultaneously.
  - The proposed dynamic convolutional layer could be integrated to any other DCNN based models.

Thanks for Watching!

Contact:

ningz@ece.ubc.ca,

Xin.Chen@nottingham.ac.uk