ISBI 2019 Challenge: Classification of Normal versus Malignant Cells in B-ALL White Blood Cancer Microscopic Images

# Neighborhood-Correction Algorithm for Classification of Normal *vs*. Malignant Cells

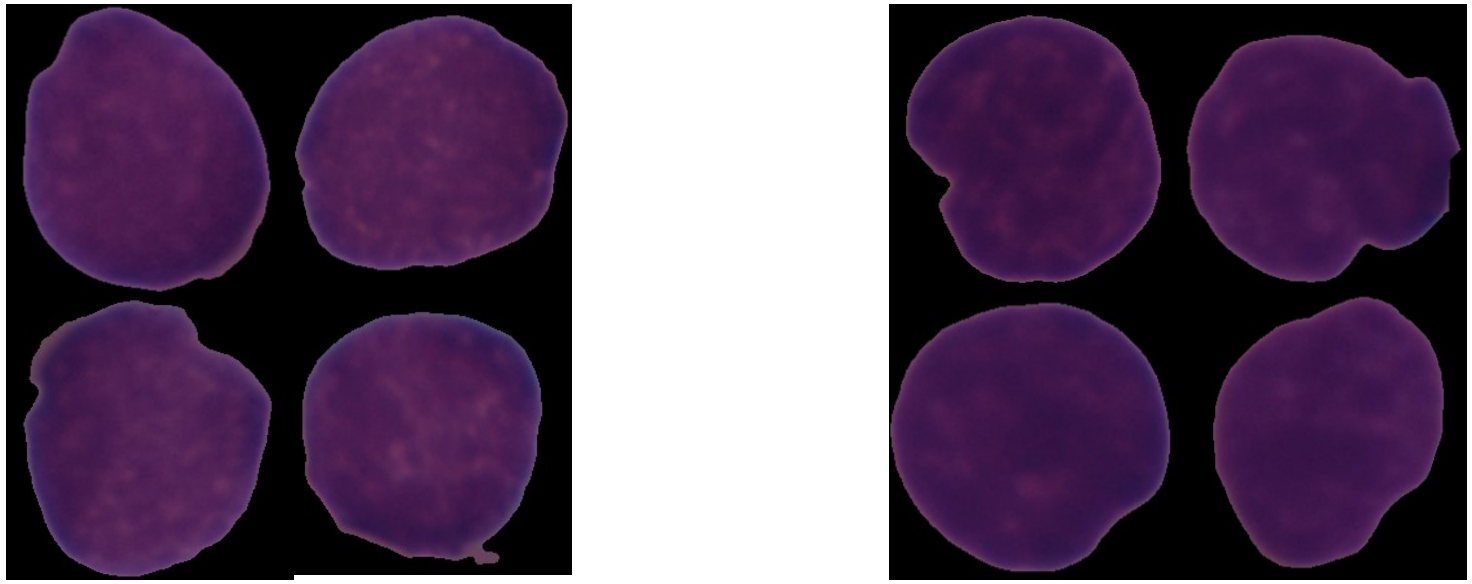
Yongsheng Pan [1, 2], Mingxia Liu [2], Yong Xia [1*], Dinggang Shen [2 *]

1. National Engineering Laboratory of China for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University (NPU), Xi'an, China

2. Department of Radiology and BRIC, University of North Carolina at Chapel Hill (UNC-CH), NC, USA

# Background and Aim

- Acute lymphoblastic leukemia (ALL): a serious health threat, mostly children with ages between 2 – 5
- Diagnosis of ALL: blood tests and bone marrow examination
- Aim: Classification of normal and malignant cells observed under a microscope for developing a cost-effective computer-aid diagnosis tool for ALL

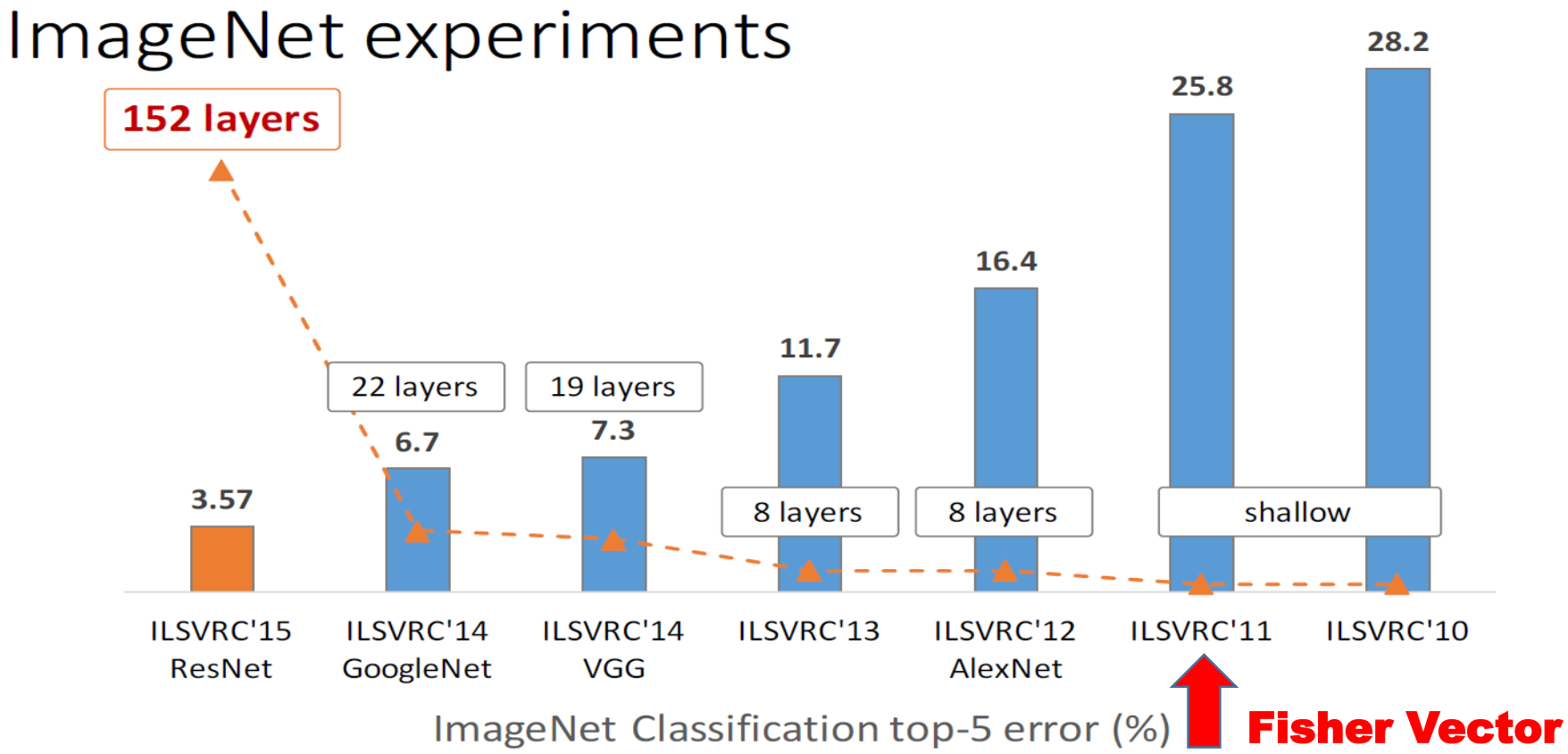


(a) Normal (b) Malignant

Fig. 1 Two cell images

# Related Work

- Fisher vector, deep convolutional neural networks, and combination




K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition. In: CVPR 2016, pp. 770-778. IEEE Press, New York. (2016)

# Related Work - Fisher Vector

- The set of local descriptors of an image: $\boldsymbol{X} = \{x \in \mathcal{R}^D; x \sim p\}$
- **Fisher Score** is the gradient of the log-likelihood $P(\boldsymbol{X}|\lambda)$

$$G_\lambda^X = \nabla_\lambda P(\boldsymbol{X}|\lambda) = E_{x\sim p} \nabla_\lambda \log u_\lambda(x) = \nabla_\lambda \int_x p(x) \log u_\lambda(x)\, dx$$

where $u_\lambda$ is the probability density function of the generative process

- **Fisher Information Matrix**

$$F_\lambda = E_{x\sim u_\lambda}[\nabla_\lambda log\, u_\lambda(x) \nabla_\lambda log\, u_\lambda(x)^{\mathrm{T}}] = L_\lambda{}^{\mathrm{T}} L_\lambda$$

- **Fisher Kernel**

$$K_{FK}(X, Y) = G_\lambda^{X^T} F_\lambda^{-1} G_\lambda^Y = \mathcal{G}_\lambda^{X^T} \mathcal{G}_\lambda^Y$$

- The **Fisher Vector** of $X$ is defined as $\mathcal{G}_\lambda^X = L_\lambda G_\lambda^X$.

# Method: Overview

- Neighborhood-correction algorithm (NCA) for classification of normal vs. malignant cells

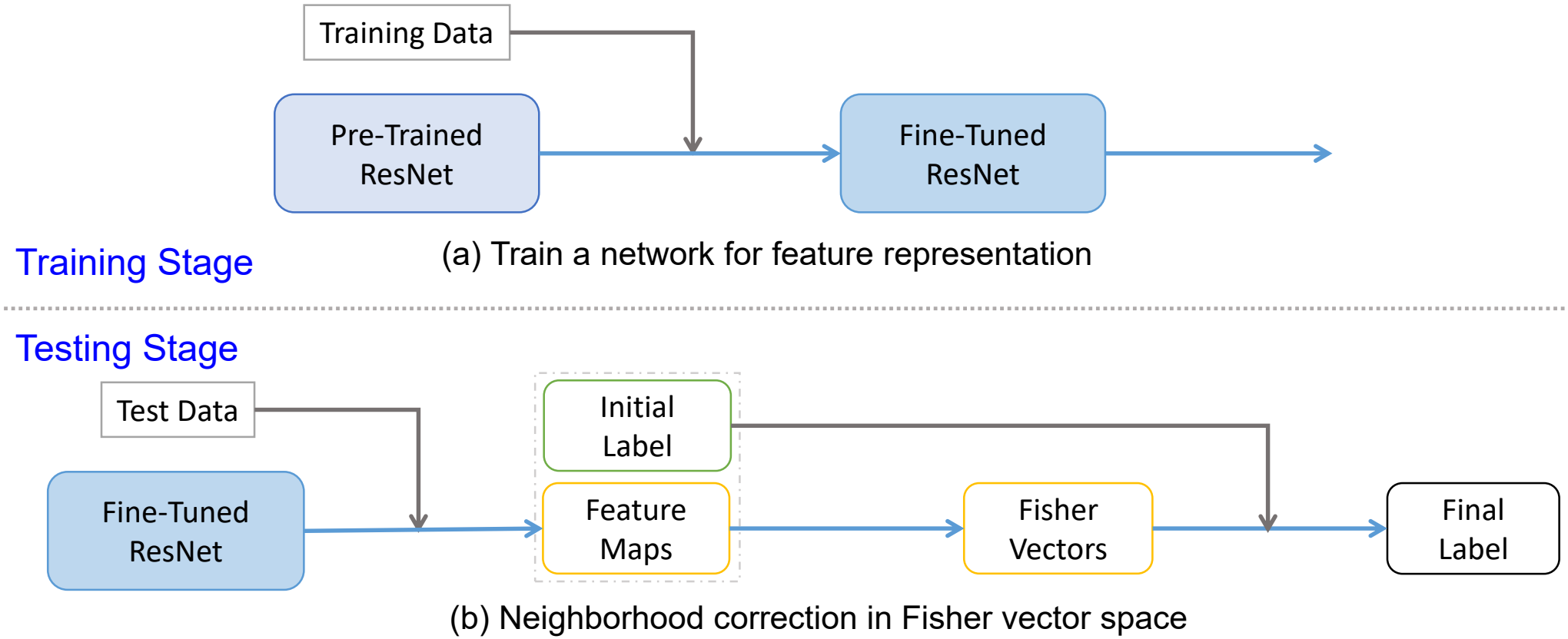



Fig. 2 Diagram of proposed NCA

# Method – Step 1 (Training)



- Data augmentation and image processing

(1) Left-right Flip

(2) Up-Down Flip

(3) Rotation

(4) Crop to 224× 224

(5) Minus mean

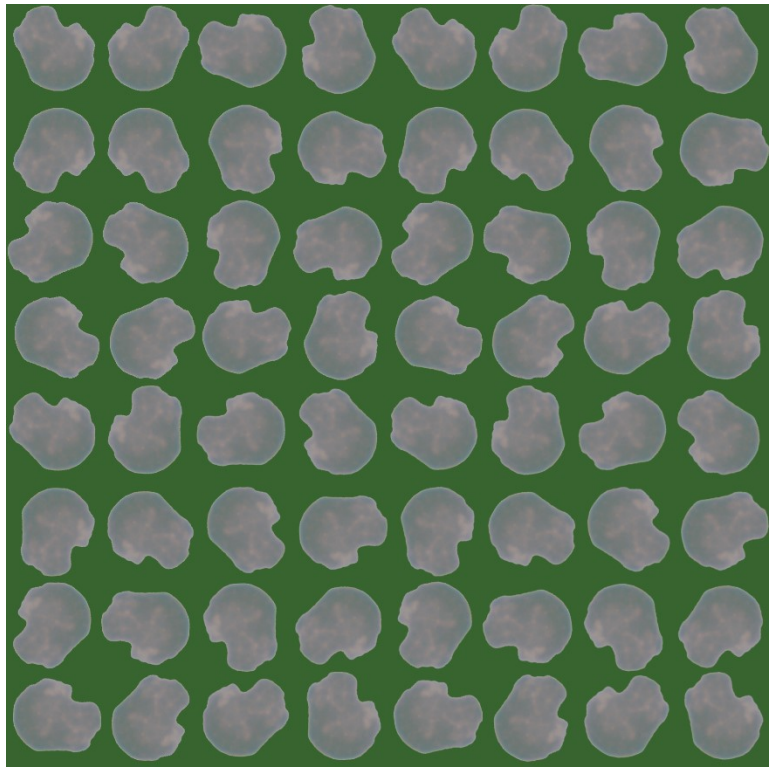


Fig. 2 Diagram of data augmentation and image processing

# Method – Step 2 (Training)

- Fine-tuning the pre-trained ResNet
  - Optimizer: SGD
  - Hyperparameters: Batch size = 32, epochs = 20, learning rate = $10^{-3}$ (first 10 epochs) / $10^{-4}$

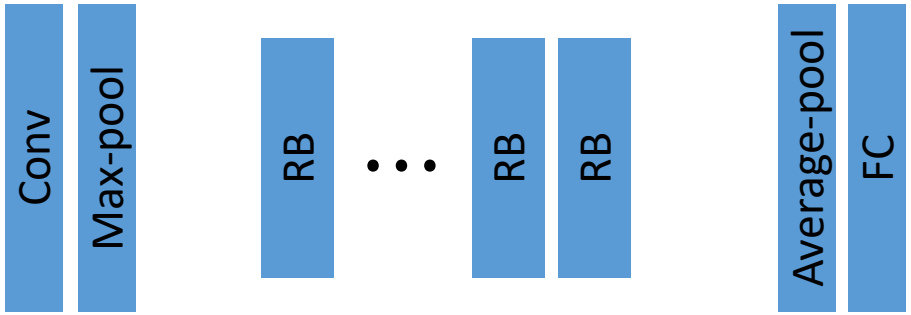



Fig.4. Diagram of ResNet (RB: Residual Block)

# Method – Step 3 (Testing)

- Applying the fine-tuned ResNet to each test image and obtaining the initial label of that image

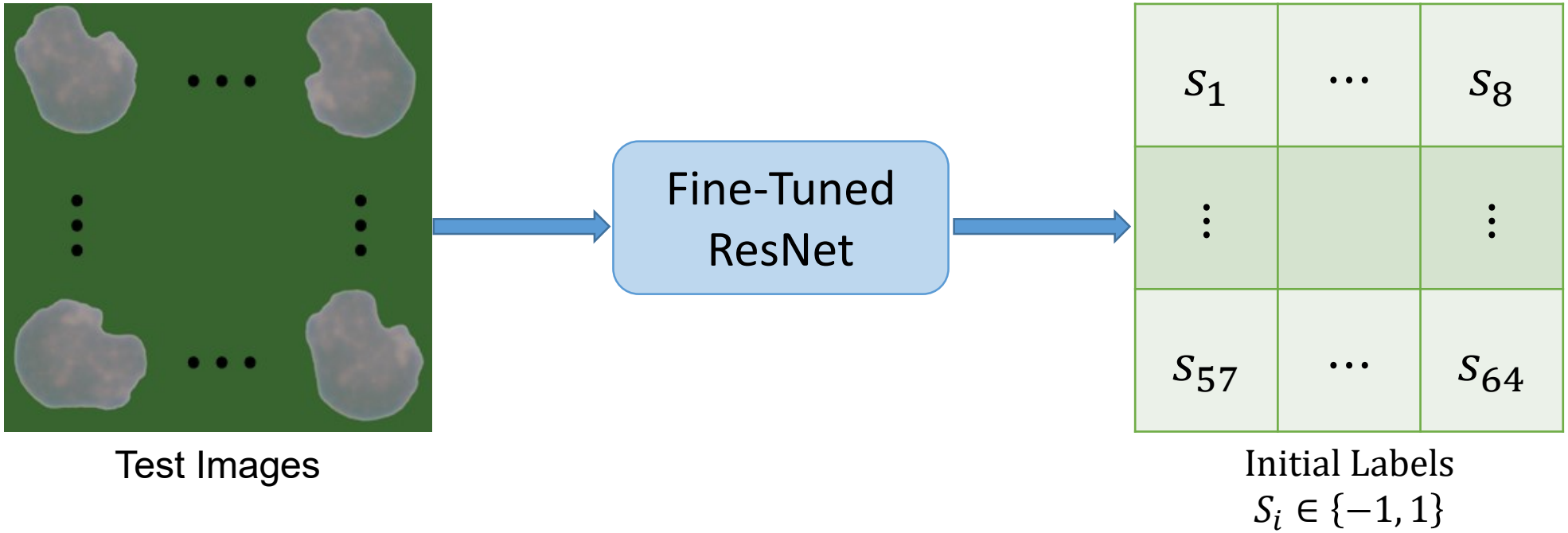
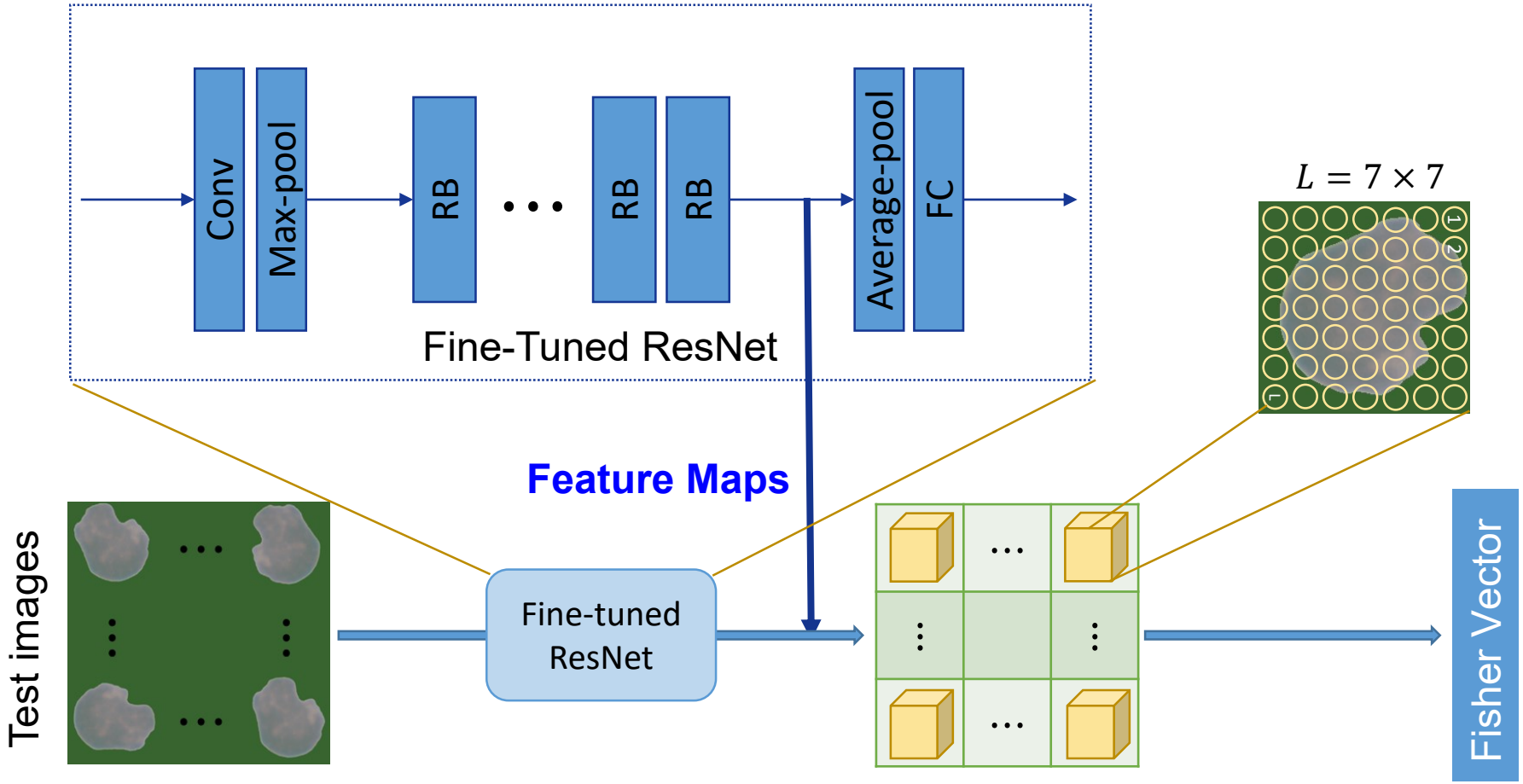



Fig. 2 Diagram of generating initial labels

# Method – Step 4 (Testing)

- Encoding the feature maps generated by the fine-tuned ResNet into a Fisher vector




Y. Pan, Y. Xia and D. Shen. "Foreground Fisher Vector: Encoding Class-Relevant Foreground to Improve Image Classification." IEEE-TIP, 2019.

# Method – Step 5 (Testing)

- Neighborhood Correction in FV Space

Let $\mathbb{B} = \{\mathcal{B}_i\}_{i=1}^{N}$ be a set of $N$ Fisher Vectors (FVs), where each FV represents a cell image. The similarity of a pair of images is

$$J(\mathcal{B}_i, \mathcal{B}_j) = \mathcal{B}_i{}^{\mathrm{T}}\mathcal{B}_j$$

Each image should be consistent with its neighbors in FV space, i.e., belong to the same category.

Suppose that $X_{i,(1)}, X_{i,(2)}, \ldots, X_{i,(K)}$ are the most similar $K$ images of $X_i$, whose initial labels are $y_{i,(1)}, y_{i,(2)}, \cdots, y_{i,(K)}$,

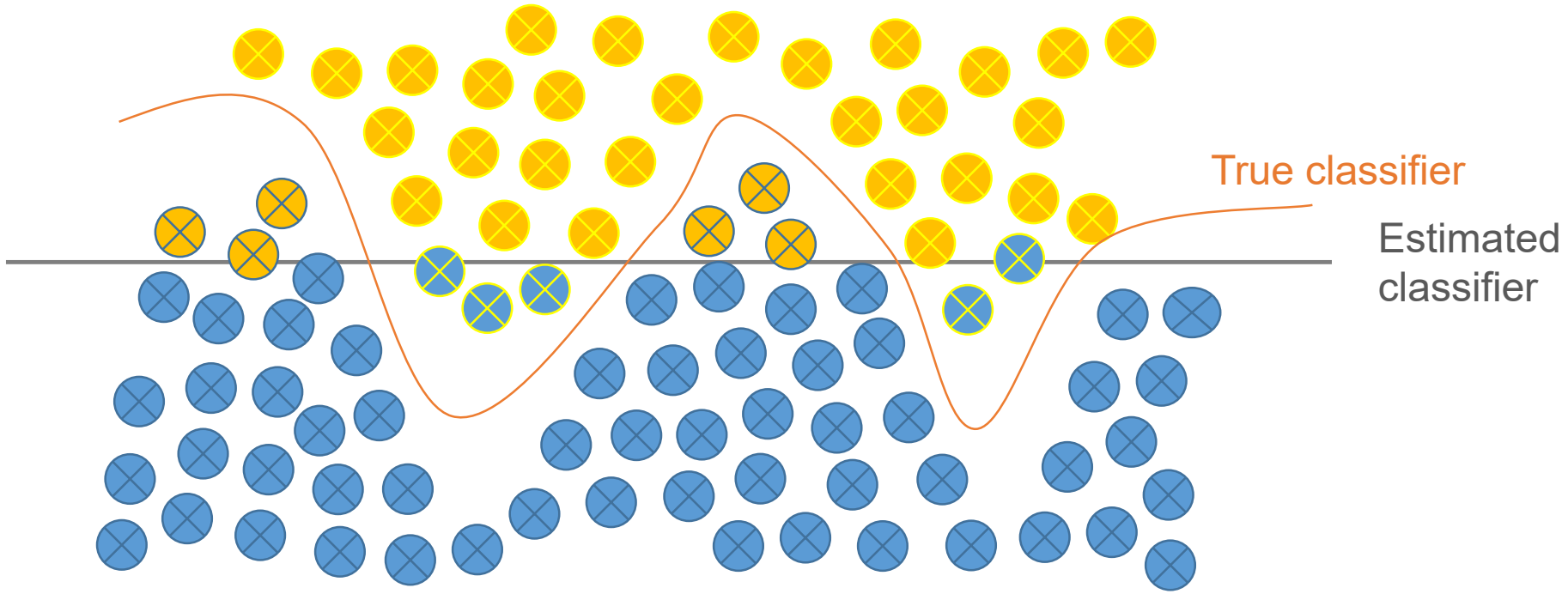
$$y_i^* = \operatorname{sign} \sum_{k=1}^{K} J(X_i, X_{i,(k)}) y_{i,(k)}$$

# Method – Step 5 (Testing)

- Neighborhood Correction in FV Space

  Most cells are correctly classified, only a small number of cells are misclassified.

  For each misclassified cell, we could probably correct it using its neighborhoods.

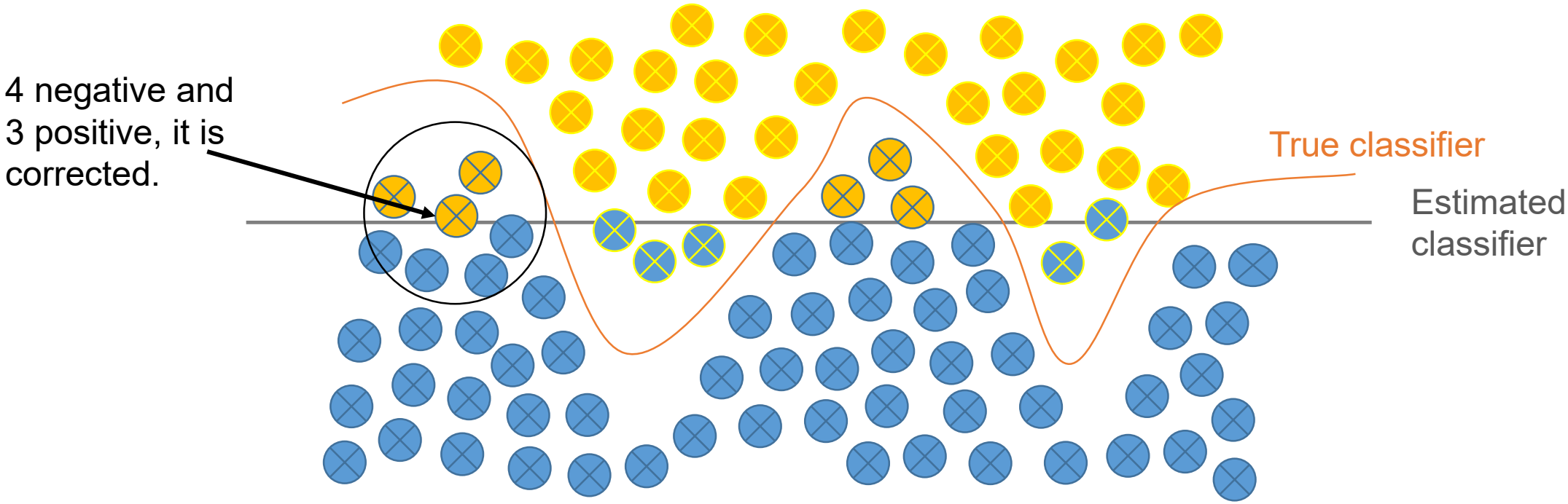


# Method – Step 5 (Testing)

- Neighborhood Correction in FV Space

Most cells are correctly classified, only a small number of cells are misclassified.

For each misclassified cell, we could probably correct it using its neighborhoods.

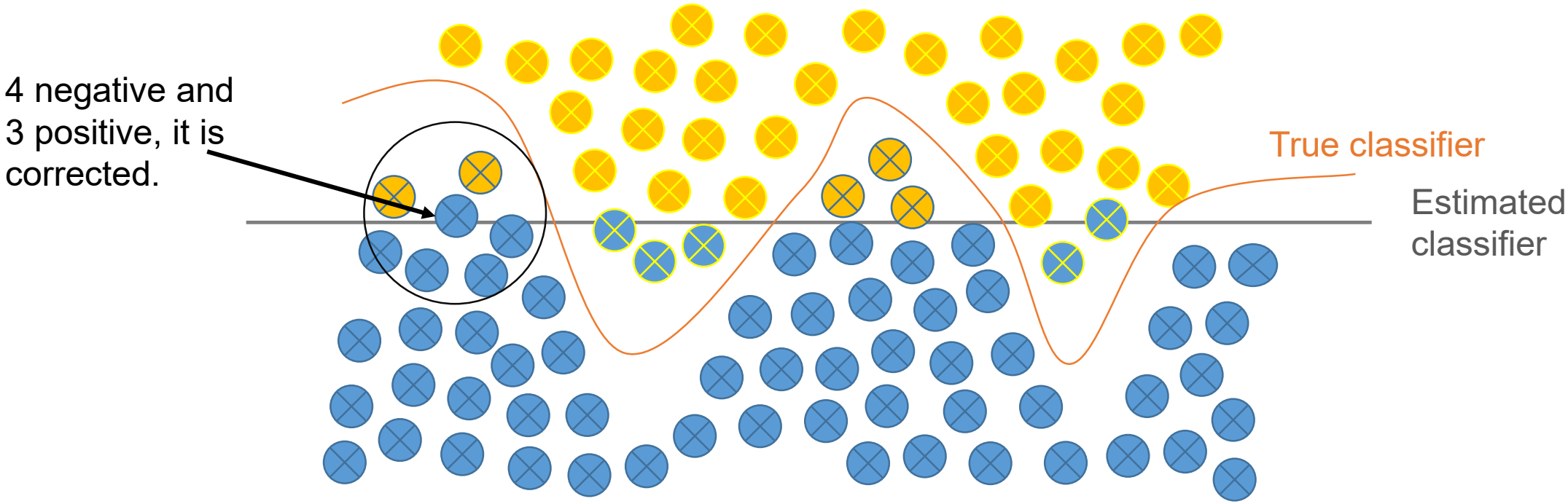


# Method – Step 5 (Testing)

- Neighborhood Correction in FV Space

Most cells are correctly classified, only a small number of cells are misclassified.

For each misclassified cell, we could probably correct it using its neighborhoods.



# Results in the C-NMC Challenge

西北工業大學 NORTHWESTERN POLYTECHNICAL UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Won the 1st Place in the Preliminary (0.9250) Test (1/60)

| # | User | Entries | Date of Last Entry | Team Name | Prediction score ▲ | Duration ▲ | Detailed Results |
|---|---|---|---|---|---|---|---|
| 1 | yspan | 177 | 01/31/19 | SAIIP-NPU & BRIC-UNC | 0.9250 (1) | 0.00 | View |
| 2 | jprellberg | 29 | 01/30/19 | | 0.8966 (2) | 0.00 | View |
| 3 | kuangrf | 174 | 01/27/19 | BUPT-CAD | 0.8955 (3) | 0.00 | View |
| 4 | yelan | 80 | 01/24/19 | | 0.8942 (4) | 0.00 | View |
| 5 | haoyuyang | 62 | 01/23/19 | Sunset Roller-coaster | 0.8929 (5) | 0.00 | View |
| ⋮ | | | | | | | |
| 59 | | | | | | | |
| 60 | | | | | | | |

# Results in the C-NMC Challenge

西北工業大學 NORTHWESTERN POLYTECHNICAL UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Won the 1st Place in in the Final Test (0.9104) (1/25)

| # | User | Entries | Date of Last Entry | Team Name | Prediction score ▲ | Duration ▲ | Detailed Results |
|---|---|---|---|---|---|---|---|
| 1 | yspan | 2 | 03/17/19 | SAIIP-NPU & BRIC-UNC | 0.9104 (1) | 0.00 | View |
| 2 | deepshad | 2 | 03/17/19 | | 0.8947 (2) | 0.00 | View |
| 3 | jprellberg | 2 | 03/16/19 | | 0.8891 (3) | 0.00 | View |
| 4 | kuangrf | 2 | 03/17/19 | BUPT-CAD | 0.8856 (4) | 0.00 | View |
| 5 | wendy | 2 | 03/16/19 | MIA Group | 0.8798 (5) | 0.00 | View |
| ⋮ | | | | | | | |
| 25 | | | | | | | |

- Submissions in the Final Test (with an improvement of 0.0416)

| METHOD | # | SCORE | FILENAME | SUBMISSION DATE | STATUS | ✔ | |
|---|---|---|---|---|---|---|---|
| Fine-tuned ResNet | 1 | 0.868851339 | ftcell_50_101res-cv0 1 2 3.zip | 03/16/2019 17:56:00 | Finished | | + |
| Proposed Method | 2 | 0.9104320738 | ftcell_50_101res-cv0 1 2 3 - 2.zip | 03/17/2019 23:52:55 | Finished | ✔ | + |

# Summary and Future Work



- Proposed a hybrid algorithm (deep learning + Fisher vector) for the classification of normal vs. malignant cells
- ResNet: feature extraction and initial label generation
- Fisher vector: Correction of inaccurate initial labels
- Future work
  - Using corrected class labels to refine ResNet
  - Improving ResNet and Fisher vector mutually and iteratively

NPU Library at Chang'an Campus