Predicting Protein Organelle Localization with Transfer Learning

Ruge Zhao,[1] Siyi Tang[2]

[1] Department of Statistics, Stanford University
[2] Department of Electrical Engineering, Stanford University

Motivation

- Advances in high-throughput microscopy images aid understanding of complex protein functions in human cells
- Accurate image classifications required for large-scale cell images provided by Human Protein Atlas (HPA) [1]
- Machine learning algorithms proven to be effective in automating the classification [2-4]

Goal

- Predict protein organelle localization labels in human cell samples
- Explore effectiveness of transfer learning

Data

- 31072 samples, 28 classes
- Each sample represented by 4 channels (RGBY)
- Each sample labeled with ≥ 1 protein organelle locations
- Imbalanced classes (e.g. 12885 vs 11)
- Train/validation/test sets: 24858/3107/3107

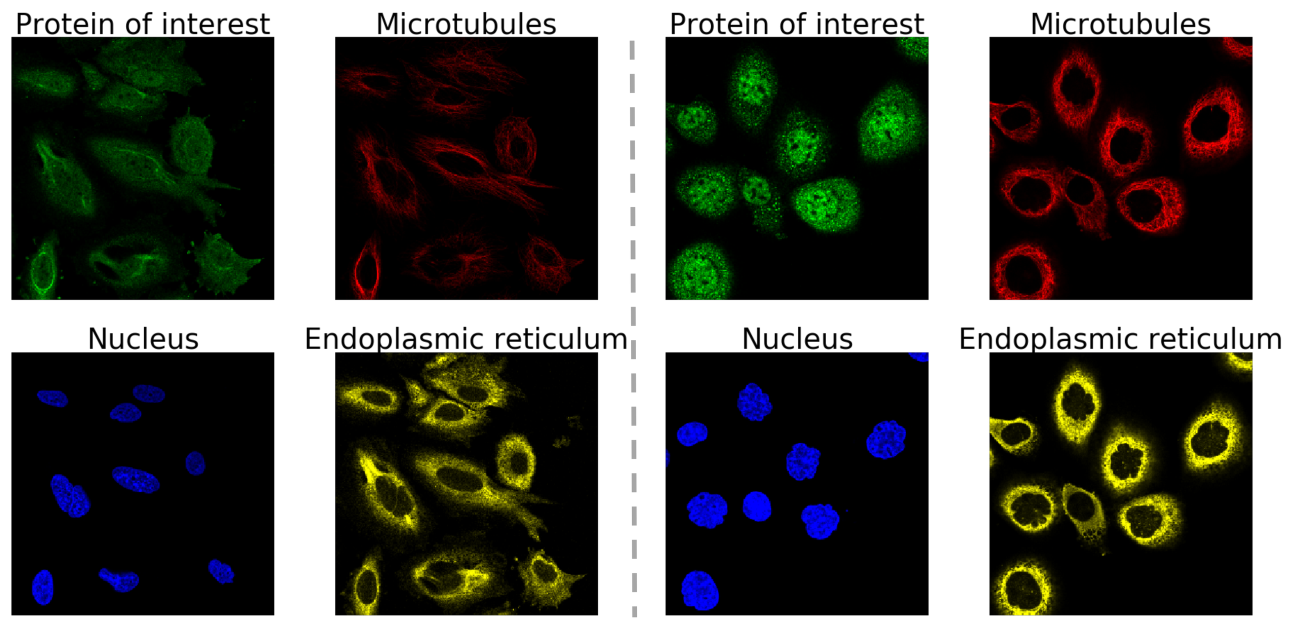

Fig. 1: Four channels of two cell samples

Preprocessing

- Standardized with respect to train set
- Data augmentation: random rotate, flip, lighting

Baseline

- 28 one-vs-all classifiers
- Logistic regression (LR) & random forest (RF)

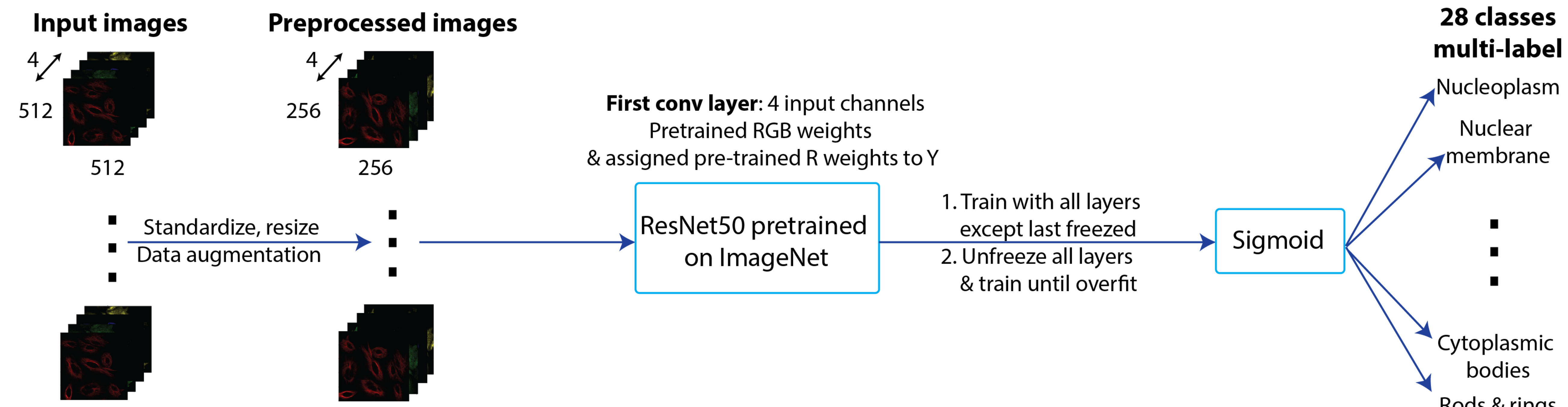

Fig. 2: Overview of preprocessing and transfer learning

Transfer learning of ResNet50

ResNet50 pre-trained on ImageNet dataset

Initializing network weights

- RGB: initialized with pre-trained weights
- Y: initialized with pre-trained R weights

Loss functions

- Binary cross-entropy (BCE) loss

$$L_{BCE}(p_t) = -\log(p_t)$$

- Focal loss [5]

$$L_{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t)$$

where $p_t = p$ if $y = 1$, else $p_t = (1-p)$;
$y \in \{0,1\}$: true label; $p \in [0,1]$: predicted prob.

- Focal loss equivalent to BCE loss when $\gamma = 0$

Training of the network

- Freeze all layers except last, train for 1 epoch
- Unfreeze all layers and train until overfit

Choosing optimal learning rate

- Picked learning rate where loss is low but still clearly decreasing

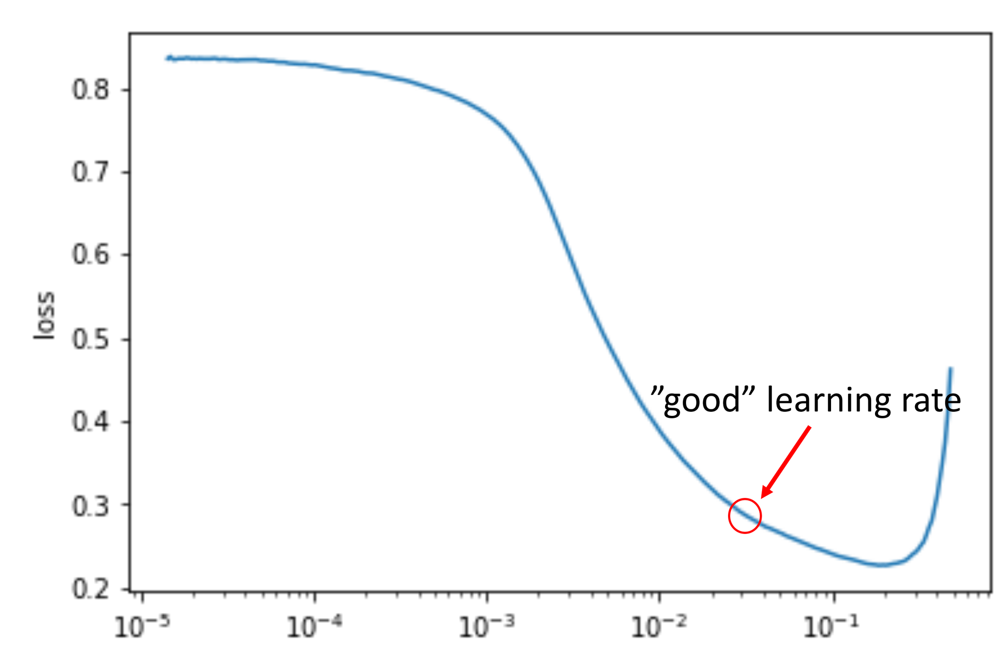

Fig. 3: Learning rate vs training loss

Differential learning rates

- Later layers need more fine-tuning to capture dataset-specific features
- Higher learning rate for later layers

Training with stochastic gradient descent with restarts (SGDR) [6]

- Learning rate annealing during each cycle, and restarts learning rate at next cycle
- Encourages to find "smooth" rather than "spiky" local minimum

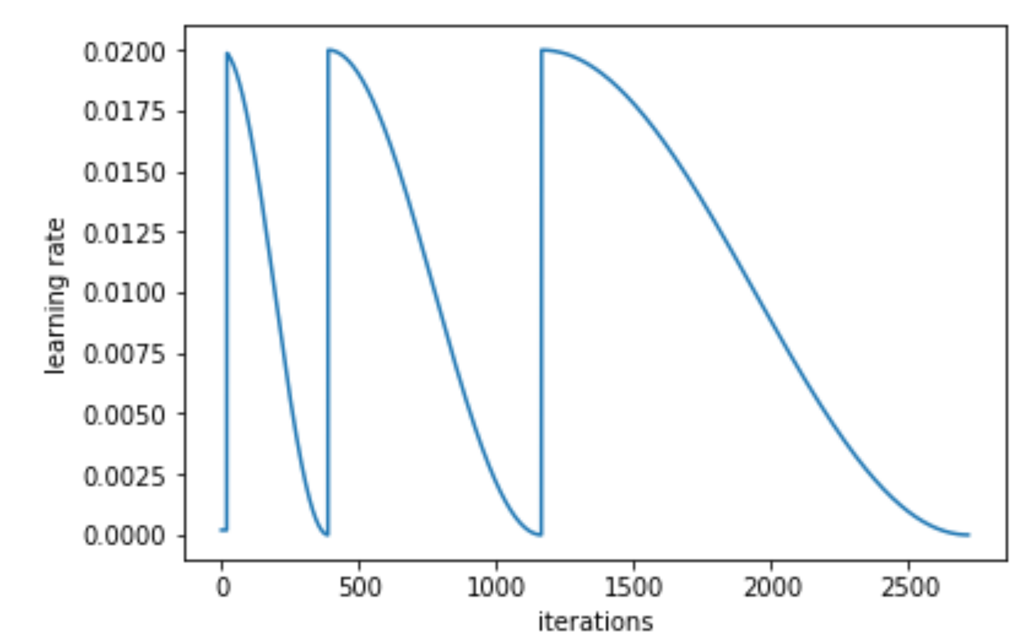

Fig. 4: Learning rate vs number of iterations during training with SGDR

Results

Baseline

- LR performed better than RF
- Best baseline model: macro F1 = 0.12

Transfer learning models

- Best model without SGDR: Focal loss with γ=0.2
- With SGDR:
  - SGDR improved macro F1
  - Focal loss and BCE loss had comparable performance.

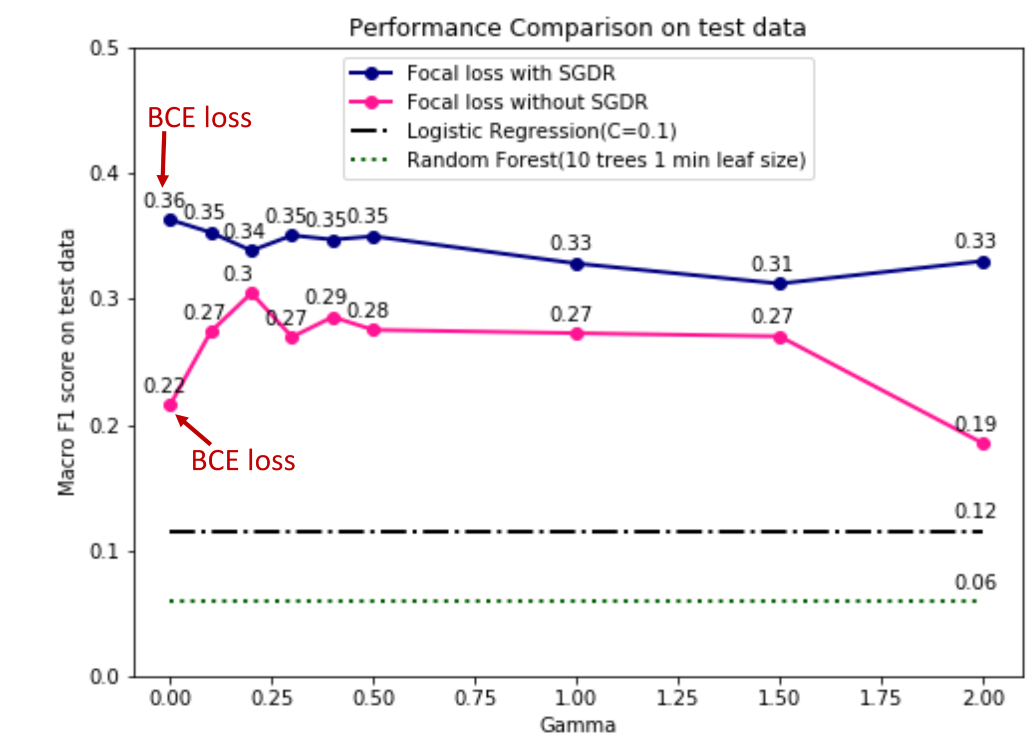

Fig. 5: Performance comparison on test data: macro F1

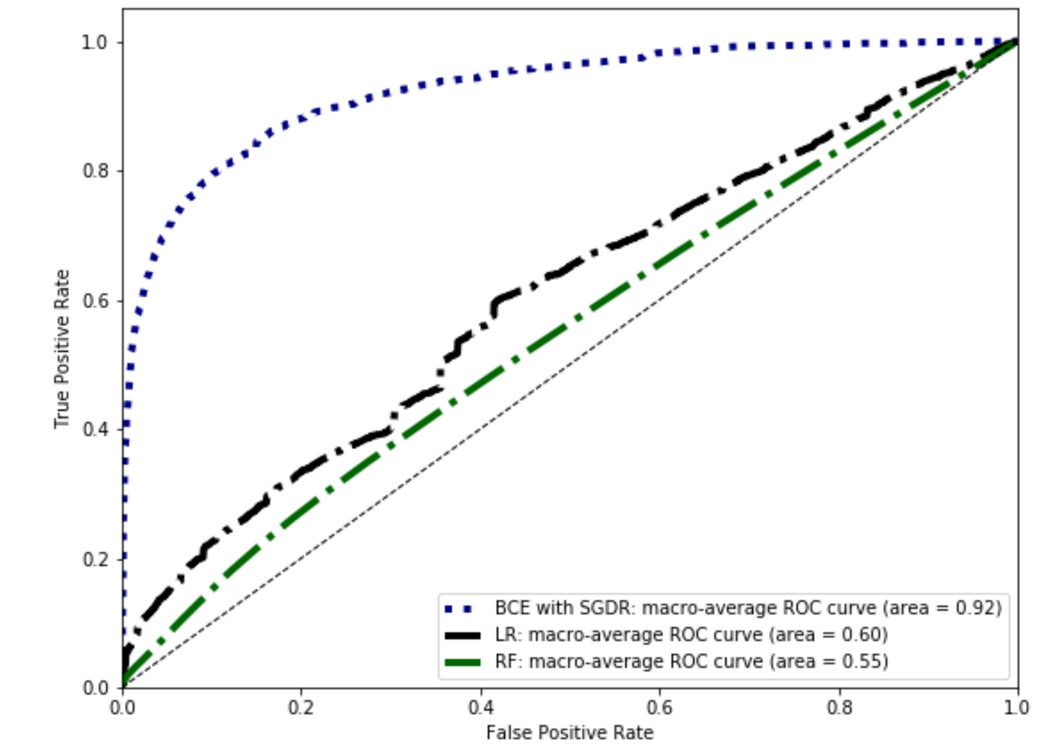

Fig. 6: Performance comparison on test data: ROC AUC

Conclusion and future work

- ResNet pre-trained on ImageNet data has learned transferable features applicable to the HPA dataset
- Transfer learning model outperformed traditional machine learning models

Future work

- Experiment with highest resolution for all models
- Reweigh / resample majority and minority classes

References

[1] D. P. Sullivan, C. F. Winsnes, L. Åkesson, M. Hjelmare, M. Wiking, R. Schutten, L. Campbell, H. Leifsson, S. Rhodes, A. Nordgren, K. Smith, B. Revaz, B. Finnbogason, A. Szantner, and E. Lundberg, "Deep learning is combined with massive-scale citizen science to improve large-scale image classification," *Nature Biotechnology*, vol. 36, p. 820, aug 2018.
[2] J. Li, L. Xiong, J. Schneider, and R. F. Murphy, "Protein subcel- lular location pattern classification in cellular images using latent discriminative models," *Bioinformatics*, vol. 28, no. 12, pp. i32–i39, 2012.
[3] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016.
[4] T. Pärnamaa and L. Parts, "Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning," *G3: Genes, Genomes, Genetics*, vol. 7, no. 5, pp. 1385–1392, 2017.
[5] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017 , 2017, pp. 2999–3007.
[6] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983v5*, 2016.
[7] Iafoss, "Kaggle kernel," https://www.kaggle.com/iafoss/pretrained-resnet34-with-rgby-0-460-public-lb, 2018.