THE UNIVERSITY OF UTAH
ELECTRICAL & COMPUTER ENGINEERING

UNIVERSITY OF UTAH HUNTSMAN CANCER INSTITUTE

HEALTH UNIVERSITY OF UTAH

MICCAI2022 Singapore

MILLanD Workshop

Stain Based Contrastive Co-training for Histopathological Image Analysis

Bodong Zhang[1,2], Beatrice Knudsen[3], Deepika Sirohi[3], Alessandro Ferrero[2], Tolga Tasdizen[1,2]

1 Department of Electrical and Computer Engineering, 2 Scientific Computing and Imaging Institute, 3 Huntsman Cancer Institute

University of Utah, Salt Lake City, UT, USA

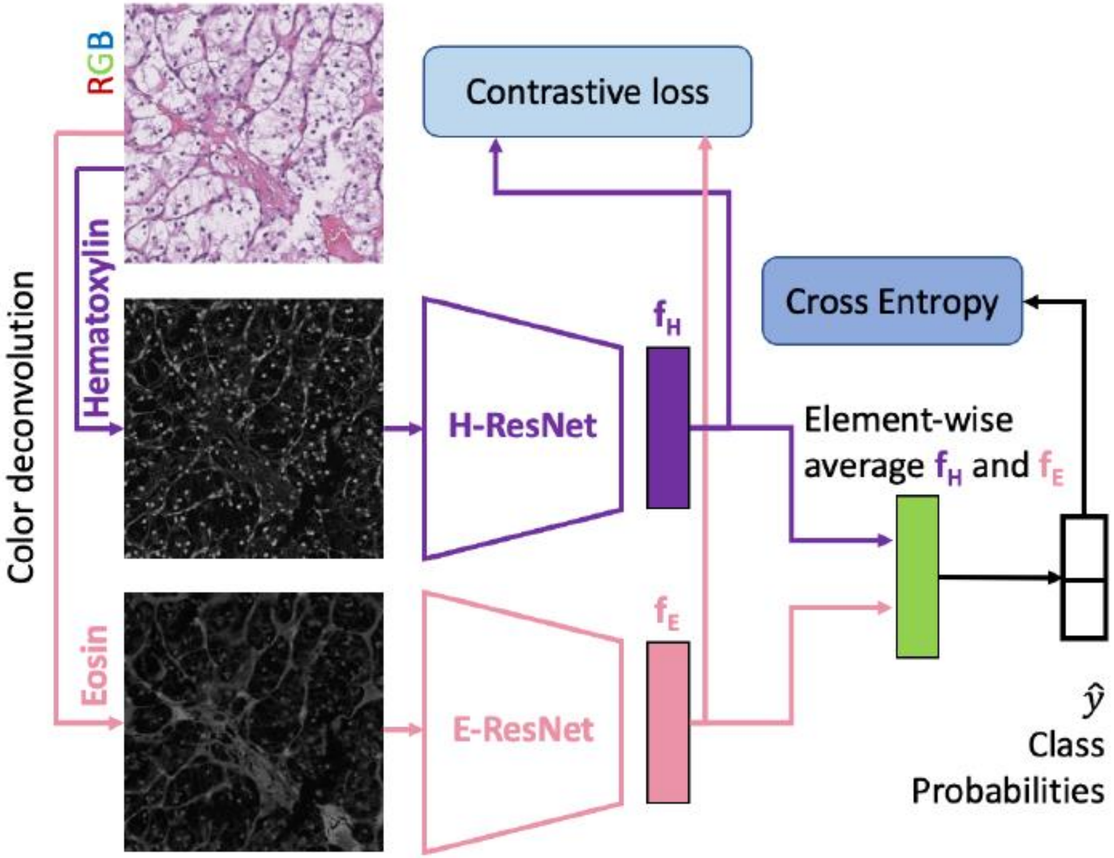
Presenter: Bodong Zhang

Introduction

- Deep learning models are widely used in histopathology image classification.
- Expert annotation at tile level for providing labels in model training has very high cost and is infeasible beyond a small number of whole slide images (WSIs).
- Semi-supervised learning (SSL) tries to utilize unlabeled data in training combined with limited amount of labeled data.
- Co-training approach to SSL achieves excellent results when multiple conditionally independent views of each sample are available, and each view is able to support accurate classification on its own.
- We explored if H&E slides can be color deconvoluted to Hematoxylin (H) and Eosin (E) stain images to fulfill co-training's view requirements.
- A novel contrastive co-training model with H and E views was tested on clear cell renal cell carcinoma (ccRCC) dataset and prostate cancer dataset.
- Our co-training model always has the best performance over other state-of-the art SSL methods in both datasets, including consistency regularization, MixMatch and FixMatch.

Stain Based Contrastive Co-training Model

- Stain separation: $\begin{bmatrix} H \\ E \end{bmatrix} = \begin{bmatrix} 1.838 & 0.034 & -0.760 \\ -1.373 & 0.772 & 1.215 \end{bmatrix} \begin{bmatrix} \log_{10} 255/R \\ \log_{10} 255/G \\ \log_{10} 255/B \end{bmatrix}$
- Contrastive loss: $\mathcal{L}_{c.t.}(x_i) = \max(\| f_H(x_i) - f_E(x_i) \|_2 - \| f_H(x_i) - f_E(x_k) \|_2 + m, 0)$
- Total loss: $\mathcal{L} = \sum_{x_j \in L} y_j \log \hat{y}_j + \lambda \sum_{x_i \in L \cup U} \mathcal{L}_{c.t.}(x_i)$



Procedure:

1 Separate original RGB H&E image into Hematoxylin (H) stain image and Eosin (E) stain image using color deconvolution.

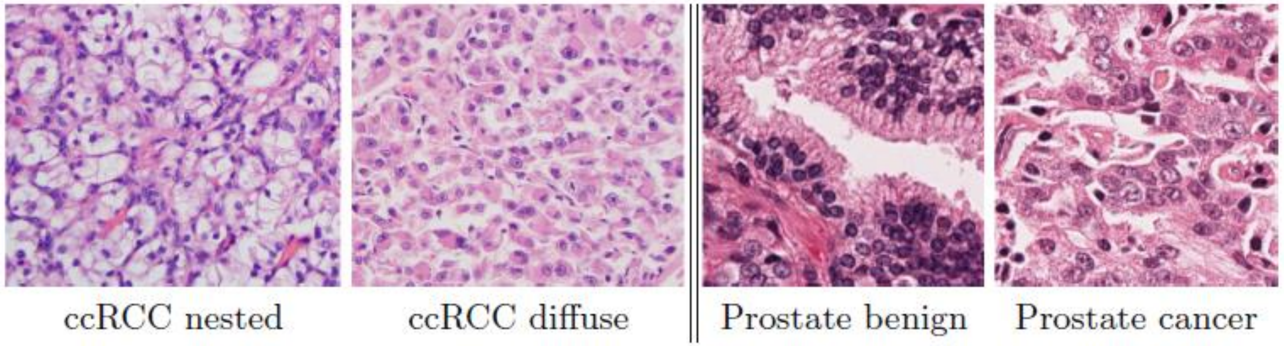
2 Generate features for H-image(F_H) and E-image(F_E) separately.

3 Calculate contrastive loss

4 Element-wise average $F_H(x_i)$ and $F_E(x_i)$, calculate class probabilities and cross entropy.

Experiment settings

- In clear cell renal cell carcinoma (ccRCC) dataset:
  - ✓Histologic growth pattern (HGP) tiles were cropped from expert annotated polygons from 53 WSIs.
  - ✓Tiles from same patient are only in same set.
  - ✓10% tiles set as labeled data from training polygons in SSL.
  - ✓Performed nested vs. diffuse (non-nested) classification and compared with other SSL models.
- In prostate cancer dataset:
  - ✓5% tiles set as labeled data from training polygons in SSL.
  - ✓Performed benign vs. cancer classification and compared with other SSL models.
- Randomly selected labeled data 5 times to calculate mean and standard deviation of classification accuracies.



Examples of ccRCC and prostate gland tiles

Experiment results

- We compared proposed co-training model with H and E views to a baseline ResNet18 model that uses RGB H&E images as input, as well as other state-of-the-art SSL methods.
- Test accuracy is recorded at epoch with best validation performance.
- The approaches were compared under two settings: using 100% of the available labeled tiles in training set for supervised learning and using only a subset (10% in ccRCC, 5% in prostate) of the available tiles as labeled data for supervised learning. Unlabeled data is also used in SSL.
- Our co-training model has the best performance in both datasets.

ccRCC Model	Test Accuracy	Prostate Model	Test Accuracy
100% label RGB ResNet	$84.8 \pm 2.4\%$	100% label RGB ResNet	$77.5 \pm 2.5\%$
100% label H/E co-train	$\mathbf{92.0 \pm 2.6\%}$	100% label H/E co-train	$\mathbf{79.1 \pm 2.0\%}$
10% label RGB ResNet	$76.9 \pm 5.9\%$	5% label RGB ResNet	$73.4 \pm 1.0\%$
10% label RGB consis	$86.8 \pm 3.3\%$	5% label RGB consis	$74.7 \pm 1.3\%$
10% label RGB MixMatch	$85.9 \pm 5.7\%$	5% label RGB MixMatch	$73.7 \pm 5.0\%$
10% label RGB FixMatch	$88.3 \pm 3.8\%$	5% label RGB FixMatch	$78.2 \pm 3.8\%$
10% label H/E co-train	$\mathbf{92.3 \pm 2.1\%}$	5% label H/E co-train	$\mathbf{78.7 \pm 1.9\%}$

Verify Co-training's assumptions

- Explore whether **each view**(H-view, E-view) is able to **support accurate classification on its own**, which is required in co-training assumption.
- Used single H-view image or E-view image as ResNet's input for testing in ccRCC dataset. (100% or 10% training samples used)
- Results show that each view is able to support classification on its own.

Model	Accuracy	Model	Accuracy
100% label H ResNet	$79.4 \pm 3.7\%$	10% label H ResNet	$73.5 \pm 4.0\%$
100% label E ResNet	$94.0 \pm 1.4\%$	10% label E ResNet	$82.3 \pm 7.0\%$

- Explore whether H-view and E-view are more independent than R, G, B channels since co-training requires more **conditionally independent views**.
- Coefficient of determination(R^2) of image mapping between various channels on ccRCC validation set was calculated to show independence between views. (Less value means more independence and less correlation)
- Results show H-view and E-view are more independent than R, G, B views.

Experiments	R^2 value	Experiments	R^2 value
$H \Rightarrow E$	0.5223	$E \Rightarrow H$	0.4613
$R \Rightarrow G$	0.8464	$G \Rightarrow R$	0.7833
$R \Rightarrow B$	0.8207	$B \Rightarrow R$	0.7713
$G \Rightarrow B$	0.8522	$B \Rightarrow G$	0.8824

Ablation study on ccRCC

- We also conducted ablation studies to separately analyze the role of the contrastive loss and the H and E channel selection in terms of classification accuracy on ccRCC.
- Omitting the contrastive loss from training while using the H and E channel inputs lowered the accuracy from 92.0±2.6% to 84.7±5.2% for 100% labeled data and from 92.3±2.1% to 78.7±8.0% for 10% labeled data.
- We ran ResNet and co-training models by taking only 2 channels from R, G, B as input with 10% labeled data in training, then calculate test set accuracy. Results of ResNet and co-training are approximately the same this time, which is expected considering the higher level of dependence among RGB channels.
- These observations suggest that the benefit of the proposed model is due to the contrastive co-training loss applied to the H and E view inputs.

Model	Accuracy	Model	Accuracy	Model	Accuracy
RB ResNet	$77.5 \pm 6.6\%$	RG ResNet	$80.2 \pm 6.4\%$	GB ResNet	$78.4 \pm 9.7\%$
R/B co-train	$78.2 \pm 4.5\%$	R/G co-train	$79.8 \pm 5.6\%$	G/B co-train	$76.6 \pm 7.3\%$

Conclusion

- The experiments show that our model outperforms all other tested state-of-the-art semi-supervised learning methods in both datasets.
- The experiments prove that the benefit of our model is due to contrastive co-training loss on hidden features combined with more independent H and E views as input.

Acknowledgements

- We are grateful for the support of the Computational Oncology Research Initiative (CORI) at the Huntsman Cancer Institute, University of Utah.
- We also acknowledge support of ARUP Laboratories and the Department of Pathology at University of Utah.
- Paper link: https://arxiv.org/abs/2206.12505
- Code: https://github.com/BzhangURU/Paper_2022_Co-training

Thanks!