

Learning to Evaluate Color Similarity for Histopathology Images using Triplet Networks

Anirudh Choudhary, Hang Wu, Li Tong, May D. Wang

Department of Biomedical Engineering, Georgia Institute of
Technology and Emory University

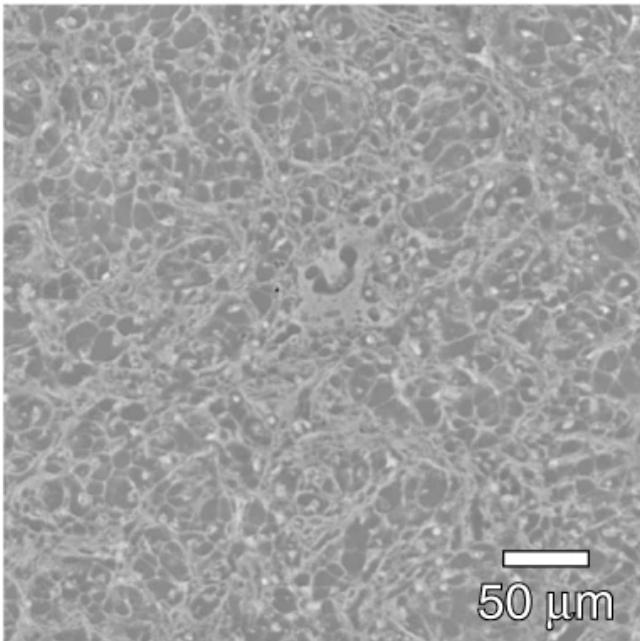
ACM-BCB 2019



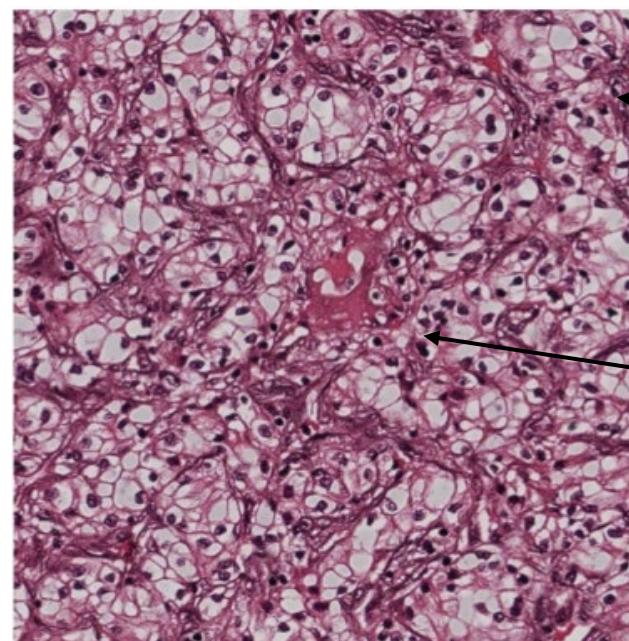
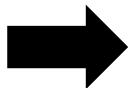
Background

Histopathology: Examination of microscopic tissue images

- Transparent tissue samples colored using chemical dyes that bind selectively
- Staining helps in better visualization and analysis of tissue morphology



Original Tissue Sample



H&E Stained

Eosin (Pink)
Fibres, Muscle Cells,
Intracellular membranes

Hematoxylin (Blue)
Nuclei

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory

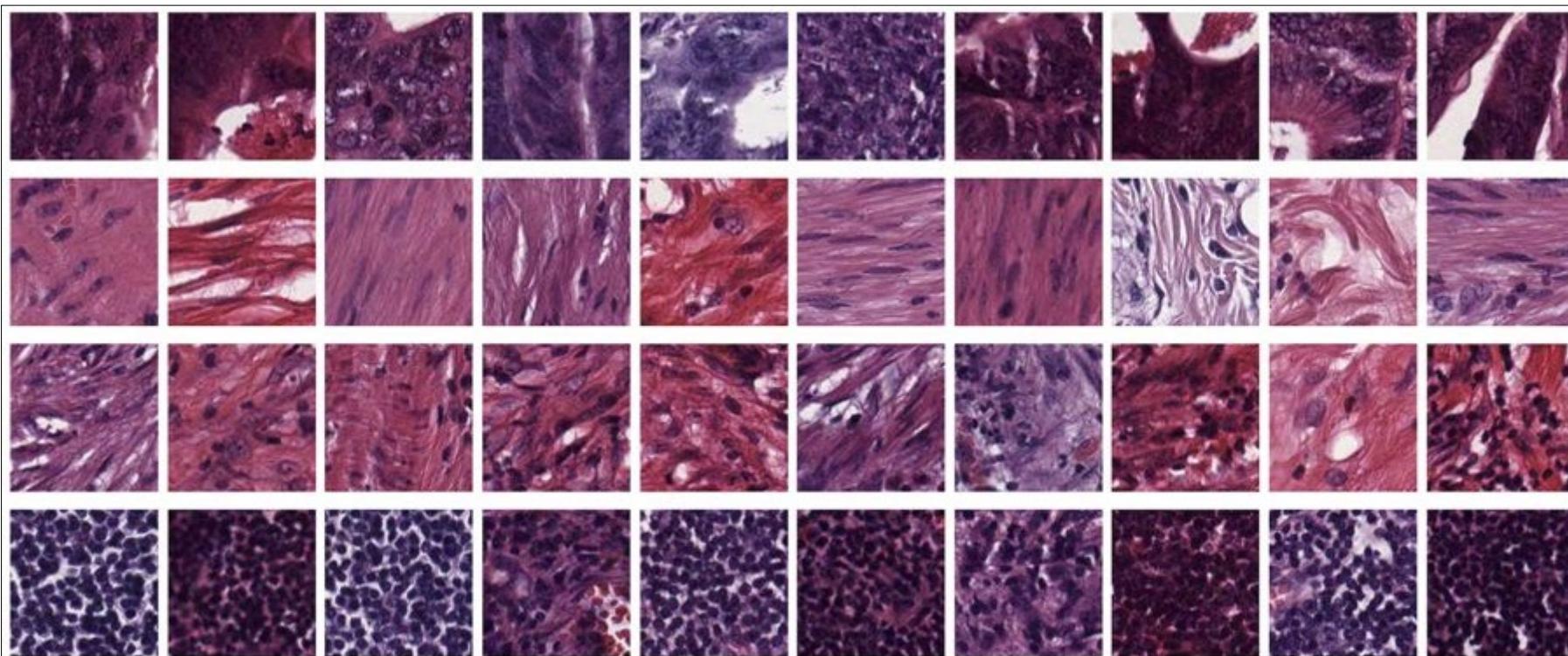


EMORY
UNIVERSITY

Background

Challenge: Stain color variance across laboratories/subjects

- Varying dye composition
- Different staining procedures
- Different scanners



Colorectal
cancer dataset

ACM-BCB 2019



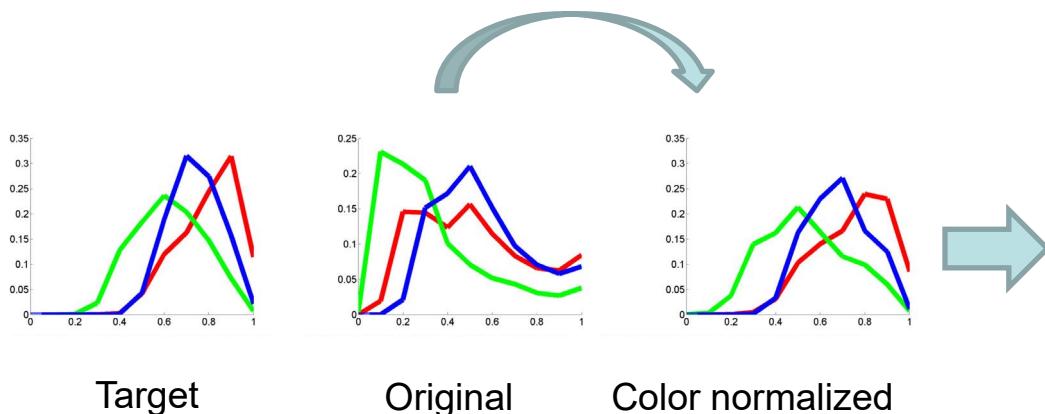
Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



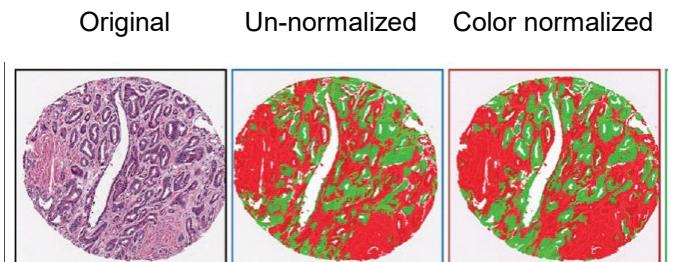
EMORY
UNIVERSITY

Stain Normalization



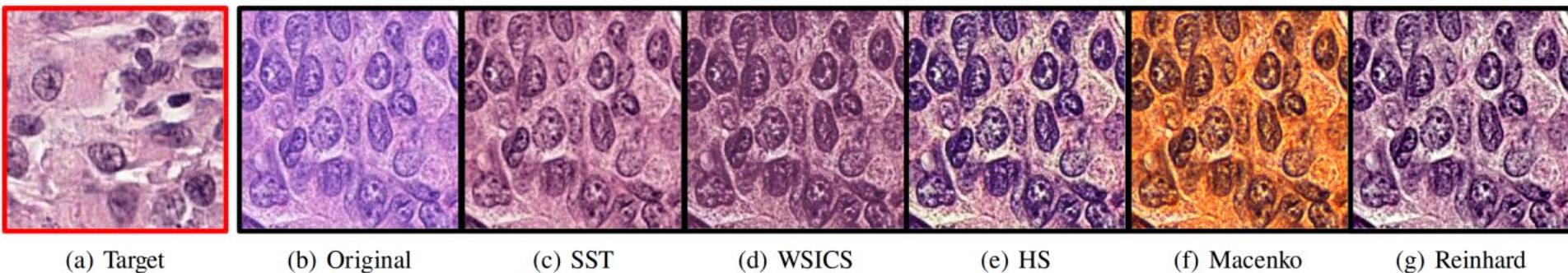
Models	Train	MITOSIS Val.	Test
Without Normalization	96.4±0.1	74.9±0.1	71.5±0.3
Grayscale	87.1±0.1	83.1±0.1	83.0±0.1
Augmentation	97.4±0.1	80.5±3.0	85.9±3.4
Color-Matching [9]	96.6±1.0	73.9±1.0	70.1±0.1
\mathcal{L}_{adv}	96.1±0.2	74.9±0.1	78.8±0.2
$\mathcal{L}_{adv} + \mathcal{L}_r$	96.1±0.2	75.0±0.3	85.9±0.1
$\mathcal{L}_{adv} + \mathcal{L}_r + \mathcal{L}_c$	95.1±0.1	76.2±0.3	90.0±0.1

Classification (Mitosis Detection)¹



Nuclei Segmentation²

Varying Performance of Stain Normalization Methods*



Quantitative evaluation of stain normalization becomes a key quality control step for downstream analysis

1. A. Bentaieb and G. Hamarneh, "Adversarial Stain Transfer for Histopathology Image Analysis," in *IEEE Transactions on Medical Imaging*, 2018
2. Sethi et. al., "Empirical comparison of color normalization methods for epithelial-stromal classification in H and E images." in *J Pathol Inform* 2016

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Quantitative Evaluation

Various metrics have been proposed/used in stain normalization studies

Color evaluation-based

Estimate color distributions for H & E stained regions
- Angle between color planes
- Normalized Median Intensity
- Coefficient of Variation

Histogram distance
(EMD, KL-Divergence)
Image L_2 distance

Perceptually motivated

Peak-Signal-to-Noise-Ratio
Structural Similarity Index (SSIM)
QSSIM, CW-SSIM

Human annotation	✓	X	X
Scalable	X	✓	✓
Color + Semantic Structure	✓	X	✓ (Limited)
Structural Changes	X	X	✓ (Limited)

Lack of standard metric : Limits scale of evaluation and benchmarking

1. Macenko et. al., 'A Method for Normalizing Histology Slides for Quantitative Analysis. Proceedings', ISBI 2009
2. Vahadane et. al., 'Structure-preserving color normalization and sparse stain separation for histological images', IEEE transactions on medical imaging, 2016.
3. Zanjani et. al., "Histopathology stain-color normalization using deep generative models", MIDL 2018
4. Zhou et. al. "Image Quality Assessment: From Error Visibility to Structural Similarity." 2004

ACM-BCB 2019



Georgia Institute
of Technology

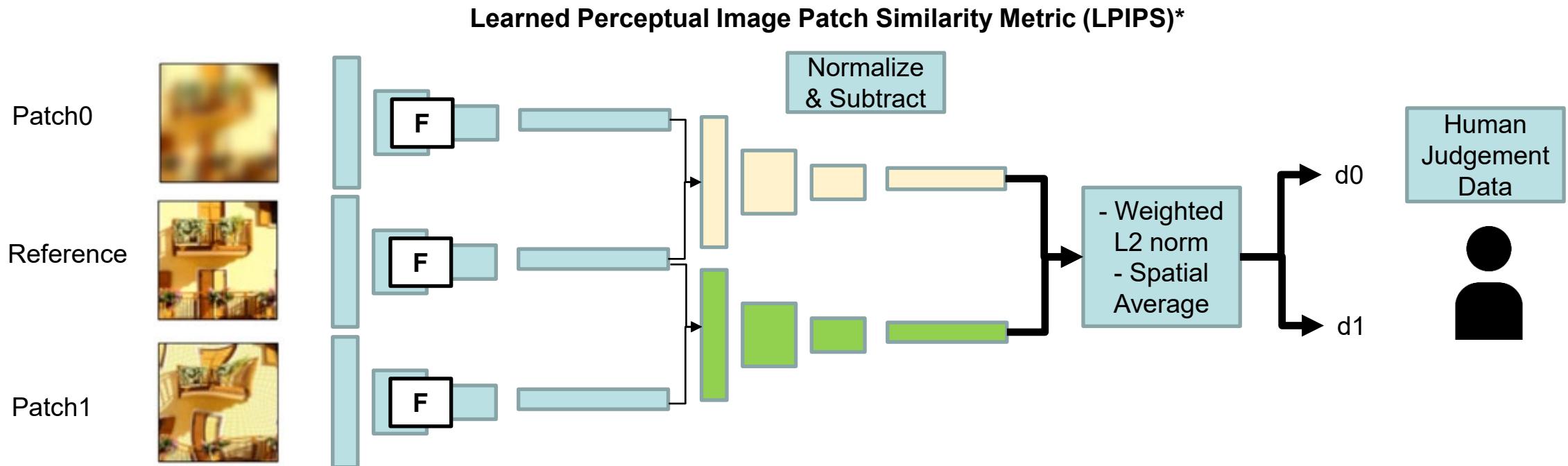
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Inspiration

Internal activations of deep convolutional networks correspond well with human perception



Challenge

Human perceptual judgement data
unavailable for histopathology images

Self-Supervised Learning

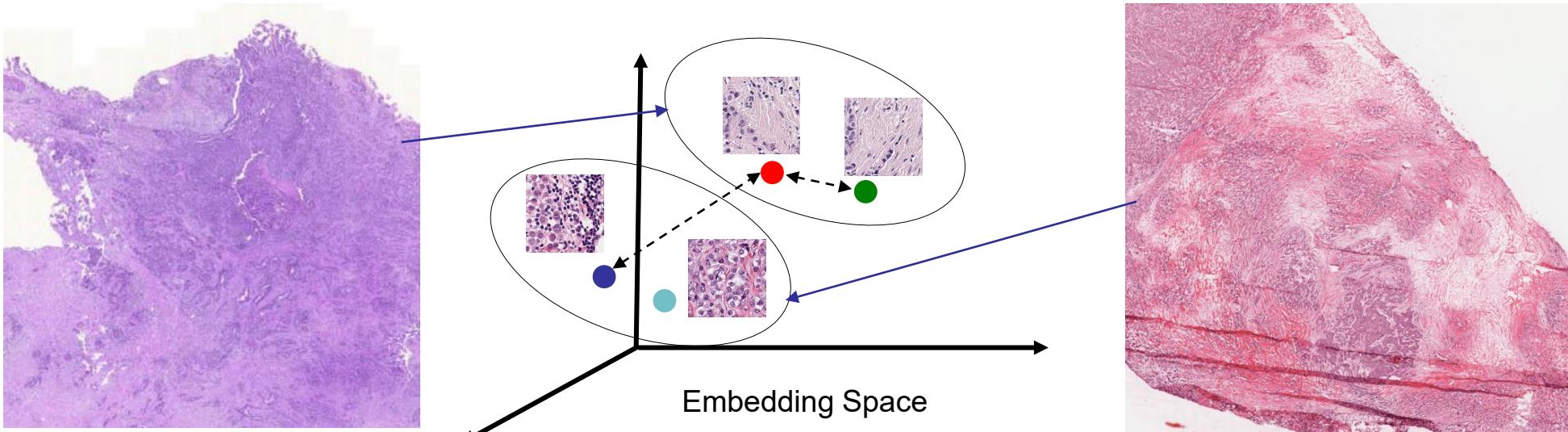
*Zhang et. al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric", CVPR 2018

ACM-BCB 2019



Self-supervised Embedding Learning

- **Objective:** Embed images into lower dimensional vectors with perceptually closer images having lower Euclidean distance in embedding space
- **Idea:** WSI samples for a patient/laboratory should be perceptually closer compared to different patients/laboratories. Each WSI represents a class.



- **Deep Metric Learning**
 - Triplet Network: Focuses on relative dissimilarity between disjoint classes

ACM-BCB 2019



Georgia Institute
of Technology

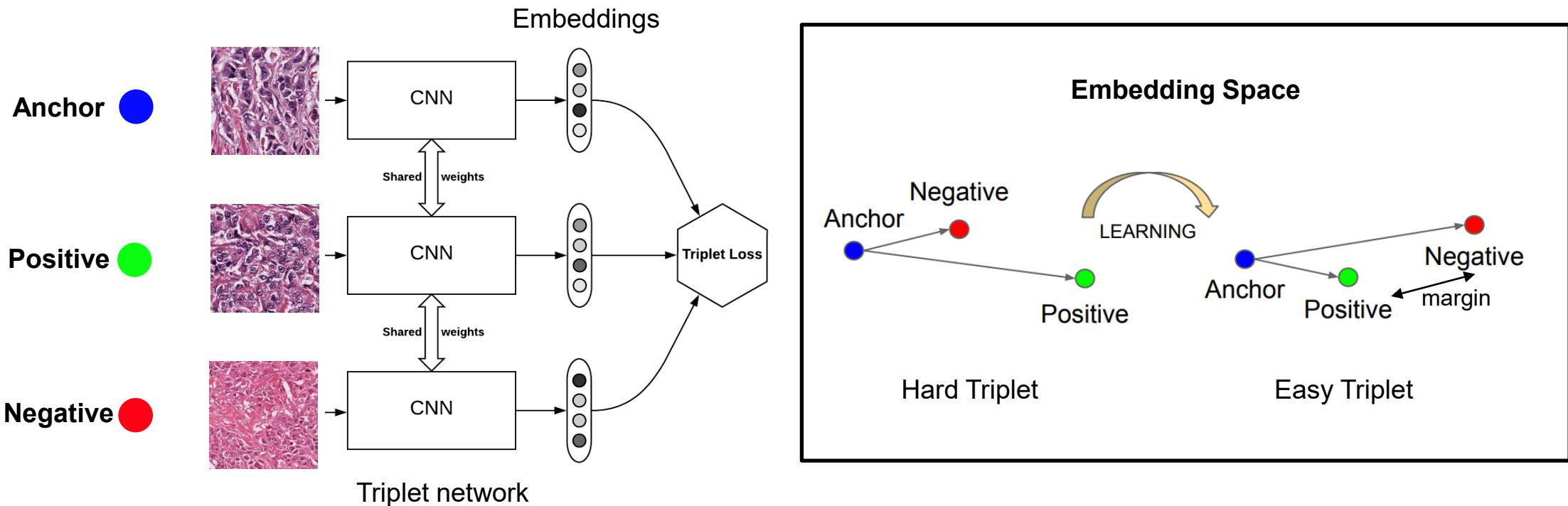
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Triplet Network

- Extensive use in human recognition and image retrieval tasks
- Doesn't enforce the embedding of each class to collapse into small clusters



Elad Hoffer and Nir Ailon. 2015. Deep Metric Learning Using Triplet Network. Lecture Notes in Computer Science (2015), 84–92

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Triplet Loss and Batch Sampling

- Triplet Loss
 - Utilize softplus-version of triplet loss

$$loss(a, p, n) = \max(d_+ - d_- + margin, 0) \quad d_+ = L_2(f_{\theta}(a), f_{\theta}(p)) \quad d_- = L_2(f_{\theta}(a), f_{\theta}(n))$$



$$\text{softplus version : } loss = \ln(1 + e^{(d_+ - d_-)})$$

- Batch Sampling (Triplet Mining)
 - Batch-Hard
 - Computes loss using hardest negative and hardest positive samples in the batch
 - Batch-All
 - Computes loss on all valid triplets forcing images from same class to be drawn closer

Alexander Hermans, Lucas Beyer, and Bastian Leibe. 2017. In Defense of the Triplet Loss for Person Re-Identification. arXiv:cs.CV/1703.07737

ACM-BCB 2019



Georgia Institute
of Technology

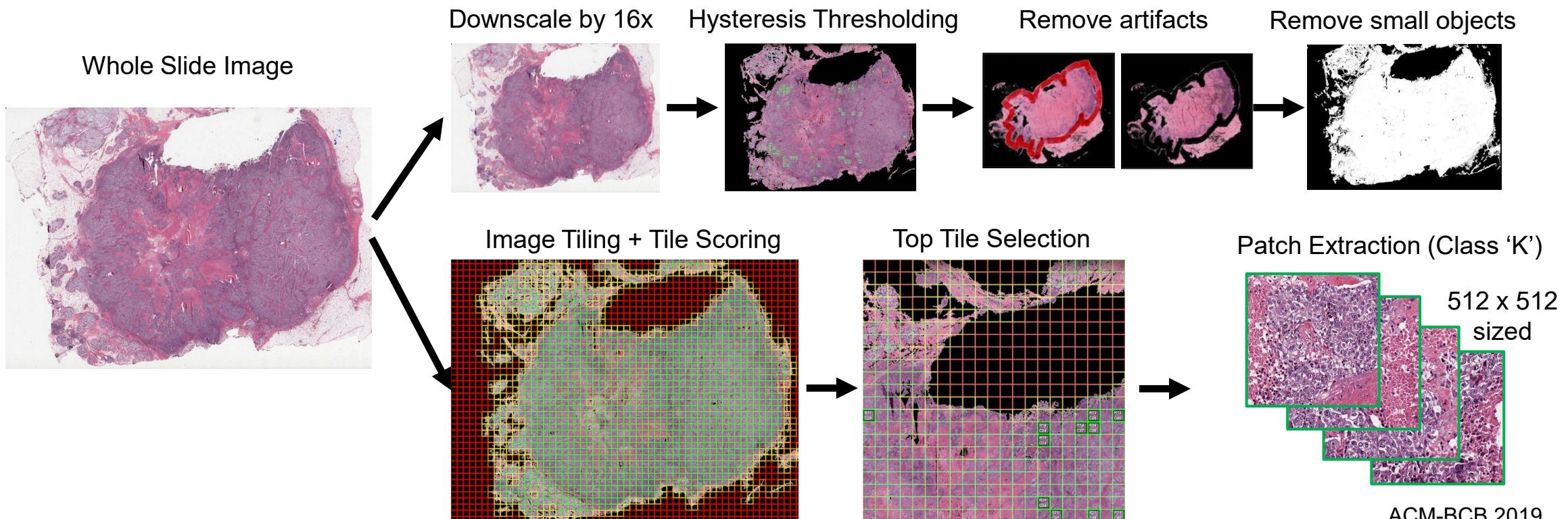
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Pre-Processing Pipeline

- Define classes: Sample upto 1000 patches from each whole slide image
- Tile selection
 - Tile Scoring: Ensures balanced representation of tissue components
 - Select tiles with more than 90% tissue



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Pre-Processing: WSI Tile Scoring

$$score = tissue_percent^2 * \log(1 + color_{factor} * saturationvalue_{factor} * quantity_{factor})$$

$$score_{0\text{ to }1} = 1.0 - \frac{1.0}{1.0+score}$$

$color_{factor}$: Favours hematoxylin stained regions over eosin stain

$saturationvalue_{factor}$: Real tissue should have broad HSV saturation and value

$quantity_{factor}$: Higher amount of tissue is better

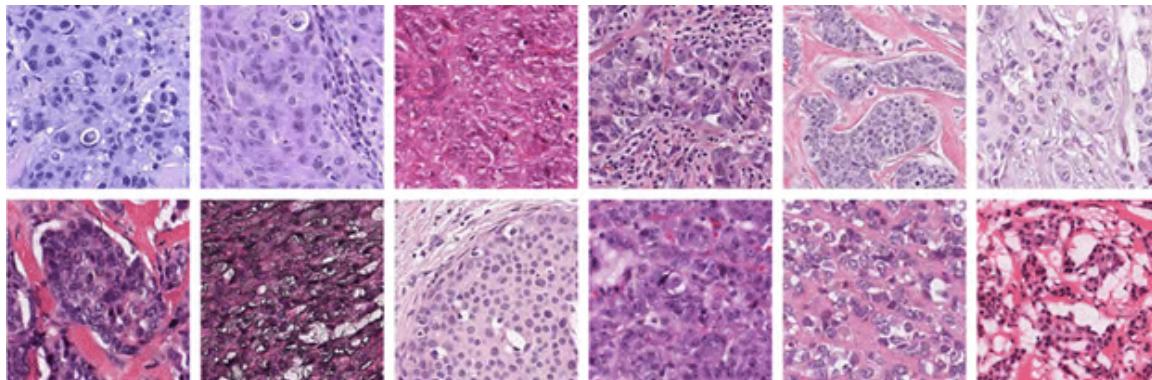
Each factor ranges between '0' and '1'



Datasets

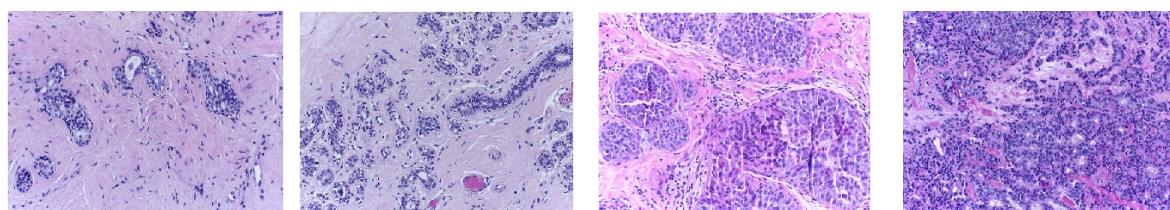
Training (210 WSI; 200,000 image samples)

Tumor Proliferation Assessment Challenge (TUPAC)



200 WSI randomly sampled (multiple laboratories & cancer condition)

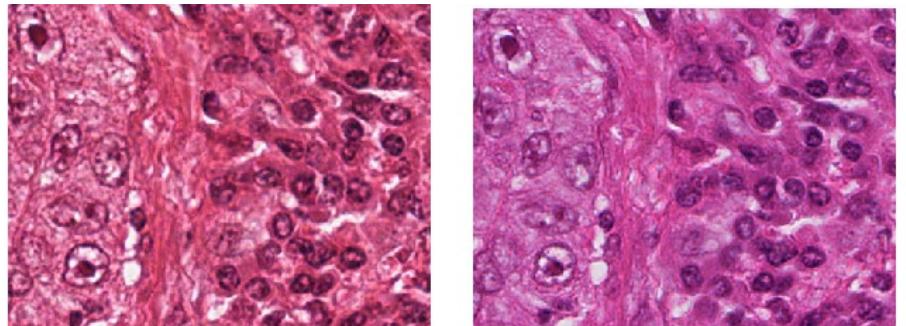
BACH(ICIAR Grand Challenge 2018)



30 high resolution breast cancer histology images

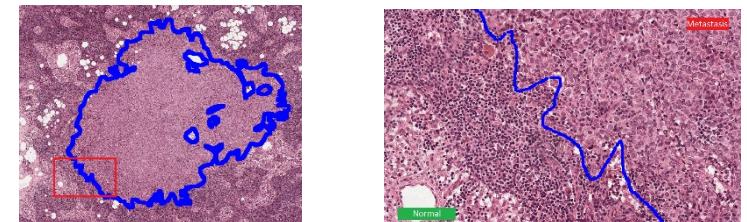
Test (37 WSI; 12,900 Samples (20% query))

MITOS – ATYPIA (ICPR 2014)



(a) Aperio scanner
Each WSI acquired using 2 microscopes (multiple patients)

CAMELYON17



Breast cancer metastases (5 labs; 3 subjects each)

ACM-BCB 2019



Georgia Institute
of Technology

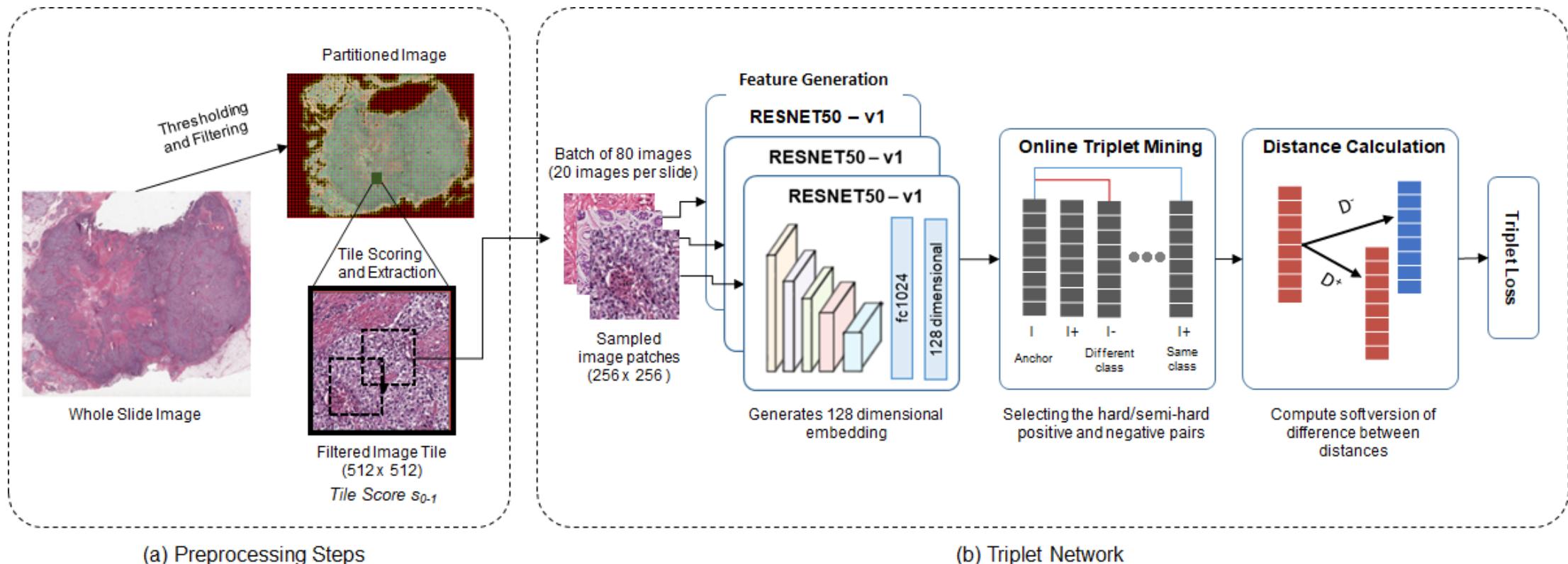
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Network Training

- Core architecture : ResNet50-v1 (pretrained)
- Input image downsampling (256 x 256) with standard data augmentation (flip, crop)
- Batch-size of 80 images (20 images per class) ; Learning rate: 3×10^{-4} ; N = 30,000 iterations



ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Evaluation

Quantitative

- Given a query image, how does our metric perform in retrieving perceptually similar images?
 - Rank @ K :
$$\frac{\text{# of retrieved items @k that are relevant}}{\text{total # of relevant items}}$$

Qualitative

- How well does our embedding capture perceptual differences across subjects/laboratories?
- How well does our embedding evaluate color normalization tasks?
- Visually, how similar are 10 nearest neighbor patches to a query image?



Retrieval Performance

Laboratory-based retrieval scores for
MITOS and CAMELYON17 datasets

Metric	Rank-1		Rank-5		Rank-10	
	MITOS	CAM	MITOS	CAM	MITOS	CAM
SSIM [10]	51.71	19.40	84.03	25.39	92.78	32.42
LPIPS [54]	47.72	20.44	72.81	54.69	80.99	72.14
TriNet-P (batch-hard)	100.00	85.02	100.00	97.13	100.00	98.18
TriNet-P (batch-all)	100.00	88.02	100.00	96.48	100.00	97.92

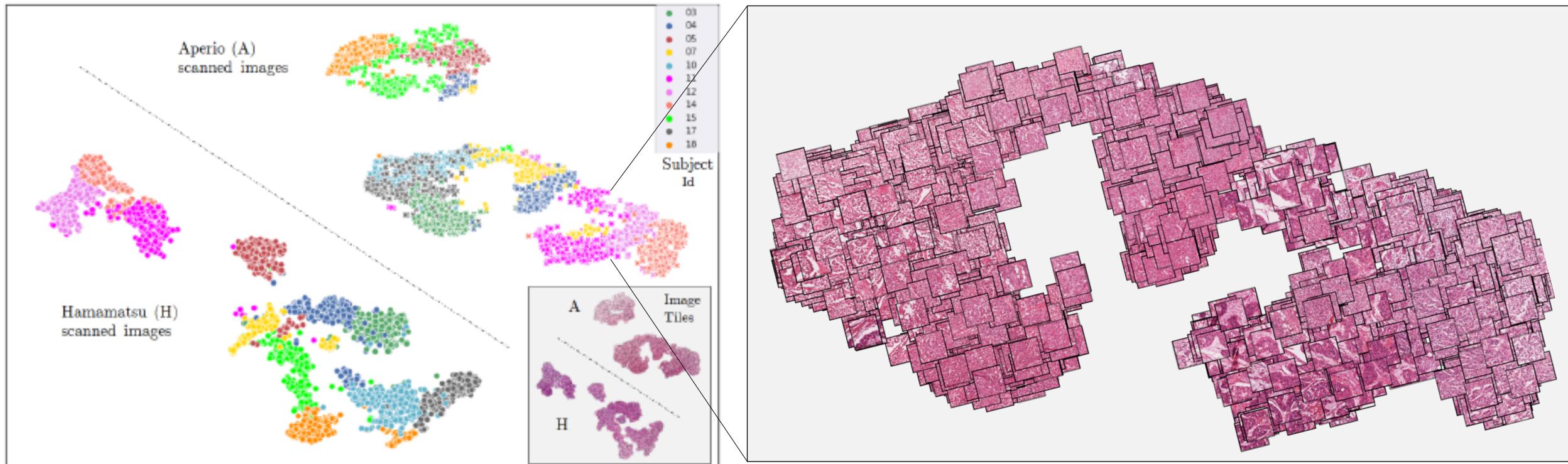
Subject-based retrieval scores for
MITOS and CAMELYON17 datasets

Metric	Rank-1		Rank-5		Rank-10	
	MITOS	CAM	MITOS	CAM	MITOS	CAM
SSIM [10]	4.37	7.16	12.74	9.64	19.77	13.28
LPIPS [54]	4.56	6.38	12.36	23.44	19.58	35.94
TriNet-P (batch-hard)	76.81	48.57	95.06	85.42	97.91	94.01
TriNet-P (batch-all)	84.03	58.59	97.53	87.63	98.67	94.66

TriNet-P outperforms unsupervised perceptual similarity metrics on image retrieval task



Visualizing Learned Embeddings



t-SNE plot : Embeddings for image patches generated from 11 subject tissues scanned using two microscopes – Aperio (A) and Hamamatsu (H) (MITOS Dataset)

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Stain Normalization (MITOS)

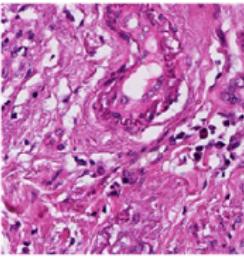
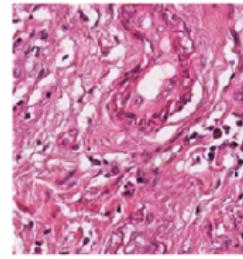
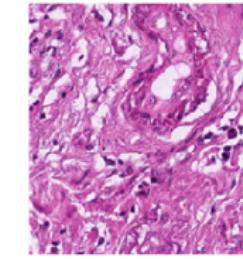
Reference	Source	Color-Normalized
		
SSIM	0.2566	0.2591
LPIPS	0.0977	0.0995
TriNet-P	0.6273	0.2549

Image embeddings capture the effectiveness of color normalization in MITOS dataset

ACM-BCB 2019



Georgia Institute
of Technology

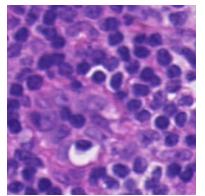
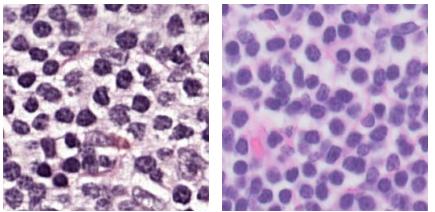
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

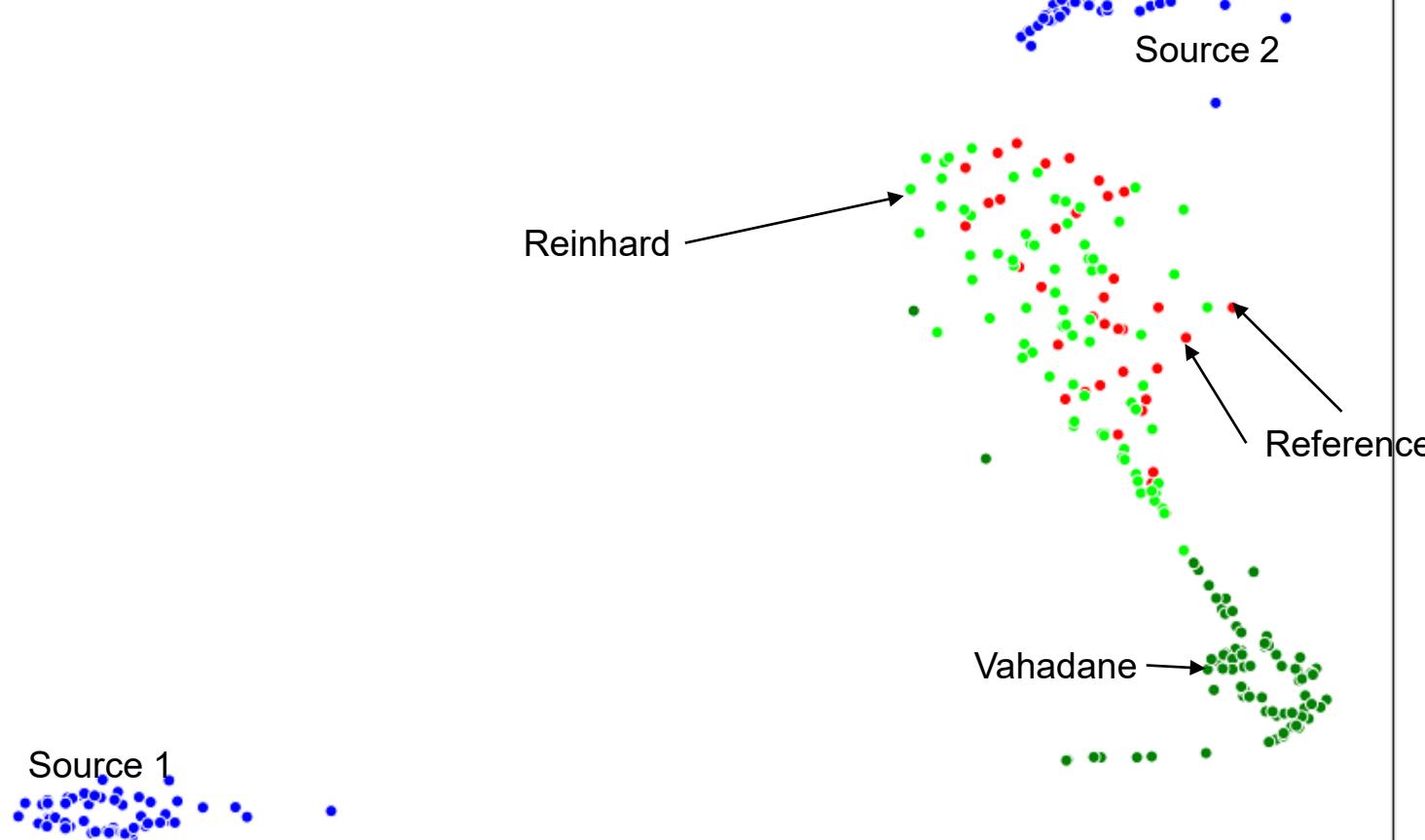
Stain Normalization (CAMELYON)

Source 1 Source 2

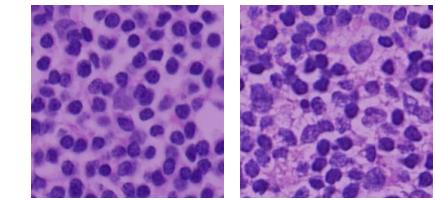
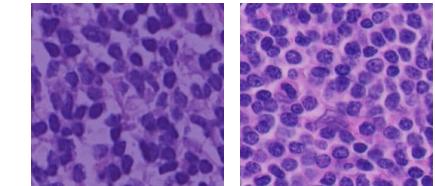


Reference

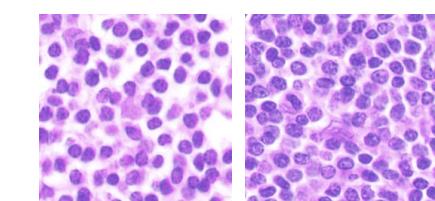
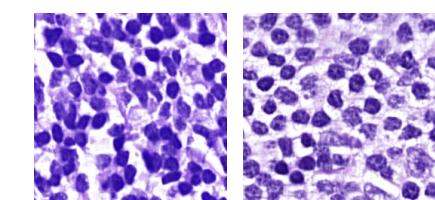
t-SNE Plot of Image Embeddings



Reinhard¹



Vahadane²



1. E. Reinhard, M. Adhikhmin, B. Gooch and P. Shirley, "Color transfer between images," in IEEE Computer Graphics and Applications, 2001

2. Vahadane et. al., 'Structure-preserving color normalization and sparse stain separation for histological images', IEEE transactions on medical imaging, 2016.

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory

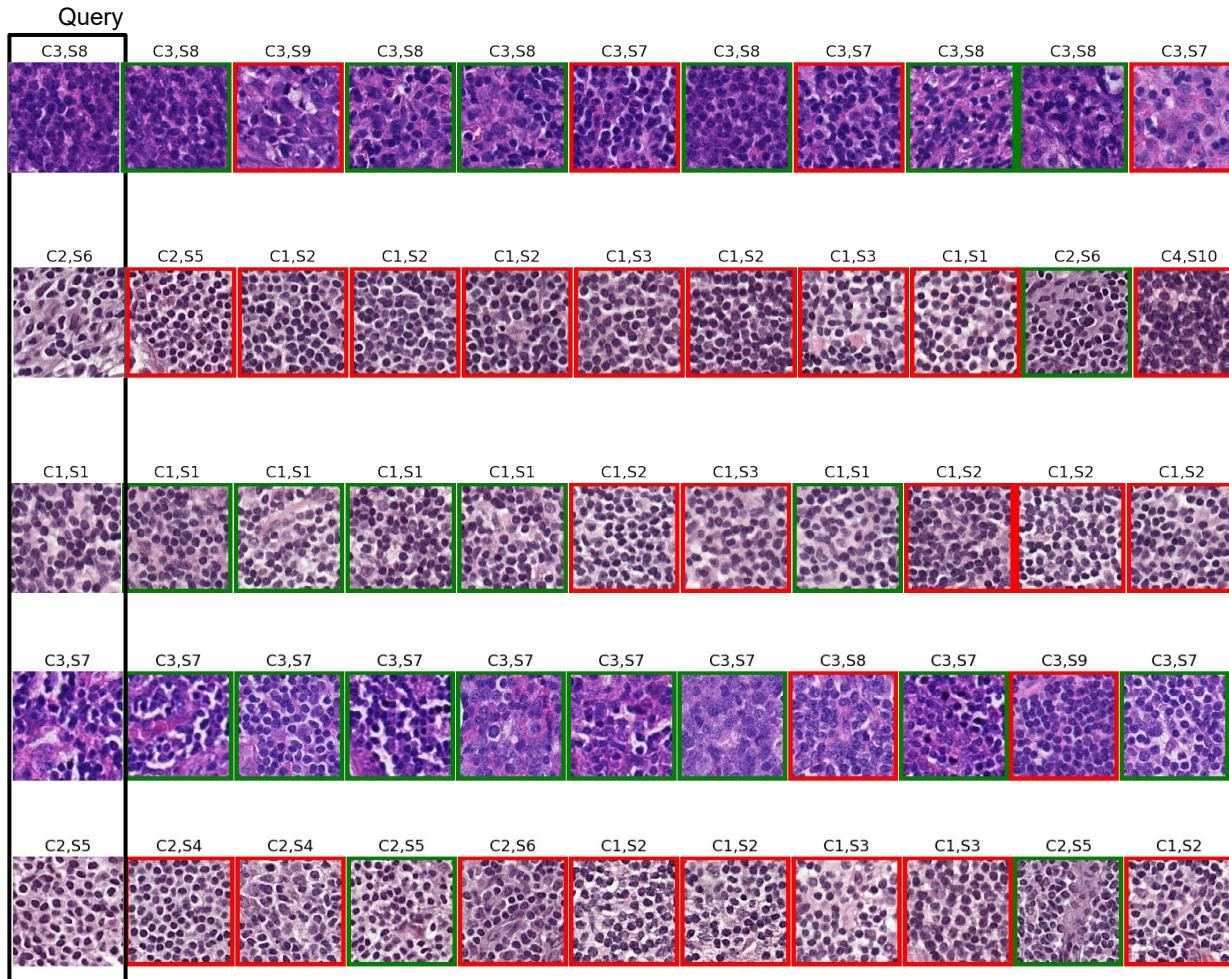


EMORY
UNIVERSITY

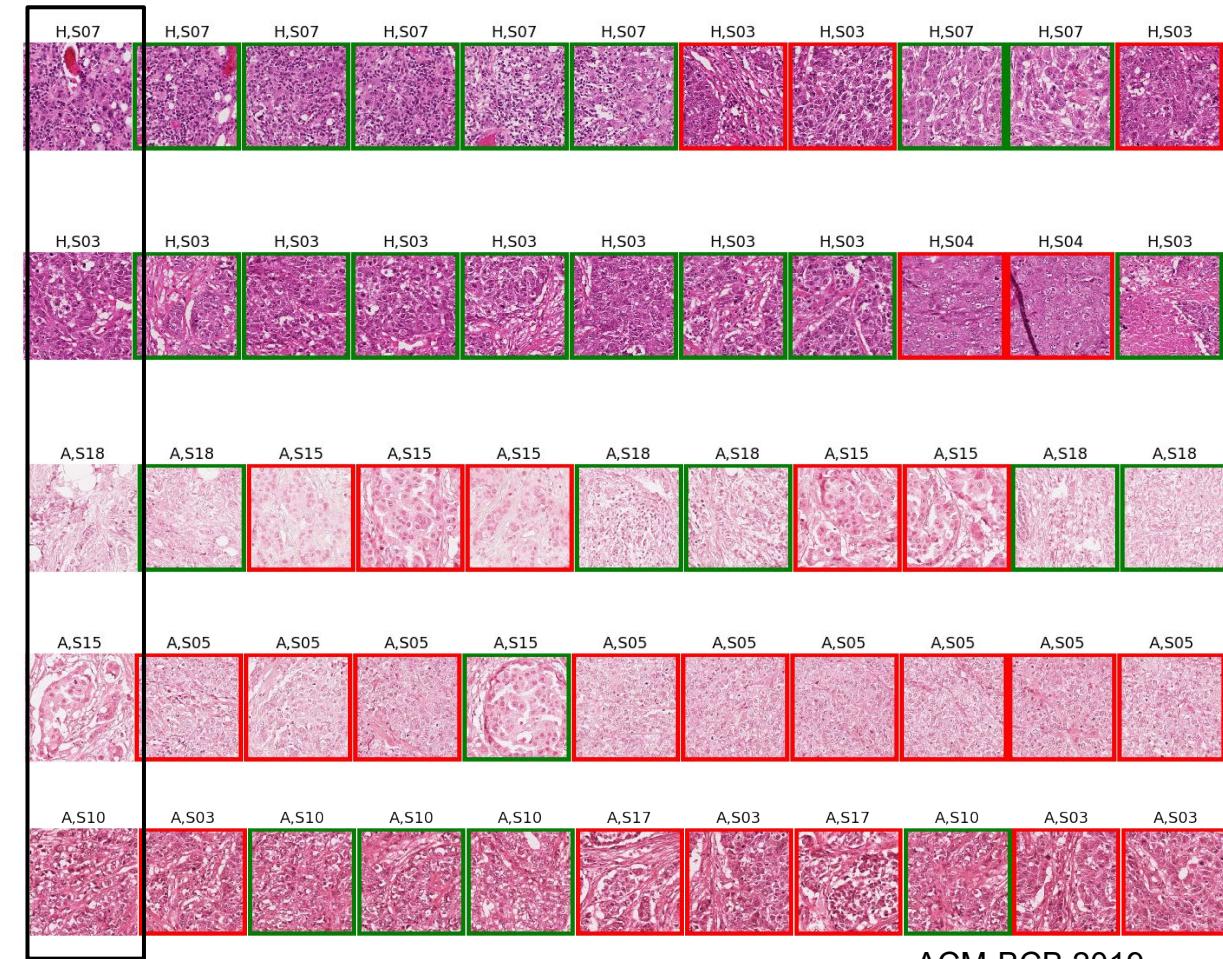
Qualitative Analysis

10 nearest patches retrieved from the test dataset for each query image

CAMELYON



MITOS



ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Summary

- Introduce **novel quantitative metric** for evaluating stain normalization based on deep embeddings
- Propose a **semi-supervised pipeline** for creating training dataset creation
- Euclidean distance margin based loss works well in capturing perceptual similarity
- Embedding distance aligns with stain normalization effectiveness
- **Future Work**
 - Incorporate into real-world pathological diagnosis pipeline
 - Improve pre-processing pipeline: Identify informative regions in whole slide image
 - Better Embeddings: Disentangled embedding
 - Loss Functions: Histogram distance-based loss, Angular loss, Siamese + Triplet Loss
 - Network Structure: Hierarchical triplet network

ACM-BCB 2019



Georgia Institute
of Technology

Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory



EMORY
UNIVERSITY

Thank You

This work is supported by funding from American Heart Association,
Giglio Cancer Research fund, Petit Institute Faculty Fellow
Fund, and Carol Ann and David Flanagan Faculty Fellow Fund

Email: achoudhary46@gatech.edu

BioMIB Lab: <https://miblab.bme.gatech.edu/>

ACM-BCB 2019