Deep Learning Towards Robustness in Medical Images

Varsha S
193079005

Outline

I. Robust Deep Learning Framework to address General Label Noise in Medical Imaging
   a) Motivation
   b) Contrastive Learning
   c) Algorithm
   d) Experiments and Results
   e) Future Work

II. Despeckling Ultrasound Images
   a) Motivation
   b) Experiments and Results
   c) Future Work

Robust Deep Learning Framework to address General Label Noise in Medical Imaging

Motivation

- Medical images are often subjectively labeled or weakly supervised, which introduces both closed-set and open-set label noise
- The low image quality due to tissue preparation or preservation artifacts also degrade the supervisory signal to a Deep Learning algorithm
- Deep learning (DL) models underperform when the quality of supervisory labels degrades as DL algorithm’s performance is largely determined by availability of labelled data
- Only few works have addressed open-set label noise and have given a unified treatment to both types of label noise

Motivation

- Medical images are often subjectively labeled or weakly supervised, which introduces both closed-set and open-set label noise
- The low image quality in such modalities due to tissue preparation or preservation artifacts can also degrade the supervisory signal to a DL algorithm
- Deep learning (DL) models underperform when the quality of supervisory labels degrades as D algorithm performance is largely determined by availability of labelled data
- Only few works have addressed open-set label noise and have given a unified treatment to both types of label noise
- *Goal* : Learn a scheme tolerant to both mislabeling and outliers

Problem Addressed

Classification of Histopathology images in presence of both ***closed-set*** *and* ***open-set*** *noise*

Problem Addressed

Classification of Histopathology images in presence of both ***closed-set*** *and* ***open-set*** *noise*

- Closed set noise : *mislabelled samples*
- Open Set noise : *out-of-distribution (OOD)* samples

Problem Addressed

*Classification of Histopathology images in presence of both **closed-set** and **open-set** noise*

- Closed set noise : *mislabelled samples*
- Open Set noise : *out-of-distribution (OOD)* samples

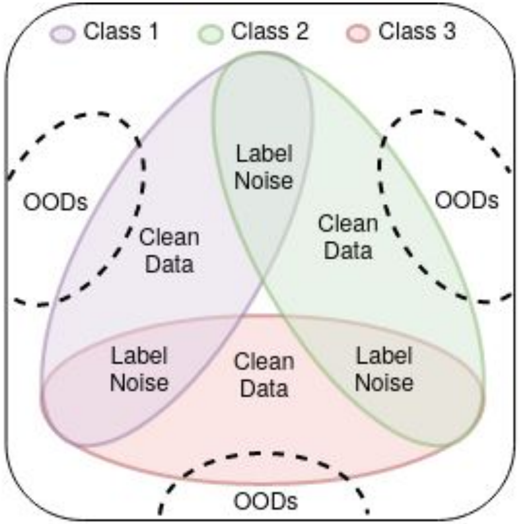


Figure : Ground truth visualisation for 3 clean classes, consisting open-set (label) noise and closed-set (OOD) noise

Proposed Method

- SimCLR based learning with three Stages
    1) Warm Up Phase
    2) Weight Calculation
    3) Classification Phase

Proposed Method

- SimCLR based learning with three Stages
    1) Warm Up Phase
    2) Weight Calculation
    3) Classification Phase

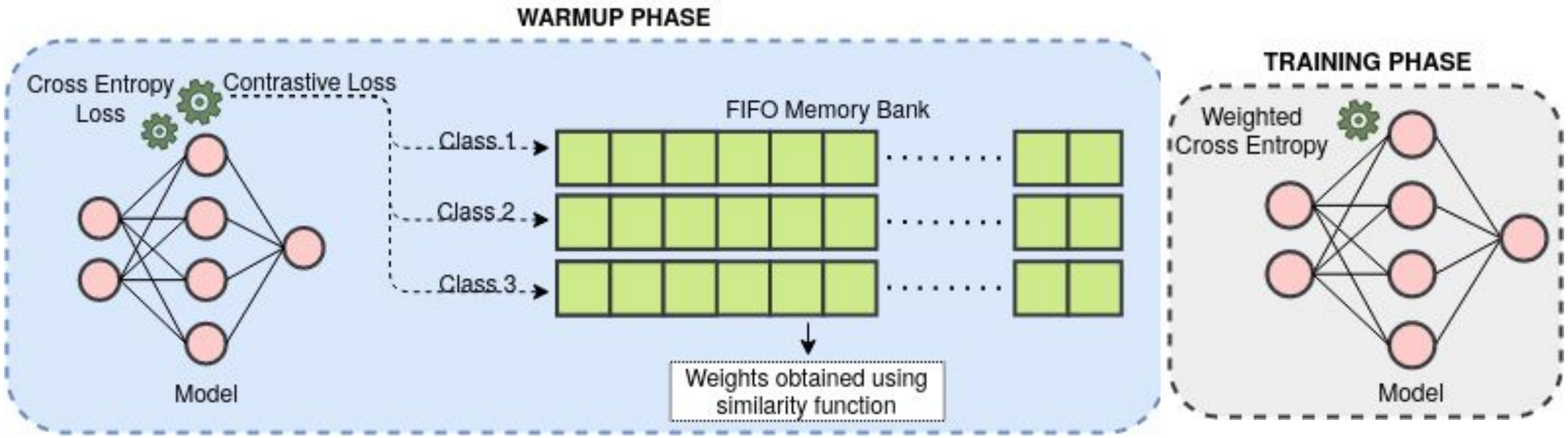


Figure : Overview of the proposed method

SimCLR : Simple Framework for Contrastive Learning

- *Contrastive Learning* :
  The main idea is to learn representations such that similar samples stay close to each other, while dissimilar ones are far apart.

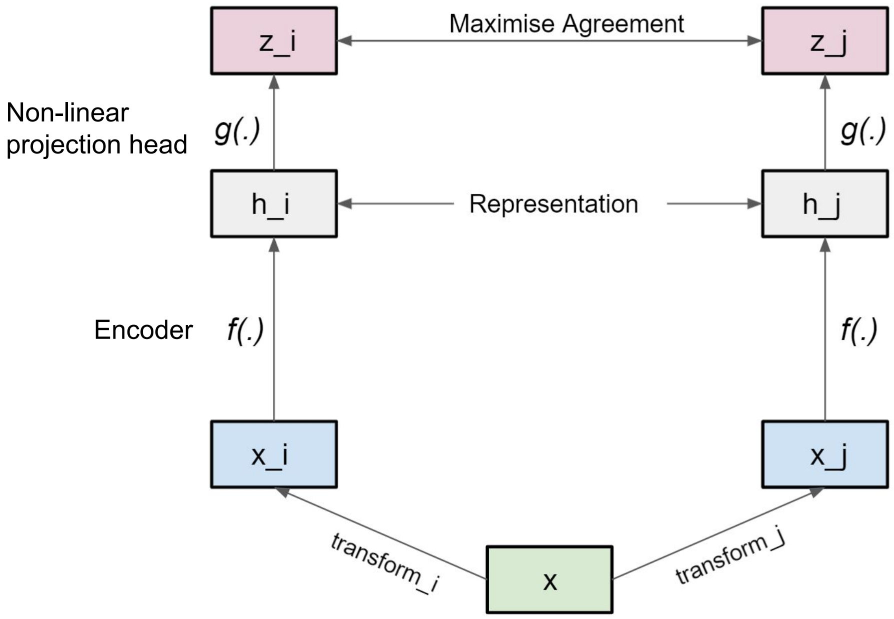


Figure : A simple framework for contrastive learning of visual representations

Inspired from - Ting Chen, et-al. "A Simple Framework for Contrastive Learning of Visual Representations". In ICML, 2020.

SimCLR : Simple Framework for Contrastive Learning

Task : maximize the similarity between the two representations z_i and z_j for the same image

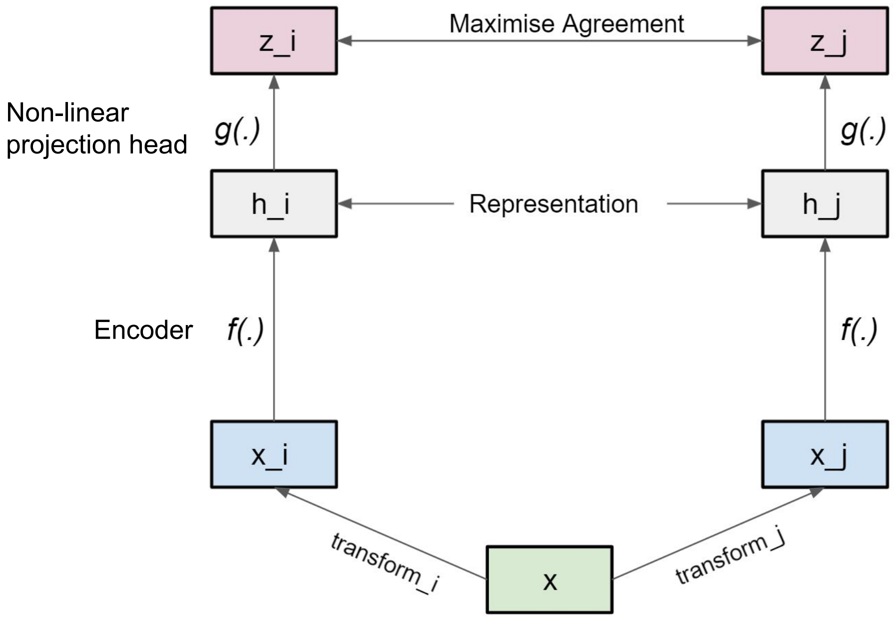


Figure : A simple framework for contrastive learning of visual representations

Inspired from - Ting Chen, et-al. "A Simple Framework for Contrastive Learning of Visual Representations". In ICML, 2020.

Contrastive loss Function

$$\mathcal{L}_{Con} = [d_p - m_{pos}]_+ + [m_{neg} - d_n]_+$$

$\mathcal{L}_{Con}$ attempts to make the distance between positive pairs d_p smaller than some margin m_{pos} and the distance between negative pairs d_n larger than some threshold m_{neg}

$[x]_+$ is $max(0, x)$
d_p - distance between positive pairs
d_n - distance between negative pairs
m_{neg} - threshold margin for negative pairs
m_{pos} - threshold margin for positive pairs

values of m_{pos} and m_{neg} are set to 1 and 0 respectively

Method : Warm Up Phase

- Histology-specific ResNet-18 SimCLR model as a backbone network
- Training with interleaved cross-entropy (to learn discriminate features) and contrastive loss training(to distinguish between clean and noisy samples)

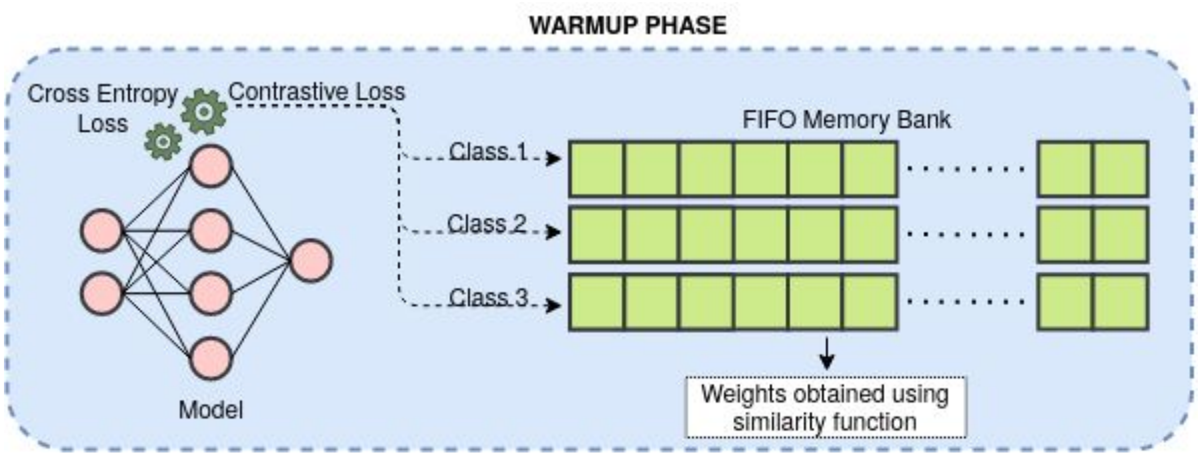
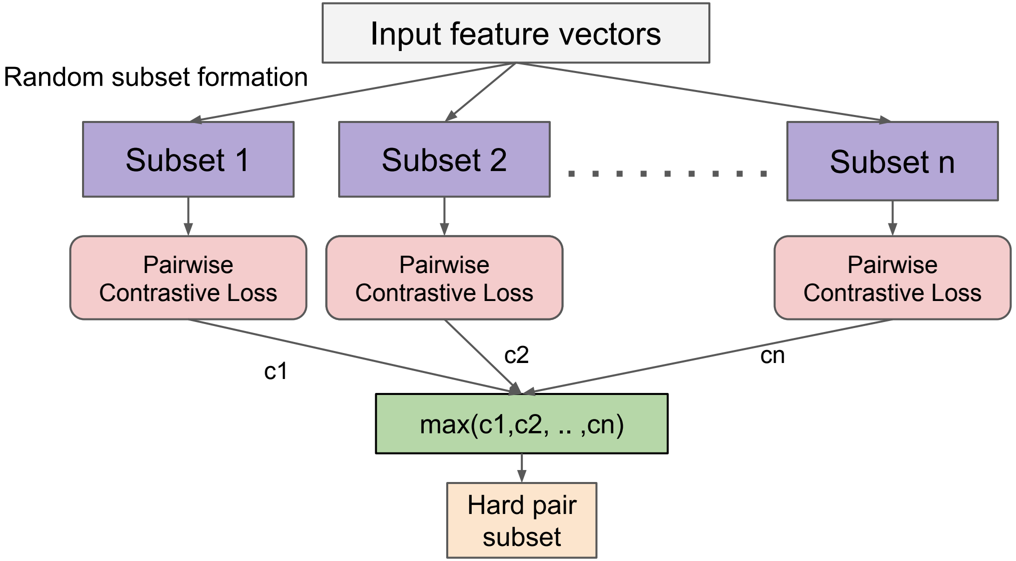


Figure : Warm Up Phase with weight calculation

Method : How is the memory bank formed



Miner function operation

Method : How is the memory bank formed

- Miner function : used to mine hard pairs
- Creates subsets of features of size “s”, finds hard pairs (subset with highest contrastive loss)
- Apart from these hard pairs, everything is added to memory bank

Method : Warm Up Phase (Memory Bank for Prototypes)

- FIFO Memory bank of fixed size per class is created by storing features of instances with least contrastive loss in a minibatch

Method : Warm Up Phase (Memory Bank for Prototypes)

- FIFO Memory bank of fixed size per class is created by storing features of instances with least contrastive loss in a minibatch

- K-medoids clustering for obtaining k prototypes per class.

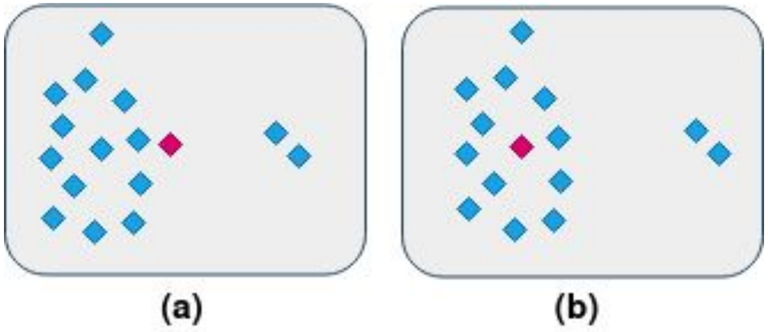


Figure : (a) shows mean in K-Means clustering, which is influenced by the outlier. (b) shows the medoid in K-medoids clustering which remains unaffected by the outliers.

Method : Weight Calculation Phase

- Weights are calculated by
  - averaging similarity scores of feature vector with all prototypes in memory bank
  - K-medoids - similarity scores of feature vector with a class prototypes of memory bank

$$S_{u,v} = \frac{u^T v}{||u||.||v||}$$

$$W_{u,v} = \frac{S_{u,v} - (S_{u,v})_{min}}{(S_{u,v})_{max} - (S_{u,v})_{min}}$$

u - feature vector of input
v - feature prototype of class corresponding to input
$S_{u,v}$ - cosine similarity between l_2 normalized u and v
$W_{u,v}$ - weights for the cross entropy loss in the range [0,1]

Method : Final Classification Phase

- Training proceeds with weight cross entropy as loss function
- Weights from previous step are used for weight cross entropy

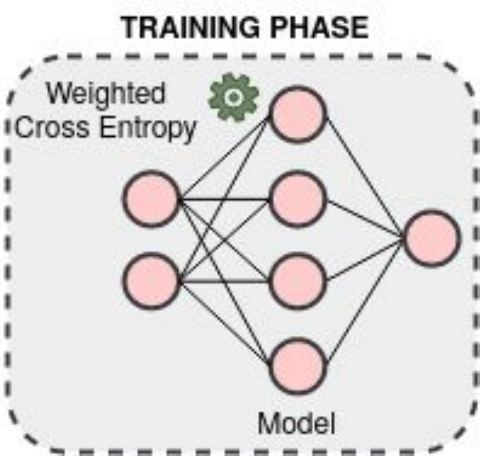


Figure : Classification with weighted cross entropy loss

Algorithm Proposed

Algorithm 1: Warm-up & Weight Calculation

Input: $\mathcal{B} = \{(x_0, y_0), (x_1, y_1)....(x_n, y_n)\}$: minibatch of size n in dataset $\mathcal{S}$;
$f(.)$: The deep learning framework;
Parameter: $\mathcal{M}$: Fixed size memory bank;
Warm-up Phase
foreach $\mathcal{B} \in \mathcal{S}$ **do**
    $\phi(x_i), \theta(x_i) \leftarrow f((x_i, y_i))$;
    $hardpairs \leftarrow miner(\phi_{\mathcal{B}}, L_{con})$;
    if $\phi(x_i) \notin hardpairs$ **then**
        $\mathcal{M}_{Y_i} \leftarrow \phi(x_i)$;
    end
    Calculate $L_{CE}(\theta(x_i), y_i)$ and update $f(.)$;
    Calculate $L_{con}(\phi(x_i))$ and update $f(.)$;
end
Calculate Weights
foreach $(x_i, y_i) \in \mathcal{S}$ **do**
    $\phi(x_i) \leftarrow f((x_i, y_i))$;
    $w_i \leftarrow cosinesimilarity(\mathcal{M}_{Y_i}, \phi(x_i))$
end

Datasets

Work Flow of Patch based Classification

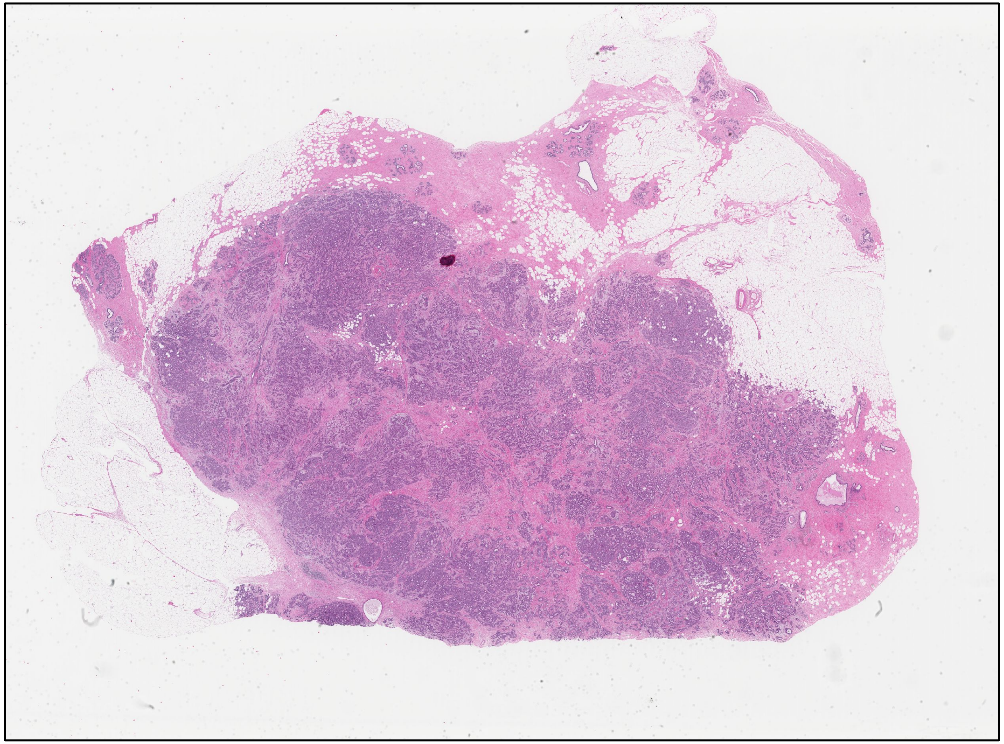


Figure : Unannotated Whole Slide image of breast cancer

Work Flow of Patch based Classification

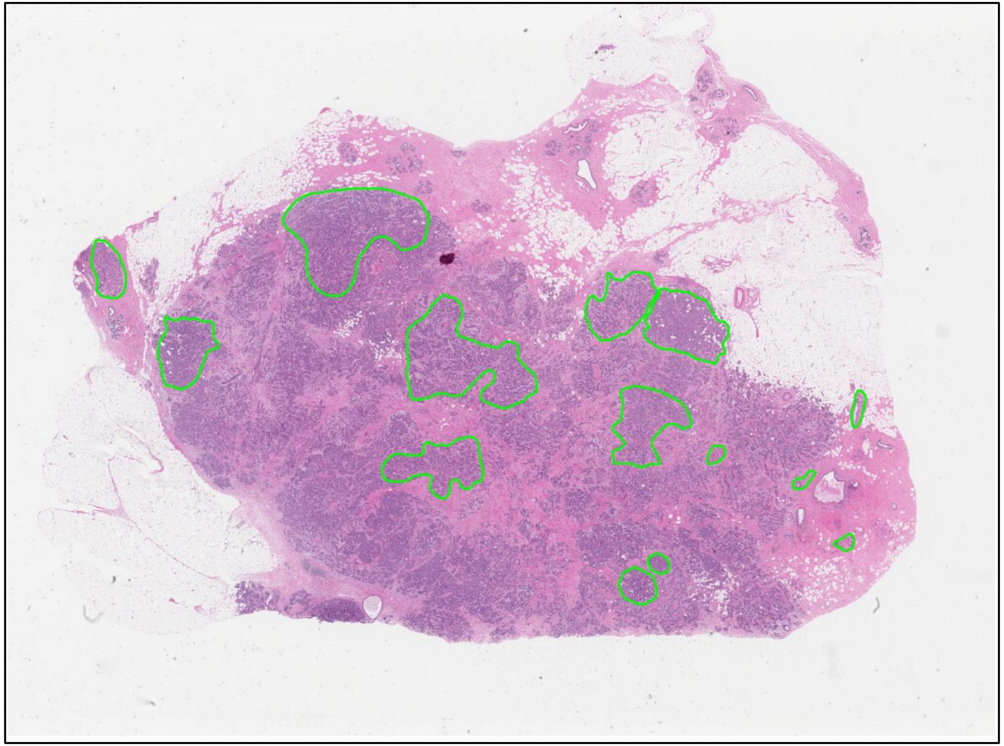


Figure : Annotated Whole Slide image of breast cancer

Dataset : BACH (breast cancer)

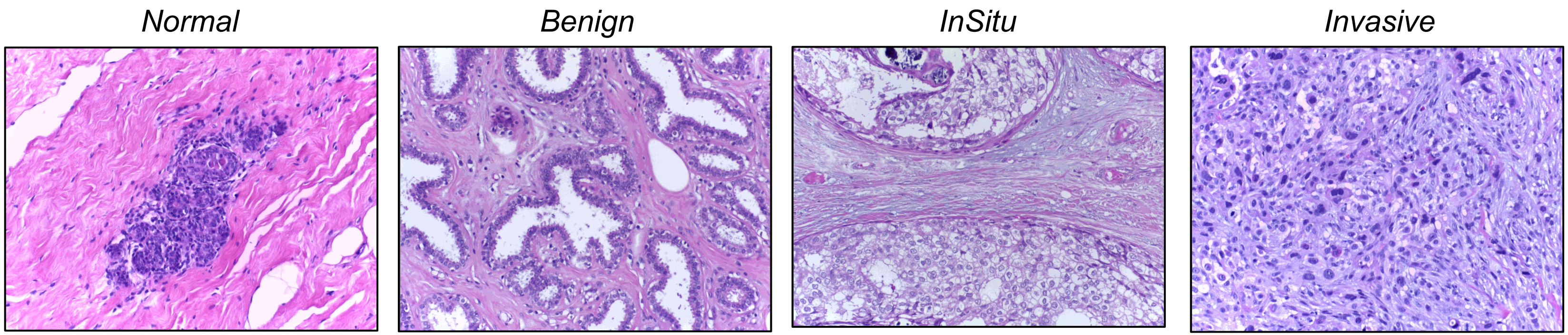


Figure : Microscopy images are labelled as normal, benign, in situ carcinoma or invasive carcinoma according to the predominant cancer type in each image

Dataset : Kather (colorectal cancer)

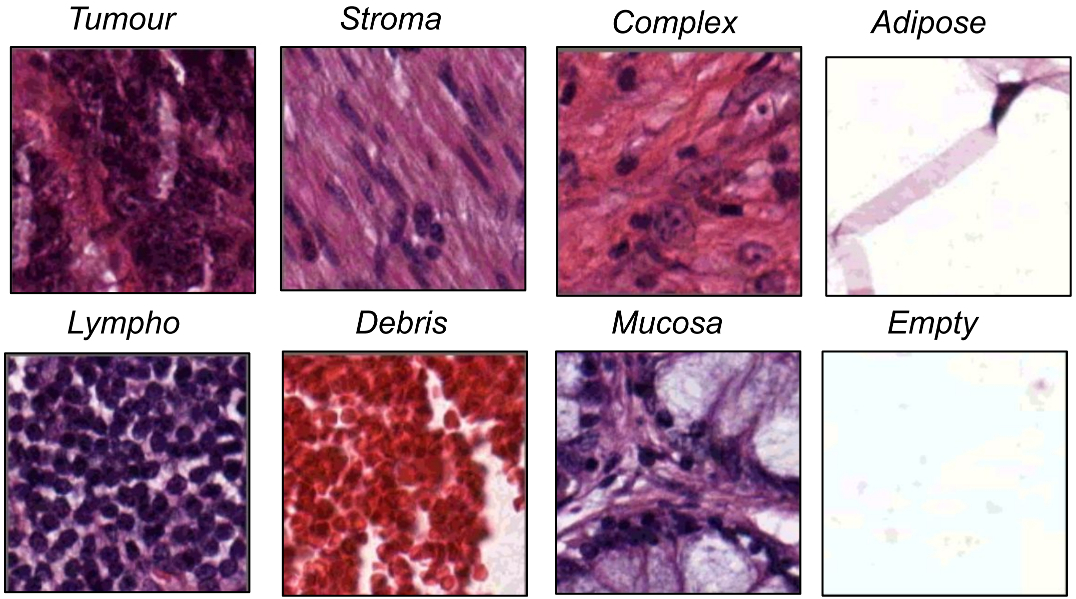
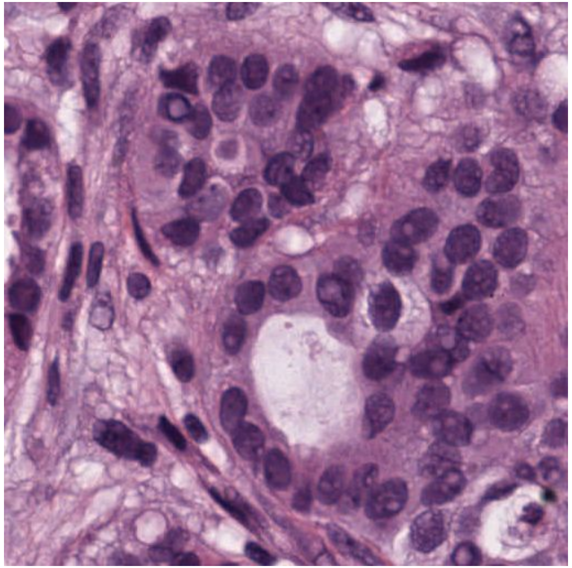


Figure : Image patches from hematoxylin & eosin (H&E) stained histological images of human colorectal cancer (CRC) and normal tissue.

Dataset : TCGA

Luminal A



Basal

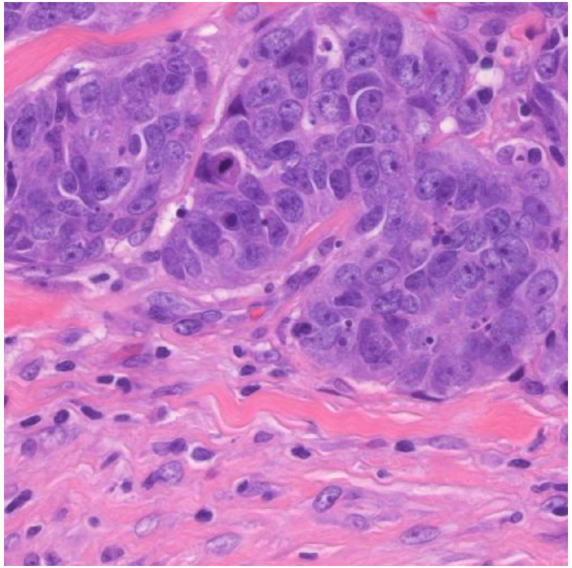


Figure : Unannotated Whole Slide image of breast cancer

Previous Approaches

- Wang et al [1] used an iterative technique to detect noisy labels using a probabilistic and cumulative local outlier factor (pcLOF) while learning deep discriminative features
- SSGCE [2] - An alternative to conventional cross entropy, that mitigates degradation due to closed-set label noise in medical images but not issues of open set.
- Chang Liu et al [3] (a metric learning framework) proposed learning robust class prototypes by using (FIFO) memory bank to aggregate clean samples based on their similarity scores

Experiments and Observations

Experiments

- Histopathology datasets used
  - BACH (breast cancer dataset)
  - Kather (colorectal cancer dataset)
  - TCGA (Basal vs Luminal A)

- Backbone architecture : ResNet-18 pre-trained (self-supervised) on a large histology dataset (self-supervised contrastive learning frameworks are widely adopted as robust initialization that can significantly improve state of art results using fewer labels)

- Learning Rate : 0.01
- Optimizer : Adam
- Data Augmentations : Color jitter, random vertical and horizontal flips

Experiment I : BACH

- *Outlier class* - InSitu (3-class classification)
- *Patch size* : 224×224
- *K-Medoids* : K is set to 3
- *Train* : 75, *Validation* : 25 images
- *Warm up phase* : 10 epochs
- *Batch size* : 8, miner size : 7
- *Memory bank size* : 300/class

Experiment I : BACH

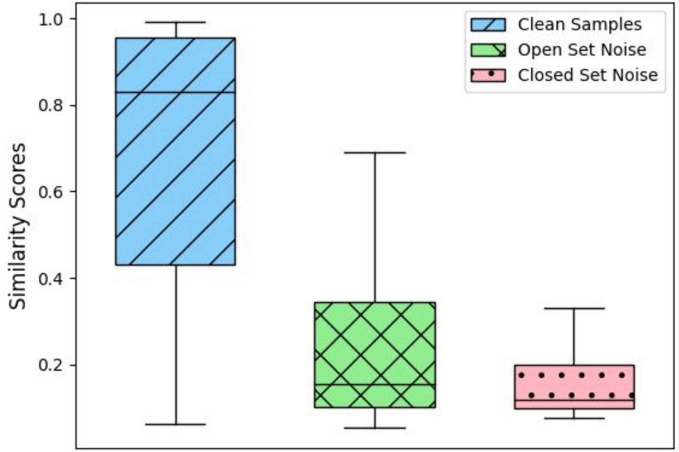
- *Outlier class* - InSitu (3-class classification)
- *Patch size* : 224×224
- *K-Medoids* : K is set to 3
- *Train* : 75, *Validation* : 25 images
- *Warm up phase* : 10 epochs
- *Batch size* : 8, miner size : 7
- *Memory bank size* : 300/class

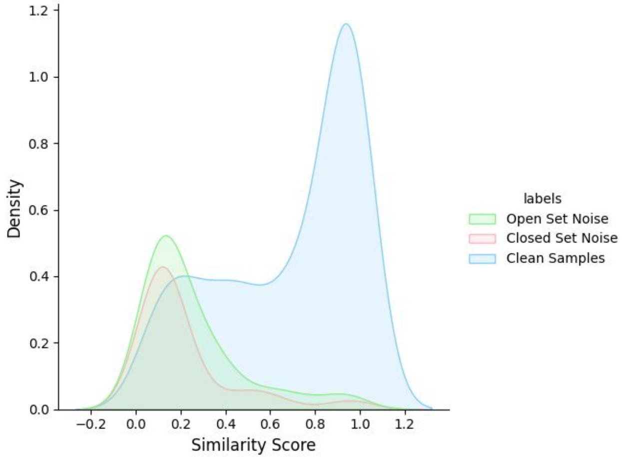
OOD Noise	15			20		
Label Noise	10	14	18	10	14	18
CE-Imagenet Pretrained	82.33±1.12	81.00±1.61	77.67±1.26	80.66±1.99	81.66±0.90	79.67±1.06
CE-SimCLR Pretrained	84.00±2.11	84.38±1.94	82.33±0.84	85.66±1.99	83.32±2.47	82.32±2.57
SSGCE	86.00±0.47	84.66±0.47	83.33±0.47	85.35±0.94	81.33±1.88	79.33±0.47
Mem-Bank (ours)	91.67±1.14	**90.33**±1.00	**89.03**±0.35	90.33±0.84	89.33±0.54	**88.66**±1.38
Mem-Bank+K-Medoids (ours)	**92.33**±0.33	90.00±0.86	88.08±2.05	**91.00**±1.26	**90.66**±0.94	88.00±0.54

Table 1 : 4-fold cross-validation classification accuracies on BACH dataset with different levels of label noise and OOD noise

Experiment I : BACH

Weights obtained in weight calculation phase clearly discriminate between clean and noisy samples





Distribution of similarity scores of samples with the memory bank obtained after warm-up phase, corresponding to clean data and noisy data

Experiment II : Kather

- *Outlier Classes* - Debris, Mucosa (6-class classification Problem)
- *Patch size* : 224×224
- *K-Medoids* : K is set to 3
- *Train* : 475, *Validation* : 75 images
- *Batch size* : 8, *Miner size* : 7
- *Warm up phase* : 10
- *Memory bank size* : 300/class

Experiment II : Kather

- *Outlier Classes* - Debris, Mucosa (6-class classification Problem)
- *Patch size* : 224×224
- *K-Medoids* : K is set to 3
- *Train* : 475, *Validation* : 75 images
- *Batch size* : 8, *Miner size* : 7
- *Warm up phase* : 10
- *Memory bank size* : 300/class

OOD Noise	30%			35%		
Label Noise	30%	40%	50%	30%	40%	50%
CE-Imagenet pretrained	89.12±1.07	88.91±0.64	87.05±0.05	89.28±0.69	89.19±0.50	85.53±0.93
CE-SimCLR pretrained	92.6±0.2	92.25±0.45	91.05±0.45	92.75±0.85	91.93±1.14	91.09±0.59
Mem-Bank (ours)	**94.08±0.26**	**92.91±0.59**	**91.68±0.32**	**92.57±0.81**	**93.71±0.32**	**92.33±0.11**
Mem-Bank+K-Medoids (ours)	92.33±0.33	90.00±0.86	88.08±2.05	91.00±1.26	90.66±0.94	88.00±0.54

Table 2 : 2-fold cross-validation classification accuracies on Kather dataset with different levels of label noise and OOD noise

Experiment III : TCGA

- *Outlier Classes* : Lobular, HER2, Luminal-B (2-class classification problem)
- *Patch size* : 512×512
- *K-Medoids* : K is set to 6
- *Train* : 90, *Validation* : 10, *Test* : 30 images
- *Batch size* : 128, *Miner size* : 32
- *Warm up phase* : 3 epochs
- *Memory bank size* : 10,000/class

Experiment III : TCGA

- *Outlier Classes* : Lobular, HER2, Luminal-B (2-class classification problem)
- *Patch size* : 512×512
- *K-Medoids* : K is set to 6
- *Train* : 90, *Validation* : 10, *Test* : 30 images
- *Batch size* : 128, *Miner size* : 32
- *Warm up phase* : 3 epochs
- *Memory bank size* : 10,000/class

	Slide-level Accuracy
CE-Imagenet	73.33
CE-SimCLR	80.00
SSGCE-Loss	83.33
Mem-Bank (ours)	80.00
Mem-Bank+K-Medoids (ours)	**86.67**

Table 3 : Basal vs Luminal A Classification accuracy percentages on 30 held out WSI

Conclusion and Future Work

Conclusion

- Demonstrated the inherent existence of open-set and closed-set noise in histology datasets
- We proposed a simple and effective method that addresses the general label noise problem in an integrated framework based on a sample weighting scheme
- Evident from mean similarity scores, our approach discriminates between clean and noisy data points after just a few training (warm up) epochs

Future Work

- Method of aggregating features in the memory bank - Use of contrastive loss instead of just adding in FIFO manner

- Re-weighting strategy - Use of non-linear functions on the similarity scores before using them with the weighted cross entropy loss can be experimented with

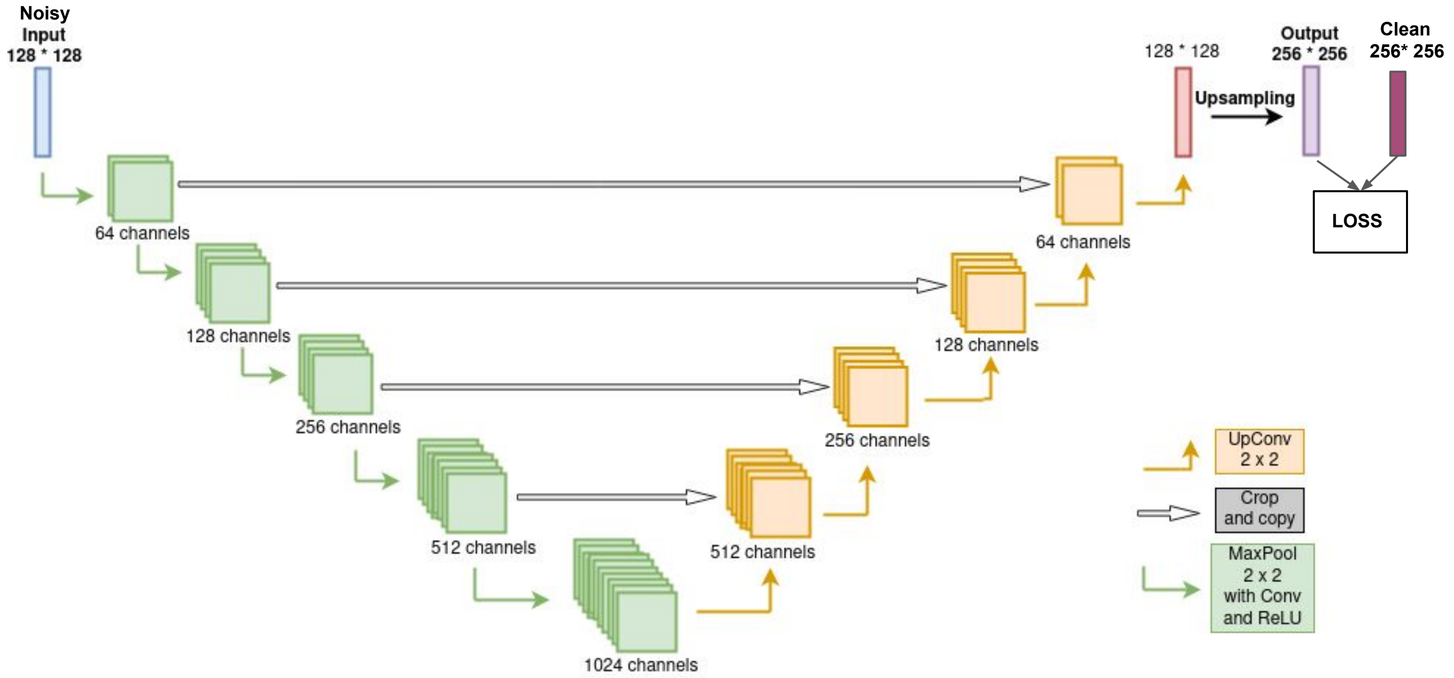
- Techniques including momentum based contrastive learning can be used to improve the embeddings further

DESPECKLING ULTRASOUND IMAGES

Motivation

- Ultrasound imaging is non-invasive, low cost - used in medical diagnosis
- Ultrasound pulses randomly interfere with objects of comparable size to the sound wavelength and the superposition of acoustical echoes produces an intricate interference pattern
- Speckle noise is an inherent property of medical ultrasound imaging
- Image resolution and contrast become reduced, affects the diagnostic value of US imaging

Method : Model Architecture



Modified U-Net Model for denoising

Experiments

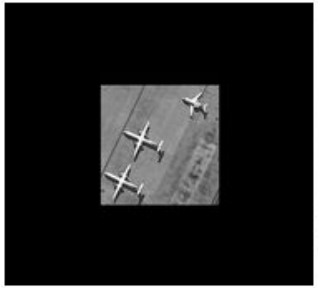
- Backbone - UNet

- Dataset - Public satellite images

- Loss functions
    i) MSE loss
    ii) L1 loss
    iii) SSIM loss

Experiments and Results

Low resolution noisy image



Mse | ssim | psnr

High resolution clean image

0 | 1 | ∞

UNet model with L1 loss

158.7 | 0.77 | 26.13

UNet model with MSE loss

191.65 | 0.69 | 25.3

UNet model with SSIM loss

280.2 | 0.78 | 23.65

Loss function	MSE Loss	SSIM Value	PSNR (dB)
L1	158.7	0.77	**26.13**
SSIM	280.2	0.78	23.65
MSE	191.65	0.69	25.3

UNet model trained with different loss function

Future Work

- Use of wavelet transforms instead of pooling layers to perform apply multiresolution analysis
- Use of different loss function for training
- Experiment with ultrasound images

Thank You !

References

1. Yisen Wang, Weiyang Liu, Xingjun Ma, James Bailey, Hongyuan Zha, Le Song, and Shu-Tao Xia, "*Iterative learning with open-set noisy labels*" . In CVPR, 2018
2. Nikhil Cherian Kurian, Pragati Shuddhodhan Meshram, Abhijeet Patil, Sunil Patel, and Amit Sethi, "*Sample specific generalized cross entropy for robust histology image classification*". In 2021
3. Chang Liu, Han Yu, Boyang Li, Zhiqi Shen, Zhan-ning Gao et al, "*Noise-resistant deep metric learning with ranking-based instance selection*" . In CVPR, 2021
4. Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton. "*A Simple Framework for Contrastive Learning of Visual Representations*". In ICML, 2020
5. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "*U-net: Convolutional networks for biomedical image segmentation*," CoRR, vol. abs/1505.04597, 2015.