Automated Gleason Grading of Whole Slide Images

Parth Dodhia (180070041) , MK Ipsit (180070032)

Introduction

- Prostate cancer : second leading cause of cancer deaths in men [1]
- Gleason grading system : prognostic predictor for patients with Prostate Cancer, identifies aggressiveness of tumors formed
- Trained pathologists perform Gleason grading manually
- Based entirely on the architectural patterns of the tumor

[1]: WHO Classification of Tumours of the Urinary System and Male Genital Organs. International Agency for Research on Cancer (IARC) (2016)

Why Deep Learning?

- Train a CNN to learn structural differences between Gleason patterns
- Remove the grading bias of a pathologist and increase generalizability
- Can drastically speed up the process of grading

Image Source :
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.bhphotovideo.com%2Fc%2Fproduct%2F1046358-REG%2Fnvidia_900_22081_0040_000_tesla_k40_gpu_accelerator.html&psig=AOvVaw09HTr98GnF9tVaB5sojgyP&ust=1592918229840000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCOj2-_3AleoCFQAAAAAdAAAAABAD

Gleason Patterns

- Benign tissues have small, well differentiated glands. As the Gleason score increases, the glands tend to become variable in size, fused , distant and less in number

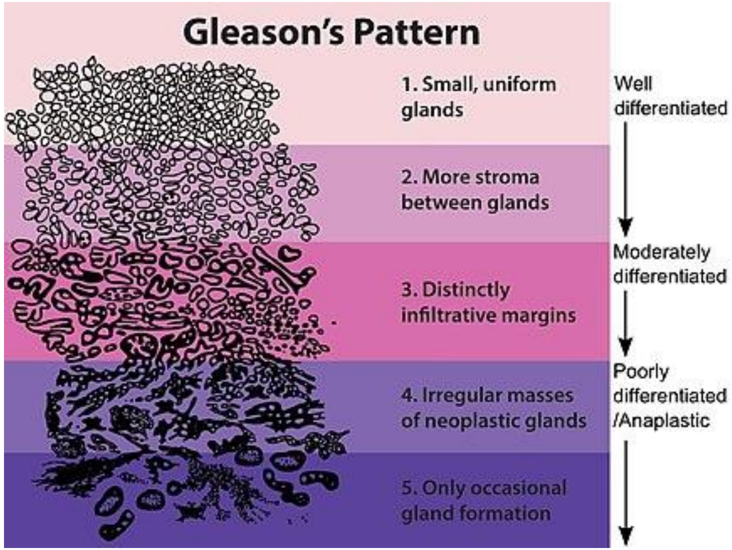


Image source : **Wikipedia**

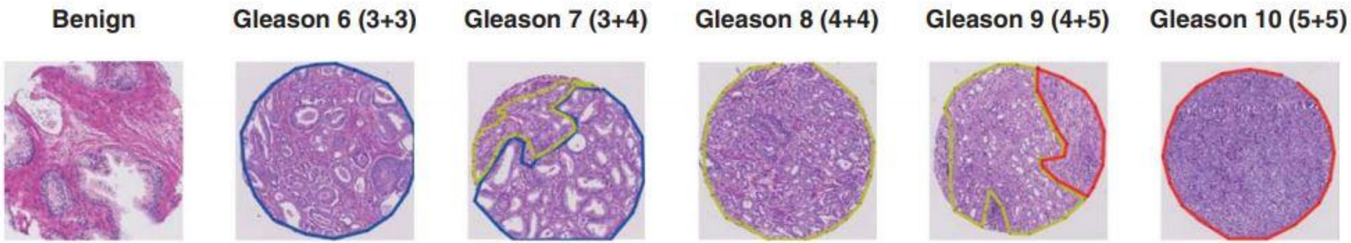


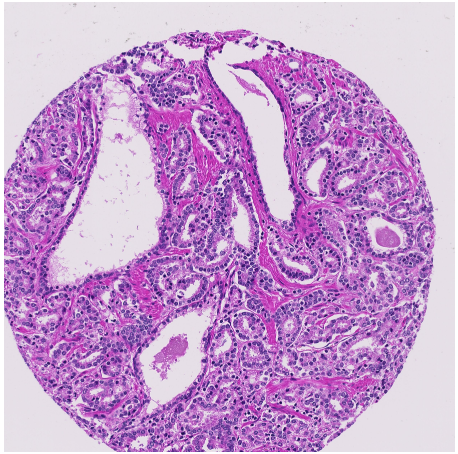
Image source : Automated Gleason grading of prostate cancer tissue microarrays via deep learning (https://www.nature.com/articles/s41598-018-30535-1.pdf)

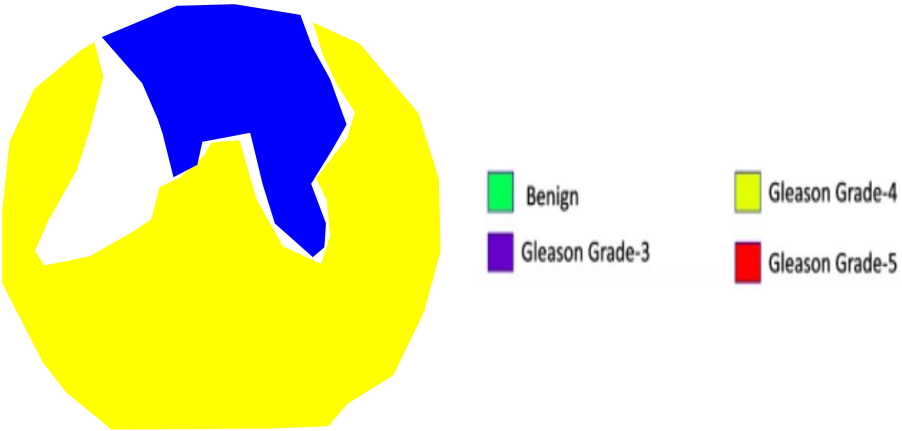
Contents

Dataset

- 641 whole slide images for training and 245 images for testing, each corresponding to a different patient (Harvard prostate cancer dataset)
- Images of size 3100x3100, with corresponding annotated ground truth images
- Image and

  Ground truth

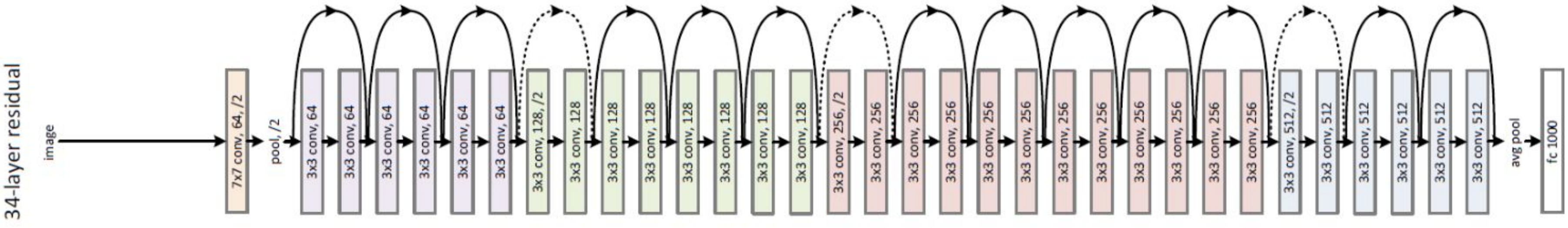




Images taken from the dataset : https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OCYCMP

Architecture

Pre trained encoder : Resnet 34

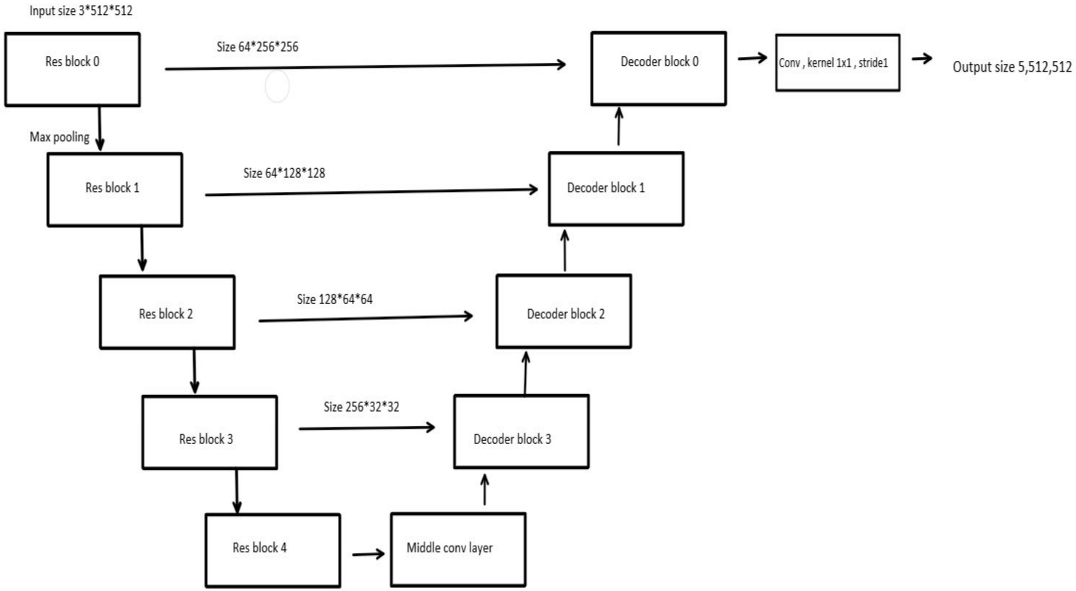


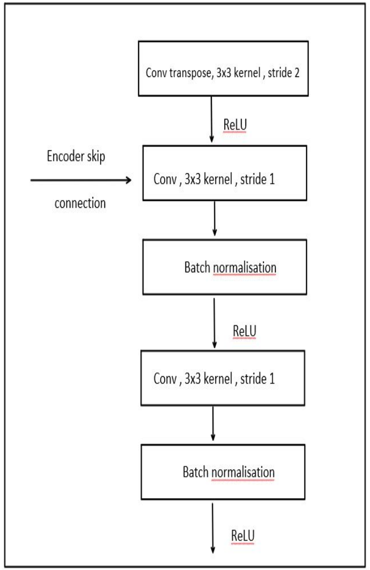
Res block 0 Res block 1 Res block 2 Res block 3 Resblock 4

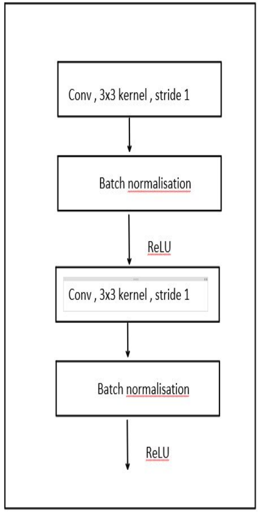
Image source : *https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035*

Architecture

- A symmetric UNet architecture for segmentation. We found pretrained encoders to work much better







Middle layer

Decoder block

Implementation

- The trainset comprised of 641 images from 4 microarrays. We took 133 images for validation and 508 images for training
- Resized image to size 512 for decent batch size. Implemented dynamic mask generation from ground truths to tackle storage issues
- Data augmentations : random horizontal and vertical flip, random affine and color jitter to tackle small trainset size and normalisation as per ImageNet mean and variance.

Loss Functions

- Cross Entropy

$$\mathscr{L}_{CE} = -\sum_{c=1}^{M} y_c \log p_c$$

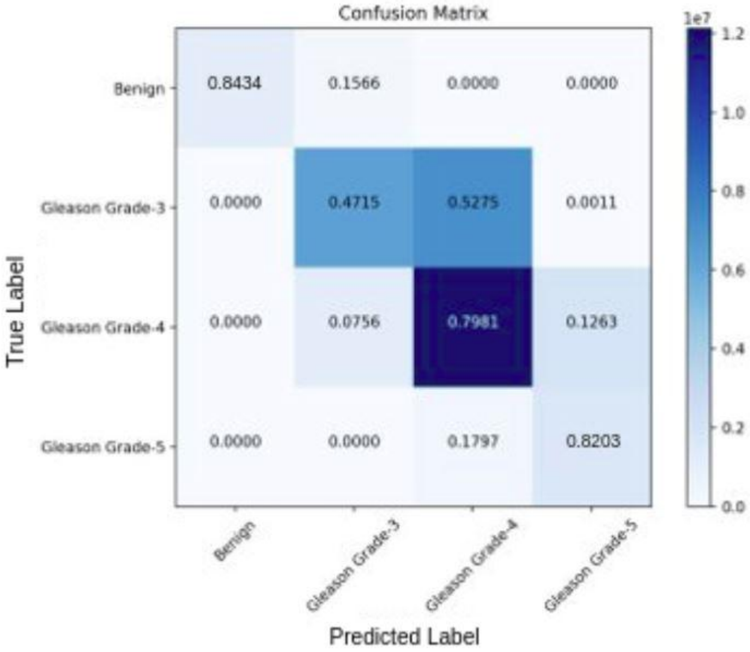
- Soft Dice

$$\mathscr{L}_{Dice}(p,q) = 1 - \frac{1}{M}\sum_{c=1}^{M} \frac{2 \times \sum_{i,j} p_{cij} q_{cij} + \varepsilon}{\left(\sum_{i,j} p_{cij}^2\right) + \left(\sum_{i,j} q_{cij}^2\right) + \varepsilon}$$

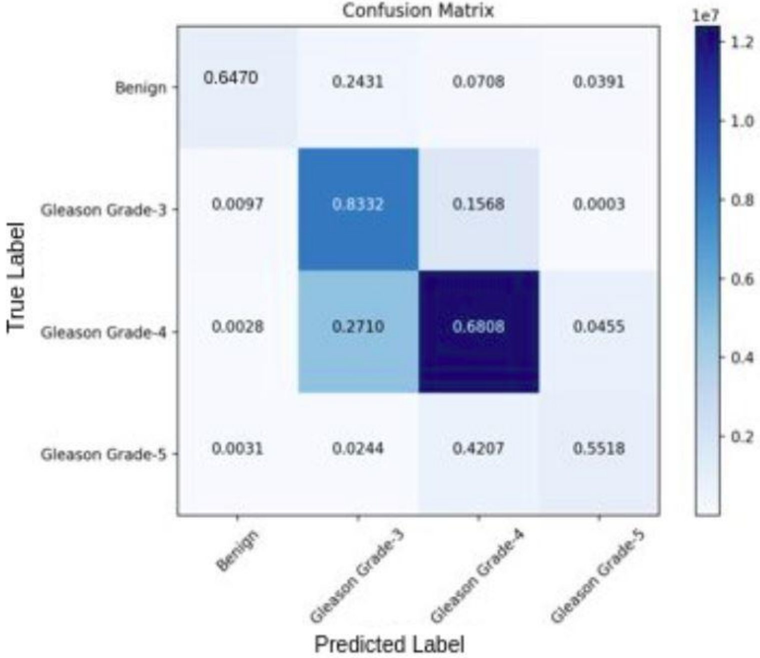
- Focal Loss

$$\mathscr{L}_{Focal} = -\sum_{c=1}^{M} y_c (1-p_c)^{\gamma} \log(p_c)$$

Results

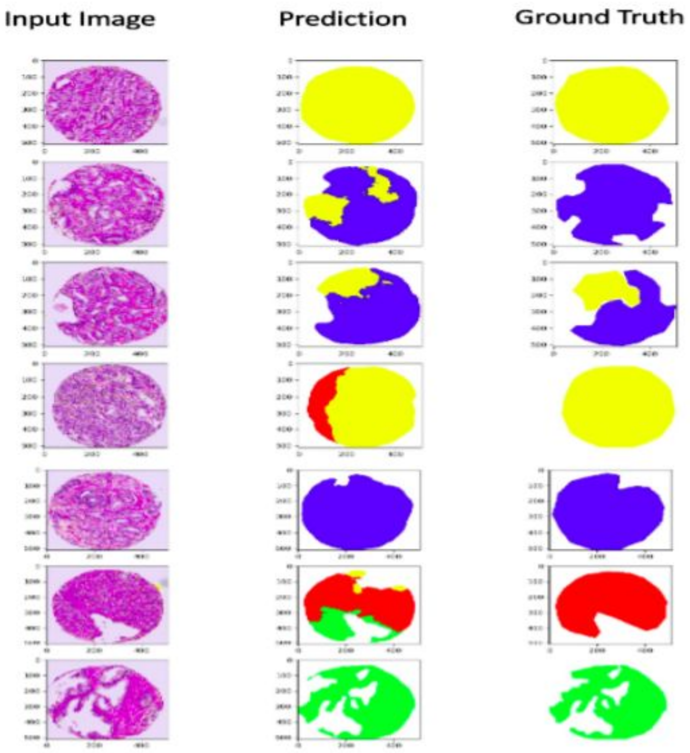


Kappa Score:0.44
Mis-classifications: 0.2667
(Among Pathologists)



Kappa Score:0.53
Mis-classifications: 0.3212
(Model vs Pathologists)

Visualisation of results



Further Experiments

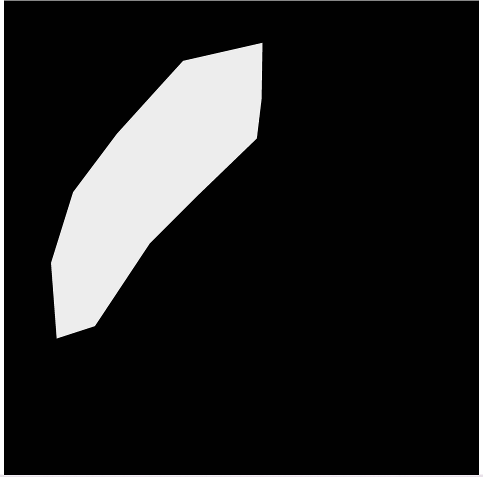
Attention Based Multiple Instance Learning

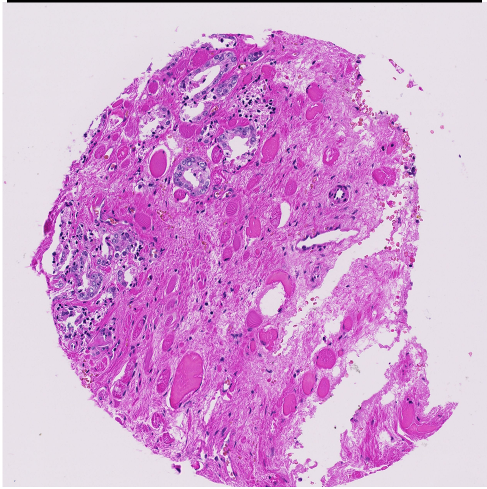
Why A-MIL?

- Medical image datasets are weakly annotated.
- Identifying key instances is useful in clinical practice.

  Eg: Targeted Therapy, Precision Medicine etc.

- Giving attention to key instances helps!!!





What is A-MIL?

- Pooling the lower dimensional embeddings by determining weights using neural networks.
- Predicting the label of a bag of instances instead of predicting for an instance.
- Advantage?
  - End-to-end training of the model using back propagation.
  - Larger weights can be used to identify key instances.
- h_k is the lower-dimensional embedding

$$z = \sum_{k=1}^{K} a_k h_k$$

where

$$a_k = \frac{exp(w^T \tanh(V h_k^T))}{\sum_{j=1}^{K} exp(w^T \tanh(V h_j^T))}$$

Further Experiments (MIL)

- We proposed this problem as a classification task and tried to use attention based multiple instance learning.
- We experimented with different image/patch sizes to make a bag.
- We used embedding based attention technique, for which ResNet18 was used as a feature extractor.
- On the right, h_i's are lower dimensional embeddings.

$$z = \sum_{k=1}^{K} a_k h_k$$

where

$$a_k = \frac{exp(w^T \tanh(V h_k^T))}{\sum_{j=1}^{K} exp(w^T \tanh(V h_j^T))}$$

MIL Architecture

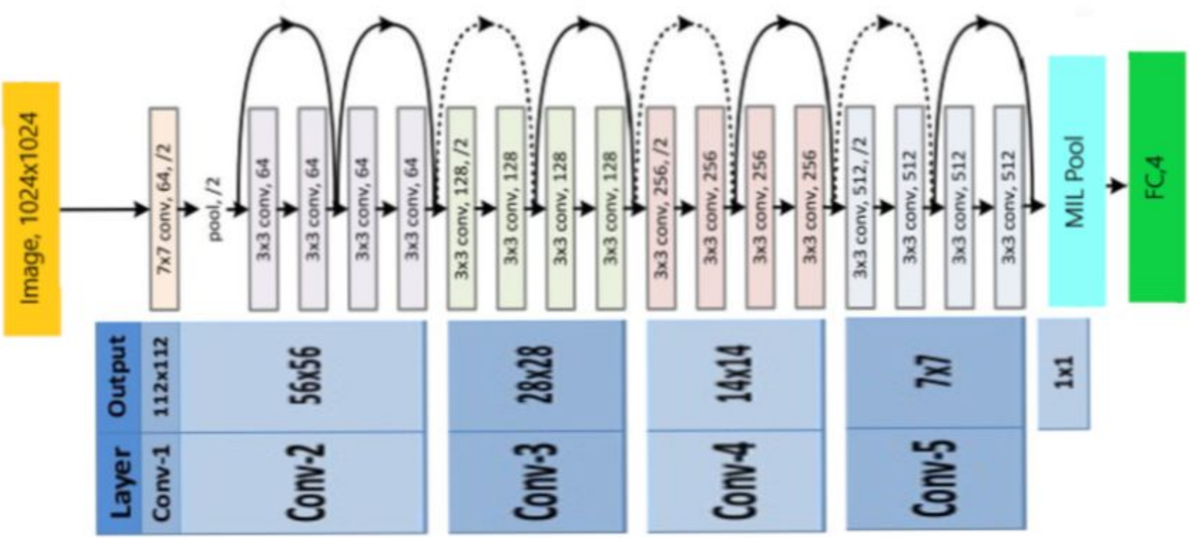


Image source:https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FArchitecture-of-ResNet-18-Figure-from-32_fig2_336278800&psig=AOvVaw3Qe3zE_MKNVWwCVujMOsya&ust=1592805961150000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCPDlzd-ekuoCFQAAAAAdAAAAABAD

MIL Pooling

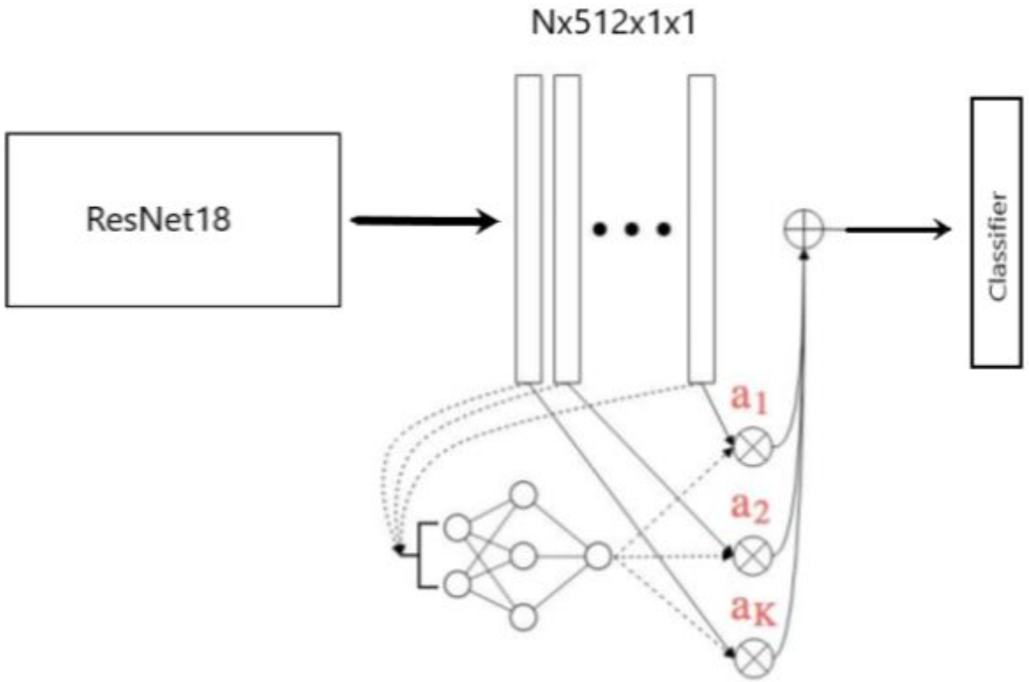


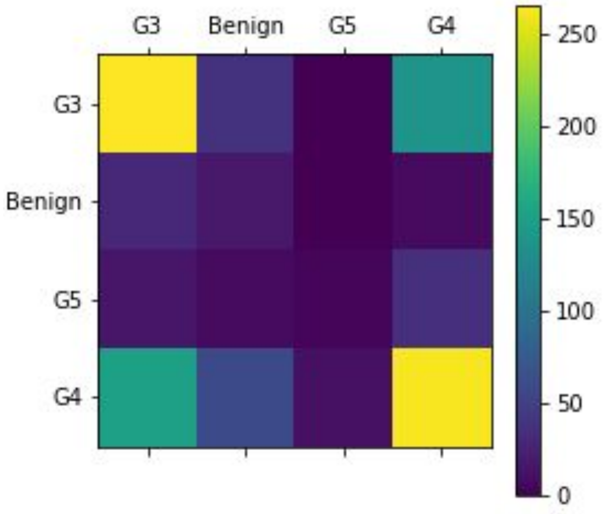
Image source:https://arxiv.org/pdf/1802.04712.pdf

Implementation

- WSI is divided into pure patches(homogeneous grade) of size 1024x1024
- Each such patch is further divided into 5 smaller patches to make a bag.
- Features are extracted using a ResNet18 backbone.
- MIL Pooling
- Bag level classifier
- Cross Entropy Loss minimization
- Data Augmentation: Random Horizontal and Vertical Flips, mean normalization using ImageNet statistics

Evaluation and Results

- Gleason 3 and Gleason 4 are being detected more accurately than others.
- The model is getting confused b/w grades 3 and 4.
- Patch-level classification accuracy on test set is 52.4%.



Conclusion and Future Work

- Multiresolution segmentation network