Method

Due to the high resolution of images (1536x2048), small patches of 300x300 were extracted from each image. Patches with greater than 90% background were excluded from training and predictions. On average 29 patches were extracted from each image. From a total of 400 images (100 per class), a total of 400*29 = 11,600 patches was extracted from the dataset. This is different from Kwok, 2016[1] where a larger patch of 1400x1400 was used which was down-sampled to 300x300 for training.

A 90/10 split was used for splitting training vs validation patches. The patches from the same case was placed in either training or validation and not both.

Model

Model was written with keras API with tensorflow 2.0 backend. A simple 9-layer convolutional neural network model was used for this attempt, without loading any pre-trained weights. Each layer has ReLU activation followed by dropout at 0.2 followed by batch normalization. Number of filters per layer are: 32, 64, 64, 128, 128, 256, 256, 512, 512. Final dense layers had 512 nodes followed by ReLU activation and batch normalization and 4 nodes followed by softmax activation. Model was compiled with categorical cross-entropy loss and Adam optimizer.

Training

Training included real time augmentation as suggested by Tellez et al, 2019[2]. imgaug python model was used with the following parameters:

```
# instantiate imgaug augmentation object
sometimes = lambda aug: iaa.Sometimes(0.5, aug)

AUGMENTATIONS = iaa.Sequential([
    iaa.Fliplr(0.5),
    iaa.Flipud(0.5),
    sometimes(iaa.Affine(
        scale=(0.8, 1.2),
        rotate=(90),
        mode=ia.ALL)),
    sometimes(iaa.ElasticTransformation(alpha=(0.8, 1.2), sigma=(9.0, 11.0))),
    sometimes(iaa.AdditiveGaussianNoise(scale=(0, 0.1))),
    sometimes(iaa.GaussianBlur((0, 0.1))),
    sometimes(iaa.MultiplyBrightness((0.65, 1.35))),
    sometimes(iaa.LinearContrast((0.5, 1.5))),
    sometimes(iaa.MultiplyHueAndSaturation((-1, 1)))
    ], random_order=True)
```

Model was trained for 300 epochs, at learning rate of 1e-4, and batch size of 32 on TeslaV100GPU. Final training accuracy = 0.92 and validation accuracy = 0.89.

References

[1] S. Kwok, “Multiclass Classification of Breast Cancer in Whole-Slide Images,” *Image Anal. Recognit.*, vol. 1, no. June, pp. 931–940, 2018.

[2] D. Tellez *et al.*, “Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology,” *Med. Image Anal.*, vol. 58, 2019.