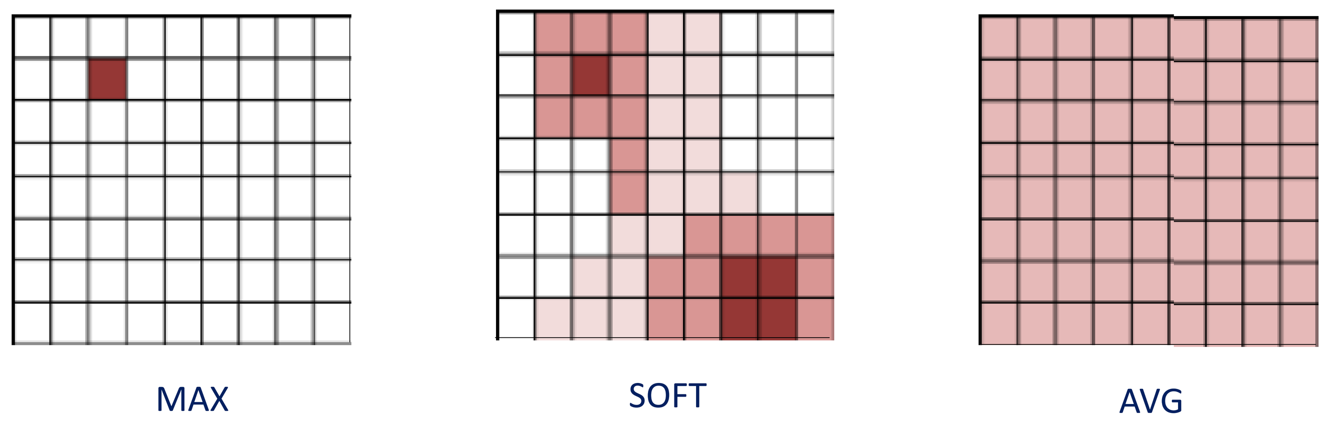
Is Soft pooling better than Max and Average pooling?
A comparative study on HEp-2 cells and Retinal image classification tasks

Nirthika R., Siyamalan M., and Ramanan A.
Department of Computer Science, University of Jaffna
{nirthika, siyam, a.ramanan}@univ.jfn.ac.lk

Abstract

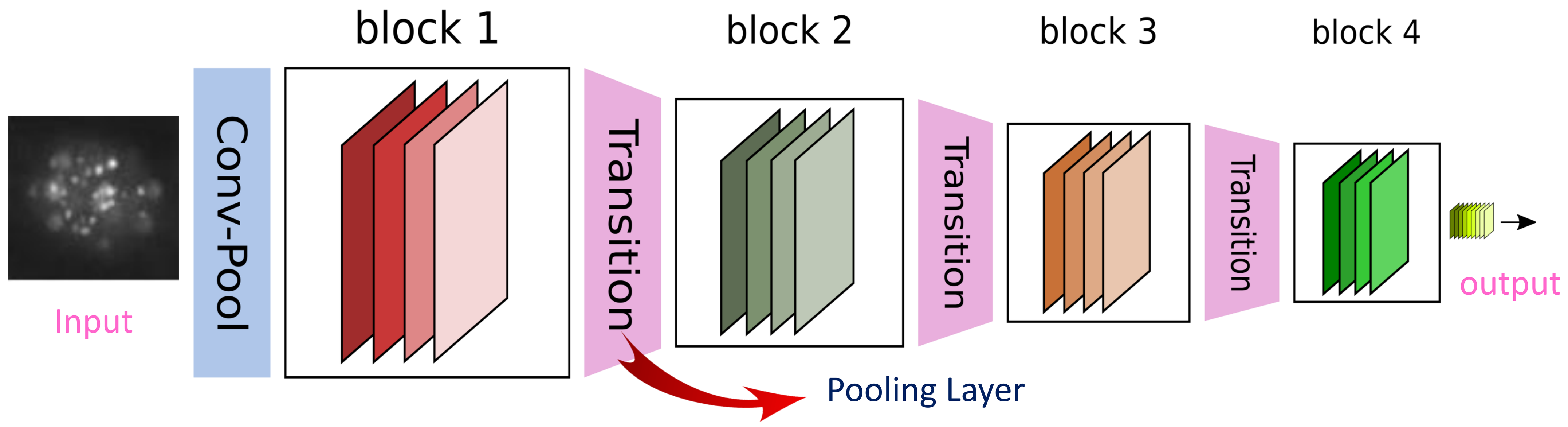
Convolutional Neural Network (CNN) has been widely used for medical image classification [1], where, pooling layers are used for down sampling the feature maps by summarizing the presence of features in local regions of the feature map. Average and Max are the widely used pooling methods. Since average pooling summarizes all the features in the feature map, background regions may dominate in the pooled representation. On the other hand, max pooling can capture noisy features as it focuses on the most activated features. To overcome this, Soft pooling have been proposed [2]. However, soft pooling has not been well explored for medical image analysis. Therefore, this work focuses on investigating its performance on two different medical image classification tasks, i.e., cell image classification and diabetic retinopathy image classification. Our experiments show that soft pooling does not produce significant improvement in performance compare to max and average pooling.

Max, Average and Soft Pooling



- **Max** pooling considers only the maximum element and ignores all the rest of the feature map. Hence, captures noisy features.
- **Average** pooling considers all the elements in the feature map. Hence, captures unwanted background information.
- **Soft** pooling captures information from a set of maximum activated elements of the feature map. Hence, can overcome the above mentioned problems with the max and average pooling.

Network Architecture



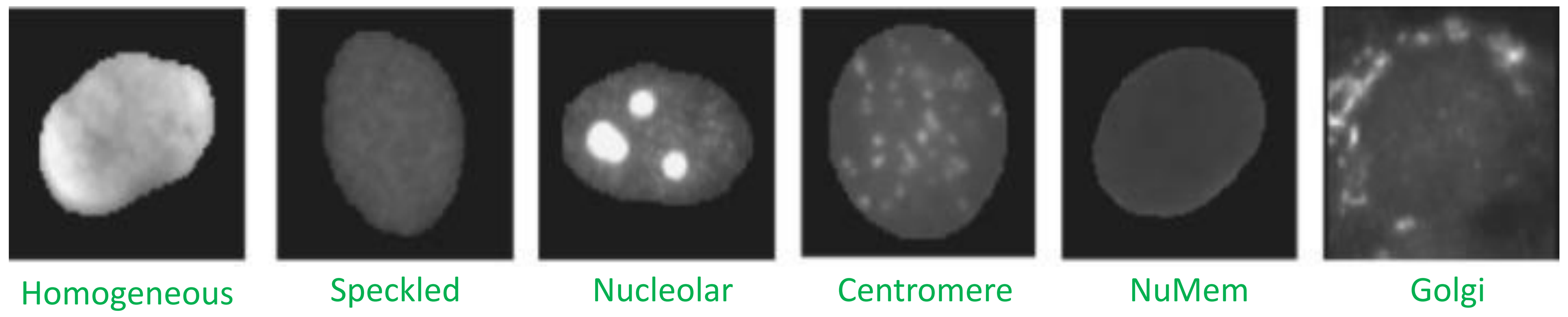
Soft pooling output over region R_i for input x:

$$f_{\text{soft}}(X) = \frac{\ln\left[\frac{1}{|R_i|}\sum_{x_j \in R_i} \exp(\lambda x_j)\right]}{\lambda}$$

where, λ is a hyper-parameter, when $\lambda \rightarrow \infty$ max pooling and $\lambda \rightarrow 0$ average pooling can be obtained.

Dataset & Experimental Setup

1. HEp-2 cell dataset [3] contains Human Epithelial type 2 (HEp-2) cell patterns from six categories



We used 15,314 images for training and 10,764 images for testing. Each experiment was iterated three times and the average and the standard deviation of mean per-class accuracies (MCA) over iterations are reported.

2. Diabetic Retinopathy (DR) dataset [4] contains images indicating the presence of diabetic retinopathy.

We used 8,837 images for training and 4,771 images for testing. Each experiment was iterated three times and the average and the standard deviation of quadratic weighted kappa over iterations are reported.

Experiments and Results

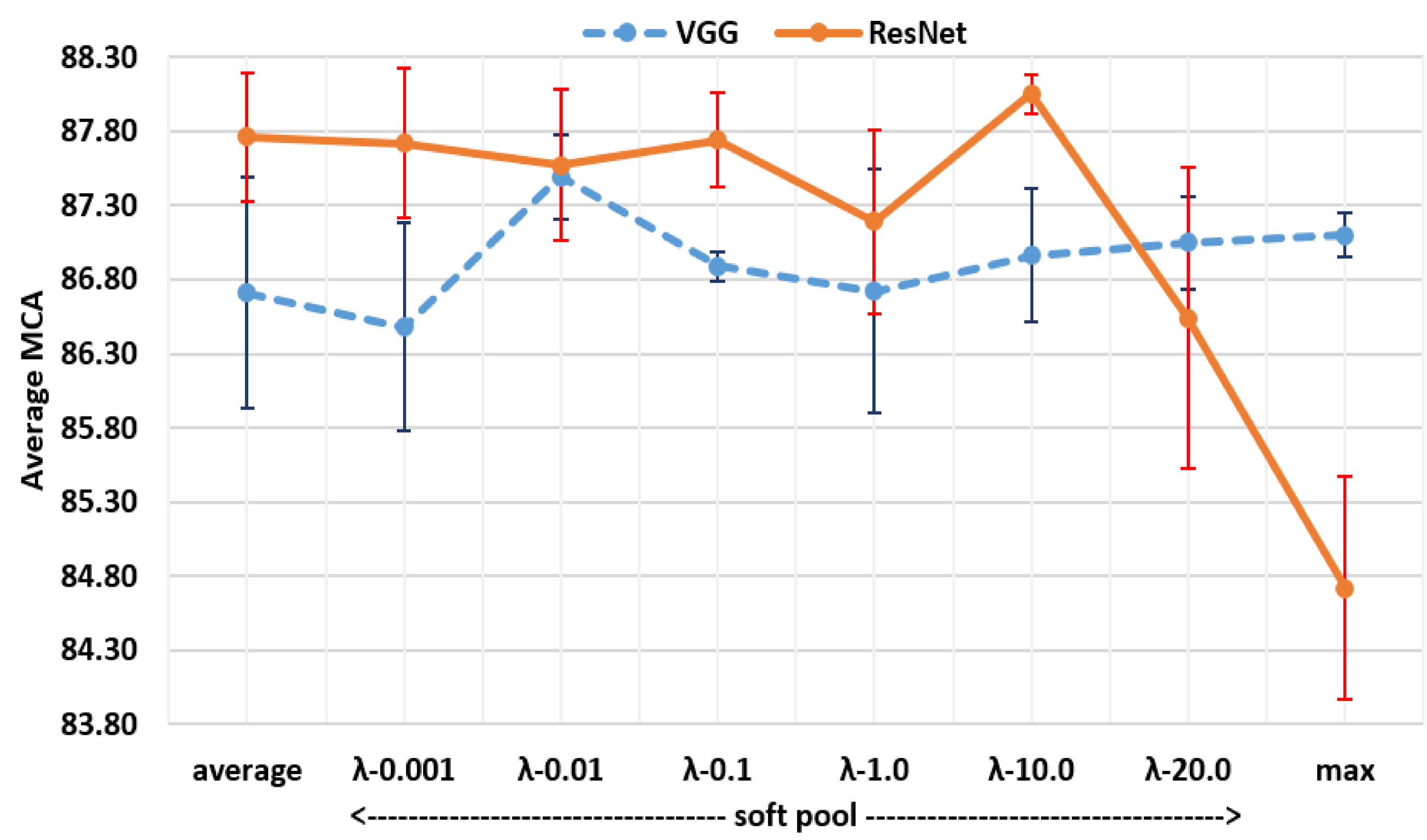


Figure 1: HEp-2 cell image dataset with different pooling strategies.

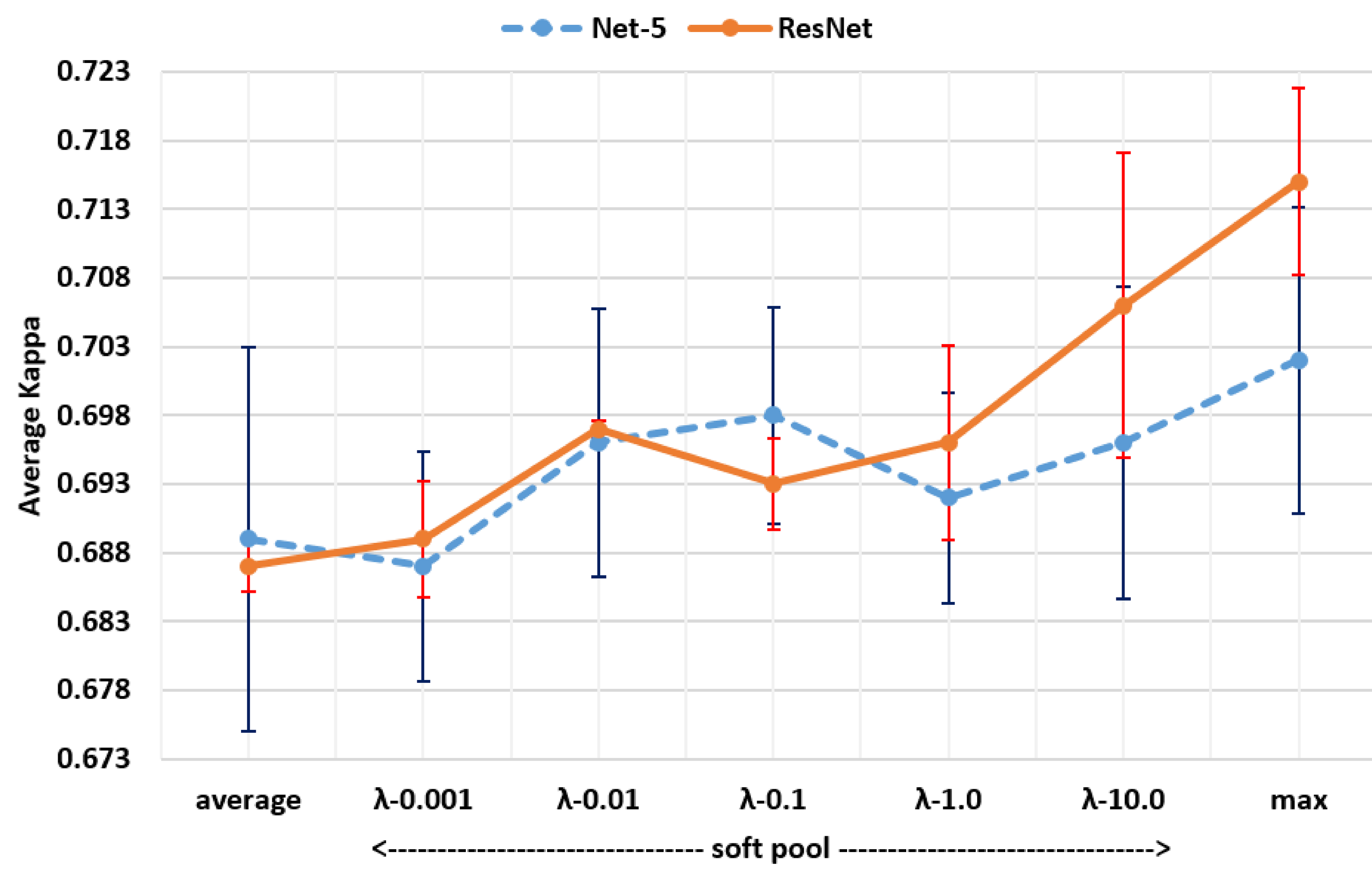


Figure 2: DR image classification with different pooling strategies.

HEp2 Cell images: For small architecture (VGG), soft pooling performs approximately same as max and average pooling, but for large architecture (ResNet), soft pooling performs approximately same as average pooling and better than max pooling. This result is expected as all the contents in the cell images are necessary to determine its class.

DR dataset: Soft pooling performs better than average pooling and worse than max pooling for both small (Net-5 [1]) and large (ResNet) networks. For both architectures max pooling performs better as the lesions usually occupy small regions in the DR images.

Conclusion

This work compares max, average, and soft pooling for HEp-2 cell and DR image classification tasks with CNN. For each of the task we used two different CNN architectures. For the DR dataset, max pooling performs better than average pooling since the lesions cover only part of the images, and for the cell dataset average pooling performs better as the images cover all the cell regions. We found that soft pooling does not perform better than average and max pooling.

References

[1] Z. Wang and J. Yang, "Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation," arXiv preprint arXiv:1703.10757, 2017.
[2] Wei, Z., *et al.*, "Building detail-sensitive semantic segmentation networks with polynomial pooling", IEEE CVPR, 2019.
[3] https://mivia.unisa.it/datasets/biomedical-image-datasets/hep2-image-dataset
[4] https://www.kaggle.com/c/diabetic-retinopathy-detection/data