Bachelor in Data Science and Engineering - 2021/2022

University Carlos III Madrid

Audio processing, Video processing and Computer vision

SERGIO AIZCORBE (**100406602**) | BERNARDO BOUZAS (**100406634**) | LS4 REPORT

Computational time and data augmentation

In order to reduce the computational time of the training phase, we have created an additional dataset object (*IsicDataset*) which iterates over the image and the label (without the mask). The default composed transformations have been replaced by Pytorch's transforms that run faster on the GPU. A new class has been implemented (*CropByMask2*), which uses PIL's crop function to crop the image based on the image mask.

The dataset provided for this practice is severely skewed, as the number of instances in the train set with label 0 (non-melanoma) is 4-6 times bigger than classes 1 and 2 (melanoma and seborrheic). Combined with different flip transformations, duplicating the training samples of these classes helps to avoid the negative effect of class imbalances and hence increase the performance of our CNNs [1].

Convolutional Neural Network architectures

As our flexibility in terms of choice of extensions and decisions is unrestricted and the computational time of parameter tuning is relatively high, we have decided to focus our approach on studying the effect of different CNN architectures on our dataset [2]. Thus, our solution proposes a comparison between the following implementations:

- **AlexNet**: the by-default-provided architecture consists of 5 convolutional layers and 3 fully connected layers [3]. Multiple filters extract desired features in an image. In each of those convolutional layers, there are many kernels of the same size.
- **GoogLeNet:** one of the best state-of-the-art deep networks [4], codenamed Inception. This allows the network to choose between multiple filter sizes in each block. It stacks these modules on top of each other, with occasional max-pooling layers designed to halve the resolution of the grid.
- **ResNet**: ResNets with various depths such as ResNet50, ResNet101 and ResNet18 use bottleneck features to improve efficiency in comparison with their predecessor CNN models [5]. We test them as they usually provide better recognition accuracy in scenarios like the one we are presented with.
- **ResNeXt:** it is structured in repetitive blocks composed of convolution and non-linear operations, but in each block operations are performed across many branches and results are aggregated together with the block input. [6]
- **VGGNet:** utilizes smaller filters of 3 × 3, compared to AlexNet 11 × 11 filter, in order to provide better features extraction from images [7]. We decided to use VGG19 over VGG16 as it is a deeper network, although slower.

Being an image classification problem, it is an ideal scenario to speed up our study using transfer learning. The proposed architectures are the most state-of-the-art networks used for this type of learning. The last layer of every network has been adapted to match the number of classes in our dataset.

Data transformations

New composed transforms have been added to the original training data loader, such as RandomHorizontalFlip and AutoAugmentation. In some cases the batch size has been reduced in order to compare more results. Therefore, we use standard and reliable techniques such as **CenterCrop, DataAugment, Resize**, **Random Horizontal Flip**, **Tensor Transformation** and **Normalization**.

VALIDATION RESULTS (batch size = 128)

	Alexnet	Googlenet	Resnet18	Resnet50	ResNeXt50	ResNeXt101	VGG-13-BN	VGG-19-BN
Default	0.818	0.83	0.819	0.83	0.811	0.802	0.775	0.805
Data Transformations	0.829	0.84	0.839	0.84	0.842	0.805	0.786	0.810
Final model	0.846	0.855	0.850	0.856	0.850	0.811	0.802	0.823

References

[1] Kassani, S. H., & Kassani, P. H. (2019). A comparative study of deep learning architectures on melanoma detection. Tissue and Cell, 58, 76-83.

[2] Brinker, T. J., Hekler, A., Enk, A. H., Berking, C., Haferkamp, S., Hauschild, A., ... & Utikal, J. S. (2019). Deep neural networks are superior to dermatologists in melanoma image classification. European Journal of Cancer, 119, 11-17.

[3] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS 2012, vol. 1, pp. 1097–1105. Curran Associates Inc., USA

[4] Szegedy, C., et al. (2015): Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 0, pp. 1–9,

[5] Lei, H., Han, T., Zhou, F., Yu, Z., Qin, J., Elazab, A., & Lei, B. (2018). A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning. Pattern Recognition, 79, 290-302.

[6] Koné, I., & Boulmane, L. (2018, June). Hierarchical resnext models for breast cancer histology image classification. In International Conference Image Analysis and Recognition (pp. 796-803). Springer, Cham.

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.