1 Experiments

1.1 Data

We make use of two ordinal datasets appropriate for deep neural networks:

- *Diabetic Retinopathy*[1]. This is a dataset consisting of extremely high-resolution fundus image data. The training set consists of 17563 pairs of images (where a pair consists of a left and right eye image corresponding to a patient). In this dataset, we try and predict from five levels of diabetic retinopathy: no DR (25810 images), mild DR (2443 images), moderate DR (5292 images), severe DR (873 images), or proliferative DR (708 images). The images are taken in variable conditions: by different cameras, illumination conditions and resolutions. These images come from the EyePACS dataset that was used in a Diabetic Retinopathy Detection competition that was hosted on the Kaggle platform. Also, this dataset was used in later works [1] and ordinal techniques (such as an ordinal cost function) were applied in order to achieve a better performance. A validation set is set aside, consisting of 10% of the patients in the training set. The images are resized to 256 by 256 pixels. Data augmentation techniques are applied in order to achieve a higher number of samples.

- *Adience*[2]. This dataset consists of 26580 faces belonging to 2284 subjects. We use the form of the dataset where faces have been pre-cropped and aligned. The dataset was preprocessed, using the methods described in a previous work [2], so that the images are 256px in width and height and pixels values follow a $(0; 1)$ normal distribution. The original dataset is splitted in 5 cross-validation folds. The training set consists of merging the first four folds together which comprises a total of 15554 images. From this, 10% of the images are held out as part of a validation set. The last fold is used as test set.

1.2 The model

A convolutional neural network (CNN) has been used for both datasets. The architecture of this CNN is presented in the Table 1.

Layer	Output shape
Conv_32_3x3	254x254x32
Conv_32_3x3	252x252x32
MaxPool_2x2	126x126x32
Conv_64_3x3	124x124x64
Conv_64_3x3	122x122x64
MaxPool_2x2	61x61x64
Conv_128_3x3	59x59x128
Conv_128_3x3	57x57x128
MaxPool_2x2	28x28x128
Conv_128_3x3	26x26x128
Conv_128_3x3	24x24x128
MaxPool_2x2	12x12x128
Conv_128_4x4	9x9x128
Dense_1_output	1

Table 1: Description of the architecture used in the experiments. For convolutional layers, Conv_N_WxH, where N is the number of filters, W the filter width and H the filter height. Stride is 1 for every convolutional layer. For max pool layers, MaxPool_SxS, where S is the pool size.

Every convolutional layer is followed by an ELU activation layer [3] and a batch normalization [4]. At the output, a Proportional Odds Model is used with different link functions (logit, probit, complementary log-log).

References

[1] J. de la Torre, D. Puig, and A. Valls, "Weighted kappa loss function for multi-class classification of ordinal data in deep learning", *Pattern Recognition Letters*, vol. 105, pp. 144–154, 2018.

[2] C. Beckham and C. Pal, "Unimodal probability distributions for deep ordinal classification", *ArXiv preprint arXiv:1705.05278*, 2017.

[3] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)", *ArXiv preprint arXiv:1511.07289*, 2015.

[4] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *ArXiv preprint arXiv:1502.03167*, 2015.

[1] https://www.kaggle.com/c/diabetic-retinopathy-detection/data

[2] http://www.openu.ac.il/home/hassner/Adience/data.html