Convolutional neural network-based ordinal regression for brain age

Ksenia Sokolova[a], Gareth J. Barker[a], Giovanni Montana[b]

[a]Department of Neuroimaging, King's College London; [b]Warwick Manufacturing Group, University of Warwick

INTRODUCTION

Understanding brain ageing has potential for determining biomarkers to be used for earlier detection of age-associated brain diseases. Nevertheless, the variability of onset ages for such diseases suggests that the effects of ageing on the human brain will also show significant differences between individuals and during human lifetime, limiting the utility of individual measures. To overcome this, we extract a **population-wide brain ageing profile** based on the classifier used to predict brain age from MRI data and identify features describing such profile, i.e **saliency maps**.

Key contributions of this work:

1. **We assess whole brain ageing effects across the human lifetime by using the Deep Embedded Clustering (DEC) method [1].**
2. **We apply the methods of vanilla [2] and guided [3] backpropagation, SmoothGrad [4], Gradient Class Activation Mapping (Grad-CAM) and guided Grad-CAM [5] to extract profile describing saliency maps.**

DATASET

- **The dataset** was collated from 34 public datasets and includes 10,878 T1-weighted MRI scans acquired in healthy subjects.
- The age range is 17 – 96 years with a mean of 42 ± 23 years (standard deviation).
- The disease state of each subject (i.e. their status as a nominally healthy control) has been inferred from the study protocols provided with the data.
- **Data preprocessing**: raw MRI scans underwent affine transformation and resampling onto MNI152 template using FSL FLIRT software [6].
- A 80%−10%−10% pseudo-random division of the dataset was used to create training, validation and testing sub-sets.

METHODS

First, the model for brain age prediction was trained. The model takes as an input structural MRI scans of size $196 \times 188 \times 232$ and is represented by a Squeeze-and-Excitation Network [7]. The model was trained using Mean Absolute Error (MAE) loss and achieveed MAE of 3.8 years.

Determining a population-wide brain ageing profile from a pretrained model using DEC

DEC learns the mapping, which is parameterized by a deep neural network, from the data space to the embedded space, in which it optimizes the clustering objective. In extracting the brain ageing profile using DEC it is hypothesized that all features defining a brain age are contained in an MRI scan, and that profile intervals can be defined by considering the similarity of such features. We wish to divide the range of K ordinal labels (samples' age labels, 17-96 years) into B intervals, which are defined by their centroids in the embedded space of input samples. We consider the profile as a set of intervals, $I = [i_1, \dots, i_B]$, where each bin contains set of ages such that $i_b = \{a_1, \dots, a_L\}$ and L is the bin length. Samples are embedded using the feature extractor part of the pretrained model. We would like to enforce data ordinality, as brain ageing is inherently ordinal in nature. Therefore, we construct the intervals of the profile in a tree-like structure. We first train a DEC model on a training set to divide the data range into 2 intervals. The "wall" defining this division is found by assigning each sample of the test set to one of the two intervals. A subset of samples from the test set is then considered for which prediction error from the classifier is smaller than a sum of the average error and its standard deviation. For each age label present in the range divided into 2 intervals a mode interval label is determined. A "wall" is set at the first age label (starting from the lowest) at which the mode interval changes from 1 to 2. The resulting intervals are then iteratively divided into two. The smallest interval is set by the MAE of the feature extractor used, rounded to the closest integer.

Constructing saliency maps describing the population-wide ageing profile

These saliency maps represent average features describing each interval of ageing. First, the subject-specific maps are generated for all the samples in the test dataset. Secondly, for producing a saliency map describing a bin, i_b, the subject-specific maps are averaged over all samples labelled with ages belonging to the bin. Thirdly, resulting map is multiplied by the MNI152 head mask in order to supress the background noise. This is possible as data pre-processing included registering each sample onto the MNI152 template.

RESULTS

Bin	1	2	3	4	5	6	7	8	9	10	
Interval	[17,22]	[23,27]	[28,39]	[40,45]	[46,51]	[52,67]	[68,71]	[72,76]	[77,85]	[86,92]	MNI152

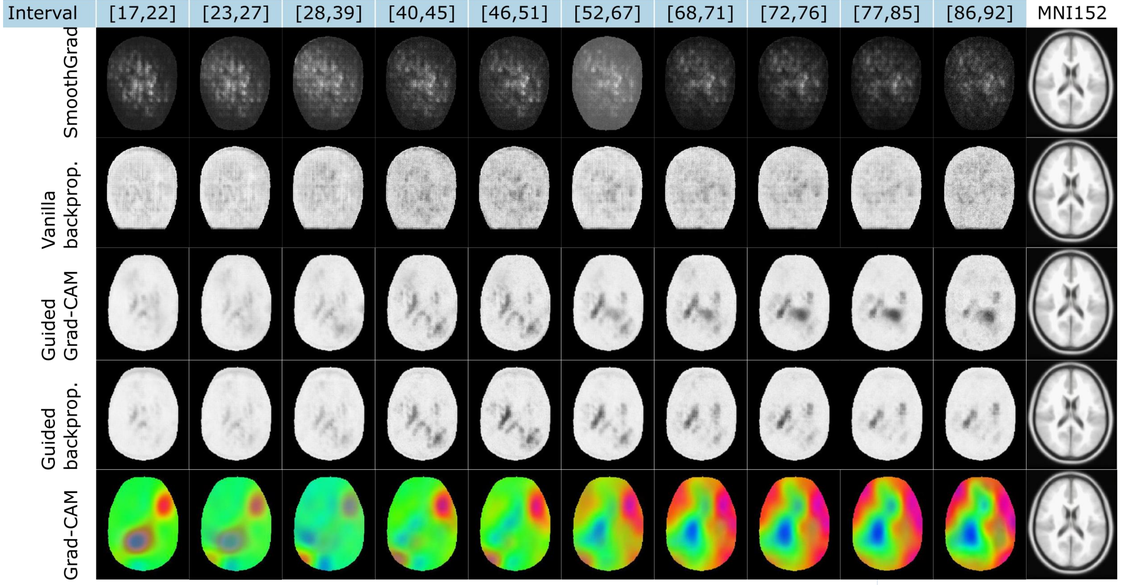

Table 1. Predicted ageing profile intervals using the DEC method and corresponding saliency maps.

REFERENCES

1. Xie et al. Unsupervised Deep Embedding for Clustering Analysis. In: 2016 International conference on machine learning, pp. 478–487, 2016.
2. Simonyan et al. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ArXiv e-prints, 2013.
3. Springenberg et al. Striving for Simplicity: The All Convolutional Net. ArXiv e-prints, 2014.
4. Smilkov et al. SmoothGrad: removing noise by adding noise. ArXiv e-prints, 2017.
5. Selvaraju et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. ArXiv e-prints, 2016.
6. Greve and Fischl. Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48(1), pp. 63 – 72, 2009.
7. Hu et al. Squeeze-and-Excitation Networks. ArXiv e-prints, 2017.

CONCLUSIONS

- The maps produced using the methods of vanilla backpropagation and SmoothGrad do not have enough sensitivity and suffer due to the noise in the data as these methods produce noisy maps even on natural images where data contrast for important features is much stronger compared to the MRI scans.
- The methods of guided backpropagation and guided Grad-CAM, which is the guided backpropagation and Grad-CAM combined, resulted in very similar maps. Both methods concentrated at the ventricles which is a known ageing-related feature and intensity increases with age as ventricles enlarge with age.
- The Grad-CAM method concentrated at ventricles in younger subjects and at brain atrophy in older ones as it occurs with age.