Age Regression from Brain MRI; Group: 54

Alvaro Prat, Théophile Sautory; {ap5915, ts4915}@imperial.ac.uk

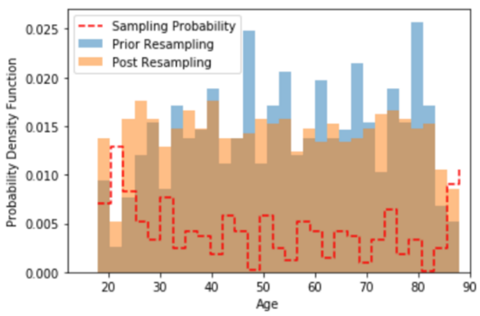
1 Summary

This report presents three machine learning (ML) methods for age regression based on 3D brain images: (i) Using volumetric features of each brain matter as input features; (ii) Performing PCA on grey matter maps to extract features to be regressed by a separate model; (iii) Using grey matter maps as input features to a convolutional neural network (CNN). The optimal results were obtained using Ridge regression on the PCA features, reaching an MAE of 5.21 for the test set with an r^2 score of 0.90.

2 Preprocessing

The courseworks' initial settings downsample the 3-D images via an image spacing of 3. In contrast we upsample this back to an image spacing of 2 to get a higher resolution, increasing potential segmentation accuracy. Indeed this increases the amount of data carried by an image batch by a factor of ∼3.4. To account for this, and to reduce redundancy during segmentation and age regression, we cropped the image, considering that ∼60% of the data is composed of black background. Cropping was also necessary as it reduced GPU memory allocation by a factor of ∼1.8. This was crucial during training of the large ResNet models proposed for image segmentation in part A and CNN age regression in part C.

Analysing the age distribution we find that several age groups are under-represented. In order to properly predict all age groups within the training range (interpolation), we resample 20% of the training set with a probability inversely proportional to the age group frequency (Figure 1 (**Left**)) and concatenate it to the original set. This produces a more uniform-like distribution. Note that during resampling we add gaussian noise to the images with 0 mean and a standard deviation proportional to the resampling probability and to the standard deviation of the feature. The latter heuristic is considered with the aim of reducing modality (and hence ease of overfitting) in the training set.



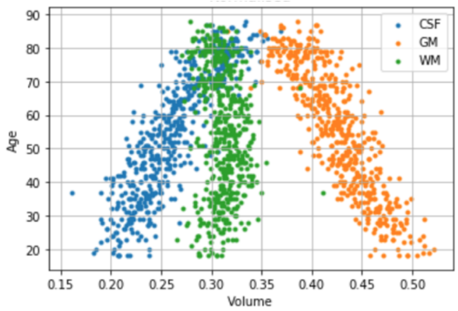


Figure 1: **(Left)** Example of the data distribution after upsampling, **(Right)** Normalised volume features from segmentation produced by the segmentator ResNet.

3 Part A

For the segmentation task we did not attempt any resampling due to the limited dataset size (47 images). Nevertheless, crop and upsampling (resolution) preprocessing steps described in section 2 were carried as this proved beneficial (Mean DICE scores improved up to 10% for GM and CSV). Due to the large feature space we propose a fully-convolutional 3-D Residual Network (ResNet)-type architecture, allowing gradients to propagate further up the hidden layer stream: 4 residual blocks are sequenced with increased channels and skip connections. No feed-forward networks (FFNNs) were used, allowing spacial awareness, resistance to over-fitting (Figure 2 **(Left)**) and exploitation of GPU computation speed boost. Both quantitative (Figure 3) and qualitative (Figure 2 **(Right)**) results are highly satisfactory, with outstanding DICE and Volume Similarity (VS) scores for all classes, with the exception of CSV: high VS and a lower DICE score suggests misclassification along the boundaries. This is expected though, considering it is a fluid.

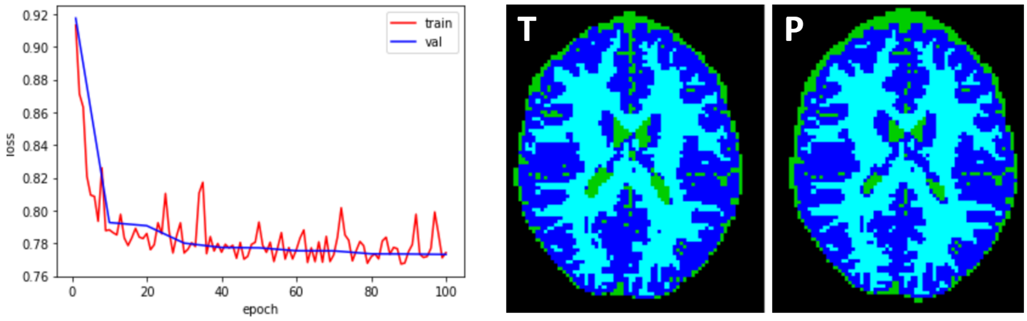


Figure 2: **(Left)** 3-D ResNet Segmentator convergence during training; **(Right)** Comparison between gold-standard **T** and predicted **P** segmentations.

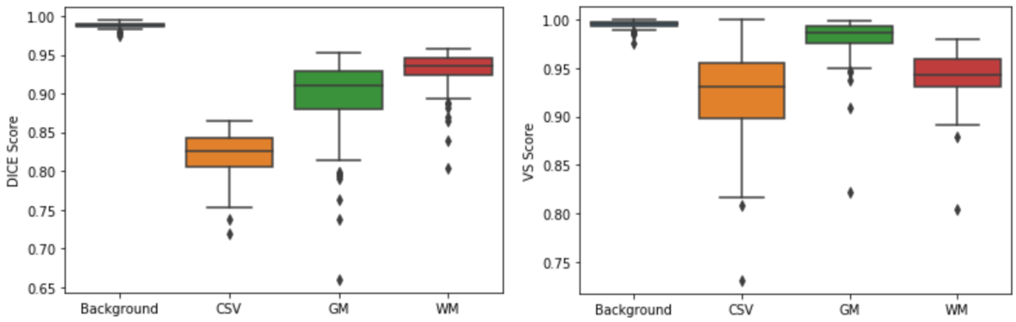


Figure 3: **(Left)** DICE and **(Right)** Volume Similarity (VS) for the 4 segmentation classes.

Figure 1 then revealed that the normalised grey matter and CSF tissue volume percentages were linearly correlated with patient age, whilst WM seemed to be independent of age. We therefore limited our exploration to volume feature sets {SCF, GM, WM} and {SCF, GM, $\emptyset$} only. The latter sets were used as inputs for 4 regressors: Random Forests, AdaBoost, Ridge Linear Regression and Support Vector Machine (SVM). A 2-fold cross-validated grid search was employed to select the best hyperparameters.

4 Part B

The grey matter maps were used as features for the age regression. The pre-processing included our custom resampling, image smoothing and downsampling. The latter steps enable to analyse the image at different scales whilst reducing the noise and compensating for errors in the voxels spatial normalisation. The Gaussian and Anisotropic diffusion functions were used to filter the images in conjunction with scale reduction factors of 2 and 4. Applying anisotropic filtering and a reduction by 4 proved to provide our optimal perfomance. The effect of downsampling by a factor of 4 and appying anisotropic diffusion is shown in Figure 4. A grid search 2-fold cross validation was then performed. The PCA was fit to the train set and then used to transform the validation set and the results are exhibited in Figure 5.

5 Part C

A similar ResNet to the segmentor model in part A was devised for age regression. The main difference lies in using larger strides in the residual blocks to reduce the dimensions of the 3D image along the hidden layer stream to 1 (regression). In this case, resampling was also considered in addition to the preprocessing steps outlined in section 3, for both cross-validation and full training. Note that during training we experienced severe overfitting. In order to deal with this we included severe dropout layers by adding FFNN stages post CNN un-folding.

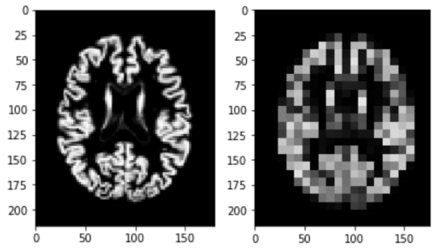


Figure 4: **(Left)** Non-processed Grey Map slice; **(Right)** Downsampled and Filtered Grey Map.

6 Age Regression Results

Figure 5 shows the cross-validated MAE performance of the selected regressors for parts A, B and C. Moreover, results at test time are displayed in Figure 6. In Part A, the optimal regressor corresponded to an rbf-kernel SVM, resulting in average MAE scores of 6.64 in cross-validation and 6.33 during testing. This result was satisfying considering the simplicity of the features. The results were further improved using PCA pre-processing and Ridge linear regression in Part B: MAE reached 5.32 on the cross validation and 5.21 on the test set. Figure 5 **(Right)** shows the results from the CNN regressor. We achieved an MAE of 6.8 on the cross validation and 6.1 on the test data. Considering the complexity of the model (+1.1M parameters) and the hyperparameter tuning perfomed, the results are insufficient: over-fitting the training set severely affected testing, where, in contrast, training MAE hovered around 2.9. Although the state-of-the-art MAE results hover around 2.7 [1], these results are gratifying considering that we trained on down-sampled images (image spacing of 2) due to time and computational constraints (GPU Cuda errors were frequently raised at image spacing of 2).

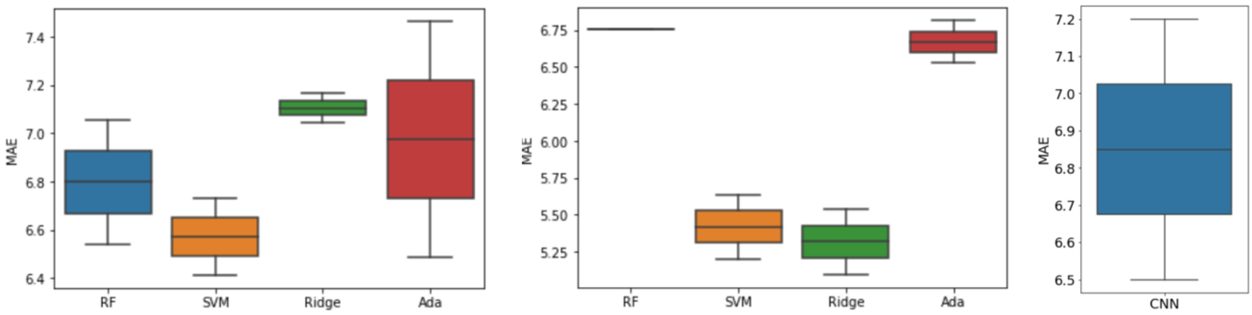


Figure 5: MAE results for **(Left)** Part A; **(Centre)** Part B; **(Right)** Part C on cross validation

Although resampling slightly reduced MAE, we found that the r^2 score increased as a consequence: outlier frequency and offset magnitude were both reduced by the increased generalisation brought by resampling. Noteworthy is that during testing of the 4 models in Part A, an MAE of 6.22 was achieved using the segmentations constructed via the ResNet segmentor. This outperfomed the models trained using the reference segmentations. This suggested that the proposed model found a more structured and interpretable pattern than that produced by subjective human segmentations.

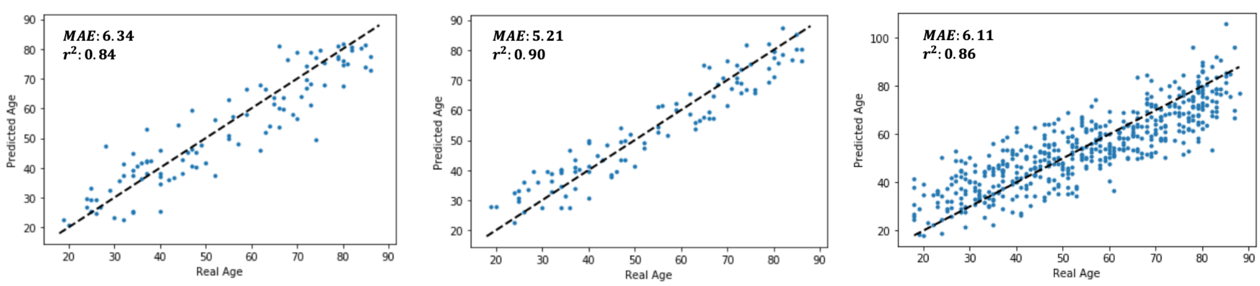


Figure 6: Scatter plots for age regression. **(Left)**: Part A; **(Center)**: Part B; **(Right)**: Part C.

7 References

[1] Beheshti, I., Nugent, S., Potvin, O., & Duchesne, S. (2019). Bias-adjustment in neuroimaging- based brain age frameworks: A robust scheme. NeuroImage: Clinical, 24, 102063.