Age Regression from Brain MRI
Group: 27

Michael Zotov, Romain Molinas, Alexandre Allani
{michael.zotov19, romain.molinas19, alexandre.allani19}@imperial.ac.uk

1 Summary

We examine three different supervised approaches to age regression from Brain MRI data. First, we identify brain tissues using a segmantation model inspired by U-Net architecture. Based on segmented images, we compute tissue volumes and use them as features for regression. The second approach is based on grey matter maps. We use Principal Component Analysis to reduce dimensionality of the maps and perform age regression on them. For the first two feature sets, we employ multiple regression models. For the last approach, we adapt a convolutional architecture inspired by SOTA models for imaging tasks. We use full grey matter maps for regression.

2 Part A

Our segmentation model was inspired by U-Net [4]. To account for 3 dimensional data, we replace the 2D-convolutional, BatchNorm and MaxPool layers with 3D counterparts. Moreover, we add *residual connections* to the convolutional blocks of network as seen in [5] We use Adam optimizer [2] with lr of $1e-3$. We train for 200 epochs with batch size of 2 due to memory considerations. The average Dice scores are [0.99, 0.83, 0.91, 0.93] for [Background, CSF, WM, GM] on validation set provided.

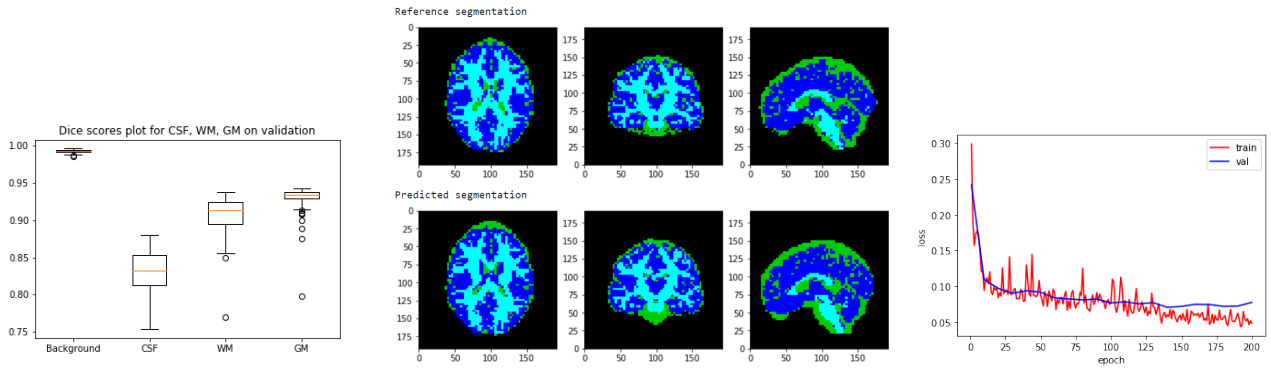

Figure 1: **Left**: Brain segmentation dice score box and whisker plot on validation set. **Middle**: Our segmentation vs ground truth. **Right**: U-Net training and validation loss

Using segmentation maps obtained through U-Net, we compute brain tissue volumes to use them as features for regression. We employ a range of regression models: SVR (Linear, Polynomial kernels), Linear Regression, Tree Regression with XGBoost and a Feed-Forward Neural Net. We obtain our best results with a FFNN with 3 hidden layers and ReLU activations: 8.15 MAE on two-fold cross-validation. We then retrain the model on full dataset to obtain 7.92 MAE on test set.

Models	Cross Val MAE	Cross Val R^2	Test MAE	Test R^2
Linear Regression	8.75	0.65	9.01	0.70
SVR with Linear kernel	8.78	0.64	8.95	0.70
SVR with Polynomial kernel	8.19	0.67	8.07	0.74
XGBoost Tree Regression	8.40	0.66	8.21	0.72
Feed Forward Neural Network	8.15	0.66	**7.92**	**0.75**

Table 1: Cross validation score comparison for Part A

3 Part B

Here we predict age using low-dimensional grey matter maps. As suggested, we down-sample the maps. We then apply smoothing filters to the images in order to reduce noise. Two types of smoothing were considered: Discrete Gaussian Smoothing and Gradient Anisotropic Diffusion. The latter provided us with better results.

We then run PCA retaining 95% of the variance. After applying the dimensionality reduction, we have 441 features per image. Based on this dataset we fit the following regression algorithms: SVR, Linear Regression and Regression Trees with XGBoost. Again, full results can be found in the appendix. We obtain best performance with Linear Regression. We apply the same procedure for training as in part A. Our best validation and test MAE are 5.93 and 5.38 respectively.

Models	Cross Val MAE	Cross Val R^2	Test MAE	Test R^2
Linear Regression	5.92	0.84	**5.38**	**0.89**
SVR with Linear kernel	5.78	0.84	6.32	0.84
SVR with Polynomial kernel	5.76	0.85	5.65	0.88
XGBoost Tree Regression	7.71	0.74	7.24	0.78

Table 2: Cross validation score comparison for Part B

4 Part C

Our approach was inspired by success of ResNet architectures on imaging tasks [1]. We reason that residual-style architectures perform well regardless of the task at hand as they allow linear maps to be learned where appropriate. We hence adapt the architecture by replacing 2D-convolutional, BatchNorm and MaxPool layers with 3D counterparts. Additionally, we make the following changes:

- Reduce number of filters output at each layer by a factor of 2.
- Increase kernel size from 3 to 7 in encoder layer and max pooling window from 4 to 8.
- Replace output layer with a linear layer for regression.

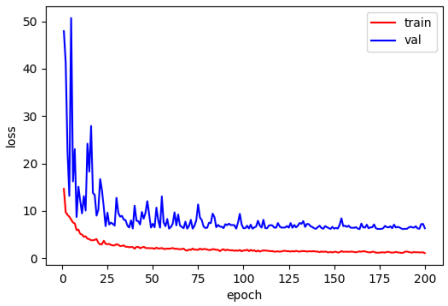

Figure 2: Average ResNet training and validation loss for two-fold cross-validation

We use AdamW optimizer [3] with lr=0.0001, batch size of 4 (memory considerations). We use MAE loss as we find it to make training more stable. We perform light hyperparameter tuning using suggested two-fold cross-validation. We then train the model for 200 epochs with two-fold cross-validation to obtain loss curves to assess the training and generalisation. Finally, we retrain the network on 400

samples, using 100 samples as a smaller validation set for early stopping. We then test our best model (obtained after 63 epochs) on the provided test set to obtain generalisation performance with low bias. The validation MAE is 6.52, while the test MAE is 4.88.

5 Age Regression Results

In our experiments, we obtain best performance of 4.88 MAE on test set with a Deep Convolutional architecture on full scale grey matter maps, with simple Linear Regression on dimension reduced maps being a close second. CNN, however, is the method that takes the longest to achieve desired results.

Figure 3: Scatter plots for age regression for three different approaches.

Models	**Cross Val MAE**	**Test MAE**	**Test R^2**
Tissue Volumes + FFNN	8.15	7.92	0.75
PCA + LinReg ion Grey Matter	5.92	5.38	0.89
ResNet on Grey Matter	6.32	**4.88**	**0.90**

Table 3: Score comparison for best models in parts A, B, C.

We notice that for all of our best models the test performance is very consistent with validation: it is of similar magnitude and decreases, as more data is available for training (250 vs 500 or 400 samples). An interesting observation is an out-of-proportion (23%) improvement for our ResNet architecture, which could suggest that while the other two models are close to their maximum capability, ResNet could further improve if more data were available (as it was trained on 400 samples only, and 100 samples of the train set were used to validate for early stopping/checkpointing). Our tissue volume approach shows high variance in errors, with some being more than double the mean error. The other two models show lower variance, with predictions contributing more evenly to the average error.

References

[1] He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (Dec. 2015). arXiv: 1512.03385.

[2] Kingma, D. P., and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (Jan. 2017). arXiv: 1412.6980.

[3] Loshchilov, I., and Hutter, F. Decoupled Weight Decay Regularization.

[4] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]* (May 2015). arXiv: 1505.04597.

[5] Zhang, Z., Liu, Q., and Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters 15*, 5 (May 2018), 749–753. arXiv: 1711.10684.