Machine Learning for Imaging – Coursework Report
Age Regression from Brain MRI

Group Leader: **Torres**

Khayle Torres, Nelson da Silva
{kt1719, nmd19}@imperial.ac.uk

1 Part A

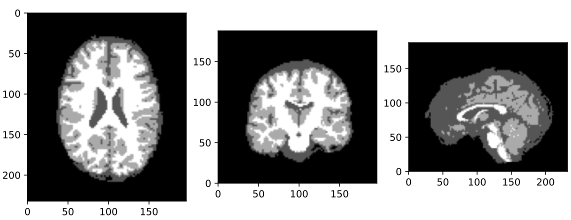
1.1 Preprocessing steps

- **Normalisation**:

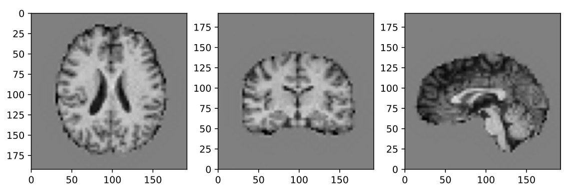
  This is normalisation of the **intensities** of the images to have a **zero-mean** and **unary standard deviation**.

- **Resampling of data**:

  This applies a transformation to the original image in order to create a new image with different pixel dimensions. The resampling used uses interpolation to create a new image with the specified dimensions and image spacing. This is to transform the originally anisotropic MRI dataset **(98x116x94)** into an isotropic dataset (either **(64x64x64)** or **(96x96x96)**).



(a) Original Image



(b) Preprocessed Image

Figure 1: Original vs Preprocessed Image

1.2 Implementation

Approach A consists of training an architecture to produce segmentation maps from input scans using 47 training images and 5 validation images. This is used to produce 500 segmentation maps, from which tissue volumes are calculated as feature vectors. These are adjusted and trained with regression methods through 2-fold cross validation to predict a patient's age from their corresponding brain segmentation. The optimal feature vector and regression method combination is evaluated on a final hold-out test set of 100 images.

1.3 Tissue Segmentation

To produce the segmentation maps, a U-Net architecture was implemented (Figure 2). U-Net consists of an **encoder** and **decoder**. The encoder refers to the first half of the network and consists of several convolution and pooling operations. The encoder convolution filters are used to extract features from the image and the pooling layers reduces the spatial dimensions of the feature maps. This allows multiple features at different spatial dimensions to be extracted by the encoder. Each feature map generated at each spatial dimension is saved for the decoder step.

The second half of U-Net (the decoder) consists of numerous transpose convolution, convolution and appending operations. The transpose convolution operations are in order to upsample the feature maps into a higher spatial dimension. This is then appended to the feature map generated at the same resolution from the encoder in order to recover spatial information lost during the downsampling process. The combined data is then convoluted together to combine information from both feature maps. This allows the map to contain good feature information whilst maintaining good resolution.

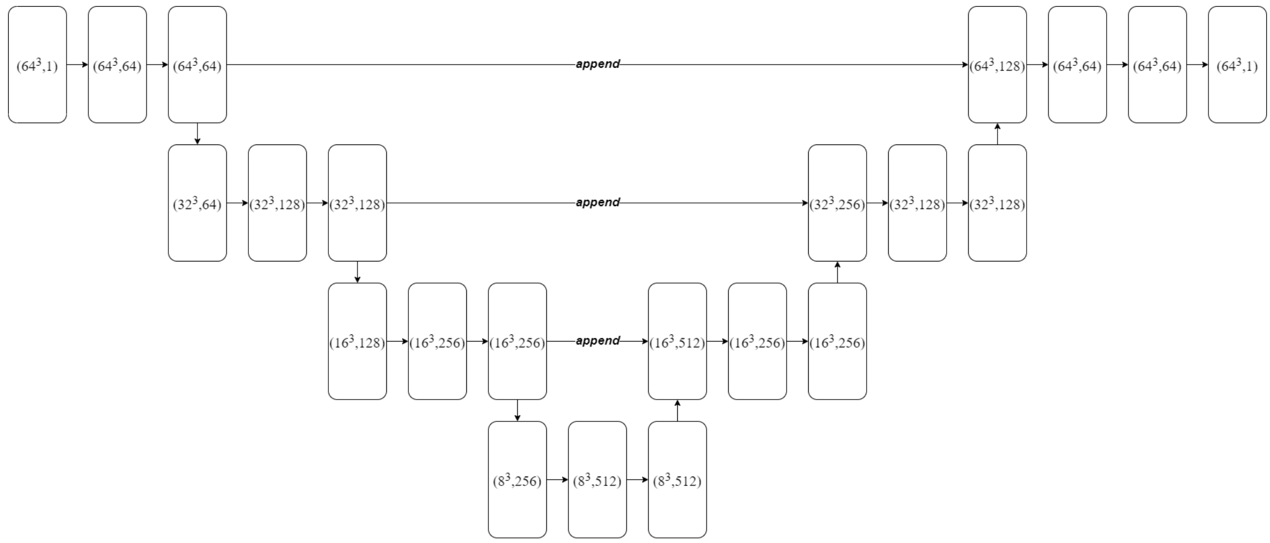
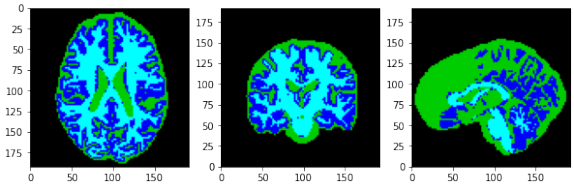
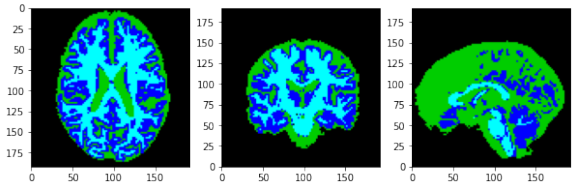


Figure 2: U-Net Diagram

Figure 3 shows an example predicted segmentation compared to it's reference. Evaluation of these generated maps was conducted through measurements of Dice Similarity Coefficient (DSC) and cross-entropy loss shown in Figure 4. The resulting DSC score distributions per segmentation map is visualised in Figure 5 and the final cross-entropy loss recorded on the validation set was 0.0657.

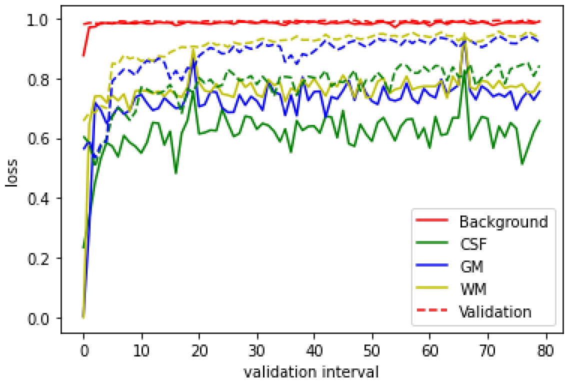


(a) Predicted segmentation example

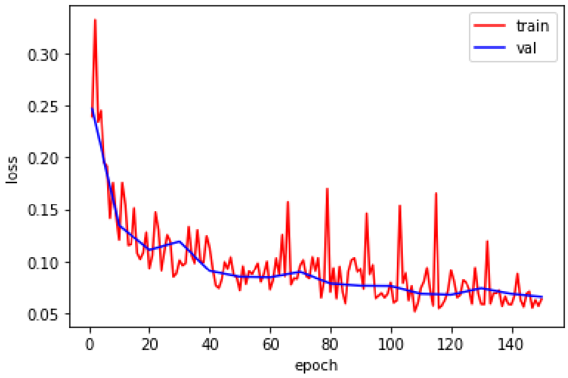


(b) Reference segmentation example

Figure 3: Predicted vs Reference segmentation masks



(a) Dice similarity coefficient



(b) Cross Entropy loss

Figure 4: Varying DSC and loss over several epochs when training

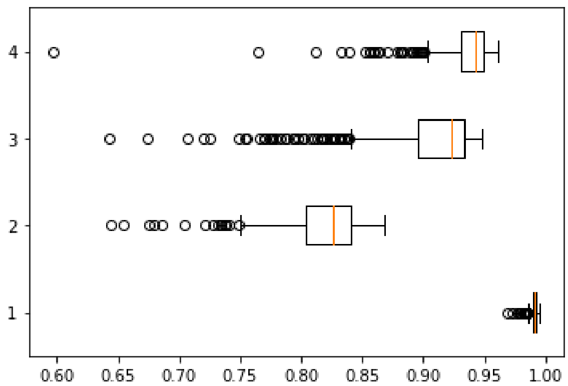
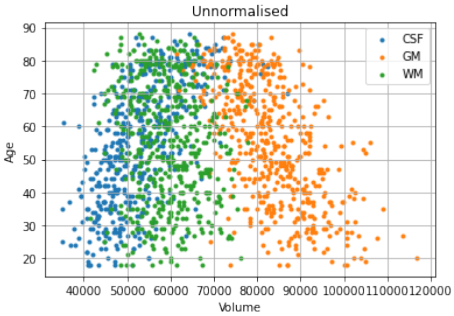


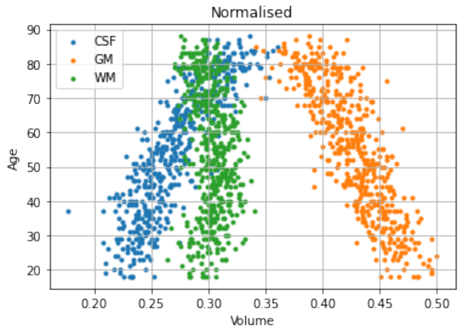
Figure 5: Box-and-whisker plot for the DSC scores of the segmentation maps produced for age regression training. 1 - Background, 2 - Cerebrospinal fluid, 3 - Grey Matter, 4 - White Matter

1.4 Age Regression

An informative feature vector must be produced to feed into a chosen regression learning method. The number of voxels belonging to the main three classes : Cerebrospinal fluid (CSF), Grey Matter (GM) and White Matter (WM), are measured and plotted to identify correlations between them and the age of the patient as shown in figure 6a. The lack of a distinguishable relationship pushes the need for normalisation where volumes are taken relative to one another which can be seen in figure 6b. These normalised values and potential combinations of them may lead to predictive models.



(a) Raw volume feature values

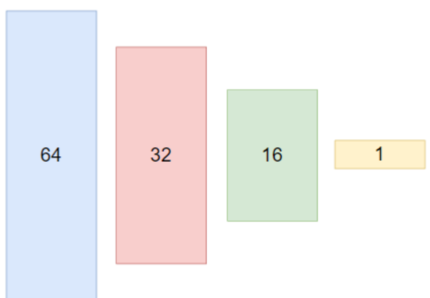


(b) Normalised volume feature values

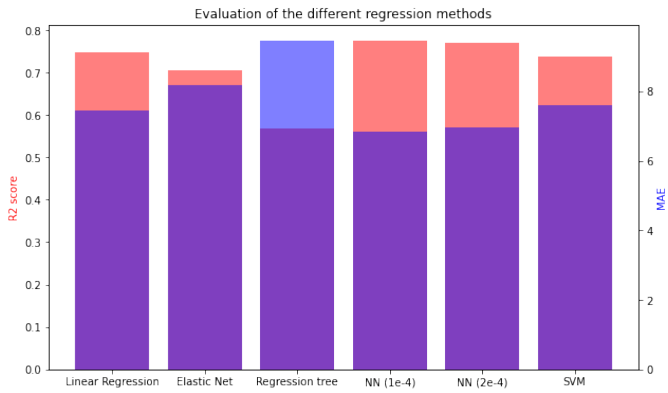
Figure 6: Volume values vs age

Different combinations of these features are evaluated in table 1 across several regression learning methods by comparing mean average error (MAE) and R^2 score produced from performing 2-fold cross validation on the training set. While volumes of CSF and GM individually display standard performances, forming combinations of these values reduces MAE and increases R^2 score, a metric that measures how much variance in the predicted age can be determined by changes in the input features. This trend is common across all regression methods yet performance begins to falter at much more complex feature vectors.

Several regression models were trained and their results are shown in table 1. When training using two-fold cross validation, a simple linear regression model and a fully connected neural network (FCNN) (shown in Figure 7a) displayed considerable results. Different levels of regularisation were looked into within these methods including Elastic Net which integrates L_1 and L_2 penalties, and the same FCNN but with an increased α parameter which encourages smaller weights. The results (Figure 2) indicate that the neural network benefits from increased regularization. Figure 7b visualises performance differences between all regression methods.



(a) Fully connected neural network, optimizer=L-BFGS



(b) Performance of different regression methods (red = R^2 score, blue = MAE)

Figure 7: Neural network structure (Left) and the metrics of the main regression methods evaluated on the test set (Right)

Feature vectors	Linear Regression	Decision Trees	NN ($\alpha = 1e-4$)	SVM
x_1	7.93, 0.720	9.95, 0.507	7.30, 0.755	7.88, 0.720
x_2	8.14, 0.690	10.56, 0.467	8.12, 0.688	8.84, 0.648
x_3	15.5, 0.0107	18.4, -0.648	16.0, -0.105	15.8, -0.0704
x_1, x_2	7.51, 0.742	9.25, 0.582	6.91, 0.760	7.60, 0.737
x_1, x_2, x_3	7.52, 0.740	9.85, 0.533	**6.87**, 0.769	8.02, 0.714
x_1, x_2, x_3 $\frac{x_1}{x_2 * x_3}$	7.30, 0.751	9.39, 0.568	7.59, 0.714	8.21, 0.701
x_1, x_2, x_1^2, x_2^2 $x_1^2 + x_2^2 + x_3^2$	7.31, 0.753	9.28, 0.588	**6.87**, **0.770**	7.988, 0.711
$x_1, x_2, x_3, x_1 * x_2,$ $x_2 * x_3, x_3 * x_1$ $x_1^2 + x_2^2 + x_3^2$	7.28, 0.756	9.46, 0.560	6.97, 0.760	7.01, 0.763

Table 1: [MAE, R^2 score] of different feature vectors with different regression methods. x_1, x_2, x_3 represent the normalised volumes of CSF, GM and WM.

Feature vectors	ElasticNet	(NN ($\alpha = 2e-4$))
x_1	7.93, 0.720	7.36, 0.751
x_2	8.14, 0.690	7.97, 0.703
x_3	15.5, 0.00107	15.4, 0.0153
x_1, x_2	8.32, 0.695	**6.75**, **0.775**
x_1, x_2, x_3	7.76, 0.728	6.94, 0.769
x_1, x_2, x_3 $\frac{x_1}{x_2 * x_3}$	8.00, 0.714	7.10, 0.748
x_1, x_2, x_1^2, x_2^2 $x_1^2 + x_2^2 + x_3^2$	7.75, 0.729	6.89, 0.769
$x_1, x_2, x_3, x_1 * x_2,$ $x_2 * x_3, x_3 * x_1$ $x_1^2 + x_2^2 + x_3^2$	7.56, 0.741	7.01, 0.763

Table 2: [MAE, R^2 score] of different feature vectors with different regularization levels

2 Part B

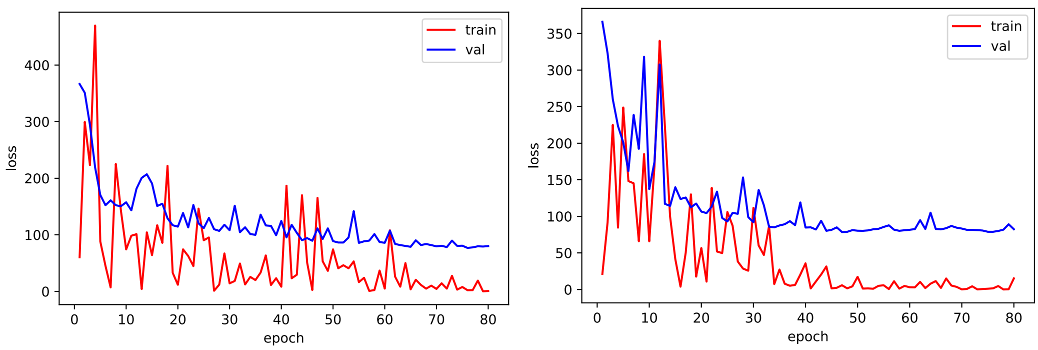
2.1 Preprocessing

Preprocessing similar to **Part A** consists of a **normalisation** and a **resampling step**.

2.2 Implementation

This approach directly predicts patient age from the input brain scan by using a CNN in a 'one-shot' learning manner. Models are evaluated in a similar way to **Part A** using 2-fold cross validation on 500 training images and then finally evaluated on 100 unseen test images. The Adam optimiser was used to allow the model to have an adaptive learning rate and improve performance. On top of this, batch normalisation and dropout are also incorporated in training to improve performance and prevent overfitting.

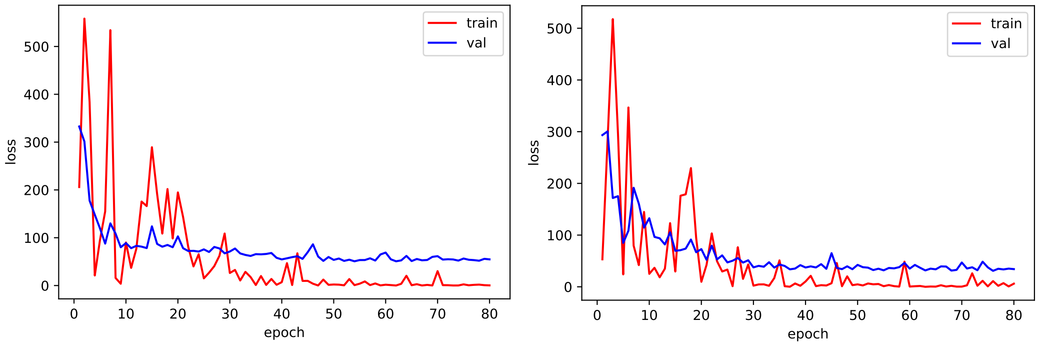
As a starting model, the **LeNet** architecture was used and modified for the use of regression. The LeNet modified model yielded a Cross Validation **MAE Score of: 7.53** and would be used as a base model reference for future models to be made. It is clear that the LeNet modified model is too simple for the data as it is designed for 32x32 image classification.



(a) LeNet Cross Validation Loss Graph (b) Final Model Cross Validation Loss Graph

Figure 8: Difference in Cross Validation Graphs between LeNet modified model and final model. Using MSE for loss.

To improve on this model, more features would need to be extracted at more spatial dimensions. Out of the several models made, the final model (Figure 10) would consist of multiple convolutional, pooling and appending operations. Each convolution step would be done in 2 parts, 1 5x5 Kernel and 2 3x3 Kernels. The outputs from the 2 parts would then be appended and then either convoluted again together or pooled. This was to allow the model to extract as many different features in as many spatial dimensions as possible. The last couple of layers consisted of dense, fully connected layers. The difference in loss functions are seen in Figures (8,9).



(a) LeNet Modified Scatter Plot (b) Final Model Scatter Plot

Figure 9: Difference in Re-Train Graphs between LeNet modified model and final model. Using MSE for loss.

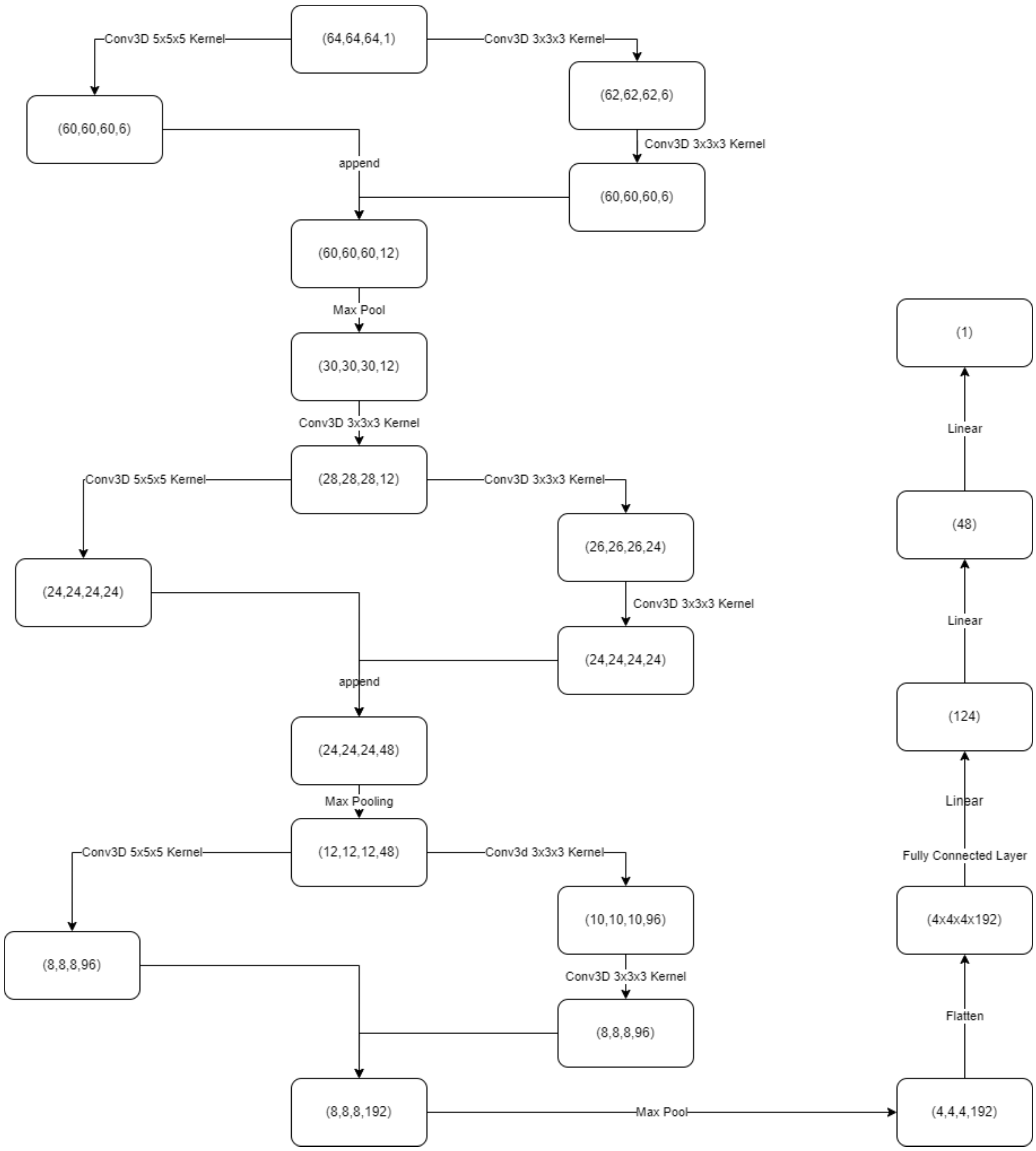


Figure 10: Part B Model Diagram for (64,64,64) size input

After some hyperparameter tuning, the new model scored significantly better than the LeNet (2). The model is then slightly modified (Figure 11) in order to take a higher resolution input image (96,96,96) which proved to perform even better (3, 4).

$$Cross\ Validation\ MAE\ LeNet\ of\ input\ (64, 64, 64):\ 7.53 \tag{1}$$

$$Cross\ Validation\ MAE\ Final\ Model\ of\ input\ (64, 64, 64):\ 6.67 \tag{2}$$

$$Test\ MAE\ Final\ Model\ of\ input\ (64, 64, 64):\ 6.73 \tag{3}$$

$$Test\ MAE\ Final\ Model\ of\ input\ (96, 96, 96):\ 5.53 \tag{4}$$

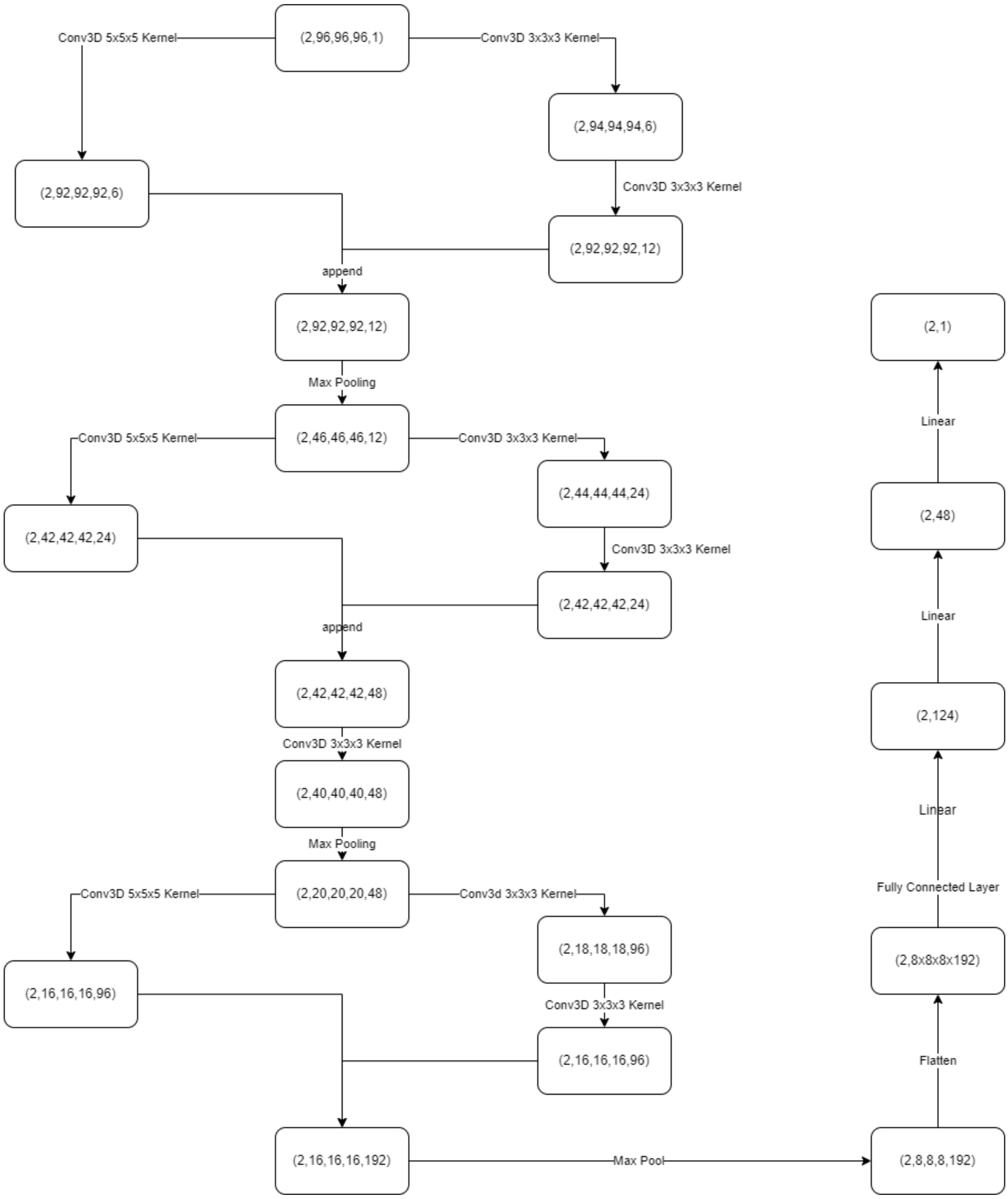


Figure 11: Part B Model Diagram Modified for (96,96,96) size input

3 Age Regression Results

3.1 Test Results

In part A, the feature vector combination of x_1 and x_2 with the FCNN ($\alpha = 2e-4$) provided the best training results thus far. This, alongside the best model in part b, were used on the test set to produce the results shown in table 3.
A visualisation of the predicted age against the actual age for each model can be seen in Figure 13 as well corresponding DSC scores for the generated segmentation in part A (Figure 12).

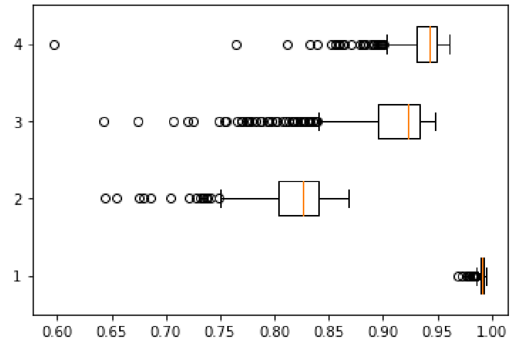


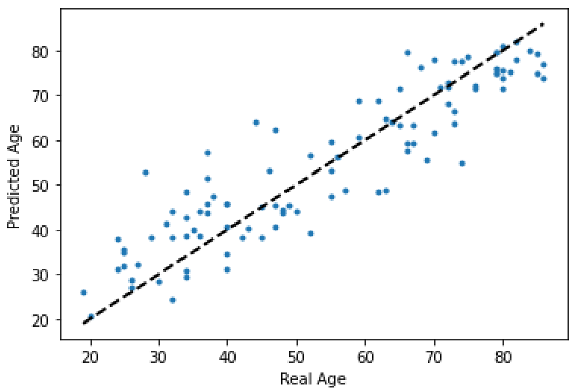
Figure 12: Part A segmentation DSC scores box-and-whisker plot

3.2 Comparison

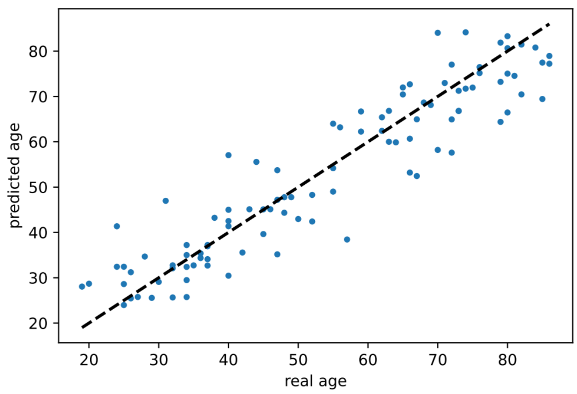
The part B model, designed in a 'one-shot' manner, exhibited superior training and test results than the model in part A (see table 3 and figure 13 for prediction results); it exceeded the R^2 score and had a much lower MAE. This may be owed to fact that the CNN trains the model end-to-end simultaneously whereas in part A, the segmentation and regression are done separately. Additionally, this separation causes information to be lost in between whereas the model in part B benefits from features from earlier layers being used throughout the network.

Model	Cross-validation training	Test
Part A	6.75, 0.775	6.70, 0.820
Part B	6.67, 0.814	5.53, 0.862

Table 3: [MAE, R^2] of each model on the final hold-out test set



(a) Part A Model Scatter Plot



(b) Part B Model Scatter Plot

Figure 13: Testing results of both models on 100 unseen test images