Alzheimer detection using deep learning

Roxana Zachman-Țîșcă
Babeș-Bolyai University, Cluj-Napoca

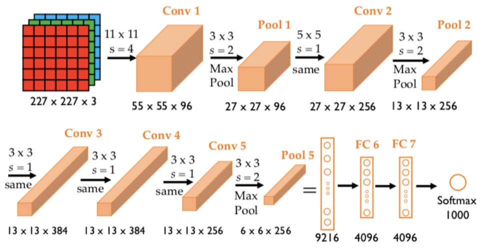
MOTIVATION

- Early detection is vital in administrating proper treatment to patients and delaying the severe symptoms of Alzheimer’s disease
- Deep learning has proven incredibly powerful in a large range of field, including healthcare

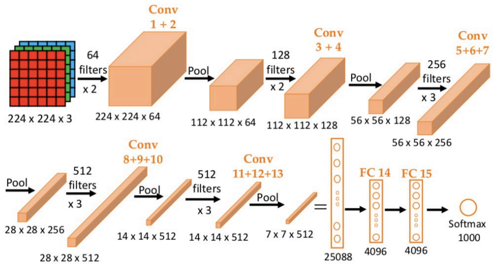
DEEP LEARNING

- We use a simple deep learning architecture for detecting AD (Alzheimer’s Disease) at various stages
- The architecture is based on the principles of the first successful CNN (convolutional neural network) architectures using in Computer Vision: **AlexNet** and **VGG**

AlexNet

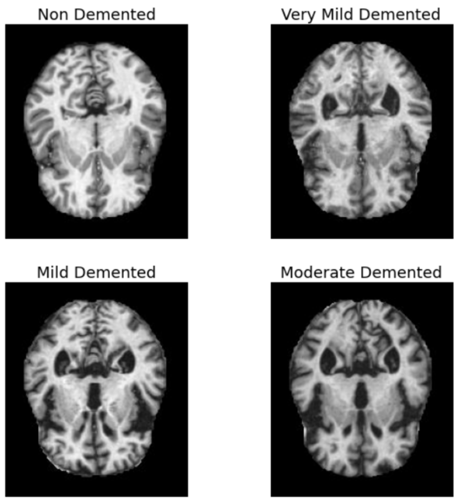


VGG



DATASET

- We used the **Alzheimer’s Dataset** from **Kaggle**, a dataset comprising of **6.4K** brain MRI scans of 4 classes: **NonDemented**, **VeryMildDemented**, **MildDemented**, and **ModerateDemented**



The images from this dataset of high quality, with resolution of **208 x 176** in all samples
Problems we found with this Kaggle dataset are that **(1)** the number of samples is quite low at **6.4K**, increasing this number would greatly help and **(2)** the dataset is pretty imbalanced. It consists of:

1. **3.2K NonDemented** samples
2. **2.2K VeryMildDemented** samples
3. **896 MildDemented** samples
4. **64 ModerateDemented** samples

We can see how we have very few samples in the later stages of the disease, and this can raise a number of issues in our model

OUR MODEL

- As already stated, we borrowed principles from the AlexNet and VGG architectures and added specific changes to create a model fit for our use case
- To be more precise, we use multiple blocks of the combination: convolutional layer which keeps the resolution of the feature maps and increases the number of feature maps **followed by** MaxPool layer which decreases the resolution of the feature maps
- Each convolutional layer is followed by a BatchNorm and then a ReLU activation function
- At the end of these convolutional layers we flatten the feature maps and add 2 fully connected layers which have the job of leveraging the learnt embeddings into class probabilities

RESULTS AND FUTURE STEPS

- The model was trained on 75 epochs with early stopping. Cross Entropy was used as the loss function and Adam as the optimizer. The results reached were of **~75%** accuracy across all **4 classes**, which shows that the model is really starting to learn patterns in the MRI scans.
- While the results are well below the state of the art (where 86% percent was hit in a similar model, and even as high as 93% in more complex architectures), it is a useful baseline model and opens the door for future improvements, such as:
    1. Using a pre-trained model and transfer learning
    2. Using a deeper model with residual connections
    3. Adding more augmentation techniques
    4. Improving the dataset
    5. Using test-time augmentation