

Feature map size selection for fMRI classification on end-to-end deep convolutional neural networks

Farahana Suhaimi *, Zaw Zaw Htike

Faculty of Engineering, International Islamic University Malaysia, Gombak, Malaysia

ARTICLE INFO

Article history:

Received 1 March 2018

Received in revised form

29 May 2018

Accepted 4 June 2018

Keywords:

Classification

Functional MRI

Deep learning

CNN

Feature map

ABSTRACT

The emergence of convolutional neural networks (CNN) in various fields has also paved numerous ways for advancement in the field of medical imaging. This paper focuses on functional magnetic resonance imaging (fMRI) in the field of neuroimaging. It has high temporal resolution and robust to control or non-control subjects. CNN analysis on structural magnetic resonance imaging (MRI) and fMRI datasets is compared to rule out one of the grey areas in building CNNs for medical imaging analysis. This study focuses on the feature map size selection on fMRI datasets with CNNs where the selected sizes are evaluated for their performances. Although few outstanding studies on fMRI have been published, the availability of diverse previous studies on MRI previous works impulses us to study to learn the pattern of feature map sizes for CNN configuration. Six configurations are analyzed with prominent public fMRI dataset, names as Human Connectome Project (HCP). This dataset is widely used for any type of fMRI classification. With three set of data divisions, the accuracy values for validation set of fMRI classification are assessed and discussed. Despite the fact that only one slice of every 118 subjects' temporal brain images is used in the study, the validation of classification for three training-excluded subjects known as validation set, has proven the need for feature map size selection. This paper emphasizes the indispensable step of selecting the feature map sizes when designing CNN for fMRI classification. In addition, we provide proofs that validation set should consist of distinct subjects for definite evaluation of any model performance.

© 2018 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Convolutional neural network (CNN) is one of the deep learning branches (Guo et al., 2016). Deep learning continues to be fundamental of many remarkable researches for mimicking human's intelligence known as artificial intelligence (Gollapudi, 2016). The enhancement over artificial neural network (ANN), which once became an important tool around 30 years ago, CNN has given platforms for many contributions in many fields including medical imaging field (LeCun et al., 2015; Greenspan et al., 2016). Various implementation of convolutional layers had been applied successfully such as Facebook language translation (Gehring et al., 2017), image captioning (Karpathy and Fei-Fei, 2015) and pneumonia detection that exceeds

expertise' level (Rajpurkar et al., 2017). Despite their dominance, they are entailed to a high computational cost partially due to the convolution layers implementation.

Deep CNN is the backbone of many achievements. The word 'deep' is referred to a much higher number of layers that CNN incorporates compared to ANN. For instance, a network is known as shallow network when five or fewer layers are constructed. A deep network, on the other hand, could have six and more layers. Rajpurkar et al. (2017) demonstrated on recently published project of detecting diseases using CNN, which 121 layers was used in their model. CNN groundwork and modification are bounded by many decisive factors, one of it being the count of layer. However, the layer count is only one of handful other concerns in building CNN due to computational cost entitlement.

Difference in image domains is one of the main challenges in this study. This leads to lack of feature map size explanation in many related studies. In the neuroimaging field, functional magnetic resonance imaging known as fMRI, is three-dimensional images

* Corresponding Author.

Email Address: farahana.ms@live.iiu.edu.my (F. Suhaimi)

<https://doi.org/10.21833/ijaas.2018.08.012>

2313-626X/© 2018 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

with high temporal resolution. It is robust to either control or non-control subjects. Greenspan et al. (2016) raised an issue regarding this data images. Most works to date are two-dimensional images analysis performance of publicly available ground-truth data. The ground-truth annotation by experts is scarce for neuroimaging, especially functional MRI dataset. Towards producing 'off-the-shelf' automatic disease diagnosis system, many publications, such as Huang et al. (2017) and Makkie et al. (2017) had addressed the need of ground-truth data.

Collecting fMRI data is known to be expensive in both cost- and time-wise. The cost of computation is increasing with the size of the data. Shortfall of expert annotations such as physician and neurologist diagnosis of each fMRI image set has doubled the challenge in classifying fMRI either on cost- or time-wise. Most relevant prior works on fMRI are focused on reducing the dimension with principal component analysis (PCA), decomposing voxels with independent component analysis (ICA), and analysing empirically by-subject using statistical parametric mapping (SPM). Mentioned challenges are countered in many of these methods by reducing the dimensions and producing own dataset including ground-truth data.

The one-task label is proposed in this paper for replacing the unavailable ground-truth data and simultaneously reducing the computation cost. The computational cost is reduced by limiting the dimension of the brain image slices. In this paper, we provide an in-depth investigation into feature map size selection for fMRI classification where an end-to-end deep CNN (also known as ConvNet) approach is investigated in this paper.

The remainder of this paper is organized as follows: the CNN related work on MRI and fMRI is tabulated and discussed. From that, the scope and approach of this paper are explained. Analysis and discussion on the proposed approach is followed. Then, the conclusion and recommended future works are mentioned.

2. Related works

This section details about existing approaches relating to classification of fMRI.

As noted by Greenspan et al. (2016), the number of publications devoted to medical imaging analysis has increased tremendously. The fMRI technology has arrived more than two decades ago (Kwong et al., 1992; Cohen et al., 2017). In which invites rapid growth on literature of understanding the human's brain. This major advance, on the other hand, has led us to various challenges in fMRI classification such as inter- and intra-subject signal variability, lengthy preprocessing due to noisy images and very high-dimensional images with low resolution. Significant CNN-based approach in MRI analysis researches is investigated comprehensively. With similar intrinsic variability of brain anatomy, we focus on studying the MRI analysis with CNN to interpret and incorporate the founding into fMRI studies.

Data treatment and data application are some of the contrasting elements between MRI and fMRI. For instance, during data preparation, augmenting the data is encouraged for MRI machine learning pre-processing step (Pereira et al., 2016; Valverde et al., 2017). This step is proposed either to increase the size of training data or to improve the overall accuracy. On the contrary, the augmentation used on the fMRI data will be ineffective (Valente et al., 2014) due to the brain anatomical maps nature.

In addition, MRI is commonly commercialized for medical diagnosis to certain parts of human bodies including the brain itself. fMRI applications, on the other hand, are applicable in areas such as diseases diagnosing, pre-surgical mapping and treating disorders that relate only to the brain maps. However, fMRI is not on the par as MRI commercialization as of date. The needs of well-versed radiologist and physician are pivotal in most of the fMRI services. For instance, HUSM, being one of the advanced hospital universities in Malaysia with owns dedicated department of Neuroscience has yet to have real services for fMRI.

Despite the differences, both technologies are quite similar in the classification methods and performance evaluation. The MRI of the brain has similar features to fMRI where both have same brain shapes or maps. Though the spatial resolution is very high compared to fMRI resolution, plenty of approaches had resolution down sampled (Li et al., 2014), reduced by-patch the MRI data (Kamnitsas et al., 2017) and used only region of interest (Zhao et al., 2017) during the pre-processing step. High spatial resolution is hugely important for physicians to diagnose the MRI but not equally important for some of computing approaches.

Furthermore, our proposed approach, the convolutional neural network, CNN has been applied in both domains. As shown in Table 1 and Table 2, significant efforts have been made to develop an effective MRI and fMRI analysis using CNN. Distinct implementation of CNN in each of the approach suggests there are many concerns in building CNN. The test data division, convolutional layer counts and feature map sizes selection are few grey areas in evaluating the CNN performance. These points will be addressed briefly in the next section.

Using 830 subjects MRI with two convolutional layers is one of the earliest CNN developments in neuroimaging (Li et al., 2014). The used of small 3D data patches as the input may have contributed to low sensitivity of the approach. With the advancement of technology of faster and bigger computation, this approach could be improved with rather higher than 10 epochs. The same feature map size is used in their approach for all its convolutional layers. Relatively, as described in the Table 1, sensitivity of CNN approach has increased in recent research publications such as illustrated by Pereira et al. (2016) and Havaei et al. (2017). Additionally, CNN is applied on other body parts of MRI too. The approach by Margeta et al. (2017) on cardiac MRI

classification shows very high sensitivity for its five convolutional layers.

Although the improvement is seen in progress throughout the years, the grey areas such as the feature map sizes and layer counts selection are mentioned vaguely. For instance, few papers stated

the sources of selecting their feature maps size. Sarraf and Tofghi (2016) adopted the GoogleNet and LeNet models for their feature map sizes. While Havaei et al. (2017), Cui et al. (2016), Burgh et al. (2017) and Zafar et al. (2017) do not present their procedure on selection of feature map size.

Table 1: Overview of papers employing CNN techniques for MRI analysis

Reference	Application; remarks	Feature map sizes	Sensitivity (average)	Accuracy
(Havaei et al., 2017)	Brain tumor; 2-phases CNN, 3 convolutional layers, 5 classes, 30 times faster computation	random sizes	0.84	-
(Pereira et al., 2016)	Brain tumor; using 3D MRI, 2 types of CNN constructions, 4 and 6 convolution layers, very small learning rate: 3×10^{-7}	twice increase of feature maps sizes for every 2 or 3 layers	0.86	-
(Zhao and Jia, 2016)	Brain tumor; automatic feature maps selection, using 2D slices, 3-cascaded CNN layers with 2 to 5 convolutional layers, random selection for test data	-	-	81.0%
(Li et al., 2014)	AD; 830 subjects, using 3D MRI patches ($15 \times 15 \times 15$), 10 epochs, learning rate is fixed to 10^{-2} , random selection for test data	same	0.77	-
(Cui et al., 2016)	Bipolar disorder; 32×32 and 13×13 MRI patches, 4 convolutional layers, random selection for test data	random sizes	-	90.8%
(Burgh et al., 2017)	ALS; 3 types of input, 4 convolutional layers, independent evaluation set	random sizes	--	84.4%
(Margeta et al., 2017)	Cardiac MRI; the input is pre-classified patches using classification forest, 5 convolution layers and decrease-by-half	decrease-by-half	0.98	-

Table 2: Overview of research papers employing CNN techniques for fMRI analysis

Reference	Application; remarks	Feature map sizes	Sensitivity (average)	Accuracy
(Zafar et al., 2017)	task-fMRI; 16 subjects, MNI template is used for normalization, ROI, not and end-to-end CNN approach, 2 convolutional layers	random sizes	-	68.6%
(Zhao et al., 2017)	3D rest-fMRI and task-fMRI; Using many steps of pre-processing, ROI is used, 2 convolutional layers, 20% random selection for test data	decrease-by-half	-	94.6%
(Sarraf and Tofghi, 2016)	Decomposed 2D AD rest-fMRI and MRI; end-to-end pipeline, base learning rate of 0.01, 25% random selection for test data	Adopted from GoogleNet and LeNet models	~1	98.8%

Nonetheless, Sarraf and Tofghi (2016) is the best in term of accuracy and sensitivity. They adopted the successful models for image recognition using 2 and 22 convolutional layers respectively LeCun et al. (1998) and Szegedy et al. (2014). Selected feature map sizes are adopted from both methods. However, MRI and fMRI data are both used for the classification purpose, where four set of MRI images were used to train the CNN. This step is hardly complied in most fMRI classification cases because MRI is not a by-product of any fMRI data acquisition. It is highly depending on the experimental setup during data acquisition.

In this paper, we present an end-to-end CNN pipeline for fMRI classification. With different configurations, we show the need of feature map sizes selection process to ensure the credibility of the founding. Although similar approach may have contributed to the state-of-art technique, lacking grey areas elaboration stain the founding. Thus, we are introducing our approach of handling this problem.

3. Approach

We present the strategy to select best feature map size of CNN fMRI. Based on previous work, six configurations of CNN are used to study the effect of selection to delineate the grey areas in neuroimaging

with deep learning perspective. The data used in this study is Human Connectome Project dataset (Van Essen et al., 2012). The dataset is readily available to download and used for research purpose. With the objective of getting concrete finding, synthetic data is not employed in this research. Proposed approach is shown in Fig. 1. The approach is explained in this section.

3.1. Data preparation

Generally, several steps of data preprocessing are considered in most of fMRI classification approaches. Brain extraction, slice timing correction, smoothing, normalization, realignment, motion correction and co-registration are some of the major steps that were applied on single subject or group analysis of fMRI data. These steps are very important to reduce noises and simultaneously prepare the data for model employment (Poldrack et al., 2011). Software packages such as SPM, FSL and AFNI are used interchangeably and side-by-side for pre-processing pipeline. However, as discussed in Eklund et al. (2016), proposed parametric cluster-wise inference to the group-analysis data by these software could inflate false positive risk during classification.

One-step of preprocessing is employed in our data group analysis to avert the problems associated with parametric statistical modelling. Normalization

is deployed that outputs pre-processed zero mean and unit variance fMRI data. This one preprocessing step is considered in light of CNN advantages and capabilities. This deep learning approach has shown the capability of recognizing various illuminated, deformed and occluded objects in their object

recognition and localization machine learning (Bengio, 2013). An intuition behind this is to avoid the smallest changes to the brain images before classification process. Hence the minimal preprocessing approach is taken.

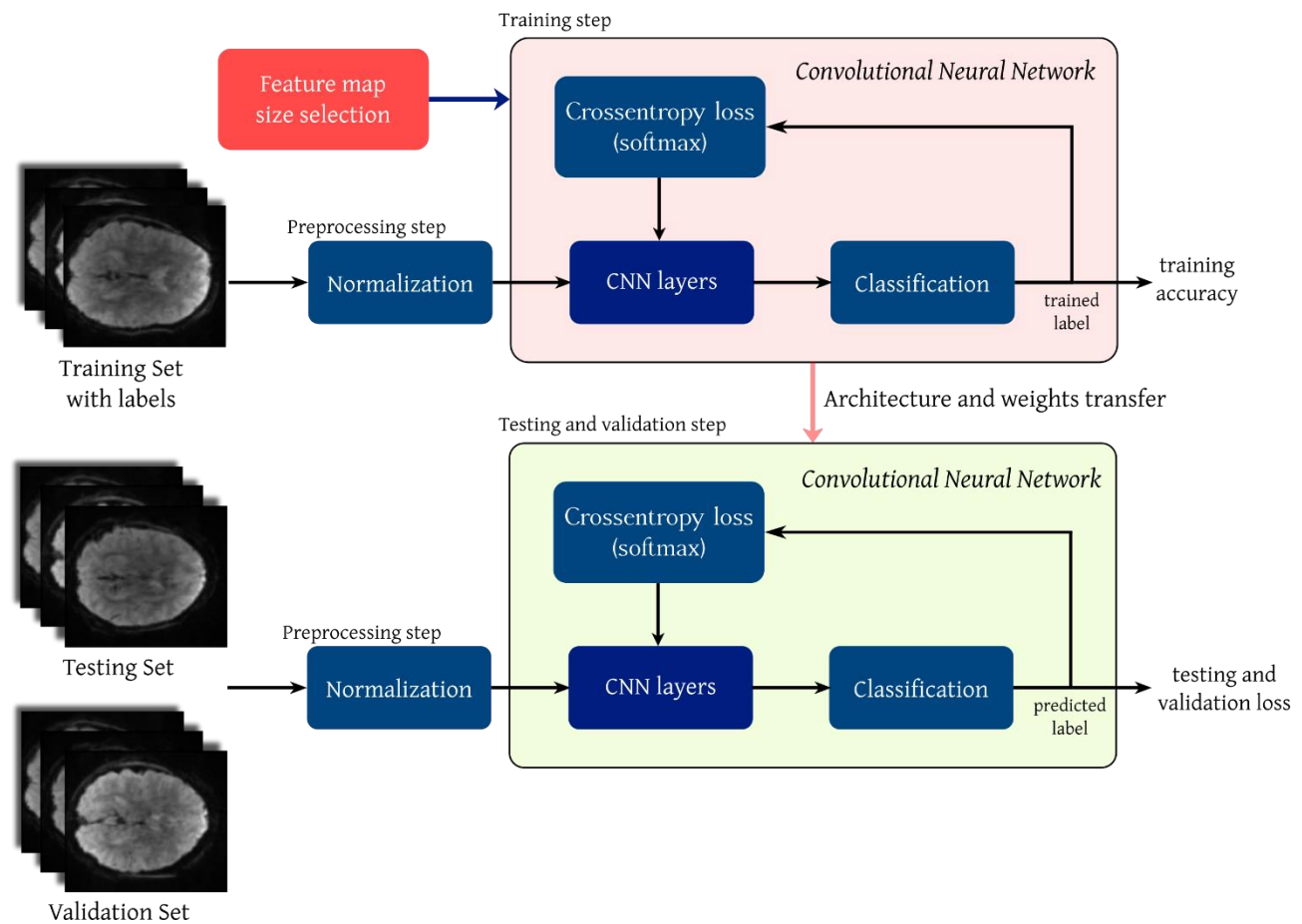


Fig. 1: Proposed architecture block diagram. The three set of data are used to validate the architecture by observing the validation loss

We are using a total of 121 subjects of task-fMRI from HCP. There are more than 1400 data subjects available in the respective server. For this study, to reduce computation cost, small portion of the data are used. To reduce overfitting chances, the 36th of axial slice had been selected which corresponds to the center of the brain horizontally. 37382 samples are acquired when emotion and motor task were chosen for this study. Though it seems too small for classification process, the samples are quite identical because it was taken from the same part for each brain and hardly classified even with human eyes. Data augmentation strategy is not deployed in this work. Uniform sampling is applied to both classes due to imbalances in class size.

Three different divisions of data are used in both classification parts as shown in Fig. 2. The 121 subjects are divided into two parts, 118 subjects' and 3 subjects' parts. This separation is intended to differentiate the training and testing set from validation or evaluation set. The 3-subject data is known as validation set. The remaining 118 data subjects are randomized separately and divided into

training and testing part using Scikit-learn (Pedregosa et al., 2011).

3.2. End-to-end CNN approach

We propose an end-to-end CNN learning approach. With the purpose to lay a foundation for basis feature map size selection, we have considered vanilla convolutional neural network implementation. A different approach of selection might be encountered for hybrid CNN such as recurrent CNN (R-CNN) and CNN with Long-Short-Term Memory (CNN-LSTM). Nonetheless, this paper is aimed to produce a simple rule of thumb for sizing the feature maps for any convolutional architecture of neural network.

Conventional neural networks consist of three main types of layers: input, hidden and output layer. Similarly, these layers are also present in CNN. Besides, there are many sub-types of layers in the convolutional neural network architecture such as dropout, max-pooling and fully-connected layers. The most significant type of layers of CNN is the convolutional layer where it differentiates the

ordinary artificial neural network to the deep learning branches. This is the layer that which feature map sizes are considered where it is placed

at the hidden layer. The main advantages of this layer are weight sharing initiators and feature location insensitive part.

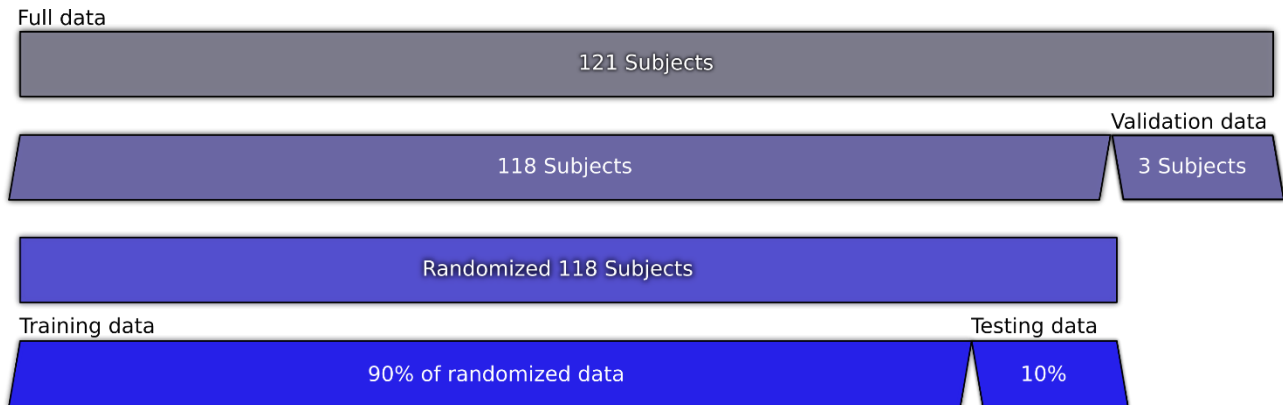


Fig. 2: Data division of 121 HCP subjects

The weight sharing is realized when the previous layer ($m - 1$) of the network is convolved with kernels and produces the next layer (m) with spatial features. Every convolutional layer has its own set of kernels (later known as weights). It is randomized with default parameter in the first stage, where there later evaluated by the loss function in the backpropagation stage. This forward (convolution processes) and backward (backpropagation with loss function) stages output the updated weights. Set of weights are updated in every iteration that satisfy the optimization equation.

In traditional neural network, weight sharing is employed for very small degree of potential. No spatial relation exists between each layer. This is because there is no sub-regions convolution with the weights of each input image. In other words, traditional neural network is updating the weights based on wide space of input image or hand-crafted weights. On the other hand, convolutions induce the low-level abstraction in the early part of the network and weights is shared to make up the higher or feature-level abstraction towards the end of the network. Furthermore, more than one feature map for each layer level is recommended to increase and multiply the feature-worth of weights that shared throughout the network. This will be discussed in the next section.

The input and output layer are the normalized fMRI and the classification labels respectively. The convolutional layer where the normalized fMRI is fed is the first hidden layer. The input layer remains the same for all the six configurations introduced in this paper. Maxpooling and dropout layers are connected among the convolutional layers. In our design experiment, each layer is stacked four times before the fully-connected (FC) layer. Classification layer is connected at the end of FC layer where the accuracy for each epoch is calculated.

Keras that wraps the Tensorflow is used in this study. Keras has been shown to be easier for those who are new to deep learning. We deployed this wrapper to get a reproducible output for other researchers. The end-to-end CNN is employed with Keras default hyper-parameters.

3.3. CNN configurations

As shown in Fig. 3, two types of CNN structure are chosen for this study; 1-step and 2-step convolution. With 1-step convolution, each layer has only one convolution input process, either the fMRI input images (as shown in the Fig. 3) or maxpooled images, where the resulted feature maps will be fed into the maxpooling layer. For the 2-step convolution, two consecutive convolution processes had been employed to convolve the products of first convolution process, where only feature maps are convolved. The stacked convolutional layers will extract more abstract feature maps compared to 1-step convolution structure (Pereira et al., 2016). As a result, edges and features are sharpening. One of the hypothesis of this research is the 2-step will produce better accuracies compared to 1-step.

The 3-by-3 kernel are slide (the convolution process) over the image to get the first feature map as depicted in Fig. 3. With the size of 45, there will be 45 set of kernels for each layer to produce distinct 45 feature maps. Three sets of each structure with 45, 90 and 180 feature maps make up for six configurations of CNN as listed in Table 3. However, similar length of structures is chosen for each configuration. For instance, when the 1-step convolution is employed, 4 sets of 45 kernels will be updated for every iteration of size 45 feature maps configuration. While for 2-step convolution structure, the kernel number is doubled to 8 sets. Thus, sharper edges come with higher number of parameters and longer computations.

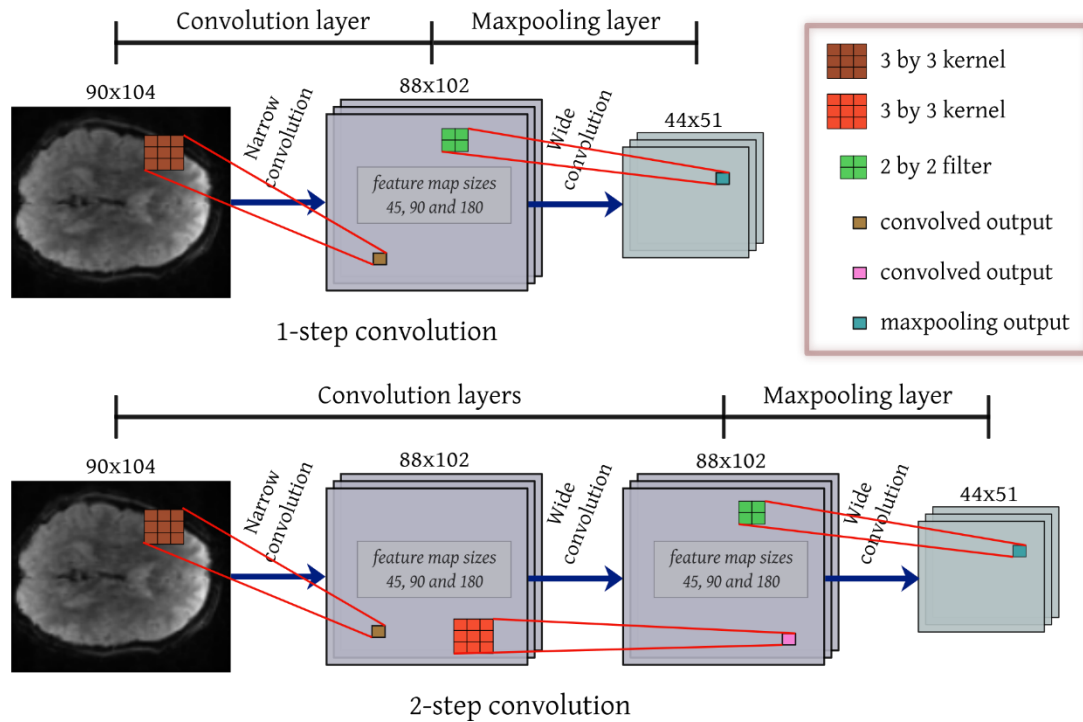


Fig. 3: Figures show different 1-step and 2-step convolution structures of the first convolution layer

Table 3: The convolutional neural network configurations; 45, 90 and 180 are sizes of the feature map products of the convolutions layers while Mp, Do, and F is the maxpooling, dropout and fully-connected layer respectively. The 'Softmax (2)' is the classification and the last layer of these CNN configurations

Name	Feature maps
CNN 1	45-Mp-Do-45-Mp-Do-45-Mp-45-Mp-F-Softmax (2)
CNN 2	45-45-Mp-Do-45-45-Mp-Do-45-45-Mp-45-45-Mp-F-Softmax (2)
CNN 3	90-Mp-Do-90-Mp-Do-90-Mp-90-Mp-F-Softmax (2)
CNN 4	90-90-Mp-Do-90-90-Mp-Do-90-90-Mp-90-90-Mp-F-Softmax (2)
CNN 5	180-Mp-Do-180-Mp-Do-180-Mp-180-Mp-F-Softmax (2)
CNN 6	180-180-Mp-Do-180-180-Mp-Do-180-180-Mp-180-180-Mp-F-Softmax (2)

Selection of hyper-parameters is critical for this study in order to reduce chances of overfitting and to reduce the computation cost due to immense volume of data. In general, the hyper-parameters are chosen based on previous studies except for learning rate. Learning rate is chosen to be very small particularly for fMRI classification. With value of 0.00001, the learning sequence of 1-slice of fMRI is expected to progress slowly. This is because we aimed at a reliable classification of high SNR fMRI images and by making CNN less prone to fast convergence.

3.4. Feature map selection

The main aspect of this paper is discussed in this section.

Each value in feature map is a product of 3-by-3 kernel and 3-by-3 input square convolution as referred to Fig. 3. For example, a narrow 90-by-104 input convolution produces 88-by-102 of each feature map. On the other hand, the wide 88-by-102 input image convolution outputs the same size feature map. The calculation is increased with the increasing of the feature map size.

Then, the feature maps are either convolved or maxpooled according to its structure. However, each convolution must pass an activation function. Rectified linear unit or ReLU is the convolutional layer activation function. This function is a non-

linearity term to differentiate the features in each feature map. It is a straightforward function; $\text{ReLU} = \max(x, 0)$ where the x is the convolved output of every convolutional layer. As compared to sigmoid function; $1/(1 + \exp\{-x\})$ and tanh function; $2\sigma(2x) - 1$, this activation has been found to accelerate gradient descent convergence (Krizhevsky et al., 2012).

The chosen 45, 90 and 180 sizes of the feature map are due to the length of the input fMRI images. The high temporal resolution input images are 90-by-104 in length. Half-length, same-length and double-length of feature map size are chosen based on many studies such as Simonyan and Zisserman (2014) and Margeta et al. (2017). The studies have shown very remarkable results in image recognition and localization.

In combating high computational cost, narrow convolution with no padding is implemented in the first hidden layer of CNN. The intuition is to reduce cost consumption due to computations because fMRI images contain black background at their borders. While for other convolution layers, same size zero-padding or known as wide convolution is used. With stride of one, every region in the input images is convolved. The product of convolution (Raschka, 2015) which also known as the features map has the weight sharing property due to the repeated convolution process.

4. Results and discussion

Results are described and discussed in this section. As show in Fig. 4, the training time is directly proportional to parameter count. These parameters are counted for various processes such as convolution, maxpooling and classification calculations. The higher the sizes of the feature map, the more the parameters to be calculated. For instance, the smaller size of kernel (3-by-3) had implied significantly reduced parameters. However, parameters in 2-step convolution are expected to be much higher than that of 1-step convolution.



Fig. 4: Training time versus parameter counts graph using Nvidia GTX1080 with Intel Xeon @3.60Hz

In this research experiment, both training and testing are set to 200 epochs for each running. Each epoch of training has increased the accuracy and lowered the loss consequently. The optimization process computes the best set of weights for each training step. The CNN 1 with 1-step convolution and half-length size of the input image has the slowest training propagation and highest training and testing loss (error) as shown in Fig. 5. While CNN 6 with twice size of image length and 2-step convolution has the fastest training convergence.

However, the rate of training step does not yield a good accuracy. In Table 4, the training and testing accuracy values are above 90% for each configuration. Though it seems that all the configurations had performed their best, evaluating the configurations will validate each of configuration performances. As shown in the Table 4, the values of validation accuracy vary from 73.96% to 99.72% as the highest accuracy. Thus, the CNN 4 has performed the best for the selected feature map according to this data.

Selecting the size of feature map for each configuration is importance because there are many factors involved, such as the uncountable shapes of human's anatomical map. MNIST and CIFAR-10 with state-of-the-art CNN approaches are getting better recognizing each image. This is because each CNN approach is required to recognize the shape of each object. On the other hand, for classifying fMRI classes, the CNN is expected to ditch and insensitive to the shape of the brain but rather sensitive to what

is inside it. The shapes of the brain are the noises that CNN should not recognize. With very low training rate, the approach is expected to reduce the effect of the noises and susceptible to overfitting.

Table 4: Accuracy values (%) on visual cognition versus Motor sensory task-fMRI of end-to-end-CNN classification for different configuration result

Name	Training Accuracy/ testing Accuracy	Validation Accuracy
CNN 1	90.56 / 95.91	83.14
CNN 2	99.04 / 100.0	73.96
CNN 3	98.98 / 99.98	74.81
CNN 4	99.52 / 100.0	99.72
CNN 5	99.85 / 100.0	98.01
CNN 6	99.78 / 100.0	83.33

The initial value of 0.693 is expected as the first epoch of each loss calculation. Using the softmax for loss calculation, the loss is the negative log probability of the correct class. The two classes' classification which 0.5 probability of each, thus $-\ln(0.5)$ is equals to 0.693.

In a nutshell, the CNN 4 shows the best configuration for this dataset and set of hyper-parameters. With 99.52% accuracy value, the 2-step convolution and same length image size of feature map size, the configuration is found to be the optimal configuration that discards the overfitting and fast convergence during training.

In addition, the training and testing accuracy values show that the 2-step convolution works better at recognizing the classes.

5. Conclusion

We have presented the effect of feature map size on the fMRI dataset classification with end-to-end deep convolutional neural network. The size plays an important role in selecting other hyper-parameters for CNN configuration and training. This is due to high computation cost of convolution process with bigger feature map sizes. With the same size of fMRI length of the input image with 2-step convolution, the size of that feature map shows promising result.

The right size of feature map is equally important as other choice of hyper-parameters such as training rate and depth of the convolution. Deeper than 2-step convolution has a trade-off between accuracy value and computational cost. One can afford to have high cost in computation to get better accuracy because of powerful graphic cards or GPU that could be stacked together.

Training CNN to classify fMRI is different to training MRI. And, to compare other domain such as handwritten dataset, speech data and other image dataset that are easily available in the net, fMRI has more challenge due to its 3D time-series nature. The differences are visualized in the training session and for choosing the best hyper-parameters. Without any confirmation that each training progress will over fit the classification approach. As such, in this paper, it is shown that a very high training accuracy does not produce a good classification.

This paper also focuses on the importance of using validation set for evaluating the CNN model. The shapes of the brains could be the high-level

feature that the CNN opt to recognize and be sensitive of. The validation set increases the confidence level of each CNN setup.

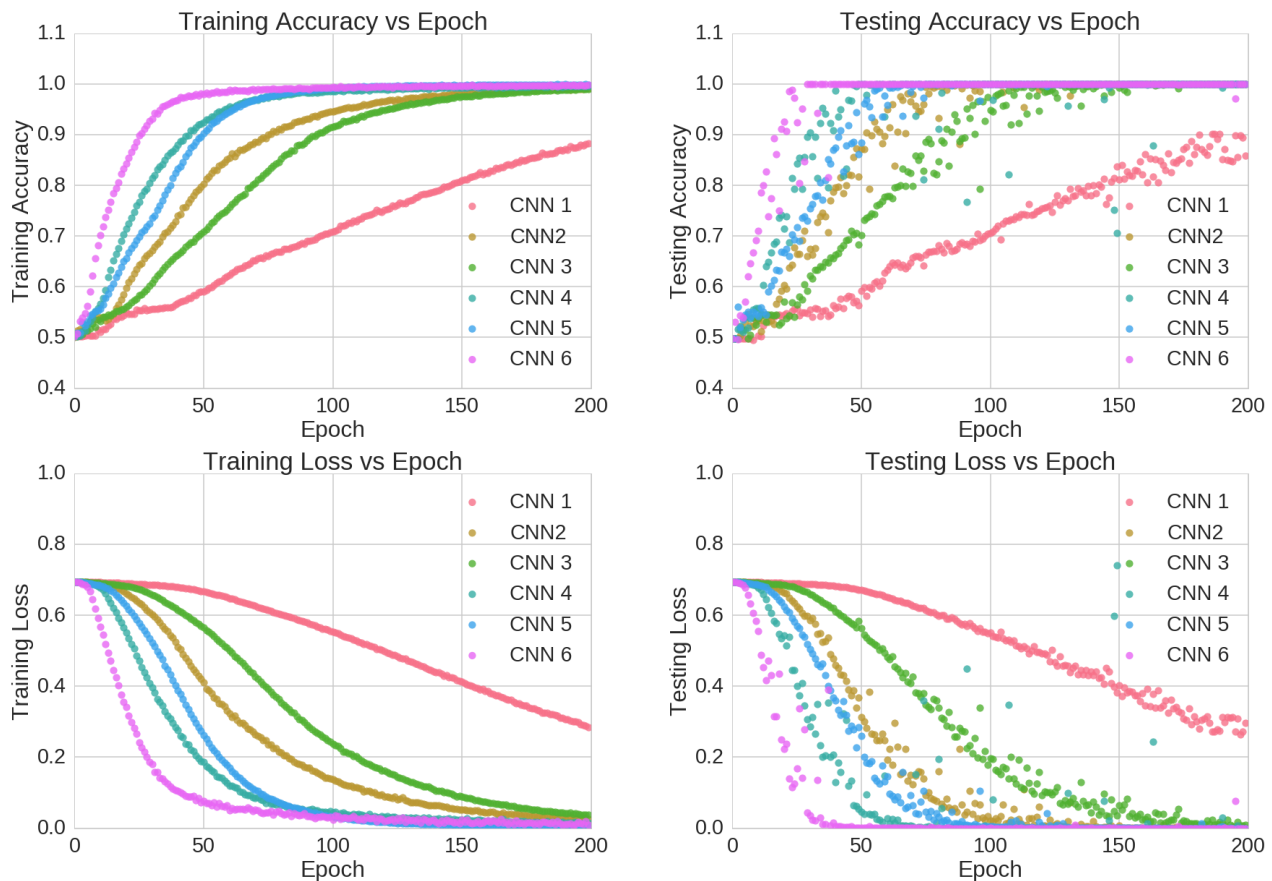


Fig. 5: Testing and training plots for different CNN configurations

Our approach has been tested on one slice of fMRI images which is the most center part of the brain. For other slices of the brain, the approach needs more than 118 subjects to test each CNN configurations. However, this research demonstrates that CNN is capable of recognizing feature inside brain maps to classify motor and emotion tasks.

It is recommended for future work to study the effect of higher dimension of fMRI images for classification. The two-dimensional convolution could be expanded to 3D convolution, while maintaining the approach of using separated subjects instead of separated randomized dataset.

Acknowledgment

The authors would like to thank the International Islamic University Malaysia for sponsoring the research under Research Initiative Grant Scheme (RIGS16-350-0514).

References

- Bengio Y (2013). Deep learning of representations: Looking forward. In International Conference on Statistical Language and Speech Processing, Springer, Berlin, Heidelberg, Germany: 1-37.
- Burgh HK, Schmidt R, Westeneng HJ, Reus MA, Berg LH, and Heuvel MP (2017). Deep learning predictions of survival

based on MRI in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, 13: 361-369.

- Cohen JD, Daw N, Engelhardt B, Hasson U, Li K, Niv Y (2017). Computational approaches to fMRI analysis. *Nature Neuroscience*, 20(3): 304-313.
- Cui Z, Yang J, and Qiao Y (2016). Brain MRI segmentation with patch-based CNN approach. In the 35th Chinese Control Conference, IEEE, Chengdu, China: 7026-7031.
- Eklund A, Nichols TE, and Knutsson H (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28): 7900-7905.
- Gehring J, Auli M, Grangier D, Yarats D, and Dauphin YN (2017). Convolutional sequence to sequence learning. Available online at: <https://arxiv.org/abs/1705.03122>
- Gollapudi S (2016). Practical machine learning. Packt Publishing, Birmingham, UK.
- Greenspan H, Ginneken B, and Summers RM (2016). Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5): 1153-1159.
- Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, and Lew MS (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187: 27-48.
- Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, and Larochelle H (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35: 18-31.
- Huang H, Hu X, Zhao Y, Makkie M, Dong Q, Zhao S, and Liu T (2017). Modeling Task fMRI Data via Deep Convolutional Autoencoder. *IEEE Transactions on Medical Imaging*.

- Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Glocker B (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36: 61-78.
- Karpathy A and Fei-Fei L (2015). Deep visual-semantic alignments for generating image descriptions. In the IEEE Conference on Computer Vision and Pattern Recognition: 3128-3137.
- Krizhevsky A, Sutskever I, and Hinton GE (2012). Imagenet classification with deep convolutional neural networks. In: Touretzky DS, Mozer MC, and Hasselmo ME (Eds.), *Advances in neural information processing systems*: 1097-1105. MIT Press, Cambridge, Massachusetts, USA.
- Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, and Turner R (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12): 5675-5679.
- LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. *Nature*, 521(7553): 436-444.
- LeCun Y, Bottou L, Bengio Y, and Haffner P (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324.
- Li R, Zhang W, Suk HI, Wang L, Li J, Shen D, and Ji S (2014). Deep learning based imaging data completion for improved brain disease diagnosis. In the *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, Quebec City, QC, Canada: 305-312.
- Makke M, Huang H, Zhao Y, Vasilakos AV, and Liu T (2017). Fast and scalable distributed deep convolutional autoencoder for fMRI big data analytics. Available online at: <https://arxiv.org/abs/1710.08961>
- Margeta J, Criminisi A, Cabrera Lozoya R, Lee DC, and Ayache N (2017). Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 5(5): 339-349.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, and Vanderplas J (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct): 2825-2830.
- Pereira S, Pinto A, Alves V, and Silva CA (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5): 1240-1251.
- Poldrack RA, Mumford JA, and Nichols TE (2011). *Handbook of functional MRI data analysis*. Cambridge University Press, New York, USA.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K (2017). CheXNet: Radiologist-Level pneumonia detection on chest X-rays with deep learning. Available online at: <https://arxiv.org/abs/1711.05225>
- Raschka S (2015). *Python machine learning*. Packt Publishing, Birmingham, UK.
- Sarraf S and Tofghi G (2016). DeepAD: Alzheimer' s Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. *BioRxiv*: 070441. <https://doi.org/10.1101/070441>
- Simonyan K and Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. Available online at: <https://arxiv.org/abs/1409.1556>
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, and Rabinovich A (2014). Going deeper with convolutions. Available online at: <https://arxiv.org/abs/1409.4842>
- Valente G, Castellanos AL, Vanacore G, and Formisano E (2014). Multivariate linear regression of high-dimensional fMRI data with multiple target variables. *Human Brain Mapping*, 35(5): 2163-2177.
- Valverde S, Cabezas M, Roura E, González-Villà S, Pareto D, Vilanova J C, and Lladó X (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155, 159-168.
- Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, and Della Penna S (2012). The Human Connectome Project: a data acquisition perspective. *Neuroimage*, 62(4): 2222-2231.
- Zafar R, Kamel N, Naufal M, Malik AS, Dass SC, Ahmad RF, and Reza F (2017). Decoding of visual activity patterns from fMRI responses using multivariate pattern analyses and convolutional neural network. *Journal of Integrative Neuroscience*, 16(3): 275-289.
- Zhao L and Jia K (2016). Multiscale CNNs for brain tumor segmentation and diagnosis. *Computational and mathematical methods in medicine*, 2016(7): 1-7.
- Zhao Y, Dong Q, Zhang S, Zhang W, Chen H, Jiang X, and Liu T (2017). Automatic recognition of fMRI-derived functional networks using 3D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*. <https://doi.org/10.1109/TBME.2017.2715281>