A deep learning based approach to Classification of CT brain images

Xiaohong W. Gao
Department of Computer Science
Middlesex University
London NW4 4BT
UK
x.gao@mdx.ac.uk

Rui Hui
Neurosurgery Centre
Navy General Hospital
Beijing, China

Abstract—**This study explores the applicability of the state of the art of deep learning convolutional neural network (CNN) to the classification of CT brain images, aiming at bring images into clinical applications. Towards this end, three categories are clustered, which contains subjects' data with either Alzheimer's disease (AD) or lesion (e.g. tumour) or normal ageing. Specifically, due to the characteristics of CT brain images with larger thickness along depth (z) direction (~5mm), both 2D and 3D CNN are employed in this research. The fusion is therefore conducted based on both 2D CT images along axial direction and 3D segmented blocks with the accuracy rates are 88.8%, 76.7% and 95% for classes of AD, lesion and normal respectively, leading to an average of 86.8%.**

Keywords — convolutional neural network. Classification, CT brain images, 3D CNN

I. INTRODUCTION

A. CT Brain Images

Due to its readiness, simple and inexpensive nature, Computerised Tomography (CT) is prevalent in nearly every hospital while presenting good quality of visual information. In addition, CT is probably the first imaging tool that was introduced into the study of human brain and has since been widely applied as the first choice to eliminate other possibilities when it comes to the diagnosis of Alzheimers' disease (AD). However, while most patients have undergone this scanning as a prelude to imaging inspection, mainly for the purpose of ruling out the other possibilities (e.g., tumour, stroke, etc.), CT data have not been implemented into the clinical diagnosis of AD due to their relatively low resolution and variations among manual measurement of certain features, such as medial temporal lobe that is associated with AD.

Today, in the UK, 800,000 people have been formally diagnosed with the condition of dementia [1]. In reality, it is estimated that 60% of people who are living with the condition go undiagnosed [2]. This is because the determination of dementia remains a convoluted process as symptoms come and go. In addition, with regard to CT brain images, specified brain atrophy is associated with not only AD but also normal ageing and cerebral vascular diseases. For example, the medial temporal lobe atrophy (MTA) together with CSF biomarkers has been demonstrated as the most important diagnostic markers for AD, which may not be specific. In addition, atrophy of hippocampus (in particular, left hippocampus), has been found in AD, which also emerges in healthy ageing adults [3]. However, by accurate measurement of atrophy factors of temporal horn ratio and suprasellar cistern ratio, it has been found that CT data can contribute significantly to the diagnosis of AD with 90.2% accuracy [3]. Therefore, CT linear measurements can be of great value in the work-up processes of AD patients.

To alleviate the considerable variations [4] may incur during manual measurements, this study is to investigate the non-supervised automatic process employing the state of the art of convolutional neural network (CNN) on the classification, segmentation and measurement of Alzheimer's data. In the first phase of this research, the classification of AD, healthy (normal) ageing and lesions data takes place, remains the focus of this paper.

B. Convolutional Neural Network (CNN)

Deep learning models refers to a class of computing machines that can learn a hierarchy of features by building high-level features from low-level ones [5, 6] , thereby automating the process of feature construction. One of these models is the well-known convolutional neural network (CNN) [6]. Consisted of a set of algorithms in machine learning, CNN comprises several (deep) layers of processing involving learnable operators (both linear and non-linear), and hence has the ability to learn a hierarchy of information by building high-level information from low-level data, thereby automating the process of information/feature construction [7]. It has been demonstrated that, when trained with appropriate regularization, CNNs can achieve superior performance on visual object recognition tasks without relying on hand-crafted features, e.g. SIFT, SURF. In addition, CNNs have been shown to be relatively insensitive to certain variations on the inputs [7].

Inspired by biological vision processes, CNNs applies a feed-forward artificial neural network to simulates variations of multilayer perceptrons where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field [8]. As a direct result, they are widely used for image and video recognition. Specifically, CNNs have demonstrated as an effective class of models for understanding image content, giving state of the art results on image recognition, segmentation, detection and retrieval.

In addition, recent advances of computer hardware technology (e.g., GPU) have propitiated the implementation of CNNs in representing images. While CNNs have lent themselves well to the computer vision field and achieved state-of-the-art results, they are built mainly for 2D images. Although several papers report the work on 2D videos [8], working on 3D images is quite a different task to a certain extent. In this study, both 2D and 3D form of CNN are elaborated to CT brain images.

This paper is structured in the followings. Section 2 entails the methodologies that are employed in this study and Section 3 presents and results. Subsequently, the conclusion is summarised in Section 4.

II. METHODOLOGY

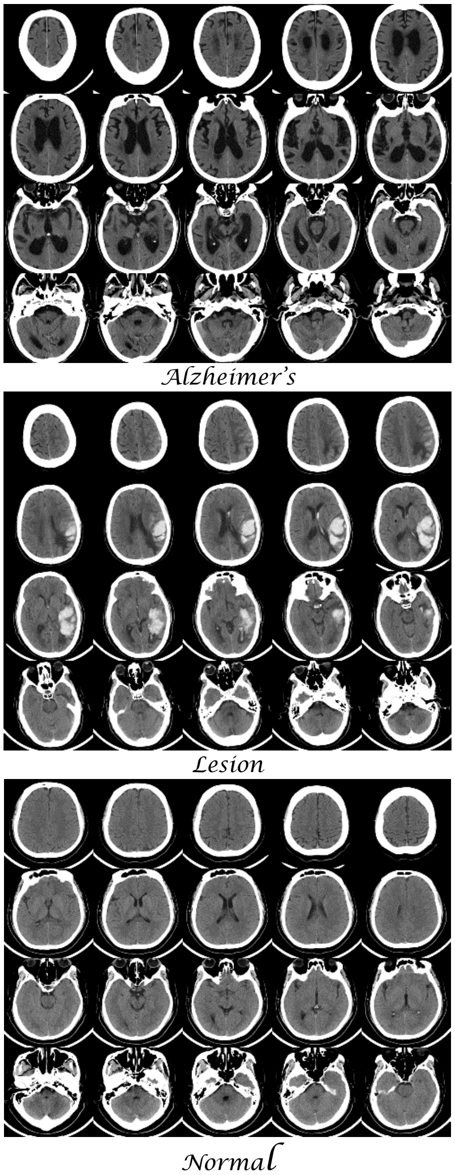
A. Data pre-processing and averaging 3D CT images

In total, 3D datasets from 282 subjects are collected and studied in this investigation, including 51 with Alzheimer's (AD), 118 with lesions and 117 being from normal healthy subjects. In our collection, CT data vary in both depth numbers of being between 16 to 33 slices and dimensions with either 512 x 512 or 912 x 912 pixels. Figure 1 depicts the montage of CT data for Alzhermer's (top), Lesion (middle) and Normal (bottom) subjects respectively.

In the 2D form, all the slices are segmented and normalised into 360 x 360 pixels. For normal and AD data, the middle 20 slices are employed from each dataset, whereas for lesion data, only slices that contain visual lesion features (e.g. tumour) are employed. As a result, although lesion datasets remain the largest among the three classes, the overall number of slices are similar to the others.

In parallel, in 3D form, each dataset is firstly registered, segmented and normalised into 200 x 200 x 20 pixels. Due to the relatively thickness between CT slices (~5 mm) in comparison with the counterpart of MR (~0.5 mm), geometric normalisation is performed to align all 3D CT images into the same space. Because CT has good structural (e.g. bone) information of the brain, rigid body geometric transformation is opted for as illustrated in Eq. (1).

$$I_{moving} = T \times I_{fixed} \qquad (1)$$



Alzheimer's

Lesion

Normal

Fig. 1. The montage view of the middle 20 slices of CT images for Alzheimer's disease (top), Lesion (middle), and Normal (bottom) subjects respectively.

Where T is the transformation matrix inclusive of translation, rotation and scaling which is to be determined, I_{moving} refers to the image to be registered and I_{fixed} the template image. Therefore to implement the determination of the parameters of T, Eq. (2) is to be minimised.

$$\sum_i (I_{fixed}(Tx_i) - s_\alpha I_{moving}(x_i))^2 \qquad (2)$$

Where x_i refers to a number of points selected from either fixed image (reference) or moving (source) images. To compensate the fact that each image might be scaled differently with reference of intensity, an additional intensity scaling factor (s_α) is added.

To implement Eq. (2), a template CT date set with relatively better aligned with 30 slices is chosen to be the fixed dataset, whereas the others are aligned to it.

After the normalisation, each 3D dataset is divided into $40 \times 40 \times 10$ boxes. Similar to 2D form, for both AD and normal data, all the boxes are applied to train the data whereas for lesion data, only boxes containing lesion contents are used.

B. The implementation of 2D and 3D CNNs

Figure 2 illustrates both 2D and 3D CNNs implemented in this study.

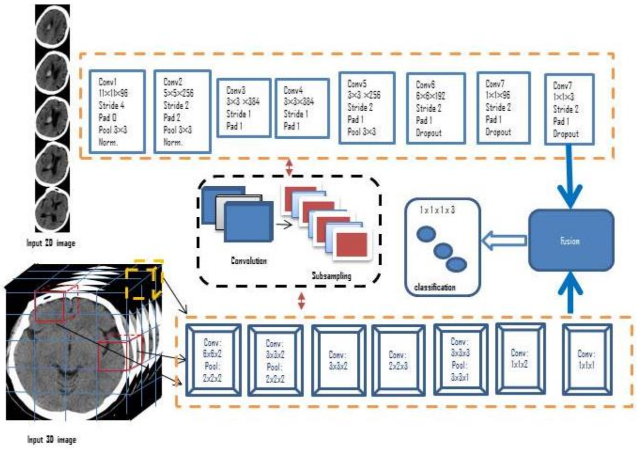


Fig. 2. The fusion of both 2D and 3D CNNs for CT images.

In this study, 2D convolutional deep learning neural network (CNN) has been applied based on MatConvNet [9] written in Matlab software, along the direction of axial of the brain.

Specifically, for training data $\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}\right)$, where image $\boldsymbol{x}^{(i)}$ is in three-dimension and $\boldsymbol{y}^{(i)}$ the indicator vector of class of $\boldsymbol{x}^{(i)}$, the feature maps of an image, namely, $w_1, \dots, w_{L,}$ will be learnt based on CNN by solving Eq. (3).

$$\underset{\boldsymbol{w}_1,\dots,\boldsymbol{w}_L}{argmin} \frac{1}{n}\sum_{i=1}^{n} \ell(f(\mathbf{x}^i; \boldsymbol{w}_1, \dots, \boldsymbol{w}_L), \boldsymbol{y}^i) \tag{3}$$

Where ℓ refers to a suitable loss function (e.g. the hinge or log loss).

To obtain these feature maps computationally, in 2D CNNs, 2D convolution is performed at the convolutional layers to extract features from local neighbourhood on feature maps in the previous layer. Then an additive bias is applied and the result is passed through a sigmoid function as illustrated in Eq. (4) mathematically.

$$\boldsymbol{v}_{ij}^{xy} = tanh\left(\boldsymbol{b}_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1} \boldsymbol{w}_{ijm}^{pq}\, \boldsymbol{v}_{(i-1)m}^{(x+p)(y+q)}\right) \tag{4}$$

Where the notations of those parameters in Eq. (4) are explained in Table 1.

TABLE 1. NOTATIONS OF PARAMETERS IN EQ. (4).

Parameter	Notation
$tanh(.)$	hyperbolic tangent function
m	index over the set of feature maps in the $(i-1)th$ layer
$\boldsymbol{b}_{ij}$	bias for the feature map f in Eq. (1).
$\boldsymbol{w}_{ijk}^{pq}$	value at the position (p, q) of the kernel connected to the k_{th} feature map
(p, q)	2D position of a kernel
P_i, Q_i	height and width of the kernel

In the subsampling layers, the resolution of the feature maps is reduced by pooling over local neighborhood on the feature maps in the previous layer, thereby increasing invariance to distortions on the inputs. A CNN architecture can be constructed by stacking multiple layers of convolution and subsampling in an alternating fashion. The parameters of CNN, such as the bias b_{ij} and the kernel weight $\boldsymbol{w}_{ijk}^{pq}$ are usually trained using either supervised or unsupervised aproaches [5, 10].

Furthermore, in order to retain CT information along z direction, 3D version of CNN is explored. In 3D CNN, the 3D convolution is achieved by convolving a 3D kernel to a box along both x-y (2D) and z directions where Eq. (4) will be extended into Eq. (5) to calculate the value at position (x, y, z) on the j_{th} feature map in the i_{th} layer.

$$\boldsymbol{v}_{ij}^{xyz} = tanh\left(\boldsymbol{b}_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i} \boldsymbol{w}_{ijm}^{pqr}\, \boldsymbol{v}_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \tag{4}$$

where Ri is the size of the 3D kernel along the z dimension, $\boldsymbol{w}_{ijm}^{pqr}$ is the $(p, q, r)_{th}$value of the kernel connected to the m_{th} feature map in the previous layer. In addition, 3D pooling is also developed to increase the degree of invariance to distortion and noise occurred in the input images.

C. Fusion of the results

In 2D form, 2D slices are applied to training the model individually, whereas in 3D form, small cubes (40 x 40 x 10)

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

are utilised. The classification therefore takes place at the level of whole dataset for each subject combine the two. In this way, for the Normal category, more than 95% of the slices (i.e., except one slice) of each dataset and all cubes (or boxes) have to be labelled as normal class since both AD and Lesion signature information usually shows on more than 1 slice. On the other hand, for the classification of either AD or Lesion, the majority vote determines the classification decision.

III. RESULTS

Table 2 gives the data numbers applied in the study. One dataset usually contains 26 to 35 slices. In total, there are 282 datasets in 3D form, whereas in 2D form, 4076 slices are applied. These data are divided into 3 groups while applying CNN, which are training, validation (to avoid over fitting) and testing datasets.

TABLE 2. THE NUMBER OF DATA SETS (IN SUBJECT NUMBER) APPLIED IN EACH PROCESS. THE NUMBERS IN BRACKETS ARE THE TOTAL NUMBER OF 2D SLICES IN THAT CATEGORY.

	Alzheimer's	**Lesion**	**Normal**	**Total**
Training **(2D frame)** **{3D boxes}**	30 (700 {775})	80 (700) {1160}	70 (1300) {1860}	180 (2700) {3795}
Validation **(2D frames)** **{3D boxes}**	12 (150) {297}	19 (120) {357}	20 (400) {540}	51 (670) {1194}
Test **(2D frames)** **{3D boxes}**	15 (150) {378}	19 (127) {483}	23 (429) {621}	57 (709) {1482}
Total **(2D frames)** **{3D boxes)**	51 (1000)	118 (947)	113 (2129)	282 (4076) {6471}

In total, 180 datasets are applied to test the developed CNN classification system containing 2700 2D images and 3795 boxes. The division of the data between each group is randomly selected. The confusion matrix for the testing is given in Table 3.

TABLE 3. THE CONFUSION MATRIX OF TESTING RESULTS OF THREE CLUSTERS.

	Alzheimer	**Lesion**	**Normal**	**Accurate (%)**
Alzheimer	13	2	0	86.7
Lesion	2	15	2	78.9
Normal	1	0	22	95.6
Average				**87.7**

In summary, the accuracy of classification of three classes are 86.7%, 78.9%, and 95.6% for Alzheimer's, Lesion, and Normal classes respectively, with the average of 87.7%.

IV. CONCLUSION

In this study, only 3 classes are considered, which are Alzheimer's, Lesion, and Normal clusters. Although the category of Lesion consists the largest dataset (N=118), not every 2D slice or 3D box contains the signature information. As a result, the lesion group has the smallest number of images with 947 slices, whilst AD and Normal groups having 1000 and 2129 images respectively in 2D form. Whilst the differences are not significant, in particular between AD and Lesion groups, the classification results appear to be in line with the number of data that each group has. For example, the normal subject group has the largest number of datasets with a total of 2129 image slices and has the highest accuracy rate of 95.6%. Therefore the direct conclusion remains being that more data will achieve better classification results.

In addition, while CT brain data are in three dimensional, the large thickness (~5mm) between slices has led to the classification of 3D CT images alone being not as accurate as that in 2D form. The fusion therefore takes place to take advantages of both 2D and 3D information, which gives better result. Also, further comparison with hand-crafted approaches, e.g., SIFT, will be conducted in the future.

ACKNOWLEDGMENT

This work constitutes part of project WIDTH that is funded by EU FP7. Their financial support is gratefully acknowledged..

REFERENCES

[1] BBC, The dementia timebomb, http://www.bbc.co.uk/science/0/21878238. Retrieved in March 2015.

[2] Dementia: https://www.gov.uk/government/news/. Retrieved in March 2015.

[3] Y. Zhang, E. Londos, L. Minthon, C. Wattmo, H. Liu, L.O. Wahlund, Usfulness of computerd Tmotgraphy linear measurements in diagnosing Alzheimer's disease, Acta Radiol, 2008, vol. 49(1), pp.91-97.

[4] A.R. Oksengaard, M. Haakonsen, R. Dullerud, K. Engedal, K. Laake, Accuracy of CT scan measurements of the medial temporal lobe in routine dementia diagnostics, International Journal of Geriatric Psychiatry 2003, vol. 18, pp.308-312.

[5] Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cyb., 1980, vol. 36, pp.193–202.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[7] Y. LeCun, F.J. Huang, and L. Bottou, Learning methods for generic object recognition with invariance to pose and lighting. In CVPR, 2004.

[8] Y. LeCun, LeNet-5, convolutional neural networks. Retrieved December 2015.

[9] S. Ji, W. Xu, M. Yang and K. Yu, 3D convolutional neural networks for human action recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, vol. 35 (1), pp.221-231.

[10] A. Vedaldi and K. Lenc, MatConvNet Convolutional Neural Networks for MATLAB , arXiv preprint arXiv:1412.4564, retrieved in December 2015.

[11] M. Ranzato, F.J. Huang, Y. Boureau, and Y. LeCun, Unsupervised learning of invariant feature hierarchies with applications to object recognition. In CVPR, 2007.

.