IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 24, 2021, accepted December 13, 2021, date of publication December 23, 2021, date of current version January 20, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3137824

Robust Hippocampus Localization From Structured Magnetic Resonance Imaging Using Similarity Metric Learning

SAMSUDDIN AHMED[1], (Member, IEEE), KUN HO LEE[2,3], AND HO YUB JUNG[4]
[1]Department of Information and Communication Technology, Bangabandhu Sheikh Mujibur Rahman Digital University, Kaliakair, Gazipur 1750, Bangladesh
[2]Gwangju Alzheimer's Disease and Related Dementias Cohort Research Center, Chosun University, Gwangju 61452, South Korea
[3]Department of Biomedical Science, Chosun University, Gwangju 61452, South Korea
[4]Computer Vision Laboratory, Department of Computer Engineering, Chosun University, Gwangju 61452, South Korea

Corresponding author: Ho Yub Jung (hoyub@chosun.ac.kr)

This work was supported by the Chosun University, in 2021.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Chosun University Hospital, South Korea (CHOSUN 2013-12-018-070).

ABSTRACT Accurate demarcation of anatomical landmarks in 3D medical imaging is a safety-critical and challenging task. State-of-the-art approaches formulate landmark localization either as a classification or as a regression problem. In this study, feature classification is performed as a verification step in a cascaded Hough regression networks (HRNs) for hippocampus localization in the structured magnetic resonance images of the brain. Global and local features of the landmarks are learned with coarse prediction and fine-tuning convolutional neural networks for coarse-to-fine localization. Siamese network was trained to learn a deep metric for verifying the roughly estimated locations. Feature verification with the Siamese network drops the outlier predictions and increase the robustness in prediction. Three-view patches(TVPs) with a size of $64 \times 64 \times 3$ are fed for rough estimation while the TVP sizes for Siamese-based verification and Hough regression network (HRN)-based fine-grained estimations are $32 \times 32 \times 3$ and $16 \times 16 \times 3$, respectively. The experiment was performed on the Gwangju Alzheimer's and Related Dementia's (GARD) cohort data set. The proposed approach demonstrated better performance with the errors of 1.70 ± 0.50 millimeters(mm) and 1.66 ± 0.49 mm for localizing the left and right hippocampi in the GARD data set. In Alzheimer's Disease Neuroimaging Initiative (ADNI) data set, the observed errors were 1.79 ± 0.83 mm and 1.55 ± 0.61 mm for localizing left and right hippocampus, respectively. Our results are comparable to those obtained by the state-of-the-art methods.

INDEX TERMS Landmark-localization, hippocampus, Hough CNN, Siamese network.

I. INTRODUCTION

Hippocampus is a key structure in the brain and its role in the learning and memory function has been intensively studied in the neuroscience field [1]. Hippocampus is also an important anatomical region in Alzheimer's disease (AD) etiology [2]–[6]. Among all of the cerebral regions, the hippocampus is one of the first affected regions in atrophy [7], [8]. Studies have shown that a significant number of patients with hippocampal atrophy developed Alzheimer's disease [9]. Visual and texture features of Magnetic resonance image (MRI) derived from the hippocampus have contributed significantly for AD diagnosis [10]. Many MRI studies [4], [11], [12] for AD suggested the computation of the shape and volume features from bilateral hippocampi for estimating the degree of atrophy of the hippocampus, which can be used as a diagnostic marker for AD. Other neurological studies [13] also found a 15%-30% percent volume reduction in the mild dementia stage of AD and the reduction of 10%-15% in the amnestic variant of mild cognitive impairment (MCI) [14].

Variation in hippocampal function, anatomy, and degeneration has been implicated in other neurological disorders such as schizophrenia and depression [15]. Changes in the morphology of the hippocampus are some of the symptoms

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

TABLE 1. GARD Data set for Hippocampus Localization.

Subjects(M/F)	MRI	Age(Average±std[range])	Education(Average±std[range])	Category
81(39/42)		73.3±8.157[49-87]	7.345±4.8689[0-18]	AD
30(20/10)		73±2.9612[66-77]	8.3±4.816[0-18]	aMCI
9(4/5)		71.8642±7.069[56-83]	7.889±6.234[0-18]	naMCI
206(98/108)		71.84951±5.3307[60-85]	8.9514±5.62[0-22]	CN

of these diseases. Therefore, the examination of this structure has consistently been of significant interest, particularly in the context of neuroimaging investigations.

Accurate location of the hippocampus is a basis requirement for automated image analysis of the region, and a significant effort has been devoted to this task by various researchers [16]–[18]. Accurate demarcation of this landmark is often important for further diagnosis by machine learning techniques. The landmarks usually provide i) useful features from the region of interest for machine learning ii) initial information for registration and segmentation and, iii) guidance for navigation and retrieval throughout the image data. Automatic hippocampus localization methods are available to avoid manual annotations which is cumbersome, labor-intensive, and requires domain expertise. Moreover, inter-observer differences and inter-subject variations for the same observer are present in manual annotations.

Automatic localization of the hippocampus in an MRI scan is not a trivial problem, and becomes even more challenging with increasing severity of anatomical atrophy. Machine learning-based studies for localization mainly focus on learning a mapping function that utilizes the MRI or image features to find landmark positions [19]–[29]. The localization problem is formulated either as a classification problem or as a regression task. Classification models [30]–[32] extract patches from a voxel and classify it as a landmark or not. However, these approaches are prone to the dataset imbalance problem. The lack of positive samples in the dataset often results in a biased classifier.

State-of-the-art regression-based approaches for localizing anatomical landmarks in 3D medical imaging are trained with both random forest and deep learning methods [33]. Using a regression-based approach implemented in an random forest framework, Pauly *et al.* [34], Glocker *et al.* [27], Criminisi *et al.* [35], and Menze *et al.* [25] proposed to learn the relative positions between the organs of interest and all of the anatomy available in the training data solely with the arbitrary scale Haar-like appearance features, i.e., the size of the appearance features and their distance from the training voxels. In a test image, the position of organs of interest is then obtained from the recognized anatomy, implemented by accumulating the relative positions of each voxel of the image to the organ of interest. Other regression-based methods have the same framework by developing a model to predict the 3D position (or displacement) from a local voxel to a target voxel by learning the non-linear relationship between these two voxel features [21], [23], [25], [27], [35]–[44].

Recently, prediction of landmarks locations directly from image inputs rather than formulating the task as a classification problem has become more popular [31]. This may be due to the significant improvements in accuracy obtained by regression models in comparison to the classification models. Regression-based models suggest exploiting predictions at different distances from the target landmark. This approach makes it possible to use any local patch for estimating a potential landmark position from given local image patches. Some methods consider global information along with local correlations. For all of these regression-based approaches, it was shown that by learning only from appearance information extracted from training data, the obtained results are robust in the presence of locally similar structures. However, due to variations in the relative positions of the anatomy used for robust prediction of landmarks, a trade-off between accuracy and robustness still exists.

A common property of all methods that use an explicit model of a geometric configuration is that after generating local predictions, appearance information is never used further in the regularization stage. The proposed approach closely follows the two-stage Hough convolutional neural network [44], [45] that considers both local and global features [44]. However, in the previous methods, the global predictions are not verified and it is not known whether the model provides the correct landmark offsets. In our approach, we have addressed the outlier predictions in global estimations with a Siamese-network [46], [47] based verification network, which provides more robust localization performance.

The rest of the paper is organized as follows: in section II, we discuss the data set used for the experiment. In section III, we present the methodology and section IV discusses on the experimental setup, ground truth preparation and the training procedure. Section V discusses the results and section VI concludes the article.

TABLE 2. ADNI Dataset for Hippocampi localization.

Subjects(M/F)	MRI(M/F)	Age(Average±std[range])	Category
20(7/13)	32/45	75.32±8.47[57-90]	ADD
18(14/4)	111/34	74.655±8.00[55-90]	MCI
22(7/15)	51/78	77.57±4.04[70-88]	CN

II. DATA SET

We have performed experiments on the Gwangju Alzheimer's and Related Dementia's(GARD) Cohort data set [48]–[50]. For comparing our findings with state-of-the-art methods, we have utilized Alzheimer's Disease Neuroimaging

Initiative (ADNI) dataset. We briefly describe the data sets in Table 1 and Table 2.

A. GARD DATASET

There are 326 MRI scans in the GARD data set acquired from 326 Korean participants at Chosun University Hospital and other Korean hospitals. The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070). All volunteers or authorized guardians for cognitively impaired individuals gave written informed consent prior to participation.

The ages of the subjects vary from 49 years to 87 years with an average age of 72.02 years and with standard deviation of 6.06 years. More than 88% of the subjects are over 65 years old. The education level ranged between illiterate to highly educated. The lowest education level was set to 0 while the highest of the same was set to 22. The mean education score was measured as 8.46 with a standard deviation of 5.42. There are four clinical categories in the data set: dementia due to Alzheimer's disease (ADD), amnestic MCI (aMCI), non-amnestic MCI(naMCI), and normal control (CN). The number of male and female participants was balanced with 162 and 164 male and female participants. The male/female ratio balance was maintained for all categories except for aMCI which is made up of 20 male subjects and only 10 female subjects. A total of 81, 30, 9 and 206 scans are found in the ADD, aMCI, naMCI and CN categories, respectively.

After the acquisition, the scans were improved non-parametric nonuniform intensity normalization i.e., N4 normalized, skull striped, and segmented into 6 tissue types and Desikan–killiany–tourville labeling protocol was applied to label 101 regions of interests.

B. ADNI DATASET

Alzheimer's Disease Neuroimaging Initiative (ADNI) data was collected from adni.loni.usc.edu. The ADNI was launched in 2003 as a public-private partnership with the primary goal of measuring the progression of mild cognitive impairment and early Alzheimer's disease from different modality of data. From the ADNI data set, we have considered 'ADNI1: Complete 3Yr 3T' data. There are 60 subjects and 351 MRIs. The average age of the participants was 75.87 years with a standard deviation of 7.078 ranges from 55 years to 90 years of age. There are three categories of the scans ADD, MCI, and CN containing 20, 18, and 22 subjects, respectively. The participants are imbalanced in terms of gender. CN category has 7 males and 15 females participants with an average age (± standard deviation) of 77.57 (± 4.04) years. The age range is 70 years to 88 years. There are 51 scans from male participants, while the number for the opposite gender is 78. There are 18 participants with MCI among which 14 are males and 4 are females with an average age of 74.655(± 8.00) ranging from 55 to 90 years. The number of scans from male and female participants are 111 and 34, respectively. Among 20 ADD subjects, 7 are males and 13 are females. The age ranges from 57 to 90 years with an average of 75.32(±8.47) years. The number of scans from male and female participants is 32 and 45, respectively. Details on these imaging protocols are available at http://adni.loni.usc.edu/methods/documents/mriprotocols/.

The raw data for MRI scans were in NiFTI format in the ADNI database. The images were MPR, grad warped, B1 non-uniformity corrected, and non-parametric nonuniform intensity normalized i.e.,. N3 Normalized and scaled. For our experiment, 3D scans were preprocessed to obtain training and testing 2D patches which will be described in the following section.

III. METHODS

The framework of our method is depicted in Fig. 1. First, we learn the deep metric for the ground truth landmark. We train a deep Siamese network to learn the hippocampus features. We have generated three-view patches (TVPs) from the vicinity of a hippocampus location (see Fig. 2.)

Then, we train a coarse prediction network (CPN) to roughly learn the displacement of the hippocampi from a spherical reference frame. At last, two fine-tuning networks for the left and right hippocampi are trained and tested to learn the displacement of the landmark from reference voxels of the rough estimation of CPN. Local features are considered for verification and fine-tuning. The rough estimation is fine-tuned by deploying another HRN that uses the local features of the landmarks. For feature verification on the rough estimation, we train a Siamese network. In the following subsection, we will discuss the modules in detail.

A. COARSE PREDICTION NETWORK

CPN learns global features to roughly estimate the offset of the landmark. To predict initial estimates of the offsets from the center of TVPs, this network takes three view slices with the size of $64 \times 64 \times 3$ from a spherical surface centered at the middle of the MRI under study. First, we start with the architecture and settings of [51]. Then, we have tweak the network by trial and error. Our CPN network consists of eight convolution layers. A max pooling layers is present after every two convolution layers. The batch normalization layer was included after every maxpooling layer (before the second convolution to the Flatten layer). The Flatten layer is followed by two dense layers. The first dense layer is followed by a dropout of 0.25. Batch normalization [52], [52]–[54] and dropout layers limits the likelihood of over-fitting.

B. FINE TUNING NETWORK

The FTN consists of five convolution layers, two max pooling layers and three batch normalization layers. The flatten layer follows two dense layers. The first dense layer is followed by a dropout of 0.25. The input to this network are TVPs of size $16 \times 16 \times 3$. A cube of size $8 \times 8 \times 8$ was assumed to be centered on the manually annotated voxel locations, and training TVP samples for FTN were generated from random offsets. This network predicts fine-tuned estimates of the offsets from the

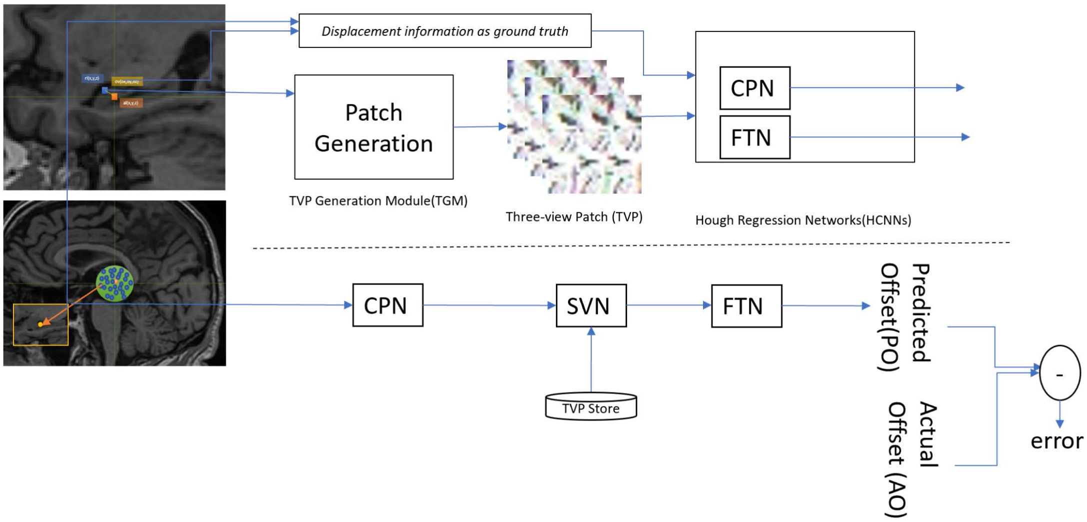


FIGURE 1. Hippocampi Localization using Hough Regression Network(HRN) and Deep Metric Learning. Patch Generation module generates three-view patches(TVPs) for coarse prediction network(CPN), fine tuning network(FTN) and Siamese verification network(SVN). CPN roughly estimates the hippocampi locations while FTN fine tunes the rough estimations that are verified with SVN.

center of TVPs. The TVPs used in this level considers local correlation between the landmark location and its neighboring voxels. The models learn the geometric correlation with its neighboring structures. The activation function used in this network follows the same pattern as CPN.

C. SIAMESE VERIFICATION NETWORK

For verifying the course location estimation from CPN, a Siamese model was trained and tested prior to deployment of the FTN to get final location. Siamese network learns the deep metric to differentiate hippocampal features and non-hippocampal features. A 2D CNN, consist of four convolution layer with two max pooling layers, was considered for the Siamese twin network. The feature map is 64 in length. The twin network learned a function that transform the input TVPs into a target space such that the Euclidean distance in the target space approximate the semantic distance between the TVPs. The learning process minimizes the contrastive loss function that ensures that the similarity metric is small for a pair of hippocampus-TVPs and large for a non-hippocampal region TVP inclusion. The CNN represents the locally distinguishing features of the hippocampus. In each channel of the twin, there are four convolution layers and one fully connected layer. There is a batch normalization layer after each convolution layer. The last layer of the twin-CNN is the Euclidean distance between the feature embedding of the two different networks.

We used leaky Relu ($f(x) = max(\alpha x, x)$) as the activation function for all of the convolution layers with $\alpha = 0.3$. In the first fully connected layer, we used tanh as the activation. In the last layer, we used leaky Relu with $\alpha = 0.9$.

The input to the network is a pair of TVPs (T_i, T_j) and a label y_{ij}. (T_i, T_j) are passed to the CNNs and each CNN work as a mapping function. The pair of TVPs yield feature representation by $F(T_i)$ and $F(T_j)$. The cost module which is the Euclidean distance operator generates the distance $\hat{y}$ between $F(T_i)$ and $F(T_j)$. At, training time the pairs of TVPs (T_i, T_j) are generated as specified in section IV-B4. At test time, the coarse predicted TVPs were compared with the TVPs stored in database constructed by the procedure mentioned in [55].

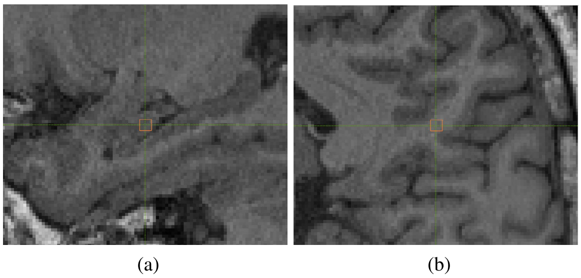

(a) (b)

FIGURE 2. An example of (a) ground-truth locations of hippocampus and (b) non-hippocampi in an MRI (viewed on the sagittal plane).

We used contrastive loss function for training the deep metric learning (DML) network. The loss function is defined in equation (1).

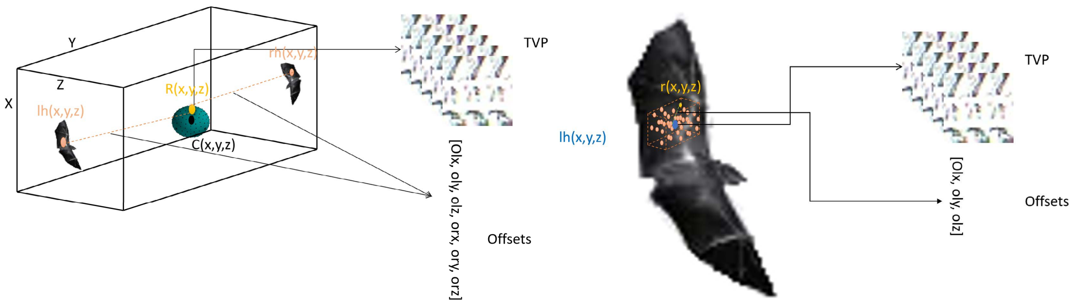
$$L(y, \hat{y}) = y\hat{y}^2 + (1 - y)[max(\lambda - \hat{y}, 0)]^2 \qquad (1)$$

Here, y is the actual distance (0 or 1) and $\hat{y}$ is the predicted distance between the input pairs. $\lambda = 2$ is used as a distance margin constraint. The constraint defines a radius in the target space around the Euclidean distance. Unlikely pairs have a contribution in the loss if their distance is within the defined margin.

IV. EXPERIMENTAL SETUP

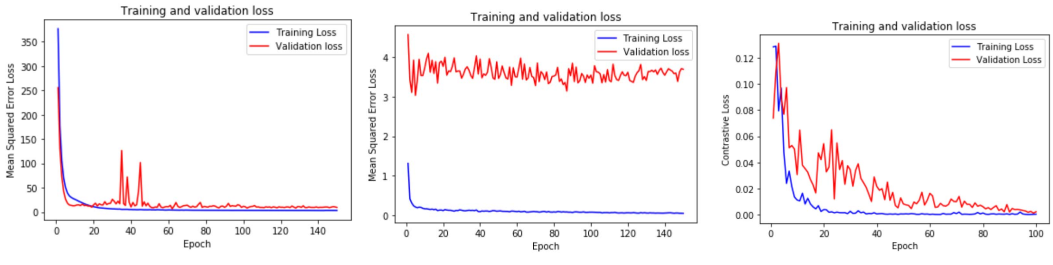
A. PLATFORM

The entire experiment was performed with a Python 3.7 environment. We used the TensorFlow GPU 1.12. Keras was used



(a) (b)

FIGURE 3. **Ground truth preparation for (a) a coarse prediction network (CPN) and (b) fine tuning network (FTN). Here, $r(r_x, r_y, r_z)$ is the reference voxels and (h_x, h_y, h_z) is manual annotations for hippocampus locations. (o_x, o_y, o_z) is an offset calculated by equation (4).**



(a) (b) (c)

FIGURE 4. **Training and validation loss of (a) Coarse Prediction Network (b) Fine-Tuning Network (c) Feature Verification Network.**

as the backend. The operating system was Windows 10. The verification network and fine-tuning model was trained on an ''Intel(R) Xeon (R) Silver 4114 @ 2.20 GHz, 10 cores and 20 logical processors with a 32 GB RAM'' machine. The GPU was NVIDIA Quadro P4000. The rest of the experiments were performed on Intel(R) Xeon (R) CPU E5-1607 v4 @ 3.10 GHz with a 32 GB RAM machine. The GPU was NVIDIA Quadro M4000. ITK-SNAP [56] was used for viewing and navigating throw the neuroimaging informatics technology initiative (NIFTI) images.

B. GROUND TRUTH PREPARATION FOR PATCH-BASED MODEL LEARNING

Each skull stripped NIfTI formatted MRI scan was intensity normalized so that the mean intensity is zero and the unit standard deviation. The intensities are normalized for each MRI individually. After intensity normalization, the offset between the reference voxel and the hippocampus voxel is calculated and used as ground truth. We trained three different network. Therefore, the reference location and TVPs (in type and size) are different for each network.

1) MANUAL ANNOTATION

The hippocampus locations for both left and right hippocampi were manually annotated by an expert. Though hippocampus covers volumes of voxels, the expert considered the location that is useful to physicians in localizing the hippocampus with or without image guided technologies. The annotations were performed in two independent runs. If the difference between the annotations is greater than two voxels, the opinion of another expert was obtained and the two closest locations were averaged. We have denoted each hippocampus location as (h_x, h_y, h_z) for both right and left hippocampus. The networks for left and right hippocampus were trained separately with respective training sets.

2) GROUND TRUTH FOR COARSE PREDICTION NETWORK

The reference locations for CPN were generated from a sphere with a radius of 8cm centered at the middle voxel of the structured MRI (provides static anatomical information). The procedure for generating the reference is illustrated in figure 3a. Let (r_x, r_y, r_z) be the reference locations randomly sampled from a sphere. Three different patches of size 64 × 64 centered at (r_x, r_y), (r_y, r_z), (r_x, r_z) in axial, sagittal and coronal views, respectively, were generated and formed a TVP and are denoted as TVP_r. For MRI, I, the corresponding patch of size $\alpha \times \beta$ at the reference voxel $r(r_x, r_y, r_z)$ is defined by:

$$\begin{aligned} view1 &= I[r_x, r_y - \frac{\alpha}{2} : r_y + \frac{\alpha}{2}, r_z - \frac{\beta}{2} : r_z + \frac{\beta}{2}] \\ view2 &= I[r_x - \frac{\alpha}{2} : r_x + \frac{\alpha}{2}, r_y, r_z - \frac{\beta}{2} : r_z + \frac{\beta}{2}] \\ view3 &= I[r_x - \frac{\alpha}{2} : r_x + \frac{\alpha}{2}, r_y - \frac{\beta}{2} : r_y + \frac{\beta}{2}, r_z] \end{aligned} \qquad (2)$$

Then, the TVP denoted as T at (r_x, r_y, r_z) is formed by

$$T = [view1 \quad view2 \quad view2] \qquad (3)$$

TABLE 3. Localization error(in millimeter) in test set of GARD and ADNI data set.

Model	ADNI		GARD	
	Left Hippocampus	Right Hippocampus	Left Hippocampus	Right Hippocampus
CPN	3.86±1.71	3.95±1.42	3.67±1.81	3.63±1.92
CPN+VN	3.26±0.40	3.01±0.44	3.13±0.46	3.04±0.52
CPN+FTN	2.17±1.76	2.27±1.37	2.13±1.88	2.05±1.36
CPN+VN+FTN	1.79±0.83	1.55±0.61	1.70±0.50	1.66±0.49

The offset to this reference voxel from manually annotated hippocampus location is denoted as (o_x, o_y, o_z), calculated below.

$$(o_x, o_y, o_z) = (h_x - r_x, h_y - r_y, h_z - r_z) \tag{4}$$

The ground truth of CPN dataset is (o_x, o_y, o_z), the offset to the hippocampus from reference point. The input feature to this CNN is corresponding TVP T at the reference point. We generated 64 TVPs from each MRI, and the training, validation, and testing set are separated by the patient ID.

3) GROUND TRUTH FOR FINE TUNING NETWORK

We have considered a cube of the size of $8 \times 8 \times 8$ centered at (h_x, h_y, h_z) of the target hippocampus position. Uniform random distribution is used to sample the reference locations from the cube. The reference position generation methods are shown in Fig. 3b.

The ground truth for the target hippocampus, (o_x, o_y, o_z) is obtained using eq. (4) with new reference positions sampled from closer $8 \times 8 \times 8$ cube. Then, we generate TVPs from the reference location and produce the corresponding data sample $\{T, (o_x, o_y, o_z)\}$. The size of the TVPs is set to $16 \times 16 \times 3$. The three view patch, T, is used to predict the correct offset (o_x, o_y, o_z) to the target hippocampus position.

4) GROUND TRUTH FOR SIAMESE VERIFICATION NETWORK

For metric learning, we generated 64 positive TVPs and 64 negative TVPs from each MRI. The positive samples were randomly produced from the $8 \times 8 \times 8$ cube centered at the manually labeled hippocampi locations. The negative samples are produced from non-hippocampi regions of the brain. To generate the negative samples, we followed the same sampling approach for coarse reference positions generation. We prepared the pairs of TVP samples $\{(T_i, T_j), y\}$. If both patches, T_i and T_j, are from the same region, then the label is 1. If they belong to different regions, then the label is 0. In all metric training samples, at least one of TVPs are from hippocampus region. 32 positives and 32 negatives pairs were generated from MRIs for training.

C. DATA SET SEPARATION

We have used TVPs as the data unit for training and validation. For testing an MRI, we have considered the predictions of the models for the TVPs of given MRI. First, training and testing were performed on the GARD data set. Then, instances of the models were utilized for the purposes of training and testing on ADNI data. We have separated the MRIs based on the subject IDs so that there are no MRIs with the same subject ID in neither of the two sets. We used 60% of the subjects for training and validation and 40% for testing. We have taken the precaution to uniformly distribute the subjects of different gender, age and category among all of the sets for better generalization.

D. TRAINING

1) SIAMESE FEATURE VERIFICATION NETWORK

The SVN network's weights were initialized with Xavier initialization while the biases were initialized with normal distribution with mean and variance 0.01 and 0.01. λ was set to 2. λ set the lower bound of the dissimilarity between the hippocampus and non-hippocampus locations of samples with different labels.

The weight was optimized with the Adam optimizer [57]. The initial learning rate was set to 0.001 with the exponential decay rate of 0.1. The mini-batch size was 32. The training phase was run for 100 epochs. The training performance is illustrated in figure 4c. The details of the process of TVP generation at training and test time is described in [55].

2) TRAINING OF HOUGH REGRESSION NETWORKS

The training and validation performances of CPN and FTN are shown in 4a and 4b of figure 4. The CPN and FTN models were trained for 250 epochs. Mean squared error (as given by equation (5)) was used as the loss function.

$$mse = \frac{\sum_{i=1}^{batchsize}(\hat{o}_i - o_i)^2}{batchsize} \tag{5}$$

Here, $batchsize = 32$ is the number of the input-output pairs in a batch and $\hat{o}_i$ and o_i are the tuples representing predicted and ground-truth offsets, respectively.

The CPN network receives a TVP and a pair of offsets to left and right hippocampus for training. The CPN yields a pair of offsets for respective hippocampus. The FTN network is trained for each left and right hippocampus. It receives a TVP and outputs an offset for the left or right hippocampus location.

For both networks, the Xavier initializer [58] was used for the weight and bias initialization. Weight was optimized by the Adam optimizer [57] with an initial learning rate of 0.01. We have used five-fold cross validation in training the networks.

V. RESULTS AND DISCUSSION

Training and validation were performed on TVPs. The test was performed on MRI. The predictions of TVPs generated from the designated reference frame were averaged for obtaining the final prediction. We have used 132 MRIs from

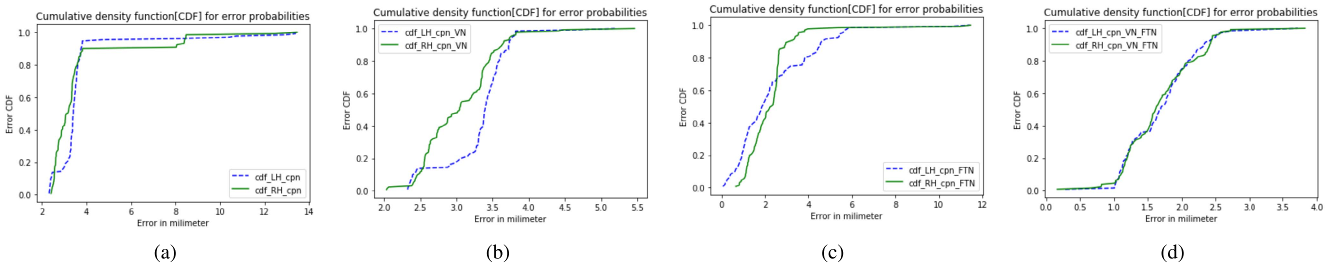


FIGURE 5. **Error cumulative density function (CDF) for (a) coarse prediction network (CPN), (b) CPN+verification network(VN), (c) CPN+fine tuning network(FTN) and (d) CPN+VN+FTN.**

TABLE 4. **Comparison of the Proposed Method with state-of-the-arts.**

Method	Landmark Name	Data set	Error
Partial Policy RL [59]	Vertibra centers	Dataset-A	2.79±1.98
Hourglass Network [60]	Knee point	Knee X-ray	2.50
Naive stacked Hourglass [61]	Optic Disk and Fovea	-	14.21
Deep CNN [32]	7 Cerebral Landmark	Head CT Image	3.45
CNN+U-Net+RNN [62]	Mitral Valve	MR	2.87
CNN+U-Net+RNN [62]	Righ Ventricular Insert Points	MR	3.64
Ensemble HCNN [44]	Left Hippocampus	GARD	2.32
Ensemble HCNN [44]	Right Hippocampus	GARD	2.25
Proposed Method	Left Hippocampus	GARD	1.70±0.50
		ADNI	1.79±0.83
	Right Hippocampus	GARD	1.66±0.49
		ADNI	1.55±0.61

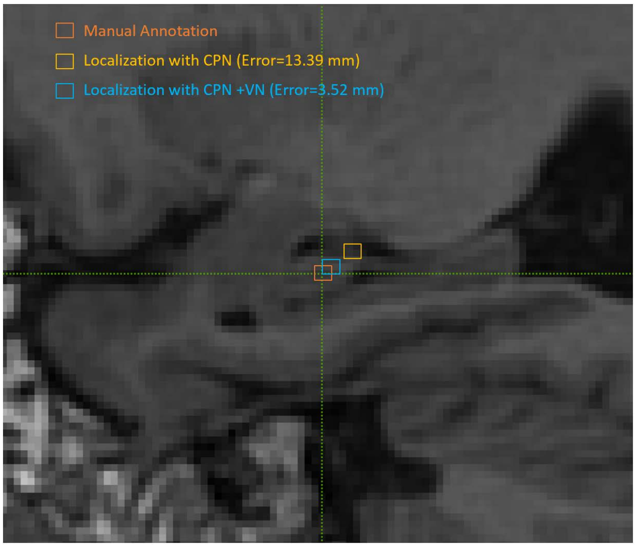

Manual Annotation
Localization with CPN (Error=13.39 mm)
Localization with CPN +VN (Error=3.52 mm)

FIGURE 6. **Error improvement with SVN; MRI 1503195 from the GARD data set is taken as an example(not drawn in scale). Error in the CPN predictions was 13.39 mm due to considering the predictions from all of the reference inputs. SVN verifies the features of the locations whether it matches with hippocampus features. If the match is not found, it drops the reference voxel and computes the location with remaining predictions. CPN+SVN provides an error of 5.52 mm;.**

GARD and 141 MRIs from ADNI for the purpose of testing. After the model was trained, the evaluation was carried out on the testing data set and the quantitative results were computed.

A. EVALUATION

To evaluate the models, we adopted the Euclidean distance between the location of manual annotation and predicted location. The output of the networks are the predicted offset $\hat{o}$. Let the reference location of the TVP under consideration be r. Adding the reference location to the offsets provides the hippocampus location.

The actual offset can be found from equation (4). If the actual offset is o_j, we can obtain the localization error from the difference between o_j and predicted offset $\hat{o}_j$. The error for each j_{th} MRI, E_j, can be computed from the Euclidean distance between the actual location of hippocampus and the predicted location averaged from numerous reference point and respective predicted offset $\hat{o}_j$. The total test error is computed by averaging over all the test set MRI errors.

B. LOCALIZATION ERROR

1) ERROR IN GARD DATASET

The error of the CPN network for localizing the left hippocampus is 3.67 mm with a standard deviation of 1.81 in the test data set. The error of the CPN network for the right hippocampus is 3.63± 1.92 mm. The CPN+VN framework shows a lower variance in the error in the test set. The error is 3.13±0.46 mm for localizing the left hippocampus. The errors are 3.04±0.52 mm for localizing the right hippocampus.

The CPN+FTN model shows an error of 2.13±1.88 mm for the left hippocampus localization. The errors for the right

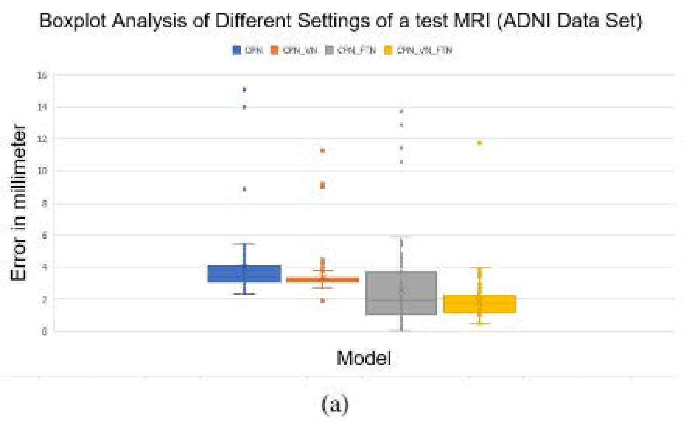


FIGURE 7. **Error Variances in different setting. Boxplot of prediction errors of CPN, CPN+SVN, CPN+FTN and CPN+SVN+FTN for localizing the left hippocampus from a) ADNI MRI (patient ID: 023_S_0926, MRI scan S20160) b) GARD MRI (ID:1503195); Total reference points for CPN is 64 for both the MRIs. While processing S20160 VN drops 17 reference points for outlier predictions and final localization was carried out by FTN with 47 TVPs. For MRI 1503195 VN drops 29 reference points for outlier predictions and final localization was carried out by FTN with 35 TVPs.**

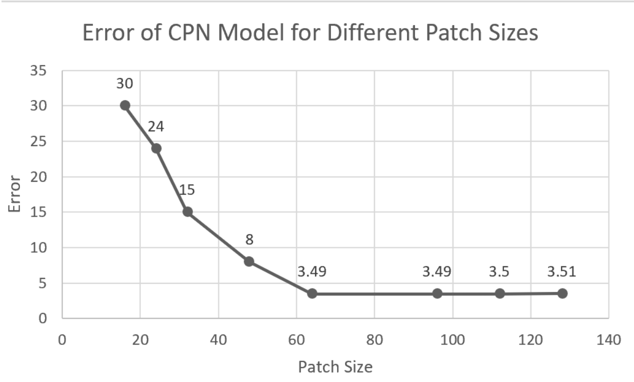


FIGURE 8. **Effects of patch size on CPN performance (observed on TVP-based analysis).**

hippocampus in the respective data set are 2.05±1.36 mm. Thus, the use of local information in the FTN networks boosted the accuracy. However, a large variance is still present in the error. The reason for the high variation in the error rate may be the contribution of the outliers predicted by CPN. These outliers were used as the reference voxel to generate the input to FTN. Using SVN to filter the locations predicted by CPN reduces the outliers. Applying FTN on the filtered predictions provides state-of-the-art accuracy. The CPN+VN+FTN provides an error of 1.70±0.50 mm for localizing the left hippocampus while the same is 1.66±0.49 mm for the right hippocampus. Table 3 reports the localization error of the models on the testing set data.

2) ERROR IN ADNI DATASET

Table 3 reports the localization error of the models on the testing set data. The error of the CPN network for localizing the left hippocampus is 3.86 mm with a standard deviation of 1.71 mm in the ADNI data set. The error of the CPN network for the right hippocampus is 3.95± 1.42 mm. The CPN+VN framework shows a lower variance in the error as demonstrated for GARD data. The error is 3.26±0.40 mm for localizing the left hippocampus, and 3.01±0.44 mm for localizing the right hippocampus. The CPN+FTN model shows an error of 2.17±1.76 mm for the left hippocampus localization and 2.27±1.37 mm for right hippocampus. The CPN+VN+FTN provides an error of 1.79±0.83 mm for localizing the left hippocampus. The error for right hippocampus localization is 1.55±0.61 mm.

C. DISCUSSION

The high variance in the error was reduced by including the SVN network to drop the outlier reference points which contributed higher error in CPN. Rather than directly estimating their locations, the model is trained to predict the displacement of the hippocampus locations from given reference locations. The TVPs on the reference point provide the global information for rough estimation. From CPN, coarse-grained hippocampus locations are calculated. We verified the feature of coarse-grained locations by SVN that makes the model robust to the outliers predicted by CPN. The filtered locations are used as the reference point for TVP generation to be used in FTN. FTN fine-tuned the rough estimation by providing more accurate local offsets. The final locations were calculated from these offsets. The method demonstrates state-of-the-art accuracy on both the ADNI and GARD data sets.

1) SVN REDUCES ERROR VARIANCE

SVN network provides robustness in localization and reduces the error variances. An examination of Table 3 shows that the errors for CPN and CPN+FTN have higher variances. CPN in GARD and ADNI data have 1.71 mm and 1.81 mm standard deviation in error for localizing the left hippocmapus. The corresponding values are 1.42 mm and 1.92 mm for localizing the right hippocampus for the same data sets. Using SVN reduces the standard deviations to 0.40 mm and 0.46 mm for localizing the left hippocampus in the ADNI and GARD data, respectively. Localization of the right hippocampus shows the standard deviations of 0.44 mm and 0.52 mm. The outliers of CPN also influence the CPN+FTN network. SVN provides robust results with smaller variance. The boxplots

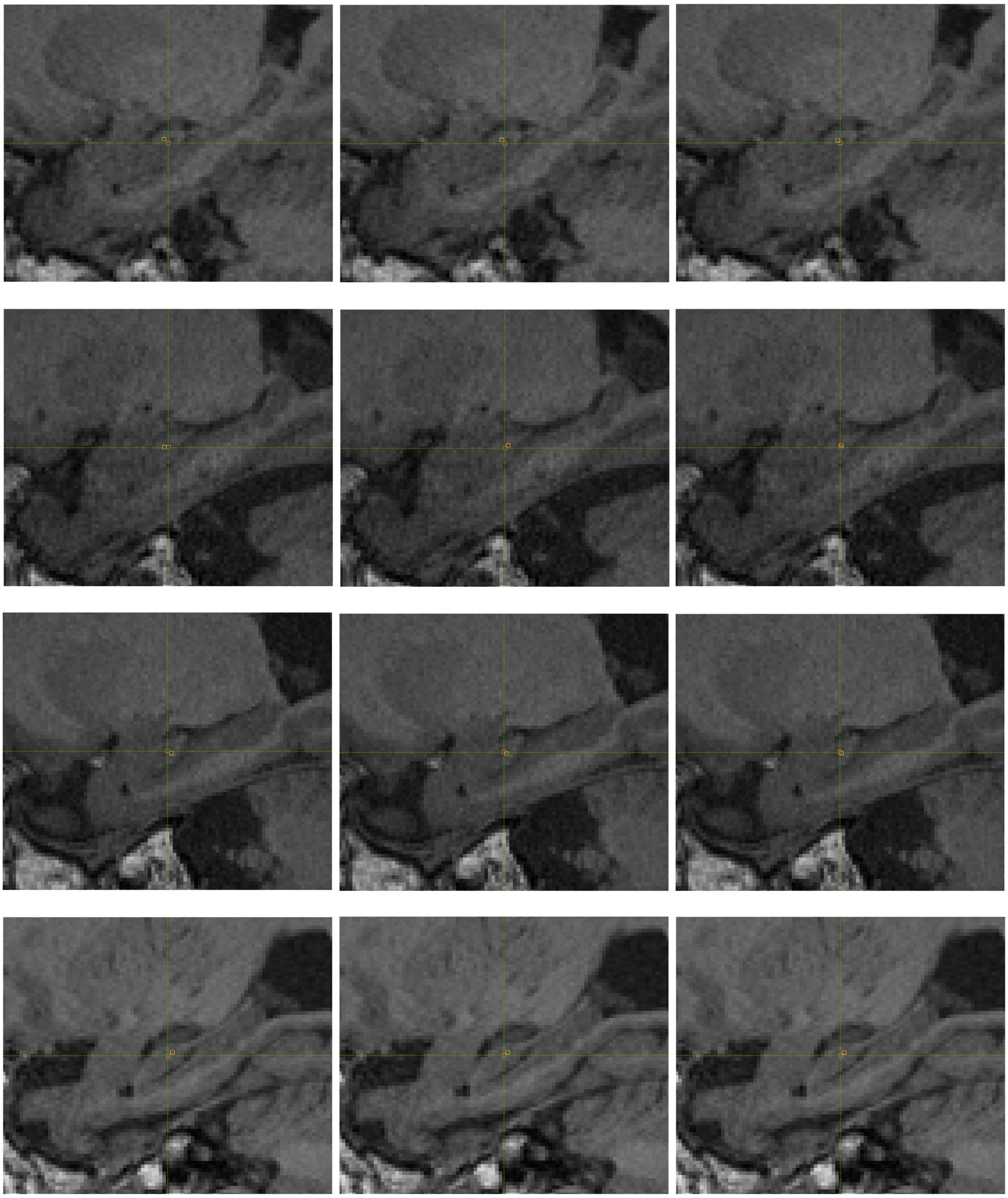


FIGURE 9. Demonstration of qualitative localization results for different MRI categories in different settings. The average case is considered for presentation. The rows from top to bottom indicate MRI instances of 17102706, 14060801, 17101603, 14062205 from the ADD, aMCI, naMCI and NC classes respectively. The columns from left to right indicate the CPN, CPN+VN and CPN+VN+FTN models, respectively. The manual annotation for 17102706, 14060801, 17101603 and 14062205 are (83, 136, 86),(85, 144, 88),(81, 154, 84), and (87, 138, 81), respectively.

in figure 7a and 7b summarizes the results, and with 95% confidence that the true medians are different when using the SVN.

Figure 7 depicts a case where the CPN prediction has a large error due to the outliers. Taking input TVP from this outlier-biased prediction influences the error in the prediction of FTN. Subsequently, FTN also provides erroneous offsets. Finally, obtaining an erroneous offset leads to larger error in the final localization. The SVN drops the outliers in coarse-grain prediction so that FTN obtains the input from the known reference frame. If we deploy SVN for feature verification, the outliers will be automatically dropped for next processing. This improves the error rate and ensures robustness. Some sample results are depicted in figure 9.

2) PATCH SIZE SELECTION

We have selected the patch size after several trials using different squared patches sizes of 16, 32, 48, 64, 96, 112, 128 (64 is chosen). For FTN, we have examined the sizes of 8, 16, and 32, and found that 16 was optimal. We found that the selected patch sizes provide a better feature for the localization in our data-set, given the ADNI and GARD MRI sizes. The details are provided in figure 8.

3) COMPARISON

Al and Yun [59] demonstrated an error of 2.79 ± 1.98 mm in localizing vertebra-centers in spine MR volumes using partial policy-based reinforcement learning (RL). For better learning, they adopted the actor-critic direct policy search method to learn the optimal agent-behavior. The partial policy-based RL algorithm was proposed for faster behavior learning.

Tiulpin *et al.* [60] and Maiya and Mathur [61] utilized the Hourglass network for localizing anatomical landmarks. 79% of the key points in knee X-ray images were located with an error of 2.5 mm in [60]. Euclidean distance of 14.21 mm was demonstrated in [61] for optic disk and fovea localization.

Zhang *et al.* [32] proposed to a deep convolutional neural network (CNN) to classify a head CT image in terms of its content and to localize landmarks. They obtained an average localization error of 3.45 mm for 7 landmarks located around each inner ear. This is better than the results achieved with earlier methods that we have proposed for the same tasks.

Zon *et al.* [62] proposed a four-stage deep learning model for localizing mitral valve points and right ventricular insert points in MR scans. They utilized CNN for Cropping, U-net for coarse grained location RNN for incorporating temporal and spatial dynamics for landmark locations. The final predictions were made based on the combination of U-net and RNN. The method demonstrated the average errors of 2.87 mm and 3.64 mm for the mitral valve points and the right ventricular insert points, respectively. Even though RNN in this approach demonstrated how to model temporal or spatial dependencies in landmark localization, Basher *et al.* [44] proposed a more accurate prediction with ensemble of two-stage Hough CNN. The model proposed by Basher *et al.* [44] did not consider the outliers predicted by CPN and thus demonstrated a lack of robustness in our experiment. We have proposed siamese verification to remove the outlier predictions in the first step. Out method provides robust accuracy. Table 4 presents the comparison of our method to the state-of-the-art methods.

VI. CONCLUSION

Siamese network along with cascaded HRNs provides robust localization performance. Our proposed pipeline demonstrated an error of 1.70 ± 0.50 mm for localizing the left hippocampus and an error of 1.66 ± 0.49 mm for localizing the right hippocampus in GARD dataset. The errors in ADNI dataset was 1.79 ± 0.83 and 1.55 ± 0.61 for localizing left hippocampus and right hippocampus, respectively. The results demonstrated a promising performance for anatomical landmark localization, specifically cerebral landmark localization in sMRI modality.

REFERENCES

[1] J. Ren, F. Huang, Y. Zhou, L. Zhuang, J. Xu, C. Gao, S. Qin, and J. Luo, "The function of the hippocampus and middle temporal gyrus in forming new associations and concepts during the processing of novelty and usefulness features in creative designs," *NeuroImage*, vol. 214, Jul. 2020, Art. no. 116751, doi: 10.1016/j.neuroimage.2020.116751.

[2] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019, doi: 10.1109/JBHI.2018.2882392.

[3] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, and M. Xu, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Mar. 2020, Art. no. 116459, doi: 10.1016/j.neuroimage.2019.116459.

[4] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of patch-based classifiers for diagnosis of Alzheimer diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019, doi: 10.1109/ACCESS.2019.2920011.

[5] A. Basher, S. Ahmed, and H. Y. Jung, "One step measurements of hippocampal pure volumes from MRI data using an ensemble model of 3-D convolutional neural network," *Korean Inst. Smart Media*, vol. 9, no. 2, pp. 22–32, Jun. 2020. [Online]. Available: http://www.koreascience.or.kr/article/JAKO202018955009108.page

[6] S. Ahmed, B. C. Kim, K. H. Lee, and H. Y. Jung, "Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0242712.

[7] R. D. Rubin, P. D. Watson, M. C. Duff, and N. J. Cohen, "The role of the hippocampus in flexible cognition and social behavior," *Frontiers Hum. Neurosci.*, vol. 8, pp. 1–15, Sep. 2014.

[8] V. Dhikav and K. Anand, "Potential predictors of hippocampal atrophy in Alzheimer's disease," *Drugs Aging*, vol. 28, no. 1, pp. 1–11, Jan. 2011, doi: 10.2165/11586390-000000000-00000.

[9] L. G. Apostolova, R. A. Dutton, I. D. Dinov, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson, "Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps," *Arch. Neurol.*, vol. 63, no. 5, pp. 693–699, May 2006, doi: 10.1001/archneur.63.5.693.

[10] J.-H. Cai, Y. He, X.-L. Zhong, H. Lei, F. Wang, G.-H. Luo, H. Zhao, and J.-C. Liu, "Magnetic resonance texture analysis in Alzheimer's disease," *Academic Radiol.*, vol. 27, no. 12, pp. 1774–1783, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1076633220300337

[11] H. C. Achterberg, L. Sørensen, F. J. Wolters, W. J. Niessen, M. W. Vernooij, M. A. Ikram, M. Nielsen, and M. de Bruijne, "The value of hippocampal volume, shape, and texture for 11-year prediction of dementia: A population-based study," *Neurobiol. Aging*, vol. 81, pp. 58–66, Sep. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0197458019301423

[12] G.-P. Peng, Z. Feng, F.-P. He, Z.-Q. Chen, X.-Y. Liu, P. Liu, and B.-Y. Luo, "Correlation of hippocampal volume and cognitive performances in patients with either mild cognitive impairment or Alzheimer's disease," *CNS Neurosci. Therapeutics*, vol. 21, no. 1, pp. 15–22, Jan. 2015.

[13] G. B. Frisoni, N. C. Fox, C. R. Jack, Jr., P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Rev. Neurol.*, vol. 6, pp. 67–77, Feb. 2010. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/20139996 and https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938772/

[14] S. A. Eshkoor, T. A. Hamid, C. Y. Mun, and C. K. Ng, "Mild cognitive impairment and its management in older people," *Clin. Intervent. Aging*, vol. 10, pp. 687–693, Apr. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4401355/

[15] K. L. Narr, P. M. Thompson, P. Szeszko, D. Robinson, S. Jang, R. P. Woods, S. Kim, K. M. Hayashi, D. Asunction, A. W. Toga, and R. M. Bilder, "Regional specificity of hippocampal volume reductions in first-episode schizophrenia," *NeuroImage*, vol. 21, no. 4, pp. 1563–1575, Apr. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811903007328

[16] R. Ugolotti, P. Mesejo, S. Cagnoni, M. Giacobini, and F. Di Cunto, "Automatic hippocampus localization in histological images using PSO-based deformable models," in *Proc. 13th Annu. Conf. Companion Genet. Evol. Comput. (GECCO)*, Dublin, Ireland, N. Krasnogor and P. L. Lanzi, Eds., Jul. 2011, pp. 487–494, doi: 10.1145/2001858.2002038.

[17] P. Mesejo, R. Ugolotti, F. Di Cunto, M. Giacobini, and S. Cagnoni, ''Automatic hippocampus localization in histological images using differential evolution-based deformable models,'' *Pattern Recognit. Lett.*, vol. 34, no. 3, pp. 299–307, Feb. 2013, doi: 10.1016/j.patrec.2012.10.012.

[18] M.-R. Siadat, H. Soltanian-Zadeh, and K. V. Elisevich, ''Knowledge-based localization of hippocampus in human brain MRI,'' *Comput. Biol. Med.*, vol. 37, no. 9, pp. 1342–1360, 2007, doi: 10.1016/j.compbiomed.2006.12.010.

[19] C. Payer, D. Štern, H. Bischof, and M. Urschler, ''Integrating spatial configuration into heatmap regression based CNNs for landmark localization,'' *Med. Image Anal.*, vol. 54, pp. 207–219, May 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841518305784

[20] M. Urschler, T. Ebner, and D. Štern, ''Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization,'' *Med. Image Anal.*, vol. 43, pp. 23–36, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841517301342

[21] D. Stern, T. Ebner, and M. Urschler, ''From local to global random regression forests: Exploring anatomical landmark localization,'' in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, Athens, Greece, vol. 9901, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. B. Ünal, and W. Wells, Eds., Oct. 2016, pp. 221–229, doi: 10.1007/978-3-319-46723-8_26.

[22] V. Potesil, T. Kadir, G. Platsch, and M. Brady, ''Personalized graphical models for anatomical landmark localization in whole-body medical images,'' *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 29–49, 2015, doi: 10.1007/s11263-014-0731-7.

[23] C. Payer, D. Stern, H. Bischof, and M. Urschler, ''Regressing heatmaps for multiple landmark localization using CNNs,'' in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, Athens, Greece, vol. 9901, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. B. Ünal, and W. Wells, Eds., Oct. 2016, pp. 230–238, doi: 10.1007/978-3-319-46723-8_27.

[24] H. Liao, A. Mesfin, and J. Luo, ''Joint vertebrae identification and localization in spinal CT images by combining short- and long-range contextual information,'' *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1266–1275, May 2018.

[25] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi, ''Vertebrae localization in pathological spine CT via dense classification from sparse annotations,'' in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, Nagoya, Japan, vol. 8150, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Berlin, Germany: Springer, Sep. 2013, pp. 262–270, doi: 10.1007/978-3-642-40763-5_33.

[26] R. Donner, B. H. Menze, H. Bischof, and G. Langs, ''Global localization of 3D anatomical structures by pre-filtered Hough forests and discrete optimization,'' *Med. Image Anal.*, vol. 17, no. 8, pp. 1304–1314, 2013, doi: 10.1016/j.media.2013.02.004.

[27] A. Criminisi, D. P. Robertson, E. Konukoglu, J. Shotton, S. D. Pathak, S. White, and K. M. Siddiqui, ''Regression forests for efficient anatomy detection and localization in computed tomography scans,'' *Med. Image Anal.*, vol. 17, no. 8, pp. 1293–1303, 2013, doi: 10.1016/j.media.2013.01.001.

[28] H. Chen, C. Shen, J. Qin, D. Ni, L. Shi, J. C. Y. Cheng, and P. Heng, ''Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks,'' in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, Munich, Germany, vol. 9349, N. Navab, J. Hornegger, W. Wells, III, and A. Frangi, Eds. Cham, Switzerland: Springer, Oct. 2015, pp. 515–522, doi: 10.1007/978-3-319-24553-9_63.

[29] Y. Duan, X. Fan, H. Cheng, and H. Kang, ''Gradient regression for brain landmark localization on magnetic resonance imaging,'' in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4013–4017.

[30] K. Xu, T. Sun, and X. Jiang, ''Video anomaly detection and localization based on an adaptive intra-frame classification network,'' *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 394–406, Feb. 2020, doi: 10.1109/TMM.2019.2929931.

[31] J. M. H. Noothout, B. D. De Vos, J. M. Wolterink, E. M. Postma, P. A. M. Smeets, R. A. P. Takx, T. Leiner, M. A. Viergever, and I. Isgum, ''Deep learning-based regression and classification for automatic landmark localization in medical images,'' *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4011–4022, Dec. 2020, doi: 10.1109/TMI.2020.3009002.

[32] D. Zhang, J. Wang, J. H. Noble, and B. M. Dawant, ''HeadLocNet: Deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs,'' *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101659. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841520300268

[33] B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum, ''ConvNet-based localization of anatomical structures in 3-D medical images,'' *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1470–1481, Jul. 2017.

[34] O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. Martinez-Möller, S. G. Nekolla, and N. Navab, ''Fast multiple organ detection and localization in whole-body MR Dixon sequences,'' in *Proc. 14th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, vol. 6893, Toronto, ON, Canada, G. Fichtinger, A. L. Martel, and T. M. Peters, Eds. Berlin, Germany: Springer, Sep. 2011, pp. 239–247, doi: 10.1007/978-3-642-23626-6_30.

[35] B. Menze, G. Langs, Z. Tu, and A. Criminisi, ''Whole-body anatomy localization via classification and regression forests,'' *Med. Image Anal.*, vol. 17, no. 8, p. 1282, Dec. 2013, doi: 10.1016/j.media.2013.09.005.

[36] O. Oktay, ''Learning anatomical image representations for cardiac imaging,'' Ph.D. dissertation, Imperial College London, London, U.K., 2017. [Online]. Available: http://ethos.bl.U.K./OrderDetails.do?uin=U.K..bl.ethos.733255

[37] O. Oktay, W. Bai, R. Guerrero, M. Rajchl, A. de Marvao, D. P. O'Regan, S. A. Cook, M. P. Heinrich, B. Glocker, and D. Rueckert, ''Stratified decision forests for accurate anatomical landmark localization in cardiac images,'' *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 332–342, Jan. 2017, doi: 10.1109/TMI.2016.2597270.

[38] N. B. Albayrak, A. B. Oktay, and Y. S. Akgül, ''Prostate localization from abdominal ultrasound images by using a two-level approach,'' in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, Antalya, Turkey, May 2017, pp. 1–4, doi: 10.1109/SIU.2017.7960487.

[39] T. Ebner, D. Stern, R. Donner, H. Bischof, and M. Urschler, ''Towards automatic bone age estimation from MRI: Localization of 3D anatomical landmarks,'' in *Proc. 17th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, vol. 8674, Boston, MA, USA, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. D. Howe, Eds. Cham, Switzerland: Springer, Sep. 2014, pp. 421–428, doi: 10.1007/978-3-319-10470-6_53.

[40] H. Y. Jung, S. Lee, Y. S. Heo, and I. D. Yun, ''Forest walk methods for localizing body joints from single depth image,'' *PLoS ONE*, vol. 10, no. 9, pp. 1–20, Sep. 2015, doi: 10.1371/journal.pone.0138328.

[41] Y.-C. Kim, Y. Chung, and Y. H. Choe, ''Automatic localization of anatomical landmarks in cardiac MR perfusion using random forests,'' *Biomed. Signal Process. Control*, vol. 38, pp. 370–378, Sep. 2017, doi: 10.1016/j.bspc.2017.07.001.

[42] D. Stern, T. Ebner, and M. Urschler, ''Automatic localization of locally similar structures based on the scale-widening random regression forest,'' in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Prague, Czech Republic, Apr. 2016, pp. 1422–1425, doi: 10.1109/ISBI.2016.7493534.

[43] S. Miao, Z. J. Wang, and R. Liao, ''A CNN regression approach for real-time 2D/3D registration,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016, doi: 10.1109/TMI.2016.2521800.

[44] A. Basher, K. Y. Choi, J. J. Lee, B. Lee, B. C. Kim, K. H. Lee, and H. Y. Jung, ''Hippocampus localization using a two-stage ensemble Hough convolutional neural network,'' *IEEE Access*, vol. 7, pp. 73436–73447, 2019, doi: 10.1109/ACCESS.2019.2920005.

[45] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, ''Automatic localization and discrete volume measurements of hippocampi from MRI data using a convolutional neural network,'' *IEEE Access*, vol. 8, pp. 91725–91739, 2020, doi: 10.1109/ACCESS.2020.2994388.

[46] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah, ''Signature verification using a 'Siamese' time delay neural network,'' *Int. J. Pattern Recognit. Artif. Intell.*, vol. 7, no. 4, pp. 669–688, 1993.

[47] G. Koch, R. Zemel, and R. Salakhutdinov, ''Siamese neural networks for one-shot image recognition,'' in *Proc. ICML Deep Learn. Workshop*, Lille, France, vol. 2, 2015, pp. 1–30.

[48] K. Y. Choi *et al.*, ''APOE promoter polymorphism-219T/G is an effect modifier of the influence of APOE $\epsilon 4$ on Alzheimer's disease risk in a multiracial sample,'' *J. Clin. Med.*, vol. 8, no. 8, pp. 1–12, 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31426376/

[49] M. N. I. Qureshi, S. Ryu, J. Song, K. H. Lee, and B. Lee, "Evaluation of functional decline in Alzheimer's dementia using 3D deep learning and group ICA for rs-fMRI measurements," *Frontiers Aging Neurosci.*, vol. 11, pp. 1–9, Feb. 2019.

[50] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, Jan. 2020, doi: 10.1007/s12021-019-09419-w.

[51] H. Fan and E. Zhou, "Approaching human level facial landmark localization by deep learning," *Image Vis. Comput.*, vol. 47, pp. 27–35, Mar. 2016, doi: 10.1016/j.imavis.2015.11.004.

[52] C. Summers and M. J. Dinneen, "Four things everyone should know to improve batch normalization," in *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–18. [Online]. Available: https://openreview.net/forum?id=HJx8HANFDH

[53] J. Yan, R. Wan, X. Zhang, W. Zhang, Y. Wei, and J. Sun, "Towards stabilizing batch statistics in backward propagation of batch normalization," in *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–17. [Online]. Available: https://openreview.net/forum?id=SkgGjRVKDS

[54] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: An empirical study of their impact to deep learning," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 12777–12815, May 2020, doi: 10.1007/s11042-019-08453-9.

[55] S. Ahmed and H. Y. Jung, "Siamese network for learning robust feature of hippocampi," *Korean Inst. Smart Media*, vol. 9, no. 3, pp. 9–17, Sep. 2020. [Online]. Available: https://www.koreascience.or.kr/article/JAKO202028662596599.page

[56] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, Y. Bengio and Y. LeCun, Eds., May 2015, pp. 1–15.

[58] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*, Sardinia, Italy, Y. W. Teh and D. M. Titterington, Eds., vol. 9, May 2010, pp. 249–256. [Online]. Available: http://proceedings.mlr.press/v9/glorot10a.html

[59] W. Abdullah Al and I. D. Yun, "Partial policy-based reinforcement learning for anatomical landmark localization in 3D medical images," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1245–1255, Apr. 2020, doi: 10.1109/TMI.2019.2946345.

[60] A. Tiulpin, I. Melekhov, and S. Saarakkala, "KNEEL: Knee anatomical landmark localization using hourglass networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Seoul, South Korea, Oct. 2019, pp. 352–361, doi: 10.1109/ICCVW.2019.00046.

[61] S. R. Maiya and P. Mathur, "Rethinking retinal landmark localization as pose estimation: Naïve single stacked network for optic disk and fovea detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1125–1129, doi: 10.1109/ICASSP40776.2020.9053177.

[62] M. van Zon, M. Veta, and S. Li, "Automatic cardiac landmark localization by a recurrent neural network," *Proc. SPIE*, vol. 10949, pp. 295–307, Mar. 2019, doi: 10.1117/12.2512048.

SAMSUDDIN AHMED (Member, IEEE) received the B.Sc. degree in computer science and engineering from the University of Chittagong, in 2010, and the M.S. degree from the Department of Computer Engineering, Chosun University, South Korea, in 2020, under the supervision of Prof. Ho Yub Jung. He is currently an Assistant Professor with the Department of Information and Communication Technology, Bangabandhu Sheikh Mujibur Rahman Digital University (BDU), Bangladesh. His research interests include machine vision, deep learning, data analysis, explainable AI, and the IoT.

KUN HO LEE received the B.S. degree from the Department of Genetic Engineering, Korea University, Seoul, Republic of Korea, in 1989, and the M.S. and Ph.D. degrees from the Department of Molecular Biology, Seoul National University, Seoul, in 1994 and 1998, respectively. He is currently an Associate Professor with the Department of Biomedical Science, Chosun University, Gwangju, Republic of Korea, where he also works with the National Research Center for Dementia. His current research interests include brain image analysis and the development of prediction model for neurodegenerative diseases based on MRI and genetic variants.

HO YUB JUNG received the B.S. degree in electrical engineering from The University of Texas at Austin, in 2002, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, in 2006 and 2012, respectively. He was with Samsung Electronics as a Senior Engineer, for two years. Since 2017, he has been an Assistant Professor with the Department of Computer Engineering, Chosun University. His research interests include computer vision, machine learning, and medical imaging.