

Symmetry-Enhanced Attention Network for Acute Ischemic Infarct Segmentation with Non-Contrast CT Images

Kongming Liang¹, Kai Han², Xiuli Li², Xiaoqing Cheng³, Yiming Li²,
Yizhou Wang⁴, and Yizhou Yu^{2,5} (✉)

¹Pattern Recognition and Intelligent System Laboratory, School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China

²Deepwise AI Lab, Beijing, China

³Department of Medical Imaging, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu, China

⁴Department of Computer Science and Technology, Peking University, Beijing, China

⁵The University of Hong Kong, Pokfulam, Hong Kong
yizhouy@acm.org

Abstract. Quantitative estimation of the acute ischemic infarct is crucial to improve neurological outcomes of the patients with stroke symptoms. Since the density of lesions is subtle and can be confounded by normal physiologic changes, anatomical asymmetry provides useful information to differentiate the ischemic and healthy brain tissue. In this paper, we propose a symmetry enhanced attention network (SEAN) for acute ischemic infarct segmentation. Our proposed network automatically transforms an input CT image into the standard space where the brain tissue is bilaterally symmetric. The transformed image is further processed by a U-shape network integrated with the proposed symmetry enhanced attention for pixel-wise labelling. The symmetry enhanced attention can efficiently capture context information from the opposite side of the image by estimating long-range dependencies. Experimental results show that the proposed SEAN outperforms some symmetry-based state-of-the-art methods in terms of both dice coefficient and infarct localization.

Keywords: Computer aided diagnosis · Infarct Segmentation · Acute ischemic stroke · Attention Mechanism · Deep learning

1 Introduction

Acute ischemic stroke is one of the leading causes of death and disability worldwide and imposes an enormous burden for the health care system [8]. The use of pretreatment neuroimaging is critical to improve neurological outcomes of patients with stroke symptoms. Compared to MRI, non-contrast head CT scan is commonly used as the initial imaging because of its wide availability and low acquisition time. To interpret early infarct signs in CT, the Alberta Stroke Program Early CT Score (ASPECTS) evaluation was proposed at the beginning in

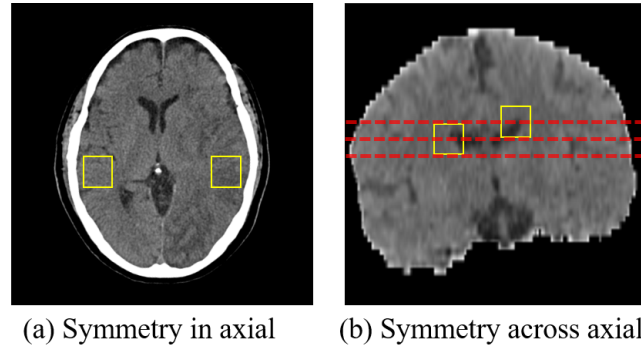


Fig. 1. The anatomical symmetry of the brain CT images. (a) and (b) show the two symmetrical patches of the image in axial view and across axial view. The dotted red line in (b) denotes the slice from axial view. Due to the rotation of patient’s head, the symmetrical landmarks may appear on different images in the axial view.

the 2000s [2] and has found increasing acceptance in clinical practice. However, ASPECTS evaluation is only an approximation of the assessment of early ischemic changes. Since the density of lesions is subtle and can be confounded by normal physiologic changes, quantitative estimation of acute ischemic infarct is challenging. In clinical practice [9], bilaterally symmetric (illustrated in Fig. 1) provides useful information for the identification of acute ischemic infarct.

Anatomical asymmetry has been utilized in previous works to localize and segment the abnormal regions for neuroimaging analysis. [12, 15] leverage the symmetry by adding extra information beyond the input image. [12] calculates the differences of each voxel by subtracting the original brain from the mirrored brain. The difference map is further used as the input to train a random forest classifier to yield lesion segmentation. [15] extracts both the original patch and its symmetric patch, and feeds them into the network simultaneously. Except for calculating the asymmetry on image-level, [3, 11, 16] propose to explore feature-level fusion of the two symmetry regions. For instance, two-branch networks (e.g. siamese network) can learn the features of left and right hemispheres and measure the difference between the features of two hemispheres to analyze abnormalities such as Alzheimer’s disease [11], ischemic stroke [3, 10] and brain tumors [16]. Even though the pixel-wise difference is widely used in previous methods, it can not efficiently exploit the bilaterally symmetric information due to the limitation of context modeling. In addition, all the above methods need the input images to be already calibrated which cannot be guaranteed in practice.

In this paper, a symmetry enhanced attention network (SEAN) is proposed for acute ischemic infarct segmentation. The proposed SEAN can automatically transform an input image into the standard space without any human supervision. The transformed image is further processed by a U-shape network that contains encoding and decoding stages. Different from the original design [13], the encoder performs 3d convolution to leverage context information of adjacent images in axial. Then, a symmetry enhanced attention module is integrated

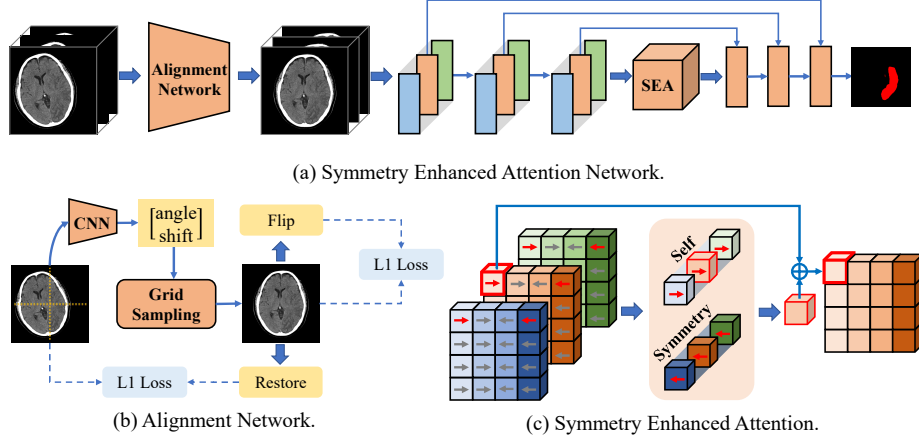


Fig. 2. Overview of our proposed network architecture.

between the encoding and decoding stages to efficiently model the anatomical symmetry. In summary, the main contributions of our paper are as follows.

1. A symmetry enhanced attention is proposed to capture both in-axial and cross-axial symmetry information by explicitly estimating the long-range dependencies.
2. A symmetry-based alignment network is proposed to transform an input image as bilaterally symmetric in axial without any human supervision.
3. We release the dataset at <https://github.com/GriffinLiang/AISD>.

2 Method

In this section, we first introduce the symmetry based alignment network in Sec. 2.1 and give the detail information of how to make the input image bilaterally symmetric in axial. Then we define the structure of the proposed symmetry enhanced attention network which can capture both the in-planar and across-planar symmetry information for ischemic infarct segmentation in Sec. 2.2. The whole pipeline is shown in Fig. 2.

2.1 Symmetry based Alignment

Since the poses of patients are arbitrary when they perform CT scans, the brain images are usually not in standard space. In order to effectively use the symmetry of the brain, we attempt to align the image to keep the region of brain in the center of the image and horizontally symmetrical. However, traditional registration based method can not be applied in clinical practice due to the high time complexity. Therefore, we proposed a Symmetry based Alignment Network

as show in Fig. 2 which can automatically align the brain images with only the information of images itself. Inspired by Spatial Transformer Networks [7], we design the symmetry based alignment network as: c2d[32,7,7]-relu-max2d[2,2]-c2d[32,5,5]-relu-max2d[2,2]-fc[3] where c2d[n, c_w, c_h] denotes a 2d convolutional layer with n filters of size $c_w \times c_h$, max2d[s_h, s_w] is a 2d max-pooling layer with the kernel size and the stride as $s_h \times s_w$, fc[n] is a fully connected layer with n units. The output of the network is interpreted as the parameters α (rotation, horizontal shift and vertical shift) of rigid transformation matrix.

Given an input volume, we define \mathbf{A}_i as the i -th slice in the axial view. During training, the output parameters α is applied to the input slice $\mathbf{A}_i^t = f_\alpha(\mathbf{A}_i)$ using parameterised sampling grid. Then we generate the flipped version of \mathbf{A}_i^t as $\tilde{\mathbf{A}}_i^t$. The total loss is designed as the following:

$$L_a = L_1(\tilde{\mathbf{A}}_i^t, \mathbf{A}_i^t) + L_1(f_\alpha^{-1}(\mathbf{A}_i^t), \mathbf{A}_i), \quad (1)$$

where $f_\alpha^{-1}(\cdot)$ denotes the inverse transformation function and L_1 denotes the L1 loss. We define the first term of Eqn. (1) as symmetry loss which is the L1 distance between the aligned image and its horizontally flipped image. The symmetry loss is based on the assumption that the difference between the image of brain and its horizontally flipped version will be minimized when the brain is perfectly aligned. In addition, we need to add constraints to the symmetry loss to avoid trivial solution where the alignment network can simply transform the brain region out of the input image. Therefore, a restoration loss is defined as the L1 distance between the restored image and the original image to learn useful transformation parameters.

2.2 Symmetry based Segmentation

The proposed segmentation network adopts the structure of UNet [13] which is mainly composed of two parts: the encoder stage and the decoder stage. Inspired by [5], we use 3D convolutions as the basic encoding block to keep the context information from adjacent images in axial view. For the decoding stage, the middle plane of input volume is retained as the target image and upsampled to the original resolution for pixel-wise labelling. We name the above network as HybridUnet. Finally, we cascade the last encoding block with the symmetry attention module. In this way, the feature representation can be enhanced by its symmetry information to efficiently assess the presence and extent of ischemic infarct.

To exploit the context information of the i -th axial image, HybridUnet takes its adjacent images $\{\mathbf{A}_{i+t} | t = -T, \dots, T\}$ as the input. The input images are firstly processed by the 3d encoder. We design the encoder block as: c3d-bn-relu-c3d-bn-relu-max3d where c3d denotes 3d convolutional layer, bn denotes 3d batchnorm layer and max denotes a 3d max-pooling layer. The encoder stage contains five encoder blocks. The output feature from the last encoder block is denoted as $\mathbf{X}_i \in \mathbb{R}^{C \times H \times W}$ for the input image \mathbf{A}_i where H and W represent the height and width of the output feature respectively and C is the number of

the output channels. The output feature \mathbf{X}_i is further processed by symmetry attention module in Sec. 2.2 and further processed by four decoding blocks with the same structure as the original UNet. We train the proposed SEAN by minimizing the combination of the generalized Dice loss and the cross-entropy loss in an end-to-end manner.

Symmetry Enhanced Attention. The symmetry information is hard to explore by the conventional operation which only processes a local neighborhood in space. Since the input position and its symmetrical position usually have long spatial interval distances, local operations need to be applied repeatedly to capture such long-range dependencies. As mentioned in [6], stacking multiple convolution operations is computationally inefficient and increases the difficulty of optimization. To compensate for the drawback of convolution operation, we propose to model the relationships between symmetrical position with attention mechanism [14]. Given an input feature map $\mathbf{X}_i \in \mathbb{R}^{C \times H \times W}$, we first divide it into $P \times Q$ partitions as below,

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i,1,1} & \mathbf{X}_{i,1,2} & \cdots & \mathbf{X}_{i,1,Q} \\ \mathbf{X}_{i,2,1} & \mathbf{X}_{i,2,2} & \cdots & \mathbf{X}_{i,2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{i,P,1} & \mathbf{X}_{i,P,2} & \cdots & \mathbf{X}_{i,P,Q} \end{bmatrix}, \quad (2)$$

where $\mathbf{X}_{i,j,k} \in \mathbb{R}^{C \times H' \times W'}$ is a subset of \mathbf{X}_i ($H = H' \times P$ and $W = W' \times Q$). Then we flip \mathbf{X}_i horizontally to generate its mirrored feature map $\tilde{\mathbf{X}}_i$. Therefore, the symmetrical partition of $\mathbf{X}_{i,j,k}$ can be denoted as $\tilde{\mathbf{X}}_{i,j,k}$.

For the input partition $\mathbf{X}_{i,j,k}$, its symmetry enhanced attention is composed of the self-attention module and the symmetry-attention module:

$$\begin{aligned} \mathbf{S}_{i,j,k}^t &= \text{Softmax}\left(\frac{\theta(\mathbf{X}_{i,j,k})^\top \phi(\mathbf{X}_{i+t,j,k})}{\sqrt{d}}\right), \\ \tilde{\mathbf{S}}_{i,j,k}^t &= \text{Softmax}\left(\frac{\theta(\mathbf{X}_{i,j,k})^\top \phi(\tilde{\mathbf{X}}_{i+t,j,k})}{\sqrt{d}}\right), \\ \mathbf{Y}_{i,j,k} &= \left(\sum_{t=-T}^T \mathbf{S}_{i,j,k}^t \cdot g(\mathbf{X}_{i+t,j,k})^\top + \sum_{t=-T}^T \tilde{\mathbf{S}}_{i,j,k}^t \cdot h(\tilde{\mathbf{X}}_{i+t,j,k})^\top \right)^\top, \end{aligned} \quad (3)$$

where $\mathbf{S}_{i,j,k}^t, \tilde{\mathbf{S}}_{i,j,k}^t \in \mathbb{R}^{N' \times N'}$ ($N' = H' \times W'$) are the similarity matrix of the self attention and symmetry attention respectively. $\theta(\cdot)$ and $\phi(\cdot)$ perform convolution operations to reduce the number of input channels to d (e.g. $d = \frac{C}{2}$) and reshape the output to $\mathbb{R}^{d \times H' \times W'}$. We use \sqrt{d} as a scaling factor for the inner product to solve the small gradient problem of softmax function. $\mathbf{X}_{i+t,j,k}$ and $\tilde{\mathbf{X}}_{i+t,j,k}$ are also fed into $g(\cdot)$ and $h(\cdot)$ to compute the new representation by convolution operations and reshape the output feature map to $\mathbb{R}^{\frac{C}{2} \times H' \times W'}$. The symmetry

enhance feature $\mathbf{Y}_{i,j,k} \in \mathbb{R}^{C \times H' \times W'}$ is reshaped to $C \times H' \times W'$ and further considered as the residual mapping of $\mathbf{X}_{i,j,k}$ to acquire the final output of the symmetry enhanced attention.

3 Experiments

In this section, we first introduce the data acquisition and evaluation indicators of our model in Sec. 3.1. Then we show the detail information of implementation in Sec. 3.2. Finally, we compare our approach with the state-of-the-art methods and conduct extensive ablation studies in Sec. 3.3.

3.1 Experiment setup

Data Acquisition. We obtain 397 Non-Contrast-enhanced CT (NCCT) scans of acute ischemic stroke with the interval from symptom onset to CT less than 24 hours. The patients underwent diffusion-weighted MRI (DWI) within 24 hours after taking the CT. The slice thickness of NCCT is 5mm. We name the above CT-MRI pairs as acute ischemic stroke dataset (AISD). 345 scans are used to train and validate the model, and the remaining 52 scans are used for testing. Ischemic lesions are manually contoured on NCCT by a doctor using MRI scans as the reference standard. Then a senior doctor double-reviews the labels.

Evaluation Metrics. To quantitatively evaluate the result of our proposed symmetry based alignment network, we compare the output transformation parameters with the human annotated rotation and offset. Specifically, a doctor annotates the beginning and end point of cerebral falx on the middle slice of each CT scan. The offset angle and the center of brain region are further calculated as ground truth. We use the average difference of the rotation angle and the offset distance in the horizontal direction between the model’s output and the ground truth for each data. As for the segmentation results, we utilize Dice coefficient to quantitatively evaluate the performance. In addition, we also calculate infarct-level evaluation metrics such as recall, precision and F1 score. To evaluate the clinical value, Pearson correlation between the estimated ASPECTS and the ground-truth is also performed.

Comparison Methods. We compare the proposed SEAN with following methods: 1) No symmetry modelling method (Unet) which is a Vanilla Unet without considering the symmetry information; 2) Symmetry modelling on image-level (Unet-IM-L1) [12] which calculates the bilateral density L_1 difference between symmetric brain regions as one of the input; 3) Symmetry modelling on feature-level by concatenation (Unet-FT-CC) which takes the original image and its flipped image as the input of Unet and concatenates the two output features from the last encoder; 4) Symmetry on feature-level with L1 distance (Unet-FT-L1) [16] which takes the distance map between the feature from the last encoder

Table 1. Quantitative results on the AIS dataset.

Network	Fusion-Level	Fusion-Type	Dice	F1	Recall	Precision
Unet	N/A	N/A	0.4588	0.5105	0.5019	0.5196
Unet-IM-L1	Image	L1	0.5035	0.5457	0.5318	0.5603
Unet-FT-L1	Feature	L1	0.5121	0.567	0.5468	0.5888
Unet-FT-CC	Feature	Concat	0.5354	0.572	0.5655	0.5786
HybridUnet	N/A	N/A	0.4952	0.5433	0.6105	0.4895
HybridUnet-IM-L1	Image	L1	0.5437	0.5992	0.5581	0.6471
HybridUnet-FT-L1	Feature	L1	0.5445	0.5982	0.5543	0.6497
HybridUnet-FT-CC	Feature	Concat	0.5577	0.6015	0.5431	0.6742
SEAN	Feature	Attention	0.5784	0.6218	0.5880	0.6597

and its flipped version as an extra information. We also implement the above methods using HybridUnet [5] as the backbone for further comparison.

3.2 Implementation Details

Our implementation is based on Pytorch framework. For data pre-processing, we truncate the raw intensity values to the range [40, 100] HU and normalize each raw CT case to have zero mean and unit variance. The Adam optimizer is used to train the model with parameters $\beta_1 = 0.9, \beta_2 = 0.99$ for 150 epochs. And we set the base learning rate as 1×10^{-4} and deploy a poly learning rate policy where the initial learning rate is multiplied by $(1 - \frac{iter}{total_iter})^{power}$ and $power = 0.9$ after each iteration.

3.3 Results and Discussions

Efficacy of Alignment Network. We evaluate the similarity of the rotation angle and the offset distance in the horizontal direction between the model’s output and the ground truth on test data (52 patients). The error of rotation angle is 3 degrees and the error of shift distance is 5 pixels. From the results, we can see that the output of the symmetry based alignment network is very close to the ground truth even though the proposed method is fully unsupervised. Besides, we register the NCCT images to the standard brain template to align the images as described in [12]. This operation cost 134 seconds per patient on average, while the time consumption of our proposed symmetry based alignment network is only 0.46 seconds per patient on average. The registration method is based on Advanced Normalization Tools(ANTS) [1], and all experiments are performed on a Linux server with Intel(R) Xeon(R) CPU E5-2697 v4 @ 2.30GHz and a NVIDIA 2080ti GPU.

Efficacy of SEAN. As shown in Tab. 1, the proposed SEAN achieves the highest performance (Dice: 0.5784, F1: 0.6218) compared to the other methods. Since the proposed SEAN benefits from leveraging both the in-planar and

Table 2. Ablation study of SEAN on AISD.

Method	Dice	F1	Recall	Precision
baseline	0.4952	0.5433	0.6105	0.4895
+ Ours (Align)	0.5281	0.5767	0.5506	0.6056
+ Ours (Align+Self)	0.5635	0.5834	0.5506	0.6205
+ Ours (Align+Self+Sym)	0.5784	0.6218	0.5880	0.6597

across-planar symmetry information, it can differentiate the infarct and the normal physiologic change more efficiently. In general, symmetry based methods observed significant improvements according to both Dice and F1. This is also consistent with doctors’ habit in clinical practice. For the effectiveness of the backbone network, HybridUnet achieves better performance than the original Unet, which demonstrates the importance of context information from adjacent images. Feature-level method outperforms image-level method, since the feature-level method is robust to the misalignment of the input image. We conduct ablation studies of SEAN and show the results in Tab. 2. The influence of the proposed alignment network and the two type of attention mechanism: self-attention and symmetry-attention is investigated. According to the results, it can be seen that the proposed alignment network improves over the baseline for a large margin. We can also see that symmetry enhanced attention yields a higher increase in both Dice and F1 comparing to only using self-attention. We also show some qualitative examples in Fig. 3.

Efficacy of Clinical Usage. To validate the clinical efficacy of the proposed SEAN, Pearson correlation between the estimated ASPECTS and the ground-truth is performed. A standard template with ASPECTS regions in the Montreal Neurologic Institute space [4] is registered to all NCCT images by performing affine transformation using ANTS. The Pearson correlation between the scores estimated by SEAN and doctor is 0.75, which further indicates the efficiency of the proposed method.

4 Conclusion

In this paper, we propose a symmetry enhanced attention network (SEAN) for acute ischemic infarct segmentation. The proposed network calibrates an input CT image and capture bilateral symmetry information by explicitly estimating the long-range dependencies. With the seamless integration, the proposed symmetry enhanced attention can be applied to any lesion segmentation task. Experimental results on acute ischemic stroke dataset (AISD) show that the proposed SEAN outperforms some symmetry-based state-of-the-art methods in terms of both dice coefficient and infarct localization. The acute ischemic stroke dataset (AISD) is published for future study.

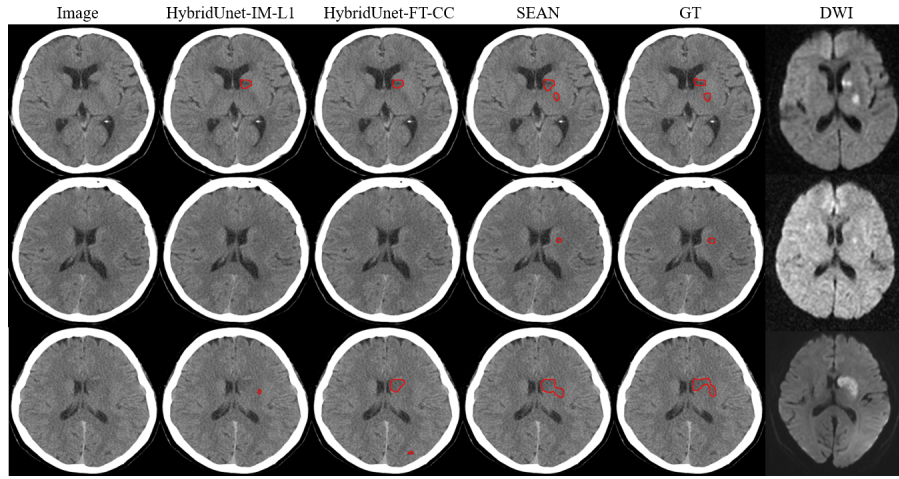


Fig. 3. Qualitative comparison of different methods.

Acknowledgments This work was supported in part by following grants, MOST-2018AAA0102004, NSFC-62061136001, the Key Program of Beijing Municipal Natural Science Foundation(7191003), and the Key Projects of the National Natural Science Foundation of China (81830057).

References

1. Avants, B.B., Tustison, N., Song, G.: Advanced normalization tools (ants). *Insight j* **2**(365), 1–35 (2009)
2. Barber, P.A., Demchuk, A.M., Zhang, J., Buchan, A.M., Group, A.S., et al.: Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *The Lancet* **355**(9216), 1670–1674 (2000)
3. Barman, A., Inam, M.E., Lee, S., Savitz, S.I., Sheth, S.A., Giancardo, L.: Determining ischemic stroke from ct-angiography imaging using symmetry-sensitive convolutional networks. In: *International Symposium on Biomedical Imaging*. pp. 1873–1877 (2019)
4. Evans, A.C., Collins, D.L., Mills, S., Brown, E.D., Kelly, R.L., Peters, T.M.: 3d statistical neuroanatomical models from 305 mri volumes. In: *1993 IEEE conference record nuclear science symposium and medical imaging conference*. pp. 1813–1817. IEEE (1993)
5. Fang, C., Li, G., Pan, C., Li, Y., Yu, Y.: Globally guided progressive fusion network for 3d pancreas segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. pp. 210–218. Springer International Publishing, Cham (2019)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Computer Vision and Pattern Recognition* (2015)

7. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: *Advances in neural information processing systems*. pp. 2017–2025 (2015)
8. Katan, M., Luft, A.: Global burden of stroke. In: *Seminars in neurology*. vol. 38, pp. 208–211. Georg Thieme Verlag (2018)
9. Khan Academy, A.A.o.C.o.N.: Diagnosing strokes with imaging ct, mri, and angiography. <https://www.khanacademy.org>
10. Kuang, H., Menon, B.K., Qiu, W.: Automated infarct segmentation from follow-up non-contrast ct scans in patients with acute ischemic stroke using dense multi-path contextual generative adversarial network. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. pp. 856–863. Springer International Publishing, Cham (2019)
11. Liu, C.F., Padhy, S., Ramachandran, S., Wang, V.X., Efimov, A., Bernal, A., Shi, L., Vaillant, M., Ratnanather, J.T., Faria, A.V., Caffo, B., Albert, M., Miller, M.I.: Using deep siamese neural networks for detection of brain asymmetries associated with alzheimer’s disease and mild cognitive impairment. *Magnetic Resonance Imaging* **64**, 190 – 199 (2019)
12. Qiu, W., Kuang, H., Teleg, E., Ospel, J.M., Sohn, I., Almekhlafi, M., Goyal, M., Hill, M.D., Demchuk, A.M., Menon, B.K.: Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced ct. *Radiology* **294**(3), 638–644 (2020)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. pp. 234–241. Springer International Publishing, Cham (2015)
14. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7794–7803 (2018)
15. Wang, Y., Katsaggelos, A.K., Xue, W., Parrish, T.B.: A deep symmetry convnet for stroke lesion segmentation. In: *IEEE International Conference on Image Processing (ICIP)* (2016)
16. Zhang, H., Zhu, X., Willke, T.L.: Segmenting brain tumors with symmetry. In: *Proceedings of NIPS Workshop* (2017)

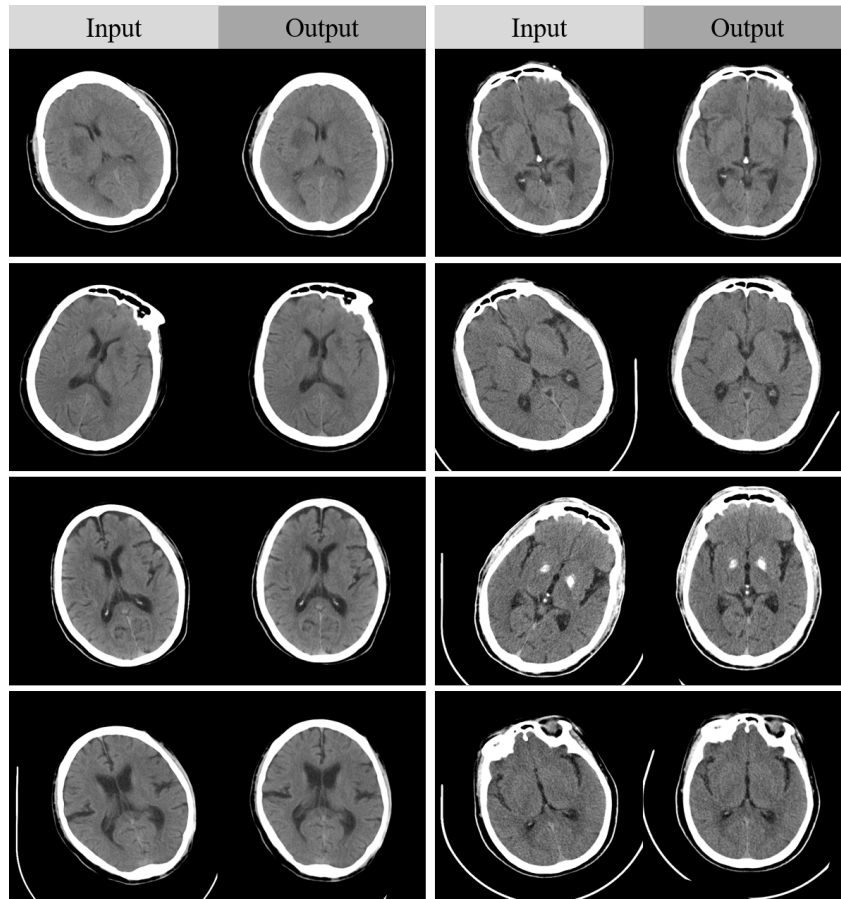


Fig. 1. Qualitative results of the proposed symmetry alignment network.

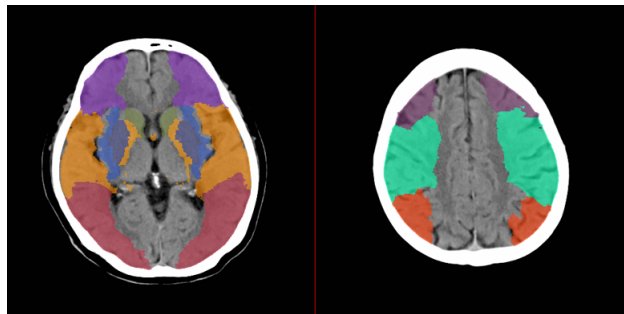


Fig. 2. ASPECTS regions are consist of 10 different parts which are symmetrical with respect to the cerebral falx.