



# Regression-Based Line Detection Network for Delineation of Largely Deformed Brain Midline

Hao Wei<sup>1,2</sup>, Xiangyu Tang<sup>3</sup>, Mingqing Zhang<sup>1,4</sup>, Qingfeng Li<sup>1,5</sup>, Xiaodan Xing<sup>1,6</sup>, Xiang Sean Zhou<sup>1</sup>, Zhong Xue<sup>1</sup>, Wenzhen Zhu<sup>3</sup>, Zailiang Chen<sup>2</sup>, and Feng Shi<sup>1</sup>✉

<sup>1</sup> Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China  
feng.shi@united-imaging.com

<sup>2</sup> School of Computer Science and Engineering, Central South University, Changsha, Hunan, China

<sup>3</sup> Department of Radiology, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China

<sup>4</sup> College of Software Engineering, Southeast University, Jiangsu, China

<sup>5</sup> School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, China

<sup>6</sup> Medical Imaging Center, Shanghai Advanced Research Institute, Shanghai, China

**Abstract.** Brain midline shift is often caused by various clinical conditions such as high intracranial pressure, which can be deadly. To facilitate clinical evaluation, automated methods have been proposed to classify whether midline shift is severe or not, e.g., larger than 5 mm away from the ideal midline. There are only limited methods using landmark or symmetry, attempting to provide more intuitive results such as midline delineation. However, landmark- or symmetry-based methods could be easily affected by anatomical variability and large brain deformations. In this study, we formulated the midline delineation as a skeleton extraction task and proposed a novel regression-based line detection network (RLDN) for the robust midline delineation especially in largely deformed brains. Basically, the proposed method includes three parts: (1) multi-scale line detection, (2) weighted line integration, and (3) regression-based refinement. The first two parts were used to capture high-level semantic and low-level detailed information to extract deformed midline, while the last part was utilized to regress more accurate midline positions. We validated the RLDN on 100 training and 28 testing subjects with a mean midline shift of 7 mm and the maximum shift of 16 mm (induced by hemorrhage). Experimental results show that our proposed method achieves state-of-the-art accuracy with a mean line difference of  $1.17 \pm 0.72$  mm and F1-score of 0.78 from manual delineations. Our proposed robust midline delineation method is also beneficial for other cases such as midline deformation from tumor, traumatic brain injury, and abscess.

## 1 Introduction

The anatomical structure of human brain consists of two cerebral hemispheres that are roughly symmetric and separated by the ideal midline (IML), which is a straight line connecting the most anterior and posterior visible points on the falx as shown in Fig. 1. However, high or unbalanced intracranial pressure (ICP) could distort the IML into deformed midline (DML), as also presented in Fig. 1. Such midline shift (MLS) can be deadly and is usually caused by conditions such as traumatic brain injury (TBI), stroke or hematoma. In fact, the degree of MLS serves as a quantitative indication for neurosurgeons to monitor and control ICP.

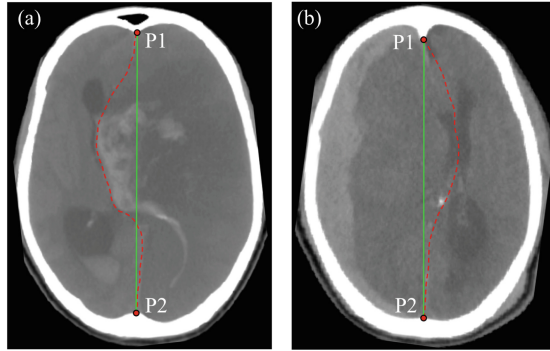
Clinically, the visual inspection of MLS on head CT images has become a widely applied practice. However, comprehensive and quantitative MLS evaluation is still challenging and time-consuming for many health care providers (i.e. emergency physicians). Thus, the computer-aided methods are of research value and practical significance given they could improve the accuracy and efficiency of this evaluation.

In the literature, limited works [1, 5, 6] have tried to facilitate this clinical evaluation. Liao et al. [5] considered the midline as three isolated segments and then utilized mathematical curves to fit these segments using local intensity symmetry. Liu et al. [6] employed a landmark detection method to locate some relatively-stable key points (i.e. falx, septum pellucidum) to build DML. Using an enhanced voigt model to simulate this shift process, Chen et al. [1] proposed a method to transform IML to DML. However, most aforementioned methods lack of robust in the CT images with largely deformed brain due to the following two aspects: (a) The midline is relatively difficult to be distinguished in CT images given the low soft tissue contrast, especially when affected by TBI or stroke; (b) The shape of DML varies a lot depending on the brain deformations.

To address such issues, we formulated the midline delineation as a skeleton extraction task and proposed a novel regression-based line detection network (RLDN) for the robust delineation especially in largely deformed brains. Our method enables end-to-end training and includes three parts: (1) multi-scale line detection; (2) weighted line integration; and (3) regression-based refinement. Specifically, we proposed a novel multi-scale feature integration strategy over whole scales to better capture high-level semantic and low-level detailed information to generate midline probability maps. Then the weighted line integration module fused these maps together to generate a final map. Finally, we proposed a regression-based refinement module to refine and thin the final map to generate the coordinates of midline. Experimental results show that our proposed method can improve the line detection accuracy and achieves the state-of-the-art performance. Moreover, to our knowledge, this is the first work employing fully-convolutional approach for midline delineation on CT images.

## 2 Method

As mentioned above, the proposed RLDN includes three parts: (1) multi-scale line detection; (2) weighted line integration; and (3) regression-based refinement,



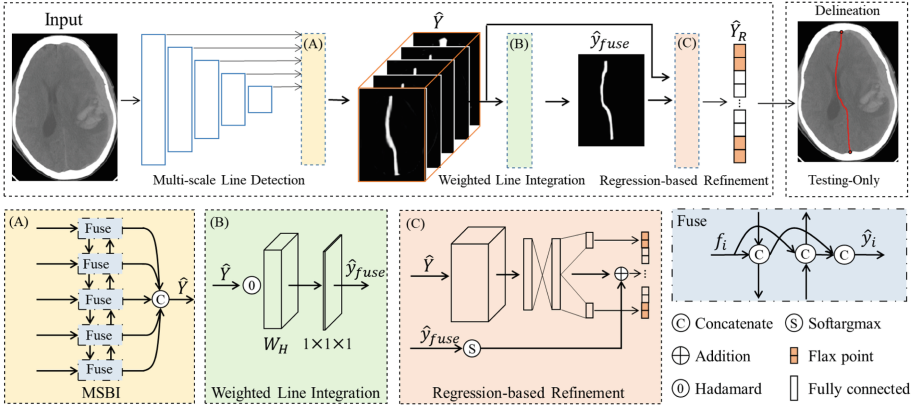
**Fig. 1.** Examples of ideal midlines (IML, green line) and deformed midlines (DML, red dotted line) on head CT images, represented as green lines and red dotted lines, respectively. The P1 and P2 represent the anterior and posterior flax points. (Color figure online)

with the architecture shown in Fig. 2. First, the multi-scale line detection network (LDN) takes a 2D CT image  $I \in R^{H \times W}$  as input, and outputs the probability map  $\hat{Y} \in R^{H \times W \times S}$  corresponding to  $S = 5$  scale levels. The loss  $\mathcal{L}_{\hat{y}_i}$  for each output  $\hat{y}_i \in \hat{Y}$  is computed with the ground truth  $Y \in R^{H \times W}$ . Second,  $\hat{Y}$  is merged by the weighted line integration module to yield better line detection result  $\hat{y}_{fuse} \in R^{H \times W}$  with the loss  $\mathcal{L}_{fuse}$  to  $\hat{Y}$ . Finally,  $\hat{Y}$  and  $\hat{y}_{fuse}$  are utilized to train the regression-based refinement module with loss  $\mathcal{L}_{regress}$  to regress  $\hat{Y}_R \in R^{(H+4) \times 1}$ , which represents the coordinates of midline and its two flax points. The following subsections elaborate each module of our method.

## 2.1 Multi-scale Line Detection Network

The proposed LDN includes the down-sampling branch and the multi-scale bidirectional integration module (MSBI). Specifically, the down-sampling branch adopts convolutional parts of VGG [8] except following different settings: (1) number of channels halved, and (2) the addition of pyramid pooling module [12] connected with the last convolution layer in the initial three conv-blocks. Moreover, it follows the side-output manner in HED [9] to output probability maps  $\hat{Y}$  including features of five different scales. Then, the MSBI module is utilized to integrate these features.

As we all known, low-level features focus more on local detailed structures, while high-level features are rich in conceptual semantic information [13]. In our line detection task, *not only* the high-level semantic information *but also* the high-resolution details are required. Thus, we propose the MSBI module inspired by [10] to integrate all scale features to overcome above *semantic vs resolution* conflict. Specifically, this module takes feature maps  $f_i$  in scale  $i$  as input and produces the same resolution feature map  $\hat{y}_i \in \hat{Y}$  as output. The MSBI consists of two directional pathways (*shallow-to-deep* and *deep-to-shallow* pathways) so



**Fig. 2.** Overview of the architecture of the proposed RLDN, which consists of three parts: (1) multi-scale line detection, where (A) denotes the multi-scale bidirectional integration module; (2) weighted line integration in (B); and (3) regression-based refinement in (C).

that features of each scale can be obtained from both semantic and detailed information.

The *shallow-to-deep* pathway starts from the first conv-block and ends with the last one. Specifically, the current scale feature  $f_i$  updates itself by encoding the learned representation  $h_{f_{i-1}}^{SD}$  of its previous scale:

$$h_{f_i}^{SD} = \begin{cases} \mathcal{F} \left( f_i, R_i \left( h_{f_{i-1}}^{SD} \right) \right), & i \geq 2 \\ f_i, & i = 1, \end{cases} \quad (1)$$

where  $h_{f_i}^{SD}$  represents the updated feature map in current scale, and  $R(\cdot)$  denotes the bilinear interpolation operation due to the mismatched size over different scale.  $\mathcal{F}(\cdot)$  represents the concatenate operation followed by  $1 \times 1$  convolution to keep the number of channels unchanged. Then, we use additional convolution and ReLU to encode  $h_{f_i}^{SD}$  to form the learned feature map  $h_{f_i}^{SD}$ .

On the other hand, the *deep-to-shallow* pathway gathers the multi-scale features from high-level to low-level. This process can be formulated as:

$$h_{f_i}^{DS} = \begin{cases} \mathcal{F} \left( f_i, R_i \left( h_{f_{i+1}}^{DS} \right) \right), & i \leq (S - 1) \\ f_i, & i = S, \end{cases} \quad (2)$$

where  $h_{f_i}^{DS}$  denotes the updated feature map. Other operations are the same as these in *shallow-to-deep* pathway.

Then, we merge the learned representations ( $h_{f_i}^{SD}$  and  $h_{f_i}^{DS}$ ) of each scale to form the fused feature representation  $M_i$ :

$$M_i = \sigma \left( \mathcal{F} \left( h_{f_i}^{SD}, h_{f_i}^{DS} \right) \right), \quad (3)$$

where  $\sigma$  denotes the non-linear activation function ReLU, and the meaning of other symbols are consistent with these in any directional pathway. Then the  $1 \times 1$  convolution and interpolation are adopted to generate the side-output prediction  $\hat{y}_i$  in the current scale with one channel.

For the each side-output  $\hat{y}_i$ , the prediction error is computed by using the weighted cross-entropy (WCE) [9] loss function:

$$\mathcal{L}_{\hat{y}_i} = -\gamma \sum_{j \in Y_+} Y_j \log \hat{y}_{ij} - (1 - \gamma) \sum_{j \in Y_-} (1 - Y_j) \log(1 - \hat{y}_{ij}), \quad (4)$$

where  $Y_+$  and  $Y_-$  denote the object and background ground-truth label sets, respectively.  $\gamma = |Y_-|/|Y|$  represents the class weight used to balance object and background.  $Y_j$  and  $\hat{y}_{ij}$  denote the label and the prediction value at pixel  $j = 1, \dots, |Y|$  for specific scale  $i$ .

## 2.2 Weighted Line Integration

The weighted line integration module receives  $\hat{Y}$  produced by LDN and generates the fused probability map  $\hat{y}_{fuse}$ . Specifically, we first construct the learnable weight  $W_H \in R^{H \times W \times S}$  and use it to do Hadamard product with the input to generate  $S$  channel maps. Then, the  $1 \times 1$  convolution is adopt to produce the fused prediction  $\hat{y}_{fuse}$ . Finally, the following loss is introduced to compute the error with the ground truth  $Y$ :

$$\mathcal{L}_{fuse} = WCE(\hat{y}_{fuse}, Y). \quad (5)$$

## 2.3 Regression-Based Refinement

The regression-based refinement module takes  $\hat{Y}$  and  $\hat{y}_{fuse}$  as inputs to regress the coordinates of midline and its two flax points. It should be noted that, since the direct line regression focuses more on the overall trend, instead of its endpoints, we decide to adopt an additional network layer to regress endpoints. Basically, this module consists of two branches: (1) the convolution branch and (2) the soft-argmax branch. The convolution branch includes three residual blocks [3] and four fully-connected layers. This branch takes  $\hat{Y}$  as input and produces a 1D vector  $\hat{Y}_R \in R^{(H+4) \times 1}$ , where the sub-vector  $\hat{Y}_{R-midline} \in R^{H \times 1}$  in the middle of  $\hat{Y}_R$  denotes the column coordinates of midline. The remaining top-2 and bottom-2 elements represent the coordinates of two endpoints.

The soft-argmax branch consists of only soft-argmax [11] layer and takes the  $\hat{y}_{fuse}$  as input to generate a  $H \times 1$  vector, which is then utilized to update  $\hat{Y}_{R-midline}$  by element-wise addition:

$$\hat{Y}_{R-midline}(i) = \hat{Y}_{R-midline}(i) + \sum_j^W \frac{\exp(\mu \cdot \hat{y}_{fuse}(i, j))}{\sum_m^W \exp(\mu \cdot \hat{y}_{fuse}(i, m))} j, \quad (6)$$

where  $\hat{y}_{fuse}(i, j)$  denotes the prediction value of location  $(i, j)$ , and  $\mu = 10$  is a hyper-parameter controlling the smoothness of the soft-argmax. Finally, the Mean Squared Error (MSE) loss function is employed to compute the error:

$$\mathcal{L}_{regress} = MSE(\hat{Y}_R, Y_R), \quad (7)$$

where  $Y_R \in R^{(H+4) \times 1}$  denotes the ground truth of the regression task.

## 2.4 Cost Function and Optimization

We simultaneously optimize all the side-outputs  $\hat{y}_i$ , the fused prediction  $\hat{y}_{fuse}$ , and the regression output  $\hat{Y}_R$  in an end-to-end way. The loss function of the whole framework is given as:

$$\mathcal{L}_{all} = \gamma \mathcal{L}_{fuse} + \xi \mathcal{L}_{regress} + \sum_i^S \lambda_i \mathcal{L}_{\hat{y}_i}, \quad (8)$$

where  $\gamma, \xi$  and  $\lambda_i$  represent the balance weights.

In the inferring phase, an input CT image  $I$  simply forwards through the aforementioned steps to generate the coordinates of midline and its two flax points. Then, we draw it into the zero-value map to get the final full-resolution midline map, as shown in Fig. 2.

## 3 Experiments

### 3.1 Data

Our dataset was derived from public CQ500 [2]. We selected all 64 midline shift cases and the same number of health subjects in this study. Specifically, a total of 5 CT slices with the largest brain area in each subject were selected to be manually delineated for the midline (total 128 subjects with 640 slices). From the subject-level, we randomly selected 100 subjects as training set and the rest 28 subjects as testing set. For pre-processing, each CT slice was resampled to uniform resolution ( $0.5 \times 0.5 \text{ mm}^2$ ), applied intensity normalization from  $(-100, 200)$  to  $(0, 1)$ , and then cropped into a patch with the size of  $400 \times 288$  to contain only the brain region utilizing a simple thresholding segmentation algorithm. Finally, we augmented training set by randomly rotating, left-right flipping, and brightness changing.

To evaluate the line detection task, we introduced two standard measures [9]: F1-score (the harmonic mean of precision and recall) when choosing threshold based on optimal dataset scale (ODS) or optimal image scale (OIS). For the evaluation of the regression task, we defined the following distance-related metrics: line distance error (LDE), max shift distance error (MSDE), anterior flax point distance error (AFDE), and posterior flax point distance error (PFDE).

### 3.2 Experimental Setting

The proposed method was implemented based on the publicly available platform Pytorch. During the training phases, the stochastic gradient descent (SGD) algorithm was used to optimize the whole network. The network weights were initialized by the Xavier algorithm and the weight decay was set to be  $1e-4$ . In Eq. (8), we set  $\lambda_i = \gamma = 1$  and  $\xi = 2$ . The remaining hyper-parameters and corresponding values were: mini-batch size (32), base learning rate ( $1e-4$ ), momentum (0.9), and maximal iteration (400). We decreased the learning rate every 200 iterations with factor 0.1. The experiments were implemented on a NVIDIA Titan Xp with 12 GB memory.

**Table 1.** The ODS and OIS F1-score on the testing dataset by different methods. Higher is better.

Method	ODS	OIS
HED [9]	0.36	0.36
SRN [4]	0.37	0.37
RCF [7]	0.43	0.44
HiFi [13]	0.40	0.40
MSB-FCN [10]	0.38	0.38
LDN (our)	0.72	0.74
<b>RLDN (our)</b>	<b>0.78</b>	<b>0.79</b>

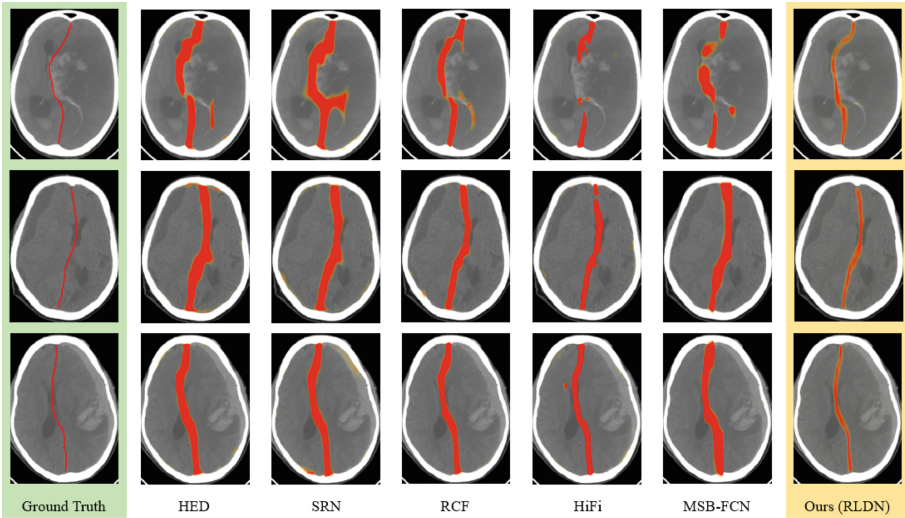
**Table 2.** The mean and standard deviation of distance error (mm) on the testing dataset. Lower is better.

Metric	VGG-16 [8]	<b>RLDN (our)</b>
LDE	5.60 (1.13)	<b>1.17 (0.72)</b>
MSDE	8.31 (3.34)	<b>2.27 (2.70)</b>
AFDE	4.42 (2.84)	<b>2.01 (1.58)</b>
PFDE	7.85 (4.46)	<b>2.47 (1.65)</b>

### 3.3 Results

Our proposed method conducts two tasks: (1) line detection task and (2) regression task. For the line detection task, we compared the result  $\hat{y}_{fuse}$  in RLDN with other five leading CNN-based methods designed for the similar skeleton extraction tasks in natural images including HED [9], SRN [4], RCF [7], HiFi [13] and MSB-FCN [10]. For the regression task, the VGG-16 [8] was used as the baseline method in our experiments.

**Visual Inspection:** Visual inspection for line detection probability maps are shown in Fig. 3. Obviously, our method achieved thinner and more accurate line detection results than all of the other methods for the severely-deformed cases (especially for the cases in the first row). This shows the superiority of our proposed method.



**Fig. 3.** Qualitative comparisons of our proposed method and the other five state-of-the-art CNN-based methods on some challenging cases. As shown, some lines by other methods are broken or quite thick, which indicate unsatisfactory results.

**Quantitative Comparison:** The performance of line detection task is shown in Table 1, where our RLDN achieved the best performance in terms of ODS or OIS. Specifically, RLDN improved the current best ODS to 0.78, mainly owing to the obtained precise line detection result, as shown in Fig. 3. Compared with LDN and RLDN in Table 1, the effectiveness of the regression task is further verified by the improved performance, e.g., 6%, and 5% F1-score performance for the line detection task. In the same way, the benefit of the line detection task to regression is also verified from the Table 2. In summary, our proposed RLDN obtained the state-of-the-art performance in the task of DML delineation.

## 4 Conclusion

In this paper, we have proposed a novel regression-based line detection network (RLDN) for delineation and measurement of largely deformed brain midline. Specifically, the algorithm was based on multi-scale line detection and weighted line integration to capture high-level semantic and low-level detailed information for extracting midline. Finally, the midline was obtained by using regression-based refinement. Comparative results demonstrated that our proposed method achieves a clear performance boost in terms of both accuracy and robustness.

**Acknowledgements.** This work was partially supported by the National Key Research and Development Program of China (2018YFC0116400).



## References

1. Chen, M., et al.: Automatic estimation of midline shift in patients with cerebral glioma based on enhanced voigt model and local symmetry. *Australas. Phys. Eng. Sci. Med.* **38**(4), 627–641 (2015)
2. Chilamkurthy, S., et al.: Development and validation of deep learning algorithms for detection of critical findings in head CT scans. arXiv preprint [arXiv:1803.05854](https://arxiv.org/abs/1803.05854) (2018)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
4. Ke, W., Chen, J., Jiao, J., Zhao, G., Ye, Q.: SRN: side-output residual network for object symmetry detection in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1068–1076 (2017)
5. Liao, C.C., Xiao, F., Wong, J.M., Chiang, I.J.: Automatic recognition of midline shift on brain CT images. *Comput. Biol. Med.* **40**(3), 331–339 (2010)
6. Liu, R., et al.: Automatic detection and quantification of brain midline shift using anatomical marker model. *Comput. Med. Imaging Graph.* **38**(1), 1–14 (2014)
7. Liu, Y., Cheng, M.M., Hu, X., Wang, K., Bai, X.: Richer convolutional features for edge detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3000–3009 (2017)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
9. Xie, S., Tu, Z.: Holistically-nested edge detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1395–1403 (2015)
10. Yang, F., Li, X., Cheng, H., Guo, Y., Chen, L., Li, J.: Multi-scale bidirectional FCN for object skeleton extraction. In: *Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 7461–7468 (2018)
11. Yi, K.M., Trulls, E., Lepetit, V., Fua, P.: LIFT: learned invariant feature transform. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9910, pp. 467–483. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46466-4\\_28](https://doi.org/10.1007/978-3-319-46466-4_28)
12. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890 (2017)
13. Zhao, K., Shen, W., Gao, S., Li, D., Cheng, M.M.: HI-FI: Hierarchical feature integration for skeleton detection. arXiv preprint [arXiv:1801.01849](https://arxiv.org/abs/1801.01849) (2018)