Incorporating Task-Specific Structural Knowledge into CNNs for Brain Midline Shift Detection

Maxim Pisov[1,2], Mikhail Goncharov[2,3], Nadezhda Kurochkina[1],
Sergey Morozov[4], Victor Gombolevskiy[4], Valeria Chernina[4],
Anton Vladzymyrskyy[4], Ksenia Zamyatina[4], Anna Chesnokova[4], Igor Pronin[5],
Michael Shifrin[5], and Mikhail Belyaev[1,2(✉)]

[1] Skolkovo Institute of Science and Technology, Moscow, Russia
m.belyaev@skoltech.ru
[2] Kharkevich Institute for Information Transmission Problems, Moscow, Russia
[3] Moscow Institute of Physics and Technology, Moscow, Russia
[4] Center for Diagnostics and Telemedicine, Moscow, Russia
[5] Burdenko Neurosurgery Institute, Moscow, Russia

Abstract. Midline shift (MLS) is a well-established factor used for outcome prediction in traumatic brain injury, stroke and brain tumors. The importance of automatic estimation of MLS was recently highlighted by ACR Data Science Institute. In this paper we introduce a novel deep learning based approach for the problem of MLS detection, which exploits task-specific structural knowledge. We evaluate our method on a large dataset containing heterogeneous images with significant MLS and show that its mean error approaches the inter-expert variability. Finally, we show the robustness of our approach by validating it on an external dataset, acquired during routine clinical practice.

Keywords: Neural networks · Midline shift · Interpretability · Confidence

1 Introduction

The brain midline can be viewed as a line on axial and coronal projections of diverse imaging modalities (Fig. 1, left). As the human brain is approximately symmetrical, the midline is straight in healthy subjects. However, various pathological conditions, such as traumatic brain injuries (TBI), stroke and brain tumors, may break this symmetry and lead to midline shift (MLS) [8].

A major number of studies show that MLS has a prognostic value for outcome prediction of various brain pathologies: level of consciousness in patients with acute intracranial hematoma [16], median survival in patients with glioblastoma
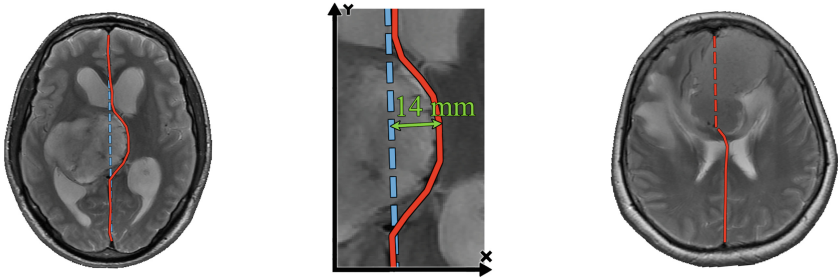

Fig. 1. Left: an axial slice from a MRI image with corresponding midline (red) and a hypothetical normal midline (blue, dashed). Center: the midline shift. Right: a dubious case with an ill-defined midline (red, dashed). (Color figure online)

multiforme [3], the outcome in patients with TBI [5]. Overall, early identification of patients with severe midline shift would assist patients management [14].

However, definitions of significant MLS vary across studies. While the 5 millimeters (mm) threshold is frequently used, other approaches are common. For example, MLS larger than 9 mm was identified in [14]; the 5 mm threshold was not justified within [5]. Such diversity is partly explained by the absence of a robust objective methodology of MLS estimation. A recent study [13] suggests that interrater variability of MLS estimation is rather high (intraclass correlation coefficients 0.72–0.89).

The importance of MLS estimation and the need for its automation was recently highlighted by The American College of Radiology Data Science Institute [10], and some promising results have already been achieved in this area (Sect. 3). In this paper we propose a novel deep learning based approach[1] for the MLS detection task. We show that combining a standard segmentation approach with task-specific structural knowledge yields results which are more accurate, compared to straightforward CNNs for regression, and also interpretable, since the key part of the method is the midline localization. Moreover, we show that our method generalizes well on highly heterogeneous data and provide a natural way of estimating its confidence.

2 Problem

We define the midline on an axial slice as a vertical curve that separates the brain hemispheres (Fig. 1, left). The midline shift for an axial slice is then defined as the maximal distance between the midline (which might be deformed) and a hypothetical normal midline (Fig. 1, center). Finally, the midline shift for a whole brain is the maximal midline shift across all axial slices where the midline is present. **The task** is to determine, for a given brain image, the midline shift as well as the corresponding axial slice on which it is manifested.

---

[1] Full code for training and inference is available at GitHub: https://github.com/neuro-ml/midline-shift-detection.

It is worth noting that in some complicated cases even professional radiologists cannot confidently determine the localization of the midline (Fig. 1, right). Taking into account such dubious cases, it is also desirable that the method for MLS detection has a means of estimating its own confidence.

3   Related Work

Most of the methods for automatic MLS estimation are computer vision (CV) based and rely on keypoints detection. The proposed approaches often have a lot of "moving parts" which makes them hard to implement and fine-tune. For example, in [9] the authors use a four-step pipeline (edge detection, morphological filtering, lines detection, rule-based filtering) just to detect the cerebral falx. Another drawback of keypoints-based methods is that they require various important regions to be present on the image, e.g. many methods can be applied only to slices that contain ventricles [1] which makes them inapplicable to cases where the midline shift is manifested on lower or higher slices.

There are also a few papers that propose deep learning methods. In [2] the authors trained an adapted a version of ResNet to classify whether there is a significant midline shift on a given slice. Another interesting approach that combines deep learning with classical CV is described in [6]. Here the authors use a U-Net [15] architecture for brain extraction, cisterns and acute intracranial lesions segmentation, while MLS detection is based on keypoints.

4   Method

A straightforward deep learning approach is to directly predict the MLS via a convolutional neural network. Following the authors of [2], we tested a ResNet-based [4] network which predicted the MLS for each axial slice of given image. The final prediction was obtained as the maximal MLS only among the slices that contained an annotated midline. However, even in such a simplified design (the model did not need to filter out the slices for which the MLS was undefined), this method yields poor results as we show in Sect. 7.

Our intuition behind this is that the midline shift is a very high-level concept: the network needs to learn to detect several keypoints located very far from each other (Fig. 1), as well as take into account their relative positions. The latter is a particularly difficult task for convolutional neural networks due to their invariance to translation.

On the contrary, the midline has visual features, like continuity and local symmetry, that are distinguishable on a smaller scale. This brings us to the idea to reduce the task of MLS prediction to the task of midline estimation: for a given slice we localize the midline while exploiting the structural knowledge about the target, then we derive the MLS from the predicted curve based on the definition given in Sect. 2. Normal midline is estimated as a straight line between prediction endpoints.
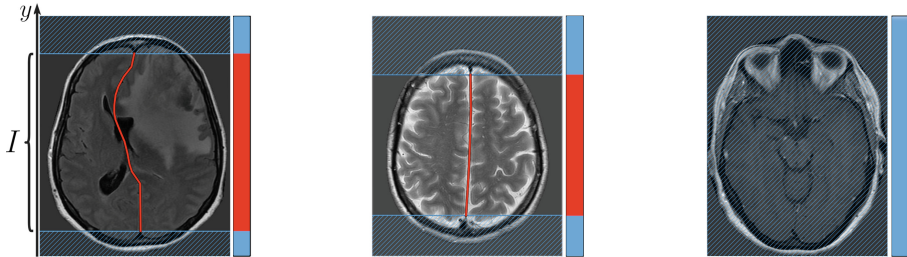
Fig. 2. The binary masks of the regions where the midline is defined (red). Note the rightmost image, for which the midline is undefined everywhere. (Color figure online)

The key structural facts are: (1) for each coordinate y there is at most one x-coordinate, which is refered as $\mathbf{midline}_y$, such that the pixel $(\mathbf{midline}_y, y)$ is situated on the midline; (2) $\mathbf{midline}_y$ exists only for y-coordinates within certain interval I on the Oy axis to which binary mask we refer as **limits** (Fig. 2).

These facts imply that our method must be capable of solving the regression problem of **mildine** estimation and the classification problem of **limits** prediction. To solve these tasks, we propose a two-headed convolutional neural network with shared input layers (Fig. 3). As loss function, we optimize a weighted combination of standard losses for regression and classification:

$$L = \lambda_1 \cdot \frac{1}{|I|} \sum_{y \in I} (\mathbf{midline}_y - \mathbf{midline}_y^{\mathrm{pred}})^2 + \lambda_2 \cdot \mathrm{BCE}(\mathbf{limits}, \mathbf{limits}^{\mathrm{pred}}),$$

where $\mathbf{midline}_y^{\mathrm{pred}}$ and $\mathbf{limits}^{\mathrm{pred}}$ are the network's predictions, BCE is binary cross-entropy.

4.1   Midline Estimation

In order to estimate the midline we adapt a segmentation approach. In a standard setting (with sigmoid activation and binary cross entropy loss) the output can be interpreted as "independent" probability of a particular pixel to be situated on the midline. In this case the **midline** is obtained after applying argmax along the Ox axis.

However, as we show in Sect. 7, significantly better results can be achieved while imposing the following constraint on the output probability map

$$\sum_x \mathbf{output}_{xy}^{\mathrm{midline}} = 1, \tag{1}$$

which follows from the structural fact (1). Next, taking into account that for any given y-coordinate the head's output represents a probability distribution, we propose to predict the midline as its expected value:

$$\mathbf{midline}_y^{\mathrm{pred}} = \sum_x x \cdot \mathbf{output}_{xy}^{\mathrm{midline}}.$$
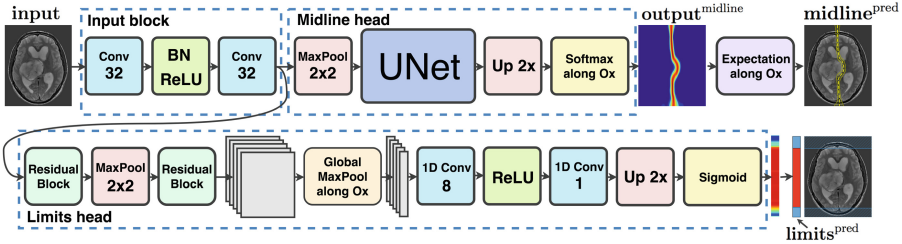
Fig. 3. Schematic representation of the proposed architecture.

The overall architecture for midline estimation is shown in Fig. 3 (top). For our experiments we chose a UNet-based [15] architecture as a de facto standard for medical image segmentation. We replaced plain convolutional layers by residual blocks [4] which are considered to improve the performance, as suggested by [11]. Also, during feature maps concatenation we use linear interpolation to make the output's shape equal to the input's shape. Finally, we apply a softmax non-linearity to the network's output along the Ox axis (instead of sigmoid), which ensures that the constraint from (1) is respected. Note that because the head's output represents a probability distribution, at inference time we can calculate various statistics based on this distribution, e.g. percentiles, which are needed to estimate confidence intervals. This is a very important aspect of our approach which gives us a natural means of estimating the model's uncertainty.

4.2   Limits Prediction

Since the proposed midline estimation approach yields $\mathbf{midline}_y^{\mathrm{pred}}$ for all y-coordinates, we need to filter out the predicted values for the regions where the midline is not defined (Fig. 2, hatched). The corresponding limits are obtained by thresholding the second head's output ($\mathbf{limits}^{\mathrm{pred}}$) and taking the convex hull.

The architecture of the second head is shown in Fig. 3 (bottom). It has the same input layers as the midline estimation network, which are followed by two residual blocks [4]. Next, a global max pooling is applied along the Ox axis in order to reduce the dimensionality of the 2D feature maps to 1D. Finally we apply two 1D convolutions followed by the sigmoid activation function.

5   Experimental Setup

At train time in all of our experiments we used Adam optimizer [7] with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999$) and a learning rate of 10^{-3}, which showed the best results on the validation set. We used equal ($\lambda_1 = \lambda_2 = 1$) weights in the final loss as we didn't notice any loss imbalance at train time.

Also, we applied a simple preprocessing in order to reduce the data variability: resampling the axial slices to a 0.5×0.5 mm pixel spacing, background removal by Otsu thresholding [12] and intensity normalization to zero mean and unit variance. Additionally, at train time we used random flips along the Ox axis as a cheap data augmentation technique.

The training was performed on batches of size 40 (which was simply determined by the amount of available GPU memory), until the validation scores reached a plateau, which happened at approx. 32000 batches. For this reason we used 32000 iterations for all our experiments.

6 Data

In our experiments we used data from two sources.

The first dataset (DS1) consists of 352 MRI series that come from a neurosurgery hospital and belong to patients with severe brain damage caused by tumors: 64% of the images have a significant midline shift (≥ 5 mm), the mean MLS is 7.8 ± 5.0 mm. The dataset was labeled by an experienced neuroradiologist (exp1) and three specialists with limited background in neuroradiology (exp2-4). Their inter- and intra-expert variability is shown in Table 2. We split this dataset using 5-fold cross-validation. For each fold, we additionally leave 8 images out the training set to form a validation set.

The second dataset (DS2) comes from an out-patient clinic and represents a homogeneous sample of 203 MRI series acquired in routine clinical practice. For this dataset only the MLS is available but not the midline itself; only 8% of images have a large MLS (≥ 5 mm), the mean MLS is 2.9 ± 1.5 mm. We use this dataset only for final models' quality assessment in a prospective fashion.

The series from both sources contain only axial slices but have various voxel spacings, ranging from $0.2 \times 0.2 \times 1$ mm to $1 \times 1 \times 5$ mm, and modalities: T1 (25%), T2 (68%) and FLAIR (7%). The images were collected using scanners from GE/Siemens and Toshiba/Siemens for DS1 and DS2 respectively.

7 Results

7.1 Midline Shift Detection

We compare the proposed method with a direct MLS regression via ResNet [4] on two tasks: (1) MLS prediction; (2) significant MLS (≥ 5 mm) detection. In order to evaluate the quality of both methods we use mean absolute error (MAE) and the area under the ROC-curve (ROC AUC) for task 1 and 2 respectively. The ROC-curve was obtained by thresholding the predicted MLS by different values (from 0 to maximal MLS magnitude). The results are presented in Table 1.

Table 1. Midline shift detection scores for various models (± std) calculated on 5-fold cross-validation.

	MAE, mm		ROC AUC	
	DS1	DS2	DS1	DS2
ResNet-152	2.92 ± 3.15	1.84 ± 1.10	0.91 ± 0.03	0.80 ± 0.04
Proposed	$\mathbf{1.54 \pm 1.98}$	$\mathbf{0.75 \pm 0.04}$	$\mathbf{0.95 \pm 0.02}$	$\mathbf{0.92 \pm 0.02}$

7.2 Midline Estimation

In order to assess the midline estimation performance we use root-mean-square error (RMSE) as well as maximal error (MAX):

$$\mathrm{RMSE}(\mathbf{midline}_y, \mathbf{midline}_y^{\mathrm{pred}}) = \sqrt{|I|^{-1} \sum\nolimits_{y \in I} (\mathbf{midline}_y - \mathbf{midline}_y^{\mathrm{pred}})^2},$$

$$\mathrm{MAX}(\mathbf{midline}_y, \mathbf{midline}_y^{\mathrm{pred}}) = \max_{y \in I} |\mathbf{midline}_y - \mathbf{midline}_y^{\mathrm{pred}}|.$$

These metrics, averaged along axial slices (MAXs, RMSEs) as well as entire brain images (MAX, RMSE), are shown in Table 2.

Table 2. Top: midline estimation metrics (± std) calculated on 5-fold cross-validation for DS1. Bottom: neuroradiologist (exp1) variability on DS1.

	MAX	RMSE	MAXs	RMSEs
Segmentation	7.45 ± 9.84	0.95 ± 0.61	2.12 ± 3.25	0.81 ± 0.64
Proposed	3.61 ± 2.62	0.79 ± 0.44	1.58 ± 1.39	0.69 ± 0.54
exp1 vs exp1	3.16 ± 2.16	0.66 ± 0.19	1.47 ± 1.05	0.62 ± 0.50
exp1 vs exp2-4	3.44 ± 2.13	0.77 ± 0.35	1.47 ± 0.97	0.66 ± 0.19

We compare our method with a naïve segmentation approach mentioned in Sect. 4.1. Note that plain segmentation performs significantly worse in terms of maximal error, which is a more important characteristic for MLS detection.
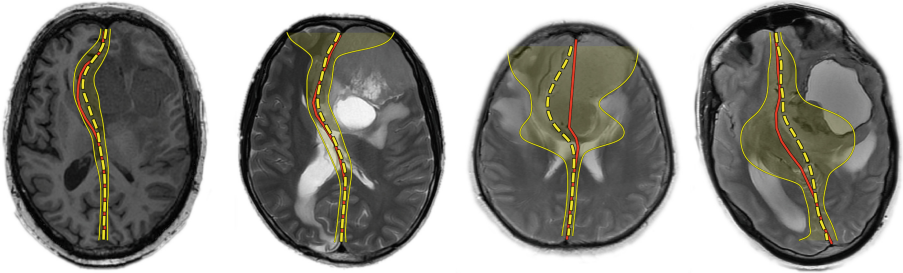
8 Discussion

Fig. 4. Ground-truth (red) and predicted (yellow, dashed) midlines with their 95% confidence intervals for 2 random samples (left) and 2 typical examples from the set of cases with the largest errors (right). (Color figure online)

Figure 4 (right) shows several examples on which our method performs poorly. Our analysis of such examples suggests that the main source of errors are some really complicated cases that even professional radiologists have doubts with, e.g. images on which the tumor is located directly in the middle of the brain, or incorrect cases with an extracerebral tumor located in the medial longitudinal fissure, e.g. falx meningioma. Note how in the areas of greatest error the model's uncertainty is much higher.

Our preliminary experiments with CT images show that the proposed method can be easily adapted to work with CT, however we require a larger dataset to support this claim, which might be the subject of our future work.

References

1. Chen, W., Belle, A., Cockrell, C., Ward, K.R., Najarian, K.: Automated midline shift and intracranial pressure estimation based on brain CT images. J. Visualized Exp. JoVE **74**, e3871 (2013)
2. Chilamkurthy, S., et al.: Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. Lancet **392**(10162), 2388–2396 (2018)
3. Gamburg, E.S., Regine, W.F., Patchell, R.A., Strottmann, J.M., Mohiuddin, M., Young, A.B.: The prognostic significance of midline shift at presentation on survival in patients with glioblastoma multiforme. Int. J. Radiat. Oncol.* Biol.* Phys. **48**(5), 1359–1362 (2000)
4. 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)

5. Jacobs, B., Beems, T., van der Vliet, T.M., Diaz-Arrastia, R.R., Borm, G.F., Vos, P.E.: Computed tomography and outcome in moderate and severe traumatic brain injury: hematoma volume and midline shift revisited. J. Neurotrauma **28**(2), 203–215 (2011)
6. Jain, S., et al.: Automatic quantification of CT features inacute traumatic brain injury. J. Neurotrauma (2019)
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Liao, C.C., Chen, Y.F., Xiao, F.: Brain midline shift measurement and its automation: a review of techniques and algorithms. Int. J. Biomed. Imaging (2018)
9. Liu, R., et al.: Automatic detection and quantification of brain midline shift using anatomical marker model. Comput. Med. Imaging Graphics **38**(1), 1–14 (2014)
10. McGinty, G.B., Allen, B.: The ACR data science institute and AI advisory group: harnessing the power of artificial intelligence to improve patient care. J. Am. College Radiol. **15**(3), 577–579 (2018)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
12. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
13. Paletta, N., Maali, L., Zahran, A., Sethuraman, S., Figueroa, R., Nichols, F.T., Bruno, A.: A simplified quantitative method to measure brain shifts in patients with middle cerebral artery stroke. J. Neuroimaging **28**(1), 61–63 (2018)
14. Pullicino, P.M., Alexandrov, A., Shelton, J., Alexandrova, N., Smurawska, L., Norris, J.: Mass effect and death from severe acute stroke. Neurology **49**(4), 1090–1095 (1997)
15. 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.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Ross, D.A., Olsen, W.L., Ross, A.M., Andrews, B.T., Pitts, L.H.: Brain shift, level of consciousness, and restoration of consciousness in patients with acute intracranial hematoma. J. Neurosurg. **71**(4), 498–502 (1989)