

Brain midline shift detection and quantification by a cascaded deep network pipeline on non-contrast computed tomography scans

Nguyen P. Nguyen¹, Youngjin Yoo², Andrei Chekkoury³, Eva Eibenberger³, Thomas J. Re², Jyotipriya Das², Abishek Balachandran², Yvonne W. Lui⁴, Pina C. Sanelli⁵, Thomas J. Schroepfel⁶, Uttam Bodanapally⁷, Savvas Nicolaou⁸, Tommi A. White⁹, Filiz Bunyak¹, Dorin Comaniciu², and Eli Gibson²

¹Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA ²Digital Technology and Innovation, Siemens Healthineers, Princeton, NJ, USA ³Computed Tomography R&D, Siemens Healthineers, Erlangen, Germany ⁴Department of Radiology, New York University, New York, NY, USA ⁵Donald and Barbara Zucker School of Medicine, Northwell Health, New York, NY, USA ⁶University of Colorado Health, Memorial Hospital, Colorado Springs, CO, USA ⁷University of Maryland Medical Center, Baltimore, MD, USA ⁸Vancouver General Hospital, Vancouver, BC, Canada ⁹Department of Biochemistry, University of Missouri, Columbia, MO, USA

Abstract

Brain midline shift (MLS), demonstrated by imaging, is a qualitative and quantitative radiological feature which measures the extent of lateral shift of brain midline structures in response to mass effect caused by hematomas, tumors, abscesses or any other space occupying intracranial lesions. It can be used, with other parameters, to determine the urgency of neurosurgical interventions and to predict clinical outcome in patients with space occupying lesions. However, precisely detecting and quantifying MLS can be challenging due to the great variability in clinically relevant brain structures across cases. In this study, we investigated a cascaded network pipeline consisting of case-level MLS detection and initial localization and refinement of brain landmark locations by using classification and segmentation network architectures. We used a 3D U-Net for initial localization and subsequently a 2D U-Net to estimate exact landmark points at finer resolution. In the refinement step, we fused the prediction from multiple slices to calculate the final location for each landmark. We trained these two U-Nets with the Gaussian heatmap targets generated from the brain's anatomical markers. The case-level ground-truth labels and landmark annotation were generated by multiple trained annotators and reviewed by radiology technologists and radiologists. Our proposed pipeline achieved the case-level MLS detection performance of 95.3% in AUC using a testing dataset from 2,545 head non-contrast computed

tomography cases and quantify MLS with a mean absolute error of 1.20 mm on 228 MLS positive cases.

1. Introduction

Midline shift (MLS), or the lateral displacement of brain midline structures, is a quantifiable imaging finding used in the diagnosis and quantification of mass effect due to space occupying intracranial lesions such as traumatic and spontaneous hematomas, tumors and abscesses [22, 1, 20]. MLS is used as an indicator of severity of mass effect, as well as an indicator of urgency of an intervention [17, 24]. Along with other clinical parameters, it has been shown to be a predictor of poor clinical outcome in patients with head injury [17] and with spontaneous intracerebral hemorrhage [24]. MLS of more than 5mm usually requires careful neurological assessment, close observation, and possible surgical intervention [10, 2]. Due to the great variability in tissue distortion based on underlying pathology and patient anatomy, accurate automated detection and quantification of midline shift is challenging. Modeling radiological features for MLS from brain anatomical structure is a high-level concept, and it is influenced by the anatomical and pathological variance across subjects, post-operations, imaging artifact and unclear landmark boundaries. A robust automated MLS detection system, which could immediately flag studies for urgent review, would be helpful to efficiently allow timely identification of patients who need

immediate care by clinicians.

MLS is clinically defined as a shift of the brain past its center line. It is typically detected and quantified by the perpendicular distance between the brain's midline and a shifted anatomical structure. One common practice is to use the maximum distance from the midline formed by the anterior falx (AF) and posterior falx (PF) to the septum pellucidum (SP), measured at the axial level of the foramen of Monro (FM) [10] as shown in Figure 1. Several approaches have been studied for computational MLS modeling from a 3D head CT scan. The symmetry-based approaches rely on the idea that the brain exhibits rough local symmetry around the midline [11, 3]. Thus, local symmetry features have been used to construct the deformed midline (DML), followed by measuring the midline shift. The construction of DML requires complex geometric settings, and that is less flexible to incorporate different image feature conditions. The landmark-based approaches presented in [21, 13] rely on anatomical markers to measure MLS distance. The best candidate markers are selected to construct the deformed midline; then, they are compared with the ideal midline for MLS quantification. Recent studies using deep learning also primarily take a landmark-based approach to estimate MLS. Jain et al. [8] applied a 3D U-Net to segment and detect landmarks in 38 computed tomography (CT) images for MLS estimation on a pre-defined range of candidate slices. The potential challenge would be less accurate determination of appropriate axial slices and the requirement of accurate in-plane orientation estimation of these axial slices to measure the perpendicular MLS distance. Pisov et al. [16] proposed a combination of deep regression and classification to predict the midline and its limits for quantification. This solution is restricted to 2D MRI slice images, and it requires line annotation (multiple points) for the regression.

In this study, we focus on developing a robust MLS detection and measurement prototype consisting of multiple computational modules for 3D CT images. 3D CT image classification and regression of landmark locations using a single deep network could be prone to overfitting and may not provide desired accuracy due to the challenges described above. To overcome these challenges, our proposed MLS estimation pipeline utilizes a cascaded approach. It first performs case-level detection, then followed by a measurement of MLS in 3D CT volumes. In the next step, we refine MLS on the 2D planes with higher resolution by selecting three consecutive axial slices instead of using a single slice to make the prediction more robust. The final MLS is quantified from the fused locations of these detected landmarks on three slices.

2. Method

Our proposed system has two major modules: MLS detection and quantification, which are performed by cascaded

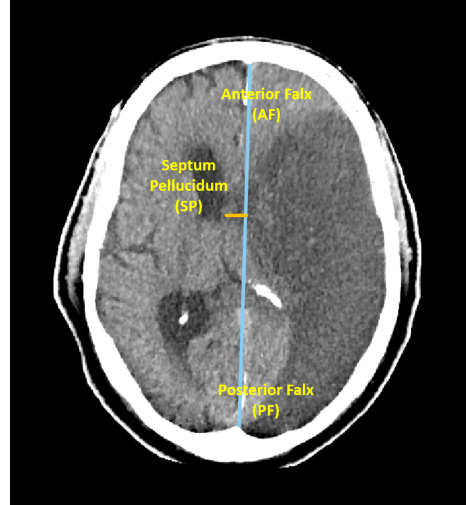


Figure 1. The clinical definition of brain midline shift (MLS) on a head non-contrast CT scan. Image Courtesy: New York University Langone Health.

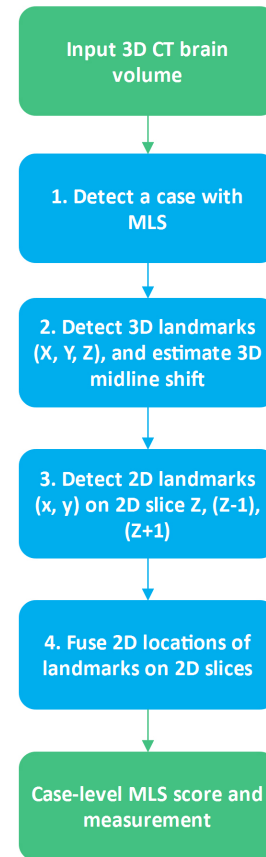


Figure 2. Midline shift (MLS) estimation pipeline.

operations using a DenseNet and U-Nets that are commonly employed in medical image analysis tasks. For MLS detection (Stage 1), the case-level MLS classification was performed by a 3D DenseNet [7] that takes a 3D CT volume

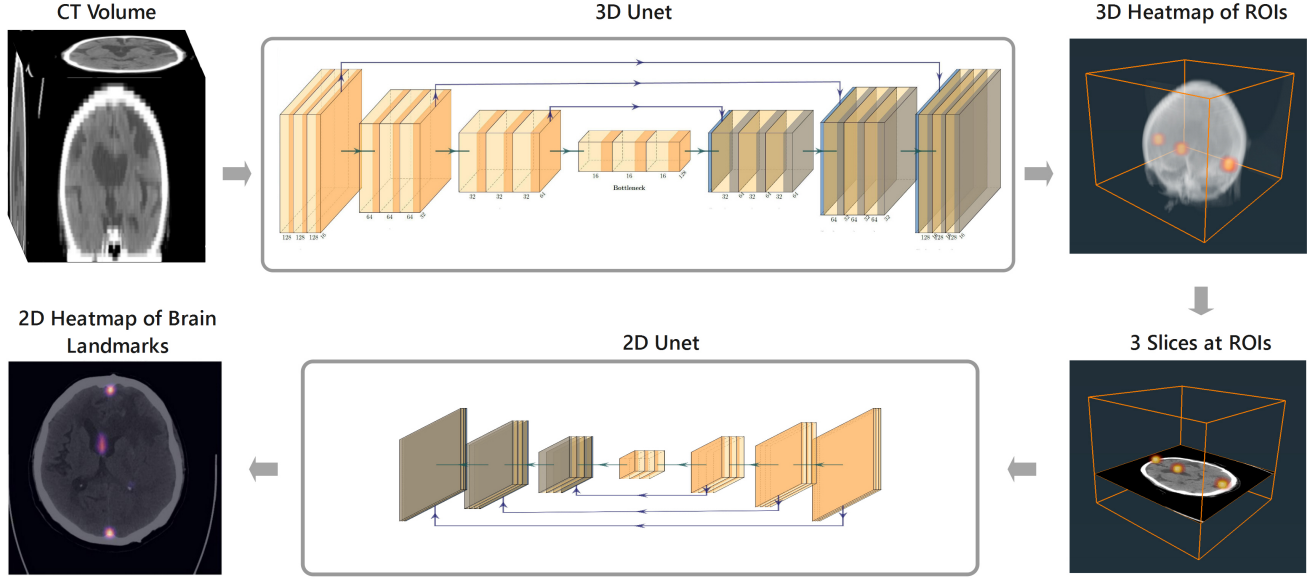


Figure 3. The cascaded 3D U-Net and 2D U-Net architecture for brain landmark heatmap prediction in Stage 2 & 3. Image Courtesy: New York University Langone Health.

as an input. MLS quantification takes a 3D CT volume with MLS detected by the previous module and consists of the next three stages: 2) The detection of 3D regions of interest (ROIs); 3) The refinement of three brain landmark points: AF, PF and SP; and 4) The fusion of detected landmarks across axial slices. For the quantification stages, we employed U-Nets [18] to predict the Gaussian heatmaps of landmark points. In stage 2, a 3D U-Net detected 3D heatmaps of ROIs, and the landmark points were estimated as the centroids of those ROIs by connected component analysis in the 3D image volume. The image slice at the Z coordinate of the second landmark SP was extracted as an FM axial level for further prediction. Stage 3 used the selected slice as an input, and a 2D U-Net predicted a 2D Gaussian heatmap to refine the location of three landmark points at higher resolution. If the prediction in the 2D heatmap deviates from the 3D prediction by a certain amount selected empirically, the 3D prediction in Stage 2 is used as a fall back option. At the FM level, SP serves as the actual midline. The ideal midline was constructed from two points AF and PF. Then the midline shift was estimated by the closest distance from SP to the ideal midline of AF and PF. Figure 2 depicts a diagram of our overall working procedure for midline shift estimation, and Figure 3 illustrates the pipeline of landmark detection in our workflow.

2.1. Stage 1: Detect a case with MLS

To classify MLS cases from normal cases, we built a 3D DenseNet [7] that takes a 3D CT volume as an input that is resampled to have image dimensions $256 \times 256 \times 64$. The 3D network uses isotropic 3D kernels to balance res-

olution and speed with deep dense blocks that gradually aggregate features down to a binary output. The network was trained end-to-end as a classification system using binary cross-entropy loss and used probabilistic sampling of the training data to adjust for the imbalance in the training dataset labels. A separate validation dataset was used for final model selection before the performance was measured on the testing set. The first layer of this network consists of convolution (kernels $7 \times 7 \times 7$)—batch normalization—ReLU and max-pooling (kernels $3 \times 3 \times 3$, stride $2 \times 2 \times 2$). The subsequent five dense blocks are isotropic with convolution (kernels $3 \times 3 \times 3$)—batch normalization—ReLU and Max-pooling (kernels $2 \times 2 \times 2$, stride $2 \times 2 \times 2$) followed by a final linear classifier. The number of filters in the first convolution layer was 16, growth rate was 5 and the numbers of layers in each block were (2, 2, 3, 4, 4).

2.2. Stage 2: Identify FM axial level, detect 3D landmarks, and estimate MLS measurement

Instead of predicting landmark locations directly, we first predicted ROIs to limit the searching range of these landmark points in 3D volume images for more efficient estimation. We built a 3D network with the U-Net architecture and the ResNet [6] backbone using the Medical Open Network for Artificial Intelligence (MONAI) framework [5]. The network’s encoder and decoder have four main layers. Each layer has 16 residual blocks [6] and instance normalization. Training target was a 3D Gaussian heatmap, which was generated from three manual landmark annotations of AF, PF and SP for each volume. We applied connected component analysis and centroid estimation on those 3D heatmaps to

detect the 3D landmark points. We then estimated the 3D midline shift by the distance between 3D point SP to the 3D line of AF and PF on the detected FM axial level. Since the Gaussian heatmap target was encoded with continuous values, we optimized this network's parameters by using a mean squared error loss function to minimize the difference between the prediction output and the training target. Figure 4 illustrates the location of detected 3D landmarks and the 3D MLS distance.

2.3. Stage 3: Detect 2D landmarks and refine MLS measurement

The 2D FM axial slice extracted from the predicted SP from Stage 2 was fed into the 2D network using the Segmentation Models PyTorch implementation [23] for location refinement at higher resolution. This network's encoder has three main layers, and each layer contains 16 residual blocks [6]. We trained the network in the same way using 2D Gaussian heatmaps. Again, we performed the connected component analysis and estimated centroids as 2D landmark points, which were used to calculate the 2D MLS estimate. The threshold for using the 3D landmarks as a fallback option was selected empirically at 2.5 mm.

2.4. Stage 4: Fuse 2D landmark points from multiple axial slices

Instead of picking a single slice to predict three landmark points, we selected three consecutive slices: the slice at SP's Z coordinate, at Z-1, and at Z+1 to make our prediction more robust. At first, the location of each landmark was calculated on each slice. After that, we fused those results into one location for each landmark using the proposed Algorithm 1. This method aims to find the location with minimum difference between the detected 2D landmarks in three slices and the detected 3D landmark based on Euclidean distance. For each landmark, two locations with the smallest distance were selected, and then we chose the location which is closer to the 3D location as the final 2D coordinate of the landmark.

3. Experiments and Results

3.1. Dataset

Anonymized non-contrast head CT volumes were retrospectively collected from six centers in the USA, Canada and India with approval from their respective ethics committees. The data set comprised 25,037 CT volumes (2,456 MLS positive) acquired from multiple manufacturers including Siemens Healthineers, GE Healthcare and Toshiba. The case-level ground-truth labels were generated by radiological reports which were provided by the contributing centers and were manually transcribed by a team of trained annotators (with at least 40 hours of training supervised by a

Algorithm 1: Fuse the 2D landmark predictions from 3 axial slices

Result: 2D locations of 3 landmark points: AF, PF and SP

for each landmark in [AF, PF, SP] **do**

1. Detect the landmark's 3D location (X, Y, Z);
2. Detect the landmark's 2D location (x1, y1), (x2, y2), (x3, y3) on 3 slices: slice at (Z), slice before at (Z-1), and slice after at (Z+1);

if The landmark's 2D coordinates are detected successfully **then**

1. Compute the distance between each pair of these 3 detected locations (x1, y1), (x2, y2), and (x3, y3);
2. Pick the smallest distance, and its two corresponding 2D locations;
3. Compute the distance of these two 2D locations to the landmark's 3D location (X, Y) in 3D volume;
4. Select the smallest distance to the 3D location (X,Y), and use its corresponding 2D location (x, y) as the 2D coordinates of the current landmark;

else

1. Extract the (X, Y) of 3D location to use as the 2D location (x, y) on 2D slice;

end

end

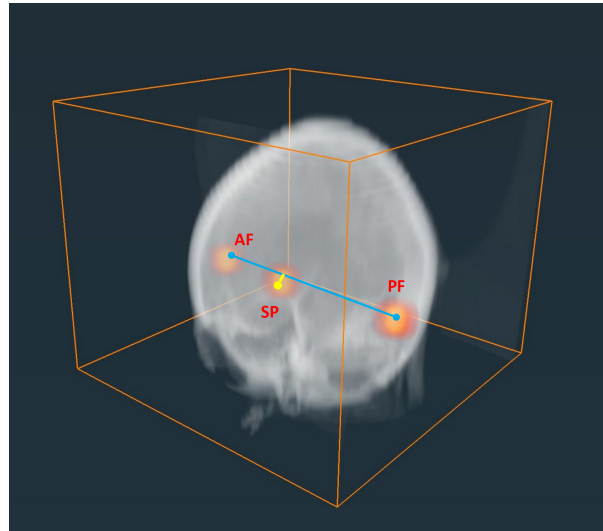


Figure 4. Prediction of MLS measurement in the 3D Gaussian heatmap visualization. Image Courtesy: New York University Langone Health.

radiologist) and reviewed by a radiology technologist (with a Bachelors in Radiology and medical imaging technologist

training with one year of training in neuroimaging) and one of three radiologist (> 5 years of experience). The manual landmark annotation was similarly performed by trained annotators, radiology technologists and radiologists for the all MLS positive cases. The data was randomly split at patient level into 19,946 training cases (2,019 MLS positives), 2,545 validation cases (232 MLS positives) and 2,546 test cases (228 MLS positives). As pre-processing steps, gantry-tilt was first corrected and then the input image was resized to the dimension of $256 \times 256 \times 64$ voxels for Stage 1 in the pipeline. In Stage 2, we resized images into the dimension of $256 \times 256 \times 48$ voxels. Stage 3 and Stage 4 used the original slice at high-resolution dimensions which were mostly 512×512 voxels. The MLS positive cases were used for training and evaluating the quantification performance (Stage 2-4). The CT head window parameters level 40 HU and width 160 HU were used.

3.2. Training

In our proposed pipeline, the first network detecting MLS cases was trained with random initialization. We used the Adam optimizer [9] with a learning rate 0.001 to minimize the binary cross-entropy loss function. Dropout [19] was applied with a rate of 0.3. For data augmentation, we applied a random mirror flip for in-plane orientations with a probability of 0.5 and in-plane random translations that were limited to 10 voxels in each dimension. We perturbed the image intensity within a random interval between $[-10, 10]$ HU. We also applied random 3D rotation limited to ± 3 degrees in each dimension. The next two U-Net models detecting landmarks were also trained from scratch. The 3D U-Net was trained with the Adam optimizer [9] and a learning rate of 0.001. The 2D U-Net was trained with the Adam optimizer [9] and a learning rate of 0.0001. In both cases, we used the mean squared error loss function to optimize the continuous heatmap output. For the quantification stages, we did not apply data augmentation on training data.

3.3. Test results and Discussion

Table 1 shows that the first network detected MLS cases with a sensitivity of 0.89 and a specificity of 0.89. The total accuracy ended up with 0.89 when using an operation point providing a balanced performance between sensitivity and specificity which was 0.13. When using an operation point 0.5 that is biased toward MLS negative cases due to disease prevalence, the classification accuracy was 0.94. Figure 5 shows the receiver operating characteristic (ROC) curve illustrating the diagnostic ability of our classifier at different thresholds. The area under the curve (AUC) of our finding was 0.95 on 2,545 CT test cases. We note that [4] had conducted MLS detection on two other datasets: Qure25k with 25k images & AUC of 0.93; and CQ500 with 500 images & AUC of 0.97. Compared to their performance re-

sults, our detection performance in AUC was higher for the equivalent-size dataset of 25k images, but lower than the AUC of 500 image dataset.

Among the false positive (FP) case-level detection error cases, there were two major error types. The first major FP type was subtle MLS, which was mostly among the 2-5 mm MLS cases. This error type was mostly attributed to ground-truth label ambiguity (will be described in more detail at the end of this section) and the limited sensitivity of our method for the subtle MLS cases. The clinical utility of flagging subtle MLS would depend on each institution's clinical workflow. The other major FP type was from the cases with distorted anatomy. The post-operative cases with shunt and surgery had deformed anatomical structures that influenced the network prediction. Also individual anatomical variants such as cavum SP and pathology such as hemorrhage were other sources of distorted anatomy. Other less frequent error types included imaging artifact, incorrect field-of-view and poor signal-to-noise ratio. For the false negative (FN) error cases, the major error type was subtle MLS. We observed fewer cases with distorted anatomy among the FN cases.

Table 1. Case-level MLS detection performance on 2,545 testing CT cases. The used operation point was 0.13 (refer to Figure 5 for the ROC curve).

AUC	Sensitivity	Specificity	Accuracy
0.95	0.89	0.89	0.89

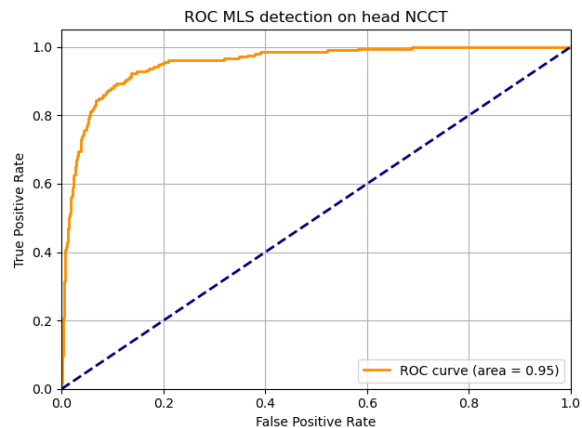


Figure 5. ROC curve for case-level MLS detection on 2,545 testing CT cases.

In the first stage of the quantification phase (Stage 2), the 3D MLS landmark prediction primarily serves as an initialization and FM axial level detection, but also gives an initial estimate of MLS. Our test result showed that the differences between the 3D MLS prediction and its corresponding tar-

get had a mean of 1.31 mm and a standard deviation of 1.34 mm.

We used the Z coordinate of the 3D point SP as an index to select the FM plane/slice containing the 2D landmark points AF, PF and SP. Figure 6 illustrates the differences between the model’s prediction and the target of the FM slice indices for each slice thickness range. The difference of slice prediction increased when the slice spacing decreased mostly due to the fact that thinner slice spacing cases can have multiple similar FM slices. The largest portion of the slice difference was 0 and 1 (89%). Specifically, in 228 positive test cases, there were 98 cases with zero slice difference (43%), 104 cases with one slice difference (46%), 16 cases with two slice differences (7%), and 10 cases with three or more slice differences (4%). Since the number of cases with a maximum difference of one slice represents 89% of the total test cases, we made the 2D refinement from three slices Z, Z-1, and Z+1. We then applied Algorithm 1 to fuse all the predicted locations of each landmark into one final 2D location. According to [14], the typical FM axial size range is 2-4mm in diameter, and our selection of these three slices mostly falls into this range.

In Stage 4 of 2D MLS refinement, the test results showed that the differences between the 2D MLS prediction and its corresponding target were a mean of 1.20 mm and a standard deviation of 1.15 mm, which is 0.11 mm smaller than the 3D estimates in the mean difference. This improvement could be attributed to the usage of axial 2D slices with higher resolution. In [8], the mean difference of 0.86 mm was reported from 38 MLS positive cases. Both the study [8] and our work’s MLS measurement errors are within the typical expert user variability which is between 1 and 2 mm, which was shown in our in-house user variability examination. These outcomes also fall in the clinically accepted precision range of 2.0 mm for landmark localization in CT cephalometric analysis [12]. Figure 7 displays the distribution of our MLS quantification error. In all positive test cases, the number of cases with a difference of MLS distance ≤ 1 mm was 130 (57%), 1 – 2 mm was 52 (23%), 2 – 3 mm was 27 (12%), 3 – 4 mm was 15 (7%), 4 – 5 mm was 3 (1%), and >5 mm was 1 (0.4%). We noted that our reference standard for these measurements in expert annotation with 1-2 mm variability could have artificially inflated the results. We also observed that our measurement was slightly over-estimated in general as shown in Figure 7.

Figure 8 shows the successful estimate cases. The landmark AF and PF were detected accurately in most cases. Although the landmark SP is more difficult to identify, our proposed method could detect it among different scales of image objects, and different distortion conditions. The landmark SP was also detected based on a rotation of 5 or 10 degrees and a skew on the left or on the right.

Our study has three key limitations. First of all, the case-

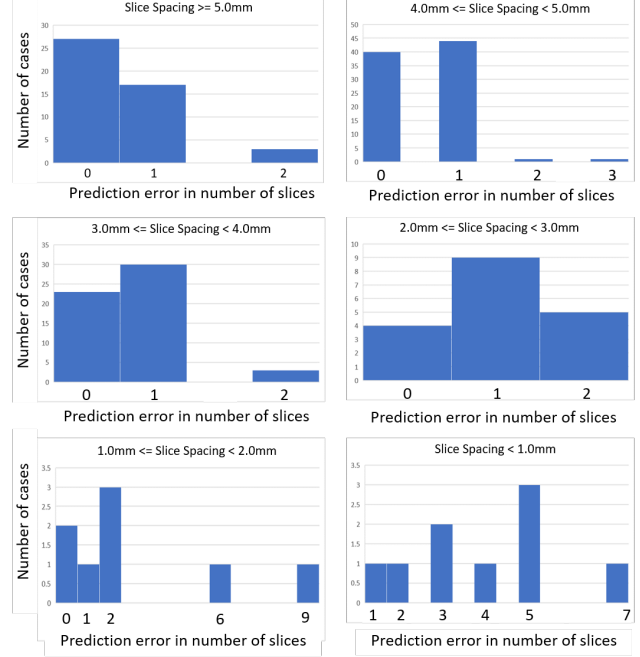


Figure 6. Differences between prediction and target in FM slice detection for each slice thickness range.

Table 2. Comparison of MLS Estimates from 3D and 2D landmarks on 228 MLS positive cases.

Quantification method	Difference from GT (mm)	
	Mean	Std
3D estimates	1.31	1.34
Final estimates with 2D refinement	1.20	1.15

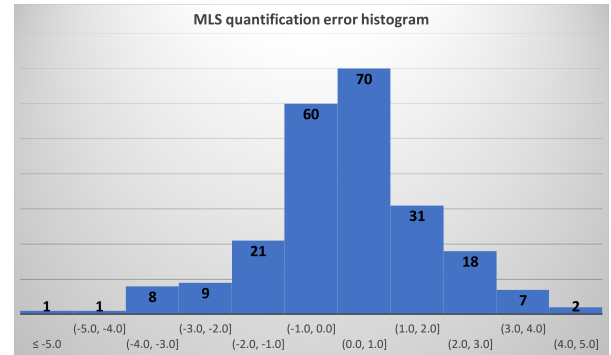


Figure 7. Histogram of MLS quantification error on positive testing cases. X-axis represents absolute error ranges between ground truth measurements and predicted measurements, and y-axis represents the number of cases for each error range.

level MLS labels were primarily created from their clinical reports. For some MLS cases (typically 3-5 mm cases) and follow-up cases in some longitudinal studies, we observed

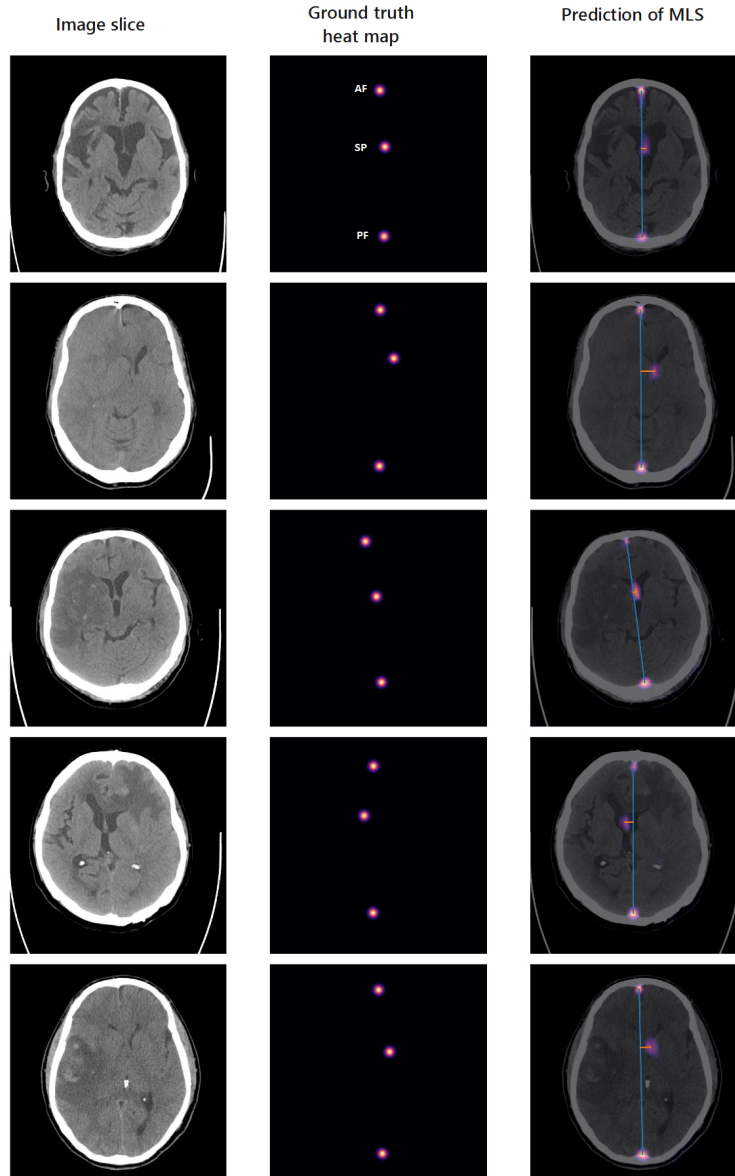


Figure 8. The examples of MLS quantification. The columns show input CT slices, target landmarks, and predicted landmarks and MLS measurements respectively. Image Courtesy: New York University Langone Health and University of Colorado Health.

that their clinical reports did not mention MLS although MLS is visually present in the corresponding CT images. Therefore, there could be some ground-truth labeling ambiguity for subtle cases in our dataset. Secondly, although our data size is comparable or larger than most prior studies, the number of MLS positive testing cases may be still limited due to the small disease prevalence. The generalization ability of our detection and quantification capability would need to be investigated further with larger testing datasets. Thirdly, data distribution/size, experiment conditions and evaluation strategies are different among the prior studies, which makes direct comparison challenging.

4. Conclusions

We successfully developed and evaluated a cascaded deep network pipeline for classifying an MLS case from negatives and detecting landmarks to quantify MLS using a large clinical dataset consisting of 3D head CT scans collected from multiple institutions. Our work avoids complex handcrafted features or complex geometric computation. For future study, in addition to identifying landmark locations from Gaussian heatmaps, we will investigate integrating the landmarks' coordinates directly into the loss function of deep neural networks to minimize their error.

Another future research objective would be to search for other U-Net like architectures [15], which may help us to delineate all the spatial features from coarse to fine layers and their level sequences. The clinical utility of the proposed method will be investigated with better clinically defined targets for assisting practicing physicians and reducing image exam turnaround time.

Disclaimer

The concepts and information presented in this paper are based on research results that are not commercially available. Future availability cannot be guaranteed.

References

- [1] Ronald HMA Bartels, Frederick JA Meijer, Hans van der Hoeven, Michael Edwards, and Mathias Prokop. Midline shift in relation to thickness of traumatic acute subdural hematoma predicts mortality. *BMC neurology*, 15(1):1–6, 2015.
- [2] M. Ross Bullock, Randall Chesnut, Jamshid Ghajar, David Gordon, Roger Hartl, David W. Newell, Franco Servadei, Beverly C. Walters, Jack E. Wilberger, and Surgical Management of Traumatic Brain Injury Author Group. Surgical management of acute subdural hematomas. *Neurosurgery*, 58(3):Suppl, Mar 2006.
- [3] Mingyang Chen, Ahmed Elazab, Fucang Jia, Jianhuang Wu, Guanglin Li, Xiaodong Li, and Qingmao Hu. 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, Dec 2015.
- [4] Sasank Chilamkurthy, Rohit Ghosh, Swetha Tanamala, Mustafa Biviji, Norbert G. Campeau, Vasantha Kumar Venugopal, Vidur Mahajan, Pooja Rao, and Prashant Warier. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*, 392(10162):2388–2396, Dec 2018.
- [5] Medical Open Network for AI. (MONAI). <https://monai.io>.
- [6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [7] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *arXiv*, Aug 2016.
- [8] Saurabh Jain, Thijs Vande Vyvere, Vasilis Terzopoulos, Diana Maria Sima, Eloy Roura, Andrew Maas, Guido Wilms, and Jan Verheyden. Automatic Quantification of Computed Tomography Features in Acute Traumatic Brain Injury. *Journal of Neurotrauma*, 36(11):1794–1803, 2019.
- [9] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv*, Dec 2014.
- [10] Chun-Chih Liao, Ya-Fang Chen, and Furen Xiao. Brain Midline Shift Measurement and Its Automation: A Review of Techniques and Algorithms. *International Journal of Biomedical Imaging*, 2018:4303161, 2018.
- [11] Chun-Chih Liao, Furen Xiao, Jau-Min Wong, and I-Jen Chiang. Automatic recognition of midline shift on brain ct images. *Computers in Biology and Medicine*, 40(3):331 – 339, 2010.
- [12] Claudia Lindner, Ching-Wei Wang, Cheng-Ta Huang, Chung-Hsing Li, Sheng-Wei Chang, and Tim F. Cootes. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. *Sci. Rep.*, 6(33581):1–10, Sep 2016.
- [13] Ruizhe Liu, Shimiao Li, Bolan Su, Chew Lim Tan, Tze-Yun Leong, Boon Chuan Pang, C.C. Tchoyoson Lim, and Cheng Kiang Lee. Automatic detection and quantification of brain midline shift using anatomical marker model. *Computerized Medical Imaging and Graphics*, 38(1):1 – 14, 2014.
- [14] Tomasz Matys, Fraser S. Brown, Fulvio Zaccagna, Ramez W. Kirolos, and Tarik F. Massoud. A critical appraisal of Monro’s erroneous description of the cerebral interventricular foramina: Age-related magnetic resonance imaging spatial morphometry and a proposed new terminology. *Clin. Anat.*, 33(3):446–457, Apr 2020.
- [15] Fausto Milletari, Nicola Rieke, Maximilian Baust, Marco Esposito, and Nassir Navab. CFCM: Segmentation via Coarse to Fine Context Memory. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 667–674. Springer, Cham, Switzerland, Sep 2018.
- [16] Maxim Pisov, Mikhail Goncharov, Nadezhda Kurochkina, Sergey Morozov, Victor Gomboleviskiy, Valeria Chernina, Anton Vladzmyrskyy, Ksenia Zamyatina, Anna Chesnokova, Igor Pronin, Michael Shifrin, and Mikhail Belyaev. Incorporating task-specific structural knowledge into cnns for brain midline shift detection, 2019.
- [17] Keith B Quattrocchi, Praveen Prasad, Neil H Willits, and Franklin C Wagner Jr. Quantification of midline shift as a predictor of poor outcome following head injury. *Surgical Neurology*, 35(3):183–188, 1991.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer, Cham, Switzerland, Nov 2015.
- [19] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [20] Yu Deok Won, Min Kyun Na, Je-Il Ryu, Jin-Hwan Cheong, Jae-Min Kim, Choong-Hyun Kim, and Myung-Hoon Han. Radiologic factors predicting deterioration of mental status in patients with acute traumatic subdural hematoma. *World neurosurgery*, 111:e120–e134, 2018.
- [21] Furen Xiao, I-Jen Chiang, Jau-Min Wong, Yi-Hsin Tsai, Ke-Chun Huang, and Chun-Chih Liao. Automatic measurement of midline shift on deformed brains using multiresolution binary level set method and hough transform. *Computers in Biology and Medicine*, 41(9):756 – 762, 2011.
- [22] Furen Xiao, Chun-Chih Liao, Ke-Chun Huang, I-Jen Chiang, and Jau-Min Wong. Automated assessment of midline shift in head injury patients. *Clinical neurology and neurosurgery*, 112(9):785–790, 2010.

- [23] Pavel Yakubovskiy. Segmentation Models PyTorch. https://github.com/qubvel/segmentation_models.pytorch, 2020.
- [24] Wen-Song Yang, Qi Li, Rui Li, Qing-Jun Liu, Xing-Chen

Wang, Li-Bo Zhao, and Peng Xie. Defining the optimal midline shift threshold to predict poor outcome in patients with supratentorial spontaneous intracerebral hemorrhage. *Neurocritical care*, 28(3):314–321, 2018.