2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

SIMULTANEOUS CLASSIFICATION AND SEGMENTATION OF INTRACRANIAL HEMORRHAGE USING A FULLY CONVOLUTIONAL NEURAL NETWORK

Danfeng Guo[⋆], *Haihua Wei*[†], *Pengfei Zhao*[⋆], *Yue Pan*[⋆], *Hao-Yu Yang*[⋆], *Xin Wang*[⋆], *Junjie Bai*[⋆], *Kunlin Cao*[⋆], *Qi Song*[⋆], *Jun Xia*[†], *Feng Gao*[⋆,‡], *Youbing Yin*[⋆,‡]

[⋆]CuraCloud Corporation, Seattle, USA
[†]Shenzhen Second People's Hospital, Shenzhen, China

ABSTRACT

Intracranial hemorrhage (ICH) is a critical disease that requires immediate diagnosis and treatment. Accurate detection, subtype classification and volume quantification of ICH are critical aspects in ICH diagnosis. Previous studies have applied deep learning techniques for ICH analysis but usually tackle the aforementioned tasks in a separate manner without taking advantage of information sharing between tasks. In this paper, we propose a multi-task fully convolutional network, ICHNet, for simultaneous detection, classification and segmentation of ICH. The proposed framework utilizes the inter-slice contextual information and has the flexibility in handling various label settings and task combinations. We evaluate the performance of our proposed architecture using a total of 1176 head CT scans and show that it improves the performance of both classification and segmentation tasks compared with single-task and baseline models.

Index Terms— intracranial hemorrhage, segmentation, classification, multi-task learning, fully convolutional network

1. INTRODUCTION

Intracranial hemorrhage (ICH), a critical disease in which bleeding occurs inside the skull, can often lead to severe disability or death. ICH can be further categorized into five subtypes based on the bleeding location: cerebral parenchymal hemorrhage (CPH), intraventricular hemorrhage (IVH), epidural hemorrhage (EDH), subdural hemorrhage (SDH), and subarachnoid hemorrhage (SAH) [1]. It has been reported that the hemorrhage location and volume are significantly associated with one-month mortality and the treatment outcome [2]. Therefore, accurate classification of ICH subtypes and quantification of bleeding volumes are critical for life saving and later recovery.

Computed tomography (CT) is an effective non-invasive diagnostic imaging approach for ICH detection [3]. Deep learning techniques have been adopted to detect and segment ICH using head CT, e.g., [4, 5, 6, 7]. However, most of the previous works have treated ICH detection and segmentation as two independent tasks. To the best of our knowledge, the first attempt of joint ICH classification and segmentation was made by Chang et al. by using a hybrid 3D/2D convolutional neural network (CNN) [6]. The feature maps obtained by 3D convolution operations were projected to 2D and the mask R-CNN architecture [8] was then used. The final 3D results had to be obtained by iterating through all slices, which led to excessive and redundant 3D convolutional operations. In addition, multiple hemorrhages presented in the same slice had to be separated and the corresponding subtypes had to be specified as inputs for the mask R-CNN architecture, which can be challenging in practice considering that mixed subtypes are often observed and regions with different types may be connected.

In this paper, we propose a unified fully convolutional network to perform ICH classification (**cls**) and segmentation (**seg**) simultaneously, referred to as ICHNet. It contains a shared encoder to extract features for both cls and seg tasks while utilizes a convolutional long short-term memory (ConvLSTM) [9] module to capture the sequential information embedded in consecutive slices. Unlike region proposal based approaches, such as mask R-CNN, that relies on instance-level subtype labels, our network can be trained with slice-level labels, offering more flexibility in practical use. We show that the joint FCN network achieves better performance for both cls and seg tasks than performing the tasks independently.

2. METHODS

2.1. Description of Tasks

Given a 3D head CT scan, our work aims at accomplishing a set of **classification (cls) tasks** and a **segmentation (seg) task**. The cls tasks include a binary ICH detection task and a subtype classification task. The binary ICH detection task is aimed at identifying the presence of ICH in each slice (slice level) and in the scan as a whole (subject level). The subtype classification task is to identify the presence of each of the

[‡]Corresponding authors

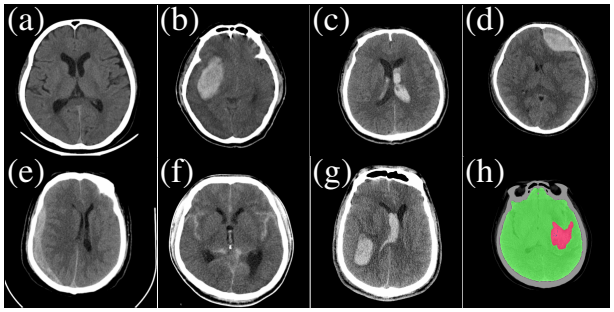

Fig. 1. Examples of head CT slices and masks. (a)-(h) Example slices of normal condition (a), CPH (b), IVH (c), EDH (d), SDH (e), SAH (f), and mixed subtypes (g). (h) An example of segmentation masks: red represents regions of hemorrhages, green for normal brain tissues and the remain for non-brain region.

five subtypes in each slice (slice-level) and in the scan as a whole (subject level). Exemplar slices of different ICH conditions are shown in Fig.1(a)-(g). The aim of the seg task is to classify each pixel in the CT image into one of three mutually exclusive classes: hemorrhage, normal brain tissue, and non-brain (including the skull and regions outside the skull), see Fig.1(h) for an example.

2.2. Proposed Multi-task FCN Architecture

The cls and seg tasks share commonalities in that both tasks require the localization of hemorrhages. To take advantage of information sharing, we propose a multi-task FCN architecture named ICHNet to jointly perform both tasks. The major components of the proposed network include an encoder for image feature extraction, a bi-directional ConvLSTM module for capturing inter-slice information, a decoder for generating segmentation masks, and a classification branch for generating class probabilities. The architecture is demonstrated in Fig. 2. The Residual Block in the architecture is the same as the two-layer residual block presented in ResNet[10]. The Conv Block consists of two 3×3 convolution layers with batch normalization and ReLU activation. For the segmentation branch, to take advantage of information from different resolution levels, we concatenate the feature maps from all four resolution levels as proposed in [11]. The final segmentation output is obtained by passing the concatenated feature maps through a 1×1 convolution block.

ConvLSTM module Head CTs are typically anisotropic and have a lower resolution along the axial dimension. To capture information embedded in consecutive slices, previous work considered using a plain RNN module and demonstrated that this approach can indeed improve classification performance[4]. However, feature maps need to be flattened before feeding into a plain RNN. The flattened feature maps are no longer useful for the segmentation task due to the lost of spatial information. To address this issue, we apply a two-layer bi-directional ConvLSTM module in our proposed architecture, whose input is a sequence of slice-level feature maps generated by the encoder. A ConvLSTM cell is similar to a plain LSTM cell except that all fully-connected operations are replaced by convolution operations; thus the spatial information is preserved. The average pooling of slice-level ConvLSTM outputs is used as the input to the classifier for subject-level binary and subtype classification.

Attention module A self-attention mechanism[12] is applied to the architecture to filter the signals encoded in the feature maps forwarded from the encoder before concatenation with the upsampled feature maps in the decoder. The attention module is useful for identifying and focusing on salient features in task-relevant regions while suppressing feature activations in less relevant regions.

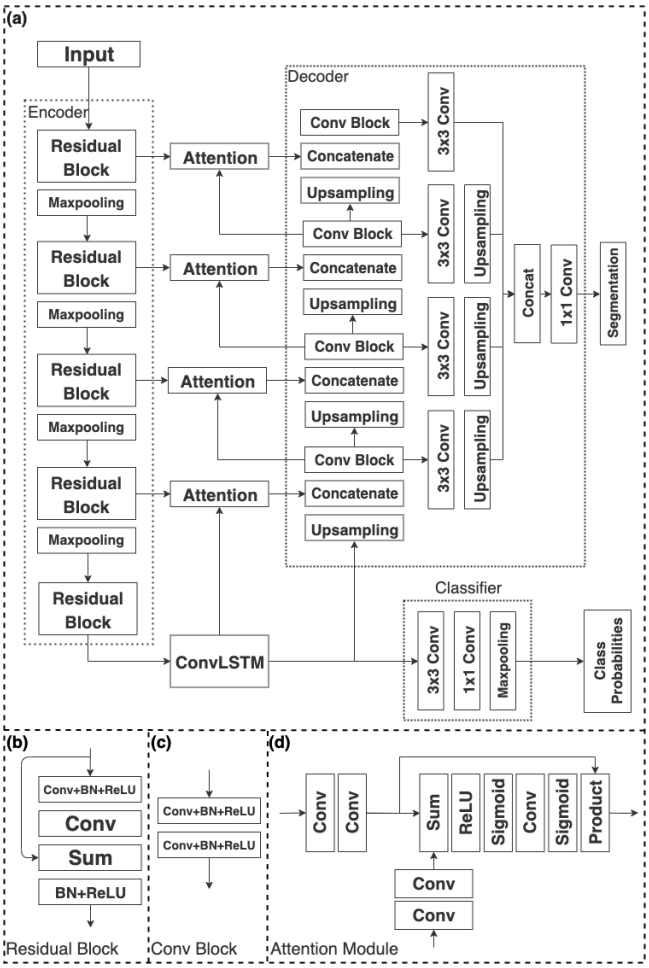

Fig. 2. ICHNet architecture. (a) The architecture consists of five components: an encoder for the generation of features, a decoder for segmentation, a classifier branch, and attention modules; (b)-(d) detailed structure of the residual block, conv block and attention module, respectively.

2.3. Loss Function

We design the loss function for our task based on two commonly used loss functions: a binary cross entropy (BCE) loss $\mathcal{L}_{\mathcal{E}}$ and a weighted multi-class soft dice loss $\mathcal{L}_{\mathcal{D}} = 1 - \frac{2\sum_{c=1}^{C} w_c \sum_{n=1}^{N} \zeta_{n,c} p_{n,c}}{\sum_{c=1}^{C} w_c \sum_{n=1}^{N} (\zeta_{n,c} + p_{n,c})}$, where $\zeta_{\cdot}$ denotes the pixel-level label, $p_{\cdot}$ denotes the model probability output, N denotes the sample size, C denotes the number of classes, and $w_{\cdot}$ is a weighting term to account for class imbalance, which is set as the inverse of the volume for each class. The set of loss functions for our cls and seg tasks include: (1) slice-level and subject-level ICH cls loss functions, $\mathcal{L}_{\mathcal{E}}^{\mathrm{sli-b}}$ and $\mathcal{L}_{\mathcal{E}}^{\mathrm{sub-b}}$, (2) slice-level and subject-level subtype cls loss functions, $\{\mathcal{L}_{\mathcal{E}}^{\mathrm{sli-m},i} \,|\, i \in \{1,\ldots,5\}\}$ and $\{\mathcal{L}_{\mathcal{E}}^{\mathrm{sub-m},i} \,|\, i \in \{1,\ldots,5\}\}$, and (3) slice-level seg loss $\mathcal{L}_{\mathcal{D}}^{\mathrm{seg}}$. A common approach for constructing a multi-task learning loss is to use a set of manually tuned weightings $\{\gamma_i\}$ to linearly combine the individual losses, $\mathcal{L}_{\mathrm{mt}} = \sum_i \gamma_i \mathcal{L}_i$. Manually tuning loss weightings can be more challenging task as the number of loss terms increase. To resolve this problem, we follow the strategy proposed in [13] to learn $\{\gamma_i\}$ by maximizing Gaussian likelihood with homoscedastic uncertainty, $\mathcal{L}_{\mathrm{mt}} \sim \sum \left(\frac{2\mathcal{L}_i}{e^{\gamma_i}} + \gamma_i\right)$.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup, Baseline Models and Evaluation Metrics

This retrospective study was approved by the ethics committees of three participating hospitals. A total of 1176 head CT scans were collected from these three hospitals, with 581 ICH patients and 595 normal subjects. Each slice in these head CT images has a size of 512×512 pixels. We divided the dataset into a random sample of 706 subjects for training, 235 for validation and 235 for testing, with details listed in Table 1.

	Subject Level			Slice Level		
	Train	Val	Test	Train	Val	Test
Total	706	235	235	21206	6707	6973
Normal	336	126	119	16349	5360	5466
ICH	370	109	116	4857	1347	1507
CPH	265	79	89	2633	674	806
IVH	127	40	34	1319	358	358
EDH	63	18	19	621	155	178
SDH	85	24	30	940	282	359
SAH	126	41	37	1275	467	464

Table 1. Subject-level and slice-level statistics for the training and testing sets.

Each slice in the head CT scans was first truncated by three different intensity ranges to amplify signals of different body parts: $[-50, 150]$ HU, $[100, 300]$ HU and $[250, 450]$ HU. These three windows were then normalized to $[0, 1]$ individually and stacked as a three-channel image. The following image augmentation procedures were applied in the training process: random shearing ($\pm 10\%$), zooming ($\pm 20\%$), rotation ($\pm 90°$), horizontal and vertical flipping, and shifting (± 20 pixels). Adam optimizer was used for parameter optimization, with a learning rate of 10^{-5}.

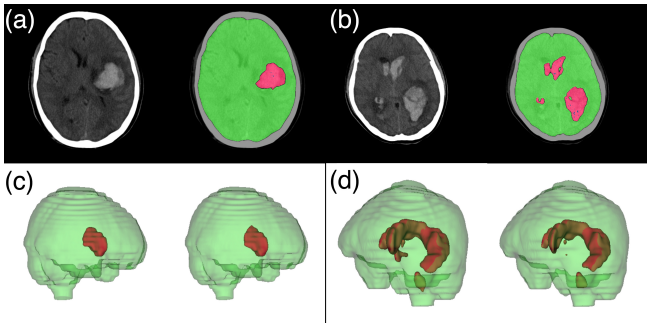

Fig. 3. Visualizations of ICHNet segmentation results. Red color denotes hemorrhages and green denotes normal brain tissues. (a)(b) 2D CT scan slices (left) and corresponding segmentation results (right). (c)(d) 3D ground truth masks (left) and segmentation results (right).

3D/2D ResNet18 [10] and 3D U-Net [14] were used as subject-level/slice-level cls task and seg task baseline models, respectively. We additionally performed ablation experiments by evaluating the performance of: (1) ICHNet for the cls tasks only without the seg branch ($\mathrm{ICHNet}_{\mathrm{cls}}$), (2) ICHNet for the seg task only without cls branch ($\mathrm{ICHNet}_{\mathrm{seg}}$), and (3) ICHNet without attention blocks ($\mathrm{ICHNet}_{\mathrm{no\_att}}$).

We used five metrics (accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC)) to evaluate models' performance for the cls tasks, and used dice coefficient as the evaluation metric for the seg task.

3.2. Results

Performance of our proposed approach as well as baseline models for cls and seg tasks is reported in Table 2 and Table 3, respectively. The proposed multi-task ICHNet generally outperforms the baseline models (3D/2D ResNet18 for cls and 3D U-Net for seg) as well as single-task ICHNet models ($\mathrm{ICHNet}_{\mathrm{cls}}$ and $\mathrm{ICHNet}_{\mathrm{seg}}$) across all metrics. It should be also noted that the single-task ICHNet model's performance is still notably better than the corresponding baseline model, indicating that ConvLSTM module can be a more efficacious approach in capturing sequential information than directly utilizing 3D convolution. For the seg task, the use of attention mechanism also brings additional performance improvement but not as substantial as adding the cls branch.

4. CONCLUSION

In this paper, we propose an end-to-end trainable multi-task fully convolution network for simultaneous classification and

			Accuracy	Sensitivity	Specificity	F1	AUC
Subject Level	ICH	ICHNet	**0.983**	**0.966**	**1.000**	**0.982**	**0.995**
		$\text{ICHNet}_{\text{cls}}$	0.949	0.956	0.941	0.949	0.980
		3D ResNet18	0.830	0.845	0.815	0.831	0.918
	Subtype	ICHNet	**0.896**	**0.779**	**0.921**	**0.728**	**0.851**
		$\text{ICHNet}_{\text{cls}}$	0.877	0.746	0.906	0.684	0.826
		3D ResNet18	0.799	0.555	0.852	0.496	0.703
Slice Level	ICH	ICHNet	**0.953**	**0.798**	**0.996**	**0.880**	**0.978**
		$\text{ICHNet}_{\text{cls}}$	0.945	0.768	0.993	0.857	0.965
		2D ResNet18	0.900	0.646	0.970	0.736	0.925
	Subtype	ICHNet	**0.929**	0.735	**0.941**	0.561	0.838
		$\text{ICHNet}_{\text{cls}}$	**0.929**	**0.737**	**0.941**	**0.563**	**0.839**
		2D ResNet18	0.904	0.540	0.928	0.412	0.734

Table 2. Comparison of algorithm performance for the classification tasks. **Bold** indicates best performance.

	Hemorrhage	Normal Brain Tissue	Non-brain Region	Multi-class
ICHNet	**0.951**	**0.983**	**0.984**	**0.956**
$\text{ICHNet}_{\text{seg}}$	0.929	0.981	0.982	0.940
$\text{ICHNet}_{\text{no\_att}}$	0.944	0.983	**0.984**	0.952
U-Net	0.919	0.958	0.960	0.919

Table 3. Comparison of algorithm performance for segmentation tasks. **Bold** indicates best performance.

segmentation of intracranial hemorrhage. Experimental results show that our architecture outperforms baseline models for both classification and segmentation tasks by a noticeable margin. In additional to performance enhancement, our proposed architecture offers more flexibility in practical use due to its compatibility with different levels of annotations. For example, even if only subject-level labels are available without pixel-level segmentation annotation, our proposed network can still be trained end-to-end while taking advantage of sequential information.

5. REFERENCES

[1] A. I. Qureshi et al., "Spontaneous intracerebral hemorrhage," *N. Engl. J. Med.*, vol. 344, no. 19, 2001.

[2] J Claude Hemphill et al., "The ICH Score: a Simple, Reliable Grading Scale for Intracerebral Hemorrhage," *Stroke*, vol. 32, 2001.

[3] J Heit, M Iv, and M Wintermark, "Imaging of Intracranial Hemorrhage," *Journal of Stroke*, vol. 19, no. 1, pp. 11–27, 2017.

[4] H Ye et al., "Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network," *European Radiology*, Apr 2019.

[5] A Majumdar et al., "Detecting intracranial hemorrhage with deep learning," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 583–587, 2018.

[6] P.D. Chang et al., "Hybrid 3d/2d convolutional neural network for hemorrhage evaluation on head ct," *AJNR*, vol. 39, no. 9, pp. 1609–1616, 2018.

[7] S Chilamkurthy et al., "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet*, vol. 392, Dec 2018.

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask R-CNN," *arXiv:1703.06870*, 2017.

[9] Jianxu Chen et al., "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," *arXiv:1609.01006*, 2016.

[10] K He, X Zhang, S Ren, and J Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385*, 2015.

[11] B Hariharan et al., "Hypercolumns for object segmentation and fine-grained localization," *arXiv:1411.5752*, 2014.

[12] O Oktay et al., "Attention u-net: Learning where to look for the pancreas," *arXiv:1804.03999*, 2018.

[13] A Kendall, Y Gal, and R Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," *arXiv:1705.07115*, 2017.

[14] O Ronneberger, P Fischer, and T Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597*, 2015.