PESA R-CNN: Perihematomal Edema Guided Scale Adaptive R-CNN for Hemorrhage Segmentation

Joonho Chang, Inchul Choi, and Minho Lee, *Member, IEEE*

***Abstract*—Intracranial hemorrhage (ICH) is a type of stroke with a high mortality rate and failing to localize even minor ICH can put a patient's life at risk. However, its patterns are diverse in shapes and sizes and, sometimes, even hard to recognize its existence. Therefore, it is challenging to accurately detect and localize diverse ICH patterns. In this article, we propose a novel Perihematomal Edema Guided Scale Adaptive R-CNN (PESA R-CNN) for accurate segmentation of various size hemorrhages with the goal of minimizing missed hemorrhage regions. In our approach, we design a Center Surround Difference U-Net (CSD U-Net) to incorporate Perihematomal Edema (PHE) for more accurate Region of Interest (RoI) generation. We trained CSD U-Net to predict PHE and hemorrhage regions as targets in a weakly supervised manner and utilized its prediction results to generate RoI. By including more informative features of PHE around hemorrhage, this enhanced RoI generation allows a model to reduce the false-negative rate. Furthermore, these expanded RoIs are aligned with the Scale Adaptive RoI Align (SARA) module based on their size to prevent the loss of fine-scale information and small hemorrhage patterns. Each scale adaptively aligned RoI is processed with the corresponding separate segmentation network of Multi-Scale Segmentation Network (MSSN), which integrates the results from each scale's segmentation network. In experiments, our model shows significant improvement on dice coefficient (0.697) and Hausdorff distance (12.918), compared to all other segmentation models. It also minimizes the number of missing small hemorrhage regions and enhances overall segmentation performance on diverse ICH patterns.**

***Index Terms*—Convolutional neural network, intracranial hemorrhage, instance segmentation, perihematomal edema, small object segmentation.**

Manuscript received 26 April 2022; revised 7 October 2022; accepted 1 November 2022. Date of publication 9 November 2022; date of current version 5 January 2023. This work was supported in part by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [22ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System] (50%) and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea grant number : HI21C1074) (50%). *(Joonho Chang and Inchul Choi contributed equally to this work.) (Corresponding author: Minho Lee.)*

Joonho Chang is with the School of Artificial Intelligence, Kyungpook National University, Daegu 41566, South Korea (e-mail: joonnozx@gmail.com).

Inchul Choi is with the NEOALI, Daegu, South Korea (e-mail: sharpic77@gmail.com).

Minho Lee is with the NEOALI, Daegu, South Korea, and also with the School of Artificial Intelligence, Kyungpook National University, Daegu 41566, South Korea (e-mail: mholee@gmail.com).

Digital Object Identifier 10.1109/JBHI.2022.3220820

I. Introduction

INTRACRANIAL hemorrhage (ICH) is a serious medical condition in which blood leaks into brain tissue because of a ruptured cerebrovascular. Its overall mortality rate is approximately 40% to 50% [1], [2], and it frequently occurs in older people. Its symptoms include severe headaches, vomiting, and paralysis, and it brings severe disabilities, such as hemiplegia, language and speech disorders, and cognitive impairment, once it occurs [3]. As a result of physical and pathological damage [4], [5], 75% of survivors are unable to return to normal life even after a year. If ICH occurs, brain tissue begins to be damaged at the time of bleeding, and it causes edema and inflammatory response to the tissue around the bleeding regions and leads to apoptosis and neurological damage, which degrades brain function and frequently results in death. Therefore, an automatic hemorrhage segmentation system is required for a fast and highly accurate diagnosis and to minimize missed hemorrhage regions [6].

Computerized Tomography (CT) scan is a widely used medical imaging technique for hemorrhage diagnosis because of its low cost and high sensitivity for detection [7]. However, scanning and analyzing procedures take a long time and require trained radiologists to analyze the obtained CT scan. It also suffers from inter-rater variability according to a radiologist due to manual delineation. To reduce such inconsistency in diagnosis, it is important to provide radiologists with reliable hemorrhage segmentation results, which can help them make a more accurate diagnosis and prevent missing any lesion areas. However, there are some challenges when segmenting hemorrhage regions in the brain, compared to other types of lesion, such as tumors. 1) Hemorrhage patterns vary in shape and size, making accurate detection difficult. 2) Additionally, numerous small hemorrhage patterns are easily overlooked but pose a significant risk to patients if missed. Moreover, 3) hemorrhage sometimes shows ambiguous patterns in CT scans due to its low contrast to the normal tissue around it. Because of these factors, small hemorrhage regions with low contrast are frequently missed and normal tissue with similar brightness to the lesion is sometimes mistakenly regarded as a hemorrhage.

To address these problems, we propose a novel method for accurate hemorrhage segmentation. Our model considers hemorrhage-induced tissue deformation, Perihematomal Edema (PHE), as an informative cue for hemorrhage regions, and designs a Center Surround Difference U-Net (CSD U-Net) to identify tissue deformed regions as candidate hemorrhage regions while learning to predict real hemorrhage. Our CSD U-Net predicts hemorrhage and tissue deformed regions and

uses them to generate region proposals. By treating both regions as Region of Interest (RoI) candidates, we obtain adaptively expanded RoIs, which aid our model in making a more accurate segmentation with a lower false-negative rate by providing more information around the hemorrhage. Also, our model addresses a fixed-size RoI alignment problem and an imbalance in RoI size distribution with Scale Adaptive RoI Align (SARA) and Multi-Scale Segmentation Network (MSSN). While conventional two-stage models [8] align various size RoIs to a single fixed-size, our model applies the SARA method to each RoI based on its scale. Because informative details are easily lost during a fixed-sized RoI aligning, especially for small-sized targets, the small hemorrhage region can also be lost with its boundary information. To prevent this problem, SARA classifies RoI into 3 groups based on its height and uses each group's representative size for alignment. Furthermore, aligned RoIs are processed with MSSN with dense connections. Each aligned RoI is provided to a corresponding segmentation network in MSSN, and all segmented results from each network are combined into a single final result. The imbalance in RoI size distribution is addressed in this approach by adopting a separate segmentation network for each scale. In experiments, our model outperforms other state-of-the-art models in terms of segmentation performance, and all of the novel components of the model are proven to be effective for accurate segmentation without missing small hemorrhages.

Our main contributions are as follows:

1) We propose a novel method for enhancing the small region detection performance of hemorrhage segmentation models. Although the performance of the state-of-the-art hemorrhage segmentation models has significantly improved, the majority of them still fail to detect small-sized hemorrhage regions. In contrast, our model can effectively detect various sizes of hemorrhage patterns and also minimize the false-negative rate of small hemorrhage regions.
2) We show that using the features of deformed tissues around the hemorrhage region, known as the PHE, can enhance the segmentation performance of a model. By using additional informative cues for hemorrhage, the proposed model's overall segmentation performance has improved. Our model trains CSD U-Net to estimate both candidate regions for hemorrhage and PHE as targets in a weakly supervised manner and provides an expanded RoI using the predicted results of CSD U-Net.
3) We minimize the loss of information in the fine scales of a feature map during RoI Align by adaptively using different feature map sizes according to each RoI's size. Instead of aligning RoI to produce a single fixed-size feature map, we classify all RoI candidates based on the size criteria and use a larger feature map for a smaller RoI to prevent information loss at the fine-scale. This SARA strategy significantly improved the segmentation accuracy, especially for small hemorrhage regions.
4) We address the imbalance problem in the number of RoIs based on RoI size in the two-stage segmentation model. ICH shows different patterns according to its region size, but the number of samples for each pattern in the given dataset is not well-balanced. Furthermore, the region proposal network (RPN) tends to generate larger RoI regardless of the target size. To address this imbalance problem, we adopt the scale adaptive RoI aligning with a separate segmentation network for each scale of RoI and combine scale-wise segmentation results at the model's final stage. These SARA and MSSN approaches enable our model to detect and localize diverse sizes of hemorrhage patterns in CT scans more accurately.

II. Related Works

A. Intracranial Hemorrhage Segmentation

Many automated segmentation models have been proposed for reliable hemorrhage segmentation and to aid radiologists and doctors in diagnosis. In machine learning-based models, researchers have used several machine learning algorithms such as decision tree, Fuzzy C-mean, random forest [9], [10], [11], and Level-Set algorithm [12] for the detection of disease. However, to apply those machine learning-based models to hemorrhage segmentation, significant domain expertise and human expert intervention are required, such as identifying and analyzing hemorrhage characteristics to reduce data complexity, and post-processing estimated results for accurate diagnosis. Furthermore, they were not robust enough to noise and scale variations, and did not have enough accuracy for practical application.

In recent years, deep learning-based models have been applied to hemorrhage segmentation and showing promising performance. Li et al. [13] used a U-Net to segment CT images, they concatenated the flipped image of the original CT slice as an input for the network to enhance the contrast between the ICH region and normal brain tissue. They also adopted adversarial training to improve the accuracy of the segmentation. Li et al. [14] modeling the hematoma expansion and variations to share the information between adjacent slices in 3D CT images. Based on the segmentation result of adjacent slices (upper, center, lower) from the U-Net, they integrated the information to the center slice and refine the final segmentation result with uncertainty estimation. Hu et al. [15] used FCN [16], which has an encoder-decoder structure called ED-Net, to address the problem that the network with more layers has lower resolution. It extracted multi-scale features to prevent information loss that comes from scale variation. Arab et al. [17] also used U-Net with a deep supervision method that integrates features from segmentation layers at different levels of the network with element-wise summation. Wang et al. [18] proposed a modified U-Net structure that adopts a multi-task semi-supervised attention-based model with a curriculum learning strategy. Their model reduced the number of layers for both the encoder and decoder by using transposed convolutions and compensated for losses from both networks. Moreover, for semi-supervised learning, they applied attention mechanism and auto-encoder structure. Compared to conventional machine learning-based algorithms, most of those models performed well, however, their segmentation accuracy on diverse size hemorrhage regions was not good enough for practical application.

B. Small Object and Lesion Detection

Small objects are difficult to detect due to their low resolution and size, and even when detected, the region details are obscured due to low contrast. Moreover, when various sizes of objects exist in an image, large objects can be the main focus of segmentation because of their rich features, making small objects segmentation worse. In an attempt to address such problems, several models have been proposed.

Mask-Refined R-CNN [19] proposed an improved version of the Mask R-CNN that accurately localizes the semantic regions at different scales. It was inspired by Feature Pyramid Network (FPN) [20] structure and combined the information from feature maps of different scales. Context union edge network (CEN) [21] used two paths, semantic stream, and edge stream, for

accurate small object and edge segmentation. The semantic stream adopted dilated ResNet-101 to extract semantic features in the global scope, and the edge stream extracted semantic edges at full resolution to preserve edge details. Then, an information exchange mechanism was used to implement each stream's shortcomings and refined the boundaries of small-scale objects.

In biomedical image applications, the lesion detection method must have high prediction accuracy for practical use because its prediction accuracy is directly related to the patient's life. However, it is not easy to obtain accurate results with conventional segmentation models, which sometimes miss small target regions. For this reason, models should be designed specifically for the medical image domain, especially when segmenting a small target, as in recent studies. ESTAN [22] was an encoder-decoder model, which adopted two encoders to extract image context information at different scales and then combines that information for accurate breast tumor segmentation. Moreover, to effectively localize small tumors, it obtained feature maps from both square and large row-column-wise kernels. Li et al. [23] extended 3D U-Net by adopting a dense connection design and a 3D multi-pooling feature fusion strategy to the encoding stage and enhancing the performance for the small volume of critical anatomical organs. Its densely connected blocks maximized the information flow for extracting low-level features, and 3D multi-pooling fusion allowed for the effective use of multi-scale and multi-level features. In infantile brains' segmentation problem of the punctate white matter lesion (PWML), Liu [24] heuristically expanded RoI to improve performance by introducing more information around the PWML. However, such a method lacked reasonable criteria for determining how much RoI should be expanded and why. In both medical and nonmedical domains, conventional models have adopted various methods to find small lesions or objects, such as reducing feature information loss through dense connection or obtaining additional features separately. However, such overheads introduced significantly more feature engineering, and even with such efforts, there is still a risk of missing small ICH regions when the target region contrast is low compared to the normal tissue around it. Furthermore, to effectively assist medical staff in their diagnosis in the actual medical field, it is highly recommended to predict all potential anomalies to minimize false negatives of prediction. For this purpose, in this article, we propose a novel two-stage hemorrhage segmentation model, PESA R-CNN. In the first stage, our CSD U-Net learns information from small and low contrast regions distinguished from normal tissue, and leverages PHE, to generate expanded region proposals for hemorrhage segmentation. Furthermore, in the second stage, our model scale adaptively aligns RoIs to prevent information loss at fine-scale and segments regions according to the given RoI size to address the imbalance imbalance of RoI size distribution. Based on these novel approaches, our model achieves state-of-the-art segmentation performance on the diverse sizes of hemorrhages compared with all other hemorrhage segmentation models.

C. Perihematomal Edema for Hemorrhage Detection

The hemorrhage causes neural cells in the vicinity of the lesion to undergo unfavorable ultrastructural changes known as PHE, which can result in neurological deterioration and dysfunction [26]. PHE represents the condition of tissues surrounding the hemorrhage region where the volume of brain tissue is increased due to abnormal increases in brain cell content [27], [28], [29]. In clinical studies, once ICH occurs, PHE starts to evolve within 6 hours of the hematoma formation and keeps growing rapidly for 1 to 3 days after the first onset [30], [31]. Its volume increases to a maximum by 7-10 days, then it gradually decreases. Statistically, PHE grows to 50% of its maximum volume within 6 hours and reaches 75% within a day [29]. Furthermore, the relation between the hemorrhage volume and the PHE volume is linear as shown in the analysis [32].

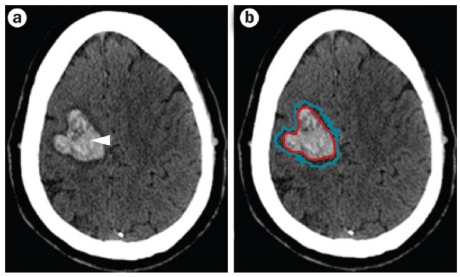

Fig. 1. Differences in tissue characteristics between PHE (purple) and ICH (red) [25].

In several studies of ICH and edema [33], [34], [35], on CT scans, hemorrhage is shown as a hyperdense region that clearly contrasts with neighboring tissue. In contrast, PHE is rather a hypodense region that is difficult to be distinguished from normal tissue [36] or other low density regions (ex. infarction) [37]. As shown in Fig. 1, the bright, highly dense region is the ICH region, and the dark area around the ICH corresponds to the PHE region [25]. Usually, PHE has the lowest tissue density than both ICH and normal tissues regions, it appears to be darker than other nearby tissue areas. However, the boundaries of PHE become less clear as it grows. Because of these properties, previous studies that attempted to measure PHE did not explicitly clarify the method for tracing PHE [38], [39], [40]. In some ICH researches [29], [34] applied midline shift (MLS) to measure PHE, but it only showed a correlation with total hemispheric swelling. Automatically measuring only PHE region is a difficult problem, however, because PHE and ICH have a high correlation, we can still leverage PHE as a useful cue for learning to detect ICH regions.

This article proposes a novel approach that leverages PHE for accurately segmenting the ICH regions from brain CT scans. Inspired by the fact that most PHE surround the ICH regions and have their own patterns, our model learns to estimate PHE regions with a weakly supervised learning method and uses the learned estimation results as RoIs to enhance the segmentation accuracy. As far as we know, no other brain hemorrhage researches have used PHE features to segment ICH.

III. PESA R-CNN

A. Model Overview

The proposed PESA R-CNN is a two-stage instance segmentation model similar to the Mask R-CNN. This model's segmentation performance is majorly enhanced by three components, CSD U-Net with pseudo PHE target, SARA, and densely connected MSSN, as shown in Fig. 2.

In the first stage of our model, there are two branches of input. In the first branch, an input CT scan is processed by CSD U-Net, which is trained in a weakly supervised manner with the pseudo PHE target labels and provides both hemorrhage regions and PHE to RPN. The RPN learns to generate RoI based on these pseudo target regions. For the weakly supervised learning of CSD U-Net, original target hemorrhage regions are expanded with dilation operations to include PHE around the hemorrhage and used as pseudo target regions for training CSD U-Net. The trained CSD U-Net prediction results are combined with original target hemorrhage regions with the 'OR' operation and used for RPN target regions. The RPN-generated expanded RoIs are classified according to their scale to form RoI groups for SARA operations and the second stage of the model.

The second input branch obtains a feature map of an input CT scan from the backbone network (ResNet-101) [41] and provides

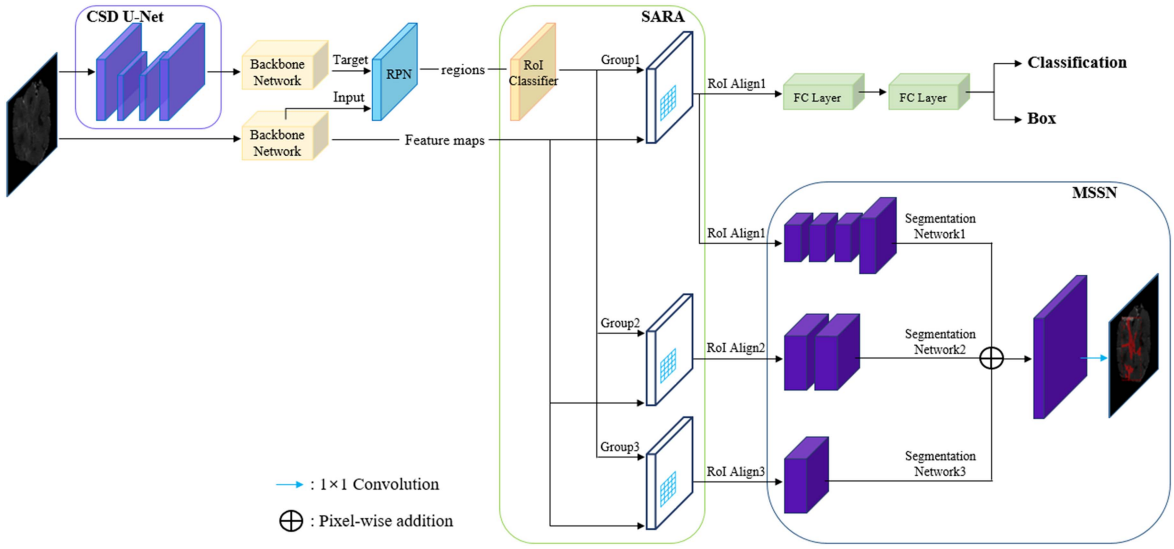

Fig. 2. The architecture of our proposed PESA R-CNN. It is mainly enhanced by CSD U-Net, SARA, and the MSSN.

it to the SARA module to assign a feature map to each RoI candidate. SARA classifies the RoI into three groups based on its scale and each group is aligned with a pre-defined feature map size larger than the original size for the small RoI region. After adaptive alignment, the aligned RoIs are provided to MSSN in the second stage. All layers in each segmentation network of MSSN are densely connected to preserve fine details of input features, and this architecture aids in reducing information loss and improving localization. The final layer of the MSSN integrates the results of all segmentation networks into a single final prediction using pixel-wise addition.

In the classification and box regression branch, the feature map of the aligned RoI is processed with Fully Connected (FC) layers, and the output feature vector is fed into the classifier to obtain the class scores. In the bounding box regression branch, the bounding box coordinate values are adjusted for each class with the FC layer as in the Mask R-CNN.

For the target loss function, we adopt a multi-task loss (L_{PESA}) of the Mask R-CNN, which combines losses from each task (classification, box regression, segmentation) and optimize them simultaneously. For the classification loss (L_{cls}), the probability distribution $p = (p_0, p_1)$ for 2 categories (hemorrhage + background) is computed for each RoI to estimate the class of the RoI (where p is Softmax output of the FC layer), and cross entropy loss is measured for the p of each RoI and the ground truth (u) to obtain the classification loss.

$$L_{cls} = -\log p_u \tag{1}$$

For the box regression loss (L_{box}), we apply the $smooth_{L_1}$ function to the target box ($v = (v_x, v_y, v_w, v_h)$) and predicted box ($t^u = (t^u_x, t^u_y, t^u_w, t^u_h)$) of each RoI, and compute the difference between the predicted box coordinates and the target box.

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if}|x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \tag{2}$$

$$L_{box} = \sum_{i \in x,y,w,h} smooth_{L_1}\left(t^u_i - v_i\right) \tag{3}$$

For the segmentation loss (L_{mask}), the mask prediction network produces a binary mask for each pixel of RoI, and binary cross entropy loss between the prediction result and the ground truth is computed to obtain L_{mask}. Finally, the multi-task loss of PESA R-CNN is represented as follows:

$$L_{PESA} = L_{cls} + L_{box} + L_{mask} \tag{4}$$

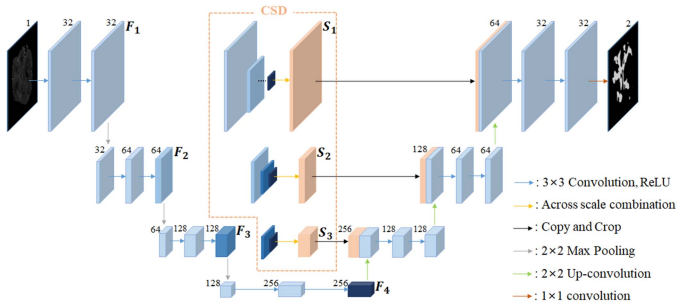

Fig. 3. The architecture of the CSD U-Net.

B. Center Surround Difference (CSD) U-Net

In clinical considerations of ICH, the hemorrhage itself can deform the brain tissue around it, or neurological deterioration frequently occurs due to nerve damage and edema. This tissue deformation, known as PHE, causes secondary damage to the patients. Based on this prior knowledge, we propose a novel CSD U-Net (Fig. 3) which enables a model to leverage PHE for hemorrhage segmentation. Although we do not have exact labels for PHE, we can adopt a weakly supervised learning approach for the coarse-grained estimation of labels.

For the weakly supervised training of PHE regions, we first obtain the initial coarse-grained pseudo labels for PHE by dilating the ground truth labels of hemorrhage. To ensure that the initial pseudo label fully covers the ground truth PHE regions, we applied the 3 times of dilation for each hemorrhage region and regarded the entire expanded region as the abnormal tissue regions, which include both hemorrhage and PHE. To refine the initial pseudo labels by learning the low-level features from the deformed tissue regions, we adopt the CSD algorithm [42] to U-Net structure and estimate the saliency of each pixel in the pseudo target regions, after training our CSD U-Net with the expanded pseudo target regions. In this weakly supervised learning process, the salient features existing in the incremented part of target regions are learned as pseudo candidate regions for PHE. We train the RPN with these refined pseudo target labels to introduce additional features from the deformed regions to the Region Proposal generation process for the segmentation of hemorrhage. Each component of CSD U-Net is illustrated in Fig. 3.

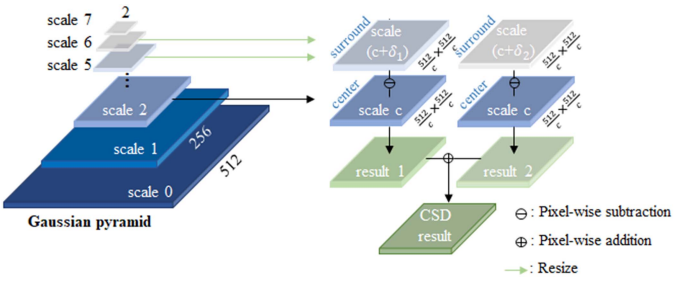

Fig. 4. Procedure of CSD algorithm.

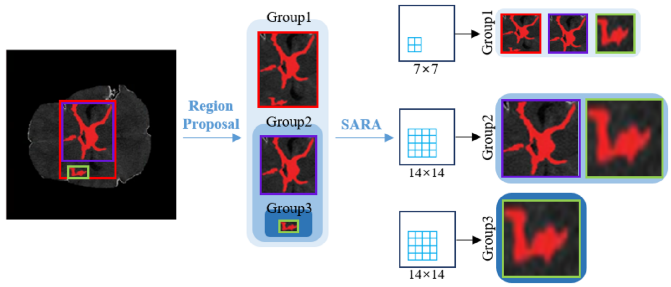

Fig. 5. Scale Adaptive RoI Align (SARA).

1) Center Surround Difference: The CSD algorithm compares 'the center' pixels in a fine-scale feature map to 'the surrounding' pixels in coarse-scale feature maps to reveal salient regions compared to their neighboring area. More specifically, it is focusing on the pixel-wise difference between fine and coarse scales of feature maps derived from Gaussian feature pyramid results. For example, as shown in Fig. 4, when we have a Gaussian feature pyramid with 8 scales, the center, $c \in \{2, 3, 4\}$, is a scale value of feature map we want to compute CSD, and the surround, s, is the scale value of feature maps coarser than the center, $s = c + \delta$, where the scale difference is set as $\delta \in \{3, 4\}$. Then, the center surround difference at $c = 2\,(I(2, 5) + I(2, 6))$ is obtained by the sum of the pixel-wise difference between the center feature map and the surround feature maps, which have resized to the same size of the center feature map.

$$CSD_c = \sum_{\delta} I(c, c+\delta) = \sum_{s=c+\delta} |I(c) \ominus I(s)| \qquad (5)$$

This process has the effect of identifying the saliency of each pixel in the feature map when it has a density difference with its neighboring points. In equation (3), $\ominus$ represents the pixel-wise difference between two feature maps.

2) U-Net: It is a well-known segmentation network for biomedical images. It is primarily made up of three structures: a contracting path (encoder), an expansive path (decoder), and skip connections. The contracting path captures the overall context of an image, and the expansive path performs accurate localization to obtain high-resolution segmentation results. This structure effectively captures contextual information from the input, but fine-scale information could easily be lost, and it reduces the overall segmentation accuracy. To prevent this problem, skip connection transfers feature details from contracting layers to expansive layers to provide better target localization.

In our CSD U-Net, we integrate the CSD algorithm into the U-Net architecture by modifying the skip connections. It applies the CSD algorithm to encoder layer feature maps and provides a CSD feature map, S_x, to each decoder layer as shown below.

$$S_x = \sum_{k=x}^{n-1} \sum_{y=x+1}^{n} |F_k \ominus F_y| = \sum_{k=x}^{n-1} CSD_k \qquad (6)$$

where n is the total number of downsampling steps, x is the step where skip connection proceeds, and F_x denotes a feature map from the last layer of x-th step in the contracting path.

With this modification, CSD U-Net can learn salient regions from an input more effectively while learning expanded target hemorrhage regions. We trained our CSD U-Net to predict both the target hemorrhage regions and all tissue deformed regions and use these prediction results as targets for RPN. While learning from these PHE included regions, RPN adaptively expands the regions for RoI candidates, and these expanded RoIs help minimize false negatives of hemorrhage regions, especially for small ones, by using more information around the hemorrhage.

C. Scale Adaptive RoI Align (SARA)

Regardless of input RoI size, every RoI in the original Mask R-CNN is aligned to only one specific size (usually 7×7) feature map. However, aligning diverse size RoIs into a small fixed-size feature map inevitably introduces information loss at fine-scale features of an image. This fine-scale feature loss affects various size RoIs, but it has more negative effects on the small-size RoI, because, compared to the large-size RoI where coarse-scale features are dominant, the small-size RoIs are mostly composed of fine-scale features. Especially for CT images of hemorrhage, losing small target regions or details of hemorrhage regions can severely degrade the performance.

To address this problem, we introduce the SARA approach for the RoI Align process. In SARA, all RoIs are clustered in 3 groups based on their size and adaptively aligned with a feature map with pre-defined sizes to minimize information loss. To begin the clustering of RoI, we sort the ground truth target boxes by height and obtain the criteria for classifying hemorrhage regions by size. For the University Hospital (UH) dataset, we divide the sorted boxes into 3 parts so that each partition has the same number of boxes, and then get the reference size. Based on the size criteria, 3 RoI groups are constructed and a larger group can be a superset of smaller ones. As shown in Fig. 5, if G_1, G_2, and G_3 are RoI groups, they satisfy $G_1 \supset G_2 \supset G_3$. G_1 includes all generated RoIs, and G_2 is composed of RoIs with a smaller height than the pre-defined medium-size (128) and a subset of G_1. Similarly, G_3 is composed of RoIs with a smaller height than 64, and it is a subset of G_2. For each of these groups, we apply a different feature map size for RoI aligning. The entire set RoI group, G_1 is aligned with a 7×7 feature map, and medium and small-sized RoI groups, G_2 and G_3, are aligned with a 14×14 feature map. By aligning medium and small-size RoI to the larger feature map than the conventional aligning size, we keep fine-scale features of small hemorrhage regions and preserve the boundary details of hemorrhage. Furthermore, the way we constructed RoI groups (larger group includes smaller one) enables the larger group to focus on scale across features, while smaller RoI groups can keep more fine-scale information. Since we adaptively align RoI with SARA, we also apply a separate segmentation network for each scale feature map in the following section.

D. Multi-Scale Segmentation Network (MSSN)

In medical image segmentation, small-sized target regions have less feature information and have a smaller number of instances than other large regions in the dataset. Furthermore, in many hemorrhage datasets, most target regions are either a large region or the collection of several hemorrhage regions, which produce large RoIs from RPN. Moreover, RPN tends to produce a larger number of large-size RoIs than smaller ones, because the larger bounding box is less sensitive to the target deviation error in terms of the Intersection over Union (IoU) measure [43].

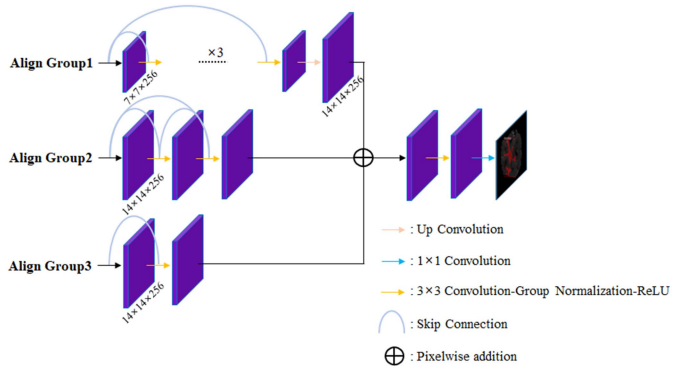

Fig. 6. Multi-Scale Segmentation Network (MSSN).

Therefore, this imbalance problem in RoI distribution degrades the segmentation performance in small-sized target regions.

To address this problem, we design a densely connected MSSN (Fig. 6). In contrast to the original Mask R-CNN, which uses only a single segmentation network for all RoI sizes, we apply a separate segmentation network for each scale's RoI group. We also adopt a different number of network layers in each segmentation network branch based on the size of the input RoI. For the deepest segmentation network, the whole RoI group G_1 is provided as an input, similar to the original Mask R-CNN segmentation network, and each of the medium-sized RoI group G_2 and the small-sized RoI group G_3 is assigned to the MSSN's medium and shallow depth networks. In the deepest segmentation network branch, the model learns contextual patterns of hemorrhage across all scales, and other shallower segmentation network branches more focus on the fine-scale features of smaller hemorrhages that can be easily missed in the deeper network. For the small-size RoI segmentation, we adopt a shallow segmentation network with two layers to preserve the pixel-wise information and obtain accurate target boundary information with rich details. The RoI group of medium-size hemorrhage is fed into a three-layer segmentation network to obtain context and fine details of the medium-size hemorrhage regions. Additionally, every layer in the segmentation networks is densely connected [44] in MSSN to strengthen the feature propagation and reuse in all segmentation networks. In the last stage of MSSN, the final segmentation result is produced by combining the results from each segmentation network with pixel-wise addition. In MSSN, an imbalance between large and small RoI groups is effectively relieved by adopting a separate segmentation network for different RoI groups and merging them with equal weights on every scale. Furthermore, the multi-scale segmentation approach enables learning finer-scale features of hemorrhage and minimizing the missing hemorrhage regions. The following experiment section shows the effectiveness of each novel component of our model and overall improved performance.

IV. Experiments

A. Dataset

a) UH dataset: The University Hospital (UH) dataset consists of 223 CT scans with multiple types of hemorrhage such as intraventricular hemorrhage, subarachnoid hemorrhage, and subdural hemorrhage, that were collected at Kyungpook National University Hospital (KNUH) between July 2014 and May 2018. The hemorrhage masks are manually labeled by radiologists using MITK Workbench (Version Nov. 2016) and confirmed by experts on brain hemorrhage. The dimensions of the CT scan is $512 \times 512 \times 30$, and the voxel size is $2 \times 2 \times 5$ mm^3. Among the 223 collected data, 176 CT slices are randomly selected for training, 20 CT slices for validation, and 27 CT slices for evaluation. This dataset was approved by the KNUH institutional review board on July 11, 2018.

b) CT-ICH physionet (PHY) dataset: The Physionet Dataset (PHY) [45], [46] collected 82 CT scans of patients with hemorrhage. Each CT scan includes about 30 slices with a slice-thickness of 5 mm and a size of 512×512. Two radiologists recorded hemorrhage types and delineated the ICH regions. It has 75 subjects in NIfTI format. From the dataset, 184 CT slices are randomly selected for training, 20 CT slices for validation, and 21 CT slices for evaluation.

c) Data pre-processing: We slice the 3D data into 2D slices, based on the manual and the code provided. For all CT slices the cranium is delineated by simple intensity-based thresholding, and only brain regions are extracted. Because datasets consist of CT slices for both ICH and normal brain, we do not manually choose the ICH layers, and all CT slices in the datasets are used for model training.

d) Target region size-wise data grouping: For the evaluation of model performance according to the hemorrhage region size, we divided the target dataset into three groups based on the height of the hemorrhage region in each dataset. The parameters for data grouping are determined by the experimental evaluation results and clinical studies [47], [48]. For the UH dataset, hemorrhage regions with a height greater than 128 are classified as a large-sized group, and regions smaller than 64 are assigned to a small-sized group. Hemorrhage regions with the height in-between the two criteria are classified as a medium-sized group. For the PHY dataset, the division boundary for the large-sized group is set to 64, and for the small-sized group is set to 32. The rest of the hemorrhages are assigned to a medium-sized group. We use the same criteria for classifying RoI groups before the SARA module in PESA R-CNN, except that the large-sized group should include all smaller RoI groups.

B. Training Details

To generate the pseudo labels from both hemorrhage datasets, we apply the dilation operation with a 3×3 kernel whose values are all set to 1, and repeat it 3 times. In the CSD U-Net, the learning rate, decay rate, drop out, and epochs are set to 0.6×10^{-6}, 0.95, 0.75, and 30, respectively. We adopt the Adam optimizer and apply 3 down sampling steps to the CSD U-Net. For the PESA R-CNN, the learning rate and the decay rate are set to 0.65×10^{-3} and 0.99, and we trained the model for 250 epochs. The batch size is set to 1 for both PESA R-CNN and CSD U-Net. For the backbone networks, we use ResNet-101. For the training of RPN, the RPN anchor scale is set to [32, 64, 128, 256, 512], which represents the height of a square anchor in pixels, and the anchor ratio is set to [0.5, 1, 2]. In addition, only 100 RoIs (positive:negative = 1:2) per image are used for training. For data augmentation, Affine transform, Gaussian blur, and Sharpening are applied. The whole model is implemented with Keras and Tensorflow library, and all experiments are performed on a desktop machine with a single NVIDIA Quadro GV100 GPU (32 GB).

C. Performance Evaluation Metrics

We evaluate the proposed model's performance using the Dice Similarity Coefficient and Hausdorff Distance, which are widely used metrics for evaluating segmentation performance.

a) Dice similarity coefficient (DSC): The Dice Similarity Coefficient measures the similarity by calculating the overlap between the prediction result and the ground truth label.

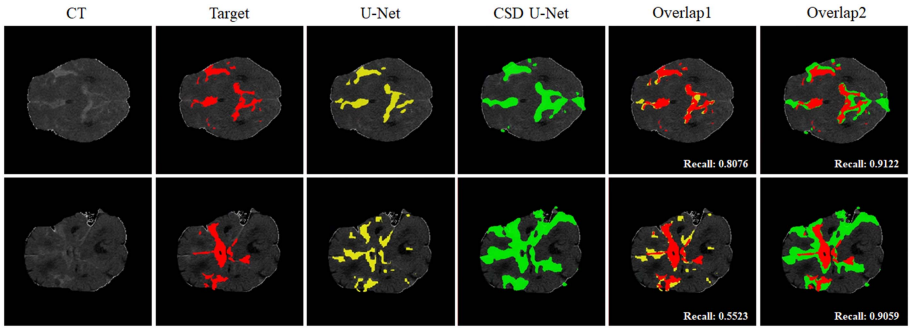

Fig. 7. Comparisons between U-Net and CSD U-Net. The 5^{th} column shows overlapped results of the target and the U-Net prediction, and the 6^{th} column shows overlapped results of the target and the CSD U-Net prediction.

The DSC is formulated as follows:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{7}$$

where P indicates the prediction result and G indicates the ground truth label. DSC ranges from 0 to 1. 1 means the labeled area was exactly the same as the predicted area and 0 means there were not overlapped.

b) Hausdorff distance (HD): The Hausdorff distance is a metric for measuring the shape difference. It selects the closest point among the ground truth label from the predicted result and calculates the distance by finding the farthest point. The HD is formulated as below:

$$HD = \max\left\{\sup_{p\in P} d(p, G), \sup_{g\in G} d(P, g)\right\} \tag{8}$$

where P, p indicates the prediction result and G, g indicates the ground truth label. Also, sup represents the supremum and $d(p,G)$ quantifies the distance from point p to G. Smaller distance values were defined as having a shape similar to the ground truth label and predicted result.

D. PHE and Hemorrhage Estimation With CSD U-Net

The purpose of CSD U-Net is to provide an approximate estimation of target labels for both PHE and the potential hemorrhage regions to enhance the overall performance. Because we do not have ground truth labels for PHE, we adopt weakly supervised learning approaches [49], [50] to estimate pseudo labels for PHE and hemorrhage regions. In the conventional pseudo label-based weakly supervised instance segmentation approaches [51], [52], pseudo labels are generated with two steps: first, localize the target instances with coarse-grained pseudo labels that are consistent with prior assumption or prediction results of pre-trained segmentation model, then propagate or refine the initial guess of pseudo labels based on pixel-wise property or a saliency estimation for each pixel. Inspired by this process, we apply a dilation operation to the ground truth hemorrhage targets to obtain the initial coarse-grained pseudo labels for PHE and hemorrhage regions. Based on the clinical definition of PHE, we only consider the tissue surrounding hemorrhage regions as potential candidates of PHE, and leverage dilated hemorrhage target labels to train the CSD U-Net to refine the pseudo labels of non-normal tissue regions (hemorrhage and PHE). To ensure initially guessed pseudo labels are fully covering the ground truth PHE regions, we apply 3×3 kernel dilation operation to the hemorrhage target for 3 times, and the total amount of target expansion is decided by the clinical studies on PHE. CSD U-Net is designed to focus on salient regions that contrast with its neighboring regions, and this property helps CSD U-Net effectively learn tissue deformed regions that are even slightly different from normal tissue regions. Although it is difficult to estimate PHE directly, CSD U-Net learns to predict the pseudo labels for both PHE and hemorrhage regions based on our weakly supervised approach and CSD algorithm. As shown in Fig. 7, CSD U-Net predicts slightly expanded regions surrounding the hemorrhage target labels, and these regions are different from merely dilated target areas. From initially expanded target regions, CSD U-Net learns to predict salient regions around hemorrhage, which may be a potential candidate for PHE. Furthermore, CSD U-Net predicts regions with more target hemorrhage than baseline U-Net results. As a result, CSD U-Net has a better chance of detecting target hemorrhage regions that were previously missed. Although the prediction results of CSD U-Net do not perfectly correspond to the ground truth hemorrhage and PHE regions, it is good enough to provide more accurate RoI regions to the RPN in the two-stage model. To support our argument, we visualize the feature discriminability of every class in Fig. 8, and also show the effectiveness of our learning-based RoI generation approach by comparing it to the simple heuristic RoI expansion scheme [24].

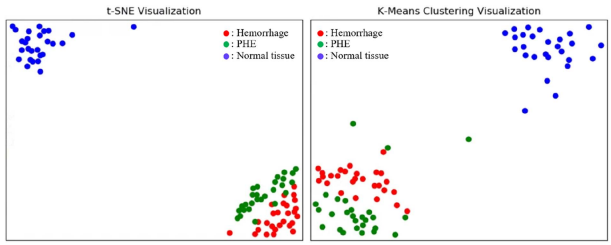

Fig. 8. Feature clustering with t-SNE and K-means algorithm.

1) Class-Separability Between Normal Tissue and Lesion: Fig. 8 shows the clustering results of features from each region (hemorrhage, PHE, and normal tissue). We classify each point of CT into one of three class labels based on the CSD U-Net prediction results after training it with pseudo target labels. The feature vector of each point is obtained from the feature pyramid of trained CSD U-Net, and every feature is clustered with t-SNE and K-means algorithms to show how much the learned lesion areas are distinguishable from normal tissue. The data used for Fig. 8 are 27 unseen samples from the hold-out test set of the UH dataset. As clearly shown in Fig. 8, feature points from hemorrhage and PHE are clustered as a single group, and their features are densely gathered to a cluster center. Also, the inter-class distance between the cluster of normal tissue (blue points) and the cluster of deformed tissue (green and red points) is far enough for clear class separability. This result shows that CSD U-Net successfully learns the common feature characteristics for both PHE and hemorrhage regions, and they clearly distinguish themselves from normal tissue regions. Furthermore, because the PHE and hemorrhage consist of almost a single group,

TABLE I
SEGMENTATION PERFORMANCE OF HEURISTIC EXPANSION ALGORITHM AND CSD U-NET

	Refined Segmentation R-CNN										CSD
k	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	
DSC	0.593 ±0.018	0.587 ±0.02	0.595 ±0.017	0.602 ±0.014	0.613 ±0.015	0.608 ±0.017	0.614 ±0.013	0.609 ±0.014	0.596 ±0.015	0.592 ±0.018	0.575 ±0.023	0.625 ±0.013
HD	30.891 ±0.824	30.989 ±0.771	32.127 ±0.808	30.588 ±0.701	28.166 ±0.624	28.862 ±0.75	27.311 ±0.727	29.022 ±0.653	31.024 ±0.849	32.426 ±0.746	34.663 ±0.812	21.826 ±0.703
Avg area of RoI	60319	62579	62987	63654	65745	66307	66848	67103	67580	67797	68111	58933

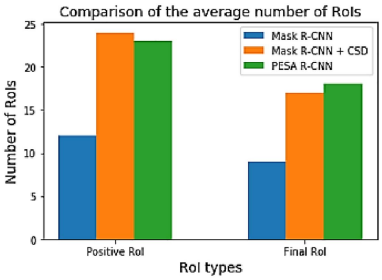

Fig. 9. Comparison of the average number of final positive RoI and RoI over threshold.

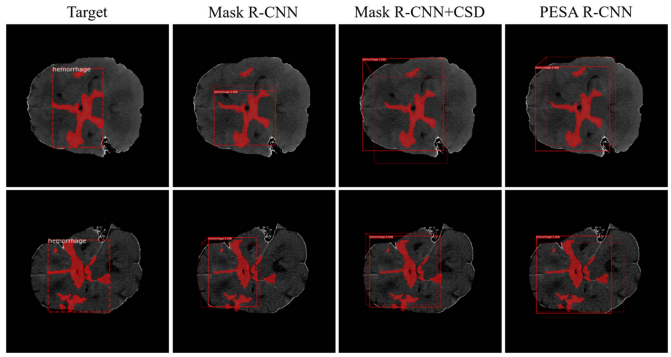

Fig. 10. RoI examples from RPN of baseline and CSD U-Net applied model.

the features from the PHE surrounding the hemorrhage can be used to improve the segmentation performance. The features from PHE not only enrich features for hemorrhage detection but also enlarge small hemorrhage regions, reducing missed hemorrhage regions. The hemorrhage patterns with low contrast or hypodense PHE regions are easily ignored and not included as RoI candidates in the conventional hemorrhage segmentation models because of their ambiguous patterns. However, CSD U-Net finds features for such ambiguous regions after learning from both PHE and hemorrhage and provides expanded RoI with more accurate localization to improve performance.

2) Learned RoI Expansion vs Heuristic RoI Expansion: We compare the segmentation performance of CSD U-Net-based Mask R-CNN with Refined Segmentation R-CNN (RS R-CNN) [24], which uses a heuristic RoI expansion algorithm to demonstrate the effectiveness of our learned RoI expansion approach. In our evaluation, we only modify the RoI generation part of Mask R-CNN to use either a pre-trained CSD U-Net or heuristic algorithm and compare the performance of two different versions of the model while linearly increasing the expansion control factor, k, of the heuristic algorithm. Table I shows the performance of the RS R-CNN as k is increased linearly from 1.0 to 2.0. According to the results, the heuristic expansion of RoI slightly improves the performance however, it is not consistently enhanced with increasing k and sometimes even worse than the baseline performance. In contrast, our CSD U-Net-based expansion method shows the best performance compared with all the cases of k, even though its average expanded RoI area far less than the expanded area of the heuristic algorithm. Therefore, our expansion method adaptively includes more meaningful areas, which correspond to PHE, and shows improved performance compared to the heuristic algorithm, which includes unrelated normal tissue regions in RoI.

3) RoI Localization Accuracy: The learned RoI expansion method provides more accurate RoI positions for segmentation by introducing feature regions of PHE around the target hemorrhage. The top-K RoI candidates in the baseline Mask R-CNN are chosen based on the objectness score and Non-Maximum Suppression (NMS). For the best prediction results, it is preferable to have more positive RoI candidates with higher objectness scores from RPN. As shown in Fig. 9, our CSD U-Net-based RoI expansion method provides far more positive RoI candidates compared to the baseline model, and even after final RoI selection, our approach has twice as many RoI candidates as the baseline model. RoI examples are shown in Fig. 10, with its targeted region. As a result, one of the best candidate RoI of Mask R-CNN does not cover all target hemorrhages and misses a small hemorrhage region. However, our CSD U-Net-based approach provides more accurate RoI localization and bounding box size, which includes every target region. When comparing the CSD U-Net applied Mask R-CNN case with PESA R-CNN, PESA R-CNN shows slightly better localization because SARA with MSSN gives more weights on small target regions during training.

TABLE II
COMPARISON OF SEGMENTATION PERFORMANCE OF THE PROPOSED APPROACH AND OTHER METHODS & ABLATION STUDY

model	UH		PHY	
	DSC	HD	DSC	HD
FCN8	0.591±0.022	44.792±1.113	0.482±0.025	106.629±1.59
FCN32	0.56±0.027	39.168±0.902	0.225±0.029	146.044±1.694
PSPNet [1]	0.491±0.025	139.52±2.807	0.257±0.019	173.934±3.337
SegNet [2]	0.594±0.035	44.259±1.108	0.469±0.042	134.34±3.35
U-Net	0.577±0.032	89.799±3.004	0.582±0.039	64.395±3.922
nnU-Net [3]	0.648±0.011	62.071±1.301	0.62±0.016	55.934±1.761
TransUNet [4]	0.627±0.021	51.388±0.53	0.61±0.022	54.191±0.981
Swin-Unet [5]	0.634±0.021	44.735±0.428	0.616±0.019	53.989±0.779
U-Net++ [6]	0.638±0.014	22.16±0.343	0.614±0.018	51.185±0.584
RS R-CNN	0.613±0.015	28.166±0.624	0.592±0.019	42.334±0.885
Mask R-CNN	0.593±0.017	31.583±0.867	0.601±0.025	21.229±1.09
M + C	0.625±0.013	21.826±0.703	0.626±0.017	18.798±0.981
M + C + SM1	0.647±0.011	22.582±0.651	0.645±0.01	18.525±0.754
M + C + SM2	0.653±0.012	19.843±0.225	0.653±0.008	17.13±0.329
M + SM3	0.644±0.015	18.895±0.194	0.626±0.015	17.631±0.25
PESA R-CNN	**0.691**±0.006	**13.027**±0.109	**0.665**±0.006	**15.225**±0.154

*M: Mask R-CNN, C: CSD U-Net, SM1: SARA + MSSN w/o dense connection SM2: SARA + MSSN with residual connection, SM3: SARA + MSSN with dense connection

E. Segmentation Performance of PESA R-CNN

1) Ablation Study on PESA R-CNN: We evaluate the contribution of each component of PESA R-CNN to the overall

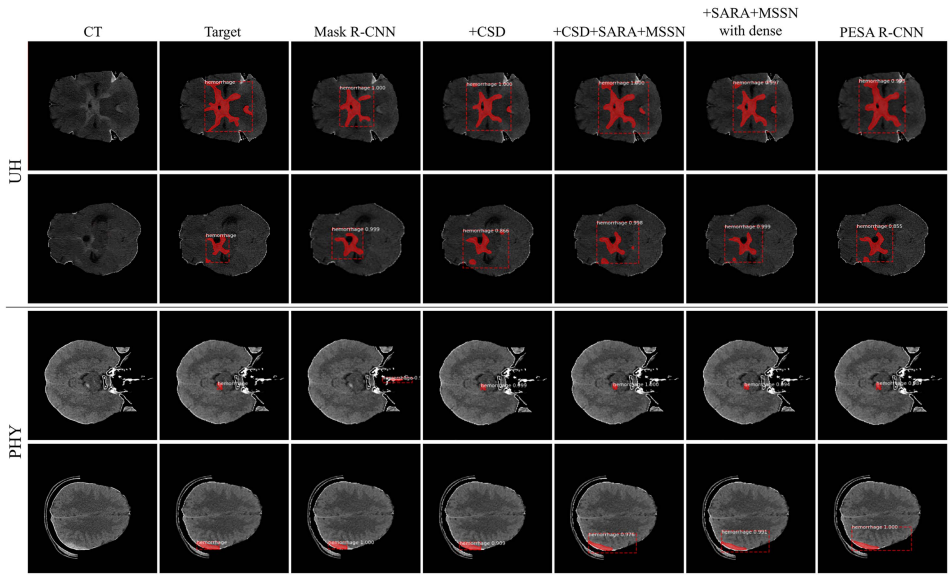

Fig. 11. Segmentation examples from ablation study of PESA R-CNN.

segmentation performance of the model. Table II shows an ablation study of the proposed model. As shown in the results, each of our novel components, provides significant performance enhancement to the baseline Mask R-CNN. The baseline shows relatively good HD contrast to its low DSC, because it has fewer false positives than false negatives for hemorrhage regions. By adding CSD to the baseline, there are significant improvements in DSC and HD because of the enhanced RoI localization. For the case of CSD + SARA and MSSN without dense connection, SARA and MSSN help to find more details in the small hemorrhage regions. Moreover, the dense connection in MSSN is especially effective for improving overall performance when it is incorporated with other components. In Fig. 11, we also show segmentation results of the ablation study. From the results in row 1 and 4, Mask R-CNN failed to include a small hemorrhage region in its predicted bounding box. In contrast, our CSD U-Net adopted model is able to detect more small hemorrhage regions, and adding SARA + MSSN provides more details of the predicted hemorrhage boundary. For the images from the UH dataset, CSD U-Net applied model detects smaller hemorrhages in which Mask R-CNN was missed, and SARA + MSSN with dense connection affect refining the segmentation result. Also, for the images from the PHY dataset, both CSD U-Net and SARA + MSSN enabled a model to localize very small hemorrhage regions of the target correctly.

2) Experiments on RoI Group Configurations: We compare the performance of the PESA R-CNN on the different RoI group configurations. First, we sort the target bounding boxes of hemorrhage based on the height of the box, then select the n ($n = 2, 3, 4$) so that we can almost equally divide the whole dataset into n groups. For example, when $n = 2$, the data is divided into two equal-sized groups based on the RoI height of 100 for the UH, 45 for the PHY dataset.

As shown in Table III, the performance of 2 group is worse than 3 group. The performance of 4 group is slightly better than 3 group, but, in the experiment, it takes too much time for training compared to 3 group. Therefore, for the practical performance, the 3 group configuration is used for all experiments in our article.

3) Comparison of RoI Size Distribution: In most fields of research [53], [54], [55], [56], various imbalance problems cause

TABLE III
PERFORMANCE COMPARISON ACCORDING TO DATA GROUPING

# of groups	UH		PHY	
	DSC	HD	DSC	HD
2 group	0.667±0.011	14.309±0.137	0.645±0.013	16.425±0.196
3 group	0.691±0.006	13.027±0.109	0.665±0.006	15.225±0.154
4 group	0.695±0.008	12.463±0.12	0.666±0.009	14.762±0.142

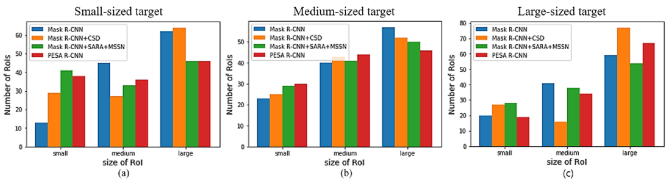

Fig. 12. Average of RoI size distribution.

performance degradation. In the two-stage segmentation model, obtaining correctly localized and scaled RoI is crucial for overall performance because the segmentation network of the two-stage model learns only from RPN-provided RoIs. Therefore, we expect RPN to generate more RoIs that are close to the target in size and location. However, in the original Mask R-CNN, the number of large-size RoIs was typically far greater than the number of small-size RoIs, and this imbalance in RoI size distribution degrades the model's segmentation performance on small or medium-size hemorrhages. To demonstrate how much each model component influences the number of RoI generated from RPN, we compared the average number of RoI for each size group for the different combinations of our model components, just as we did in the ablation study. As shown in Fig. 12, RPN in the Mask R-CNN always produces larger RoIs than smaller ones, regardless of the target size. However, if we apply our novel model components, the imbalance is alleviated according to the target region size. For the small-sized target case, CSD U-Net has the effect of increasing the number of small-size RoI by adaptively expanding the small-size RoI candidates. For SARA + MSSN, it increases the small-size RoI and decreases the number of large-size RoI from distribution. By adding dense connections to MSSN, PESA R-CNN has the additional effect of fine-tuning the RoI distribution for optimal performance. For

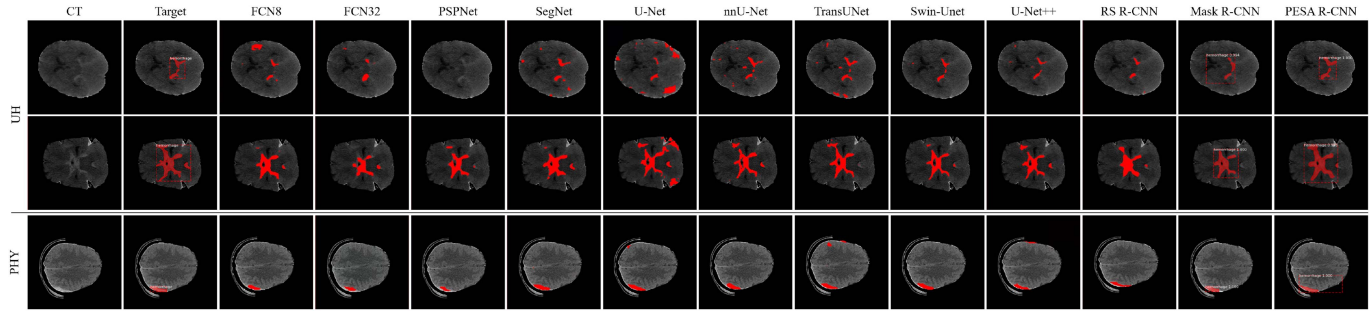

Fig. 13. Segmentation examples of diverse size hemorrhages obtained from PESA R-CNN and other models.

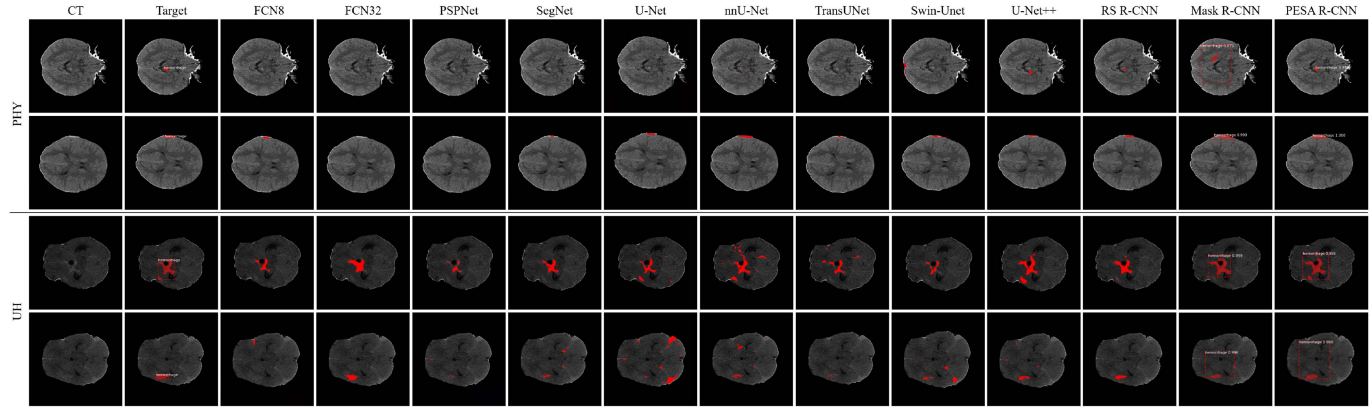

Fig. 14. Segmentation examples of small-sized hemorrhages obtained from PESA R-CNN and other models.

the medium-sized target, the overall RoI distribution is adjusted so that the number of medium-size RoI is increased while the number of large-size RoI is decreased. In the large-sized target case, CSD U-Net increases both small and large-size RoI, but as more components are added, overall RoI distribution is adjusted so that there are more large-size RoIs than the original Mask R-CNN case.

4) Comparison With Other Segmentation Methods: We evaluate the segmentation performance of the proposed model and other state-of-the-art segmentation models in Table II. The UH dataset consists of diverse sizes of ICH, and the PHY dataset consists of many small-size ICH. From this result, FCN8 and FCN32 show relatively better HD scores than U-Net, which means they have low false-positive rates. PSPNet and U-Net have a high HD because of their high false positives. Compared to other U-Net variants, U-Net++ shows a large improvement in HD because adopting the dense connection preserved more fine-scale information. The R-CNN-based models show lower HD than other U-Net-based models. The overall results show that our PESA R-CNN outperforms all other models on both datasets. Fig. 13 shows examples of segmented CT images from all evaluated models. In the first row, the prediction results of FCN8, FCN32, and SegNet are inaccurate and even regard normal tissues as hemorrhage (high false-positive rate). U-Net++ predicts regions close to a target shape, but still shows false positives. The results of TransUNet and Swin-Unet show much worse performance than some CNN-based models, because our dataset is not large enough for the training of Transformer-based networks [63]. For the PHY dataset, most other models show incorrect localization and some false positives. In comparison to these models, PESA R-CNN accurately segments every target region for both large and small cases and does not miss small ICHs without introducing false positives.

TABLE IV
COMPARISON OF SMALL-SIZED HEMORRHAGE SEGMENTATION PERFORMANCE OF PROPOSED APPROACH AND OTHER METHODS

model	UH		PHY	
	DSC	HD	DSC	HD
FCN8	0.588±0.024	48.965±1.207	0.538±0.029	96.299±1.161
FCN32	0.544±0.03	46.277±1.314	0.399±0.032	142.257±1.793
PSPNet	0.475±0.028	85.79±2.11	0.424±0.031	287.685±3.402
SegNet	0.577±0.041	47.536±1.215	0.443±0.053	140.814±3.518
U-Net	0.596±0.034	74.976±3.485	0.575±0.045	94.982±4.258
nnU-Net	0.637±0.012	65.912±1.713	0.61±0.018	63.611±2.07
TransUNet	0.628±0.02	56.952±0.642	0.602±0.023	58.919±1.135
Swin-Unet	0.629±0.017	44.884±0.577	0.609±0.019	56.743±0.874
U-Net++	0.632±0.015	26.545±0.391	0.613±0.02	52.855±0.742
RS R-CNN	0.601±0.019	32.769±0.836	0.606±0.023	47.214±1.089
Mask R-CNN	0.583±0.021	25.957±1.074	0.591±0.033	35.827±1.319
PESA R-CNN	**0.729**±0.008	**7.486**±0.121	**0.653**±0.014	**18.761**±0.221

5) Segmentation Performance on Small-Sized Hemorrhage Regions: We evaluate all models for the pre-defined small-sized hemorrhage group to show the proposed model's effectiveness on small-sized hemorrhage regions. Table IV shows the performance of various models on the small-sized hemorrhage groups. For the UH dataset, most other models show slightly worse or similar performance than the entire hemorrhage group results which represents their weakness on the small-sized hemorrhage segmentation problem. Even for Swin-Unet and TransUNet, their attention mechanisms show no benefit for the small-sized hemorrhage segmentation. Also, most U-Net-based models show worse performance on the small-sized hemorrhage segmentation. However, our model significantly outperforms all other models in both DSC and HD. Moreover, it is even better than the results of using the entire UH dataset. For the PHY dataset, many models miss small ICH and show severely poor

performance in HD. However, PESA R-CNN shows better performance than all other models with a large margin, especially for HD, demonstrating its major performance enhancement in small hemorrhage segmentation. Fig. 14 shows examples of small-sized hemorrhage segmentation results from PESA R-CNN and other state-of-the-art models. Most other models fail to detect the very small and low contrast target in the first row. Even some models, Mask R-CNN and U-Net++, for example, detect unrelated normal brain regions as hemorrhage. In contrast, the proposed model detects and segments target hemorrhage regions without error. FCN32 and PSPNet fail to find any hemorrhage in the second row. SegNet and U-Net++ detect only a part of the target region. However, PESA R-CNN correctly localizes and segments the target region. For the example of the UH dataset, the target region in the third row consists of two small hemorrhages, and among them, a relatively large hemorrhage region is segmented by all evaluated models, however, FCN8, FCN32, and PSPNet predict a large number of false positives and false negatives, and their predicted hemorrhage region boundaries differ significantly from the targets. Furthermore, most other models fail to segment a smaller hemorrhage region. The target of the fourth row also has very low contrast and a small-size region. Therefore, most of the other models can't segment any hemorrhage regions and predict normal tissue as hemorrhage. However, in contrast to those results, PESA R-CNN correctly detects and segments the target hemorrhage regions without any missing small regions. Also, the boundary and shape of the predicted region are closest to the target region. Especially for small-size hemorrhages, it shows the most accurate result without introducing many false positives or false negatives. From these experiments, we demonstrate the strength of our PESA R-CNN on segmenting diverse size hemorrhages and also show that PESA R-CNN can learn to discriminate low contrast small hemorrhage regions from normal tissue regions, which previously failed in many models.

V. Discussion

In this study, we experimentally demonstrate that introducing PHE-induced features significantly improves the overall performance. Additionally, we demonstrate that scale adaptively aligning RoI and applying MSSN are highly effective for diverse size hemorrhage segmentation, especially for small-size ICH. In numerous pre-clinical and clinical studies, it has been demonstrated that PHE is a promising surrogate indicator of secondary brain injury after ICH. Despite its importance, directly measuring PHE is difficult due to its low clarity on CT scans. Therefore, instead of aiming to segment PHE regions accurately, we approximately estimate potential candidate regions of PHE and use features from those regions to clarify the RoI of ICH. Although this approach does not accurately segment PHE, its hypodense tissue pattern can provide an additional informative cue for identifying ICH.

We introduce CSD to U-Net architecture to emphasize the salient regions that contrast with neighboring regions, and devise CSD U-Net. We train it to learn patterns from both ICH and PHE in a weakly supervised manner and use its prediction results for obtaining expanded RoI. The experiments clearly demonstrated that the PHE is a useful cue for accurately segmenting small and low contrast hemorrhage regions. During this research, we discovered that target boxes in the hemorrhage dataset mainly focused on the detection of ICH in a CT scan. Thus, even if there are several small-size hemorrhages in a CT scan, they do not tend to be labeled as a separate bounding box. This type of labeling convention can cause small-size regions to be easily lost during region proposal or the RoI Align process. We will conduct additional research to determine how much the labeling scheme affects the accuracy of hemorrhage segmentation.

VI. Conclusion

In this article, we propose PESA R-CNN, a novel, highly accurate hemorrhage segmentation model. In our model, we devise a CSD U-Net to learn the PHE region around the hemorrhage and use its predicted regions for expanding RoI. This expansion method provides extra feature regions for enhancing the localization of small-sized hemorrhages. Furthermore, to prevent missing small hemorrhage regions or boundary details during the segmentation process, we scale adaptively align RoI candidates with SARA, and adopt MSSN according to each aligned RoI size. In our experiments, each novel component of PESA R-CNN effectively enhanced the overall performance of the baseline model, and also minimized the missed small hemorrhage regions while correctly detecting other diverse size hemorrhages. As a future work, we will further optimize SARA based on the RoI size distribution, so that the model can segment more effectively with respect to input target sizes.

References

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

[2] D. Parker Jr., D. H. Rhoney, and X. Liu-DeRyke, "Management of spontaneous nontraumatic intracranial hemorrhage," *J. Pharm. Pract.*, vol. 23, no. 5, pp. 398–407, 2010.

[3] J. Aronowski and X. Zhao, "Molecular pathophysiology of cerebral hemorrhage: Secondary brain injury," *Stroke*, vol. 42, no. 6, pp. 1781–1786, 2011.

[4] D. Klebe, D. McBride, J. J. Flores, J. H. Zhang, and J. Tang, "Modulating the immune response towards a neuroregenerative peri-injury milieu after cerebral hemorrhage," *J. Neuroimmune Pharmacol.*, vol. 10, no. 4, pp. 576–586, 2015.

[5] P. Ren, B.-C. Wang, Y.-Z. Wang, S.-L. Hao, T.-W. Guo, and X.-F. Li, "Evaluating tensile damage of brain tissue in intracerebral hemorrhage based on strain energy," *Exp. Therapeutic Med.*, vol. 16, no. 6, pp. 4843–4852, 2018.

[6] G. Xi, R. F. Keep, and J. T. Hoff, "Mechanisms of brain injury after intracerebral haemorrhage," *Lancet Neurol.*, vol. 5, no. 1, pp. 53–63, 2006.

[7] H. Bhadauria, A. Singh, and M. Dewal, "An integrated method for hemorrhage segmentation from brain CT imaging," *Comput. Elect. Eng.*, vol. 39, no. 5, pp. 1527–1536, 2013.

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[9] V. A. Convertino et al., "Use of advanced machine-learning techniques for noninvasive monitoring of hemorrhage," *J. Trauma Acute Care Surg.*, vol. 71, no. 1, pp. S25–S32, 2011.

[10] S. Keerthana and K. Sathiyakumari, "Brain stroke segmentation using fuzzy c-means clustering," *Int. J. Comput. Appl.*, vol. 154, no. 4, pp. 26–30, 2016.

[11] M. Scherer et al., "Development and validation of an automatic segmentation algorithm for quantification of intracerebral hemorrhage," *Stroke*, vol. 47, no. 11, pp. 2776–2782, 2016.

[12] C.-C. Liao, F. Xiao, J.-M. Wong, and I.-J. Chiang, "A multiresolution binary level set method and its application to intracranial hematoma segmentation," *Computerized Med. Imag. Graph.*, vol. 33, no. 6, pp. 423–430, 2009.

[13] L. Li et al., "Deep learning for hemorrhagic lesion detection and segmentation on brain CT images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1646–1659, May 2021.

[14] X. Li, G. Luo, W. Wang, K. Wang, Y. Gao, and S. Li, "Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 3, pp. 1140–1151, Mar. 2022.

[15] K. Hu et al., "Automatic segmentation of intracerebral hemorrhage in CT images using encoder–decoder convolutional neural network," *Inf. Process. Manage.*, vol. 57, no. 6, 2020, Art. no. 102352.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[17] A. Arab et al., "A fast and fully-automated deep-learning approach for accurate hemorrhage segmentation and volume quantification in non-contrast whole-head CT," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.

[18] J. L. Wang, H. Farooq, H. Zhuang, and A. K. Ibrahim, "Segmentation of intracranial hemorrhage using semi-supervised multi-task attention-based U-Net," *Appl. Sci.*, vol. 10, no. 9, 2020, Art. no. 3297.

[19] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation," *Sensors*, vol. 20, no. 4, 2020, Art. no. 1010.

[20] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[21] Y. Chong, X. Chen, and S. Pan, "Context union edge network for semantic segmentation of small-scale objects in very high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6000305.

[22] B. Shareef, A. Vakanski, P. E. Freer, and M. Xian, "Estan: Enhanced small tumor-aware network for breast ultrasound image segmentation," *Healthcare*, vol. 10, no. 11, 2022, Art. no. 2262.

[23] X. Li, Z. Gong, H. Yin, H. Zhang, Z. Wang, and L. Zhuo, "A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images," *Neural Netw.*, vol. 124, pp. 75–85, 2020.

[24] Y. Liu et al., "Refined segmentation R-CNN: A. two-stage convolutional neural network for punctate white matter lesion segmentation in preterm infants," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2019, pp. 193–201.

[25] S. Urday et al., "Targeting secondary injury in intracerebral haemorrhage–perihaematomal oedema," *Nature Rev. Neurol.*, vol. 11, no. 2, pp. 111–122, 2015.

[26] K.-Y. Wang et al., "Ultrastructural changes of brain tissues surrounding hematomas after intracerebral hemorrhage," *Eur. Neurol.*, vol. 74, no. 1/2, pp. 28–35, 2015.

[27] N. Li et al., "Association of molecular markers with perihematomal edema and clinical outcome in intracerebral hemorrhage," *Stroke*, vol. 44, no. 3, pp. 658–663, 2013.

[28] J. M. Gebel Jr. et al., "Natural history of perihematomal edema in patients with hyperacute spontaneous intracerebral hemorrhage," *Stroke*, vol. 33, no. 11, pp. 2631–2635, 2002.

[29] S. Urday et al., "Measurement of perihematomal edema in intracerebral hemorrhage," *Stroke*, vol. 46, no. 4, pp. 1116–1119, 2015.

[30] M. K. Mittal and A. LacKamp, "Intracerebral hemorrhage: Perihemorrhagic edema and secondary hematoma expansion: From bench work to ongoing controversies," *Front. Neurol.*, vol. 7, 2016, Art. no. 210.

[31] Y. Chen, S. Chen, J. Chang, J. Wei, M. Feng, and R. Wang, "Perihematomal edema after intracerebral hemorrhage: An update on pathogenesis, risk factors, and therapeutic advances," *Front. Immunol.*, vol. 12, 2021, Art. no. 740632.

[32] H. Arima et al., "Significance of perihematomal edema in acute intracerebral hemorrhage: The interact trial," *Neurology*, vol. 73, no. 23, pp. 1963–1968, 2009.

[33] C. Venkatasubramanian et al., "Natural history of perihematomal edema after intracerebral hemorrhage measured by serial magnetic resonance imaging," *Stroke*, vol. 42, no. 1, pp. 73–80, 2011.

[34] W.-J. Peng et al., "The risk factors and prognosis of delayed perihematomal edema in patients with spontaneous intracerebral hemorrhage," *CNS Neurosci. Therapeutics*, vol. 25, no. 10, pp. 1189–1194, 2019.

[35] N. Ironside, C.-J. Chen, D. Ding, S. A. Mayer, and E. S. Connolly Jr., "Perihematomal edema after spontaneous intracerebral hemorrhage," *Stroke*, vol. 50, no. 6, pp. 1626–1633, 2019.

[36] D. Aksoy et al., "Magnetic resonance imaging profile of blood–brain barrier injury in patients with acute intracerebral hemorrhage," *J. Amer. Heart Assoc.*, vol. 2, no. 3, 2013, Art. no. e000161.

[37] H. Kim, N. J. Edwards, H. A. Choi, T. R. Chang, K. W. Jo, and K. Lee, "Treatment strategies to attenuate perihematomal edema in patients with intracerebral hemorrhage," *World Neurosurg.*, vol. 94, pp. 32–41, 2016.

[38] G. Appelboom et al., "Volume-dependent effect of perihaematomal oedema on outcome for spontaneous intracerebral haemorrhages," *J. Neurol., Neurosurgery Psychiatry*, vol. 84, no. 5, pp. 488–493, 2013.

[39] L. H. Sansing, E. A. Kaznatcheeva, C. J. Perkins, E. Komaroff, F. B. Gutman, and G. C. Newman, "Edema after intracerebral hemorrhage: Correlations with coagulation parameters and treatment," *J. Neurosurgery*, vol. 98, no. 5, pp. 985–992, 2003.

[40] B. Volbers et al., "Semi-automatic volumetric assessment of perihemorrhagic edema with computed tomography," *Eur. J. Neurol.*, vol. 18, no. 11, pp. 1323–1328, 2011.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[42] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[43] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, *arXiv:2110.13389*.

[44] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[45] M. Hssayeni, M. Croock, A. Salman, H. Al-khafaji, Z. Yahya, and B. Ghoraani, "Computed tomography images for intracranial hemorrhage detection and segmentation," *Intracranial Hemorrhage Segmentation Using Deep Convolutional Model. Data*, vol. 5, no. 1, 2020.

[46] M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial hemorrhage segmentation using a deep convolutional model," *Data*, vol. 5, no. 1, 2020, Art. no. 14.

[47] P. Daverat, J. Castel, J. Dartigues, and J. Orgogozo, "Death and functional outcome after spontaneous intracerebral hemorrhage. A prospective study of 166 cases using multivariate analysis," *Stroke*, vol. 22, no. 1, pp. 1–6, 1991.

[48] R. Ahmed et al., "Predictors of in-hospital mortality for intracerebral hemorrhage: A hospital-based study in pakistani adults," *J. Stroke Cerebrovascular Dis.*, vol. 10, no. 3, pp. 122–127, 2001.

[49] X. Li, M. Kan, S. Shan, and X. Chen, "Weakly supervised object detection with segmentation collaboration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9735–9744.

[50] B. Zhang, J. Xiao, Y. Wei, M. Sun, and K. Huang, "Reliability does matter: An end-to-end weakly supervised semantic segmentation approach," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 12765–12772.

[51] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, "Simple does it: Weakly supervised instance and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 876–885.

[52] J. Ahn, S. Cho, and S. Kwak, "Weakly supervised learning of instance segmentation with inter-pixel relations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2209–2218.

[53] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance problems in object detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3388–3415, Oct. 2021.

[54] A. Jalali and M. Lee, "Atrial fibrillation prediction with residual network using sensitivity and orthogonality constraints," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 407–413, Feb. 2020.

[55] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, 2002.

[56] A. Jalali, S. Kavuri, and M. Lee, "Low-shot transfer with attention for highly imbalanced cursive character recognition," *Neural Netw.*, vol. 143, pp. 489–499, 2021.

[57] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[58] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[59] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[60] J. Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[61] H. Cao et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.

[62] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Berlin, Germany: Springer, 2018, pp. 3–11.

[63] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.