IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 2, 2020, accepted March 25, 2020, date of publication April 6, 2020, date of current version April 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985671

Cascade U-ResNets for Simultaneous Liver and Lesion Segmentation

XUE-FENG XI[1], LEI WANG[1], VICTOR S. SHENG[2], (Senior Member, IEEE), ZHIMING CUI[1], BAOCHUAN FU[1], AND FUYUAN HU[1]

[1]School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215000, China
[2]Computer Science Department, Texas Tech University, Lubbock, TX 79409, USA

Corresponding author: Victor S. Sheng (shengli.sheng@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61728205, Grant 61673290, Grant 61672371, Grant 61876217, Grant 61876121, and Grant 61750110534, in part by the Innovative Team of Jiangsu Province under Grant XYDXX-086, in part by the Jiangsu Development Project under Grant BE2017663, in part by the Science & Technology Development Project of Suzhou under Grant SYG201817, and in part by the Jiangsu Higher Education Natural Science Foundation under Grant 19KJB520054.

ABSTRACT In recent years, several deep learning networks are proposed to segment 2D or 3D bio-medical images. However, in liver and lesion segmentation, the proportion of interested tissues and lesions are tiny when contrasting to the image background. That is, the objects to be segmented are highly imbalanced in terms of the frequency of occurrences. This makes existing deep learning networks prone to predict pixels of livers and lesions as background. To address this imbalance issue, several loss functions are proposed. Since no researches are having made a comparison among those proposed loss functions, we are curious about that which loss function is the best among them? At the same time, we also want to investigate whether the combination of several different loss functions is effective for liver and lesion segmentation. Firstly, we propose a novel deep learning network (cascade U-ResNets) to produce liver and lesion segmentation simultaneously. Then, we investigate the performance of 5 selected loss functions, WCE (Weighted Cross Entropy), DL (Dice Loss), WDL (Weighted Dice Loss), TL (Teverskry Loss), WTL (Weighted Teversky Loss), with our cascade U-ResNets. We further assemble all cascade U-ResNets trained with different loss functions together to segment livers and lesions jointly on the liver CT (Computed Tomography) volume. Experimental results on the LiTS dataset[1] showed our ensemble model can achieve much better results than every individual model for liver segmentation.

INDEX TERMS
Data imbalance, deep learning, ensemble learning, lesion segmentation, liver segmentation, medical image segmentation.

I. INTRODUCTION

Liver and lesion segmentation are to delineate a liver and its lesions in medical images (such as CT, MRI, PET images). In computer-aided detections and diagnoses, precise automatic segmentation of the liver is meaningful, but manually delineating liver outline for millions of medical image slices is time-consuming. Automatic liver segmentation is one of the most difficult tasks in computer vision because of the diverse shapes of livers and low contrast with nearby tissues. Many algorithms have been proposed to cope with liver and lesion segmentation, such as region-based methods, thresholding, graph-cut, machine learning, and so on.

In recent years, applications of deep learning on medical image analysis are soaring. Particularly, deep convolutional neural networks can learn high-level features automatically and give reasonable output. Different from common image classification, liver and lesion segmentation is a pixel-level classification task, where a classification model needs to assign a label to each pixel and output the same size mask.

Long *et al.* [11] did some adjustments on VGG-16 by replacing the fully-connected layers in VGG-16 with deconvolution layers, which can return feature maps to the original size of an image by deconvolution operations. This is the first work to make pixel-level predictions. U-Net [24] is also based on a fully convolution neural network (FCN) with a

The associate editor coordinating the review of this manuscript and approving it for publication was Mufti Mahmud.

[1]https://competitions.codalab.org/competitions/17094#

good performance on biomedical image segmentation. This is because it introduces lateral connections to the structure. The lateral connections make U-Net able to capture context information and precise localization with fewer training images.

However, CT slices are a time sequence data. The information extracted from 2D images is not sufficient for accurate segmentation. In order to use the time axis of CT sequences, Ben-Cohen *et al.* [2] and Vorontsov *et al.* [34] stacked three adjacent slices as input to improve the performance of segmentation. Sun *et al.* [31] utilized three different phases of CT images to train a multi-channel FCN. Another problem in liver and lesion segmentation is that the location and the shape of lesion vary hugely in different slices. Some researchers used an extra model to perform precise lesion segmentation. Vorontsov *et al.* [34] proposed a cascade structure to combine two FCNs to get more precise results. Its first FCN finds the liver area firstly, and then its second FCNs can focus on the liver area to detect lesion.

Except for the problems mentioned above, the data imbalance problem should also be considered. The proportion of a liver and its lesion is tiny, contrasting to images background. That is, objects (a liver and its lesion) to be segmented are highly imbalanced in terms of the frequency of their occurrences. This phenomenon is obvious in FIGURE 1.

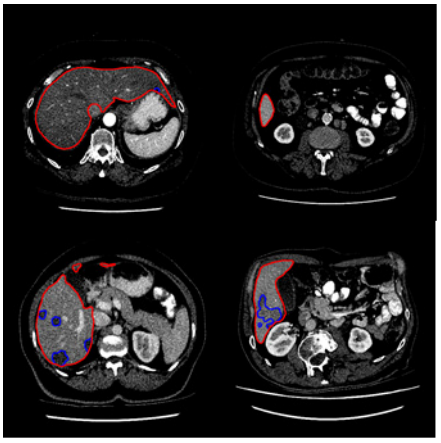


FIGURE 1. Liver and lesion segmentation. The red and blue lines in the above images denote the outline of a liver and its lesions respectively, the lesion may absence in some CT slices.

Several strategies have been proposed based on loss functions to solve this problem. Re-weighting is a frequently used strategy by assigning high weights on rare classes to offset the negative impact of the imbalanced distribution of samples. Both reweighted cross entropy and generalized dice [4] adopt this strategy. Similarity-based methods, such as Dice loss, Jaccard loss, and Tversky loss, are originally used to evaluate image segmentation since these methods are not related to the size of the segmented object. That is, they are resistant to unbalanced data.

In this paper, we try to develop a novel deep learning network for liver and lesion segmentation. We will alleviate the impact of the imbalanced problem in the liver and lesion segmentation by investigating the performance of different loss functions, and further, improve the performance of our novel deep learning network using ensemble learning. In general, the main contributions of this paper are as follows: (i) We propose a novel network (cascade U-ResNets) to use a cascade structure to produce liver and lesion segmentation respectively. (ii) Then, we investigate the performance of different loss functions on the liver and lesion segmentation with our proposed model. (iii) We further assemble all cascade U-ResNets trained with different loss functions together to produce joint segmentation. Our experimental results show that the ensemble model produces the best performance in terms of multiple different evaluation measurements.

II. RELATED WORK

A. FULLY CONVOLUTION NEURAL NETWORKS

As shown in FIGURE 2, there are three classes in the liver and lesion segmentation task: a background class in grey, a liver class in red, and a lesion class in green. For liver and lesion segmentation, the goal of a learning model is to classify each pixel in the CT image to one of the three classes (liver, lesion or background). That is, a learning model will assign a label for each pixel in the input image and the output size will be the same as the input. Otherwise, the input and output can't match.

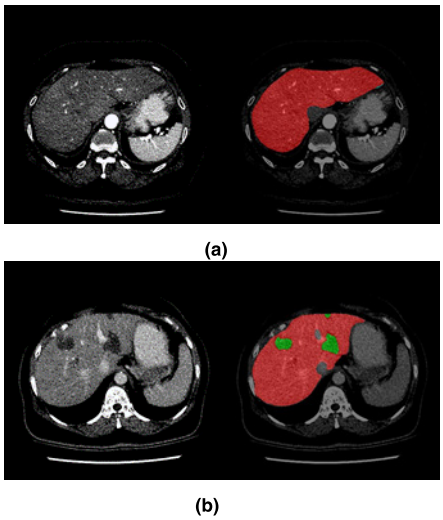


FIGURE 2. An example of semantic segmentation.

Several famous convolutional frameworks had been proposed, such as AlexNet [19], VGG-Net [27], GoogLeNet [28], ResNet [15], and so on. However, there are two main obstacles for them to perform semantic segmentation. As we know that there are max-pooling layers inherent in convolutional networks, and max-pooling layers decrease the resolution of an input image. The final feature map will become 16 or 32 times smaller than the original size of an input image after 4 or 5 pooling operations. But the output we need for medical image segmentation is the same size image as that of the input image. That is, we need to train a model to

return a final feature map to the same size as the input image. The second obstacle is that the fully-connected layer at the bottom of the above-mentioned networks destroys the spatial information of the feature map extracted from each image because the fully-connected layer flats the features extracted from an image to a vector.

To address two obstacles above, Long *et al.* [11] proposed the Fully Convolution Neural Network (FCN), which is a network composed of all convolutional layers. FCNs abandon the fully connected layer on VGG-Net and replace it with deconvolution (i.e., transpose convolution) layers, which can be seen as a transpose operation of normal convolution. The deconvolution layer can up-sample the feature map extracted from the original image. Following this strategy, a ''Pixel-to-Pixel'' segmentation can be achieved.

There exist a few neural networks following FCN's way for medical image segmentation, such as U-Net [24], which introduced the inter-skip connection to enhance the feature learning of its decoder. Except for using the deconvolution layer, SegNet [3] used the 'max-pooling indices' to perform non-linear upsampling. This eliminates the need for learning to upsample. Li *et al.* [13] and Vorontsov *et al.* [34] used the nearest neighbor interpolation to up-sample the feature map extracted from the input image to avoid checkerboard artifacts created by deconvolution. In the vanilla FCNs [11], extracted features only flow layer by layer from the top to the bottom. In comparison, U-Net adds a lateral path from its down-sampling phase to its up-sampling phase. The lateral path makes features flow from a previous layer to a latter layer easier. Its effectiveness has been demonstrated in medical image segmentation.

In a normal semantic segmentation, an input image is usually a 2D image. However, CT imaging produces a 3D sequence, which contains cross-sectional (tomographic) images. The 2D FCNs neglect the volumetric context in the CT sequence, so they can't capture the inter-slice information between slices. Some works were proposed to directly segment 3D volume. V-Net [22] is a 3D FCNs, which can directly handle 3D medical images. However, it suffers from the heavy computation cost and the limitation of GPU memory resources. Therefore, the tradeoff between the inference time and the segmentation accuracy should be considered. Li *et al.* [13] proposed a hybrid structure, which is composed of a 2D FCN and a 3D FCN. This structure can effectively extract intra-slice and inter-slice features, and also reduce the computation complexity.

We will develop a novel deep learning network to take the advantages of both U-Net and ResNet for liver and lesion segmentation. We intend to perform the liver and lesion segmentation simultaneously and allow information communications between liver and lesion segmentation so that we can achieve better performance for both liver and lesion segmentation.

B. DATA IMBALANCE IN SEMANTIC SEGMENTATION

Designing a good loss function can help build a better model. One of the commonly used loss functions in sematic segmentation is cross entropy. However, objects to be segmented in different images always vary hugely. Cross entropy does not consider data imbalance. We can expect that it could fail to detect tiny objects, like lesions in a liver. For solving the data imbalance problem, Christ *et al.* [6] assigned a weight to each class, which is the reciprocal of the proportion of pixels belonging to the class (i.e. $w_i = 1/n_i$, where n_i denote the number of pixels in class i). However, the proportion of lesion pixels is usually less than 1% in an image. In this manner, the reciprocal of the lesion occurrence probability will be very tiny. Christ *et al.* [7] further proposed a method to calculate the weight following, where N denotes all the total number of pixels in an image.

The similarity based loss functions are frequently used in medical image segmentation [8], [22], [27], [30]. Milletari *et al.* [22] introduced the dice coefficient score (DSC) as a loss function to solve the data imbalance problem, which is originally used to evaluate the performance of segmentation. Since the value of dice is unrelated to the number of object pixels, the dice is not affected by the proportion of each class. Cai *et al.* [8] employed another segmentation criterion Jaccard coefficient as a loss function. In medical image segmentation, recall is more important. Since the dice loss ignores the difference between false positives and false negatives, Salehi *et al.* [29] proposed Tversky loss function based on Tversky similarity index (Tversky, 1977) to make the difference between false positives and false negatives into consideration. Specifically, the Tversky loss function adopts two parameters α and β to make a tradeoff between false positives and false negatives. For multi-class segmentation, Sudre *et al.* [30] used re-balanced properties of generalized dice overlap [4] to enhance detecting rare classes.

We will investigate the performance of popular loss functions for liver and lesion segmentation with our proposed deep learning network.

III. CASCADE U-RESNETS

FIGURE 3 shows the structure of our proposed deep learning network cascade U-ResNets for liver and lesion segmentation, which contains two U-ResNets. Briefly, we employ a cascade structure to segment livers and their lesions

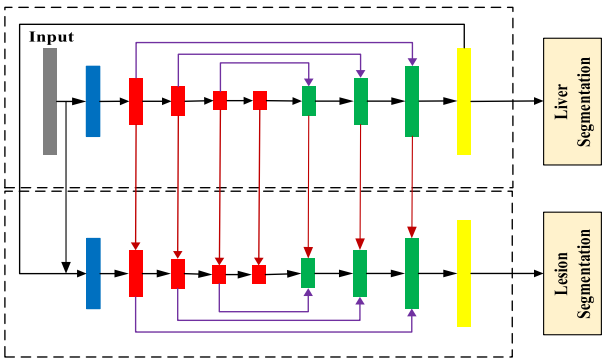


FIGURE 3. The architecture of cascade U-ResNets.

simultaneously, since those two tasks are correlated. The liver lesion is abnormal tissue within the liver and we consider that conducting them together will reduce the false positive of lesion segmentation. The correlation of liver and lesion segmentation is also a reason why we utilize two same structures for those two tasks. First, the above U-ResNet shown in FIGURE 3 will segment the liver from a given medical image (an input image). Second, the input image will be concatenated with liver segmentation results, and then fed to the below U-ResNet shown in FIGURE 3 to delineate the lesion area from the segmented liver. Finally, two networks will be trained in an end-to-end manner.

A. U-RESNET

Our proposed U-ResNet is inspired by U-Net and ResNet. The structure of U-ResNet is shown in TABLE 1. It is comprised of two phases, i.e., a down-sampling phase and an up-sampling phase. In the down-sampling phase, four layers (i.e. Res-block1_x, Res-block2_x, Res-block3_x, Res-block4_x) are used to extract features from input images. Each Res-block layer is comprised of a series of the residual block, which contains two 3 × 3 convolutions. The input of Res-block is added to the output as ResNet did. In this way, the loss can quickly backpropagate to the early layer. The structure of the Res-block is shown in FIGURE 4. Each time, the feature map is transformed from one layer to another layer. The size of it will be halved and we employ a 3 × 3 convolution with stride 2 to do this work. The shrink of the resolution from an input image to its final feature map is called 'output stride'. Although setting the output stride as 32 can reduce the model's inference time, the output stride is commonly set as 16 in semantic segmentation for denser feature extraction [9].

TABLE 1. The architecture of U-ResNet

Feature map	Layer	Liver	Lesion
		Structure	Structure
112×112	Conv1	7×7×96 , stride2	7×7×96 , stride2
56×56	Res-Block1_x	$\begin{bmatrix}3\times3\times24\\3\times3\times48\end{bmatrix}\times6$	$\begin{bmatrix}3\times3\times24\\3\times3\times48\end{bmatrix}\times3$
28×28	Res-Block2_x	$\begin{bmatrix}3\times3\times48\\3\times3\times96\end{bmatrix}\times9$	$\begin{bmatrix}3\times3\times48\\3\times3\times96\end{bmatrix}\times4$
14×14	Res-Block3_x	$\begin{bmatrix}3\times3\times96\\3\times3\times192\end{bmatrix}\times18$	$\begin{bmatrix}3\times3\times96\\3\times3\times192\end{bmatrix}\times9$
14×14	Res-Block4_x	$\begin{bmatrix}3\times3\times192\\3\times3\times384\end{bmatrix}\times12$	$\begin{bmatrix}3\times3\times192\\3\times3\times384\end{bmatrix}\times6$
28×28	Trans Conv1	2×2×192	2×2×192
28×28	Res-Block5_x	$\begin{bmatrix}3\times3\times96\\3\times3\times192\end{bmatrix}\times3$	$\begin{bmatrix}3\times3\times96\\3\times3\times192\end{bmatrix}\times3$
56×56	Trans Conv2	2×2×96	2×2×96
56×56	Res-Block6_x	$\begin{bmatrix}3\times3\times48\\3\times3\times96\end{bmatrix}\times3$	$\begin{bmatrix}3\times3\times48\\3\times3\times96\end{bmatrix}\times3$
112×112	Trans Conv3	2×2×48	2×2×48
112×112	Res-Block7_x	$\begin{bmatrix}3\times3\times24\\3\times3\times48\end{bmatrix}\times3$	$\begin{bmatrix}3\times3\times24\\3\times3\times48\end{bmatrix}\times3$
224×224	Trans Conv4	2×2×3	2×2×3
224×224	Conv2	1×1×1 ,sigmoid	1×1×1 ,sigmoid

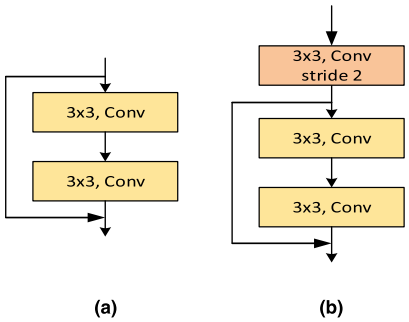


FIGURE 4. The architecture of Res-Block.

The up-sampling phase includes three Res-block layers and four transpose convolution (Tran-Conv) layers. A trans-Conv layer is comprised of ''BN (Batch Normalization) - ReLU - Transpose Convolution''. When a feature map goes through it every time, its feature map size is doubled. The transpose convolution will learn to up-sample the feature map with trainable parameters. Note that the structure of the Res-block layers in the up-sampling phase is the same as that of the Res-block layers in the down-sampling phase. In some CT slices, the size of the liver and its lesion is very tiny, which is hard to detect. Image Pyramid [1] is a common method to enrich the image information by mapping a raw image to different scales, which can endow a learning model able to detect variable-sized objects. However, using the output of Image Pyramid as the input will obviously increase the inference time. Feature Pyramid Network (FPN) [33] uses different sized feature maps, which is inherent in the network to detect different scale objects. We followed the strategy of FPN, and have different sized feature maps outputted by each Res-block layer in the up-sampling phase to predict the segmentation result in parallel. The prediction produced by each Res-block layer will be merged with a 1 × 1 convolution layer to generate final results.

B. SKIP-CONNECTIONS

In our cascade U-ResNets, we introduce the intra-network and the inter-network skip connections to reuse the information inherent in the two U-ResNets. They are denoted by purple lines and dark-red lines respectively in FIGURE 3.

The intra-network skip connections (purple lines) from the down-sampling phase to the up-sampling phase is employed to concatenate the feature maps with the same size in two phases. U-Net has already demonstrated the effectiveness of this intra-connection in medical image segmentation. It endows a deep learning network to capture the context information during training.

In addition to the intra-network skip connections of U-ResNet, we also employed the inter-network skip connections between two U-ResNets to enhance both liver and lesion segmentation. As shown in FIGURE 3, features extracted by the above U-ResNet will flow to the below U-ResNet through this inter-path (dark-red lines). From one side, these inter-network skip connections help the above U-ResNet ''inform'' the below U-ResNet where the liver is by enhancing information flow. On the side, the below U-ResNet can also provide feedback to the above U-ResNet to correct the liver location through the inter-path, since the loss can backpropagate to the former layer in the first U-ResNet easier by inter-connection. It will update the parameters in the first U-ResNet to produce a better liver segmentation or produce less loss. The inter-network skip connections between two U-ResNet make two models share the features in real-time.

As shown in FIGURE 3, the intra-network skip connections concatenate the feature map directly on their channel, while the inter-network skip connections firstly let the feature map go through a 1×1 convolution operation, and then concatenate the features on their channel. As we mentioned before, the two U-ResNets have different tasks. We expect that the convolution operation in the two U-ResNets can learn some features that are much suitable for lesion segmentation. Besides, we also expect that the inter-network and intra-network skip connections can help the two models convergence faster and produce better segmentation results.

C. LOSS FUNCTIONS

Since a loss function can impact the performance of deep learning, we try to choose a proper loss function for our cascade U-ResNets by investigating the effectiveness of five popular loss functions on liver and lesion segmentation. These five loss functions can be roughly divided into three categories: reweight-based loss functions (i.e., reweighed cross entropy (WCE)), similarity-based loss functions (i.e., Dice loss (DL), Tversky loss (TL)) and integrated loss functions (i.e., Generalized Dice loss (GDL), Generalized Tversky loss (GTL)).

Before we provide the definitions of each loss function, notations used in the following formulations are defined here. i denotes the index of the channel in a ground truth image, j denotes the index of a pixel in the image, l denotes the total number of classes in a segmentation task, N denotes the total number of pixels in an image, p denotes the probability outputted by a classification model, and g denotes the ground truth. Then, we use $p_{ij}$to present the probability of pixel j belonging to class i, and g_{ij} to present the value of pixel j in channel i of the ground truth image. If pixel j in the input image belongs to the class 0, the value of $g_{0j=1}$ while $g_{1j}, g_{2j,\ldots}, g_{lj,=} 0$.

ReWeighed Cross Entropy (WCE): The re-weight strategy erases data imbalance by adding weights to rare classes. WCE is one of the frequently used loss functions based on this manner, which is defined as follows. We choose it as a benchmark.

$$WCE = \sum_{j=0}^{N} \sum_{i=0}^{l} w_i g_{ij} \log(p_{ij}) \tag{1}$$

where w_i denotes the class weight. In this paper, all w_i is defined as follows.

$$w_i = \frac{N - n_i}{n_i} \tag{2}$$

where n_i presents the number of pixels belonging to class i.

Dice Loss (DL) and Generalized Dice Loss (GDL): A similarity-based loss function has a similarity criterion, which is independent of the size of objects, so the loss of a small object will not be affected by its background. Dice Score Coefficient (DSC) is a criterion for evaluating the overlap between a segmentation result and its corresponding ground truth. Milletari *et al.* [22] proved its performance when DSC is adopted as a loss function. Since we need to segment the liver and its lesions from medical images, we have two foreground classes. Including one background class, which presents other organs or tissues, there are three classes (multi-class) in our image segmentation task. For multi-class segmentation, multi-class DSC can be expressed as follows.

$$DL = 1 - \frac{2 \sum_{i=0}^{l} \sum_{j=0}^{N} p_{ij} g_{ij}}{\sum_{i=0}^{l} \sum_{j=0}^{N} p_{ij}^2 + \sum_{i=0}^{l} \sum_{j=0}^{N} g_{ij}^2} \tag{3}$$

For highly unbalanced data, Generalized Dice loss [4] combines the re-weight strategy with dice loss for further improvement, which is defined as follows.

$$GDL = 1 - \frac{2 \sum_{i=0}^{l} \sum_{j=0}^{N} w_i p_{ij} g_{ij}}{(\sum_{i=0}^{l} \sum_{j=0}^{N} w_i p_{ij}^2 + \sum_{i=0}^{l} \sum_{j=0}^{N} w_i g_{ij}^2)} \tag{4}$$

Tversky Loss (TL) and Generalized Tversky loss (GTL): In medical diagnoses, the false negatives are much less tolerable than false positives. Therefore, a variation of dice loss called Tversky loss (TL) is proposed. Two parameters α and β are introduced to adjust the ratio between false negatives and false positives. The formula of TL is defined as (5), as shown at the bottom of the next page.

Where $p_{\tilde{ij}} = 1 - p_{ij}$ and $g_{\tilde{ij}} = 1 - g_{ij}$. Inspired by (4) and (5), we propose Generalized Tversky Loss (GTL) based on re-weight and the Tversky loss, which can be defined as (6), as shown at the bottom of the next page.

IV. EXPERIMENTS

A. EXPERIMENTAL DATASET AND ENVIRONMENT

In this section, we give more details about the experiment dataset and the implementation environment. The experiment data we choose is the LiTS dataset, which contains 131 liver CT volumes for training a model and 70 CT volumes for testing the model. We randomly select 120 volumes from 131 volumes to train our model and validate our model on the remaining 11 volumes aiming to adjust its hyper-parameters. Finally, we use the 70 testing volumes to evaluate our model. The CT slices in the dataset involve two foreground classes

(i.e., liver and lesion) and one background class comprised of unrelated organs. The resolution of each CT slice is 512×512.

Since the limitation of the computation capacity of our experimental environment, we first resize each image to 224 × 224 and train our model on the resized images for 70 epochs with a batch size 16 with an initial learning rate 1e-3. Then, we finely tune our model on the original size images for 20 epochs with a batch size 6 with an initial learning rate 1e-4. The learning rate decayed is also employed, so the learning rate will be updated according to the formula:

$$lr = lr * (1 - \frac{iterations}{totalIterations})^{0.9} \tag{7}$$

Parameters α and β in the Tversky loss function are set as 0.3 and 0.7 by following [29].

We implemented our model in the Tensorflow framework and run our experiments on Ubuntu 16.04 with CPU i7 8700K CPU, NVIDIA GTX 1080Ti GPU and 12G memory.

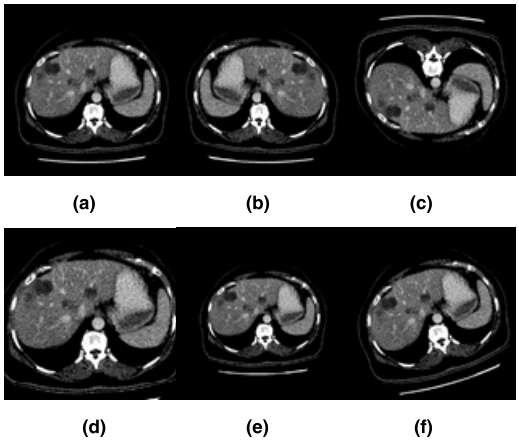


FIGURE 5. Data augmentation; (a) Raw image; (b) Horizontal flip; (c) Vertical flip; (d) Zoom in; (e) Zoom out; (f) Rotation.

B. DATA PRE-PROCESSING AND DATA AUGMENTATION

Before training, some preparations should be made on the original images. First, we truncated the pixel value of CT slice into [−200, 250] to remove the unrelated organs from the images. Then, since even experienced expert radiologists still need to combine the above and the below slice to determine the liver or its lesion area from a current slice, we take the above and the below slice of the current one into consideration and stack three consecutive 2D slices as the model's input for each slice.

For data augmentation, we randomly vertically or horizontally flip each medical image with 50% change, scale images between 0.9 and 1.1 with 50% chance, and rotate images up to 15 degree with 20% chance. FIGURE 5 shows a series of results generated by the data augmentation from a CT slice. All the data augmentation operations are combined based on their own probability during training.

C. EVALUATION METRICS

According to the evaluation metrics of liver segmentation challenges [23], we selected three evaluation measures from all metrics to evaluate our sliver and lesion segmentation results, i.e., Volumetric Overlap Error (VOE), Relative Volume Difference (RVD), and Dice coefficient.

For VOE, the smaller the value, the better the performance of the model is. For RVD, the smaller the absolute value, the better the performance of the model is. Note that the value of 0 for RVD doesn't imply that the ground truth and the segmentation result is identical. For this reason, RVD can't be the only criterion for segmentation evaluation [14]. For Dice, a greater value indicates better segmentation in terms of Dice.

V. RESULTS

Our experimental results are shown in TABLE 2. From TABLE 2, we can first see that the reweight-based loss function (i.e., WCE) is not a good loss function here. It performs the worst and has a big gap to other loss functions. We can also see that all similarity-based loss functions (i.e., DL and TL) perform much better than WCE, even though the WCE loss takes the data imbalance into consideration. Between the two similarity-based loss functions (i.e., DL and TL), TL performs slightly better than DL. Between the two integrated loss functions (i.e., GDL and GTL), our Generalized Tversky function (GTL) performs slightly better than GDL. In addition, we can also observe that GDL performs slightly better than the Tversky function in terms of VOE and Dice, and GDL performs slightly worse than DL.

Our lesion segmentation results of the five different loss functions are shown in TABLE 3. Again, TABLE 3 shows that WCE has the worst performance. The other four loss functions perform much better than WCE. Between the two similarity-based loss functions (i.e., DL and TL), DL outperforms TL in terms of all three measurements, which is completely opposite on the liver segmentation. Between the two integrated loss functions (i.e., GDL and GTL), GDL outperforms GTL on the lesion segmentation, which is also completely opposite on the liver segmentation.

$$TL = 1 - \frac{\sum_{i=0}^{l}\sum_{j=0}^{N} p_{ij}g_{ij}}{\sum_{i=0}^{l}\sum_{j=0}^{N} p_{ij}g_{ij} + \alpha\sum_{i=0}^{l}\sum_{j=0}^{N}(p_{ij}g_{\tilde{ij}})^2 + \beta\sum_{i=0}^{l}\sum_{j=0}^{N}(p_{\tilde{ij}}g_{ij})^2} \tag{5}$$

$$GTL = 1 - \frac{\sum_{i=0}^{l}\sum_{j=0}^{N} w_i p_{ij}g_{ij}}{\sum_{i=0}^{l}\sum_{j=0}^{N} w_i p_{ij}g_{ij} + \alpha\sum_{i=0}^{l}\sum_{j=0}^{N} w_i(p_{ij}g_{\tilde{ij}})^2 + \beta\sum_{i=0}^{l}\sum_{j=0}^{N} w_i(p_{\tilde{ij}}g_{ij})^2} \tag{6}$$

TABLE 2. Evaluation of liver segmentation (the best is in bold, and the second best is in italic.)

Methods	VOE	RVD	Dice (%)
WCE	0.162	0.891	87.6
DL	0.127	**0.001**	93.1
GDL	0.129	*-0.007*	92.9
TL	0.122	-0.045	93.3
GTL	*0.116*	-0.056	*93.8*
Ensemble	**0.095**	0.021	**94.9**

TABLE 3. Evaluation of lesion segmentation (the best is in bold, and the second best is in italic.)

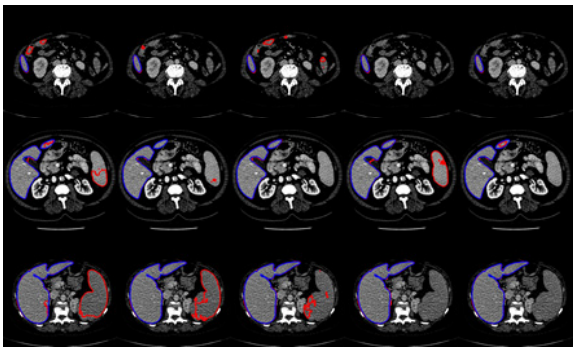
Methods	VOE	RVD	Dice (%)
WCE	0.488	0.551	63.9
DL	*0.386*	*-0.126*	*74.6*
GDL	0.393	**-0.124**	73.9
TL	0.415	-0.213	72.7
GTL	0.448	-0.288	69.6
Ensemble	**0.379**	-0.159	**75.2**

Since DL and TL perform well on liver segmentation and lesion segmentation respectively, we want to take advantage of different loss functions and make them complement each other. Therefore, we ensembled models trained with DL, GDL, TL, and GTL to jointly decide the final segmentation. Simply, we added the probability maps outputted by each model together, and then the average value was used to determine the liver and the lesion area in the liver CT volume.

The segmentation results of the ensemble approach are shown in the last row of TABLE 2 and TABLE 3 respectively. From TABLE 2 and TABLE 3, we can observe an obvious improvement in both liver and lesion segmentation in terms of VOE and Dice. The Dice of liver and lesion segmentation got 0.9% and 0.6% increment respectively, compared with the second-best in TABLE 2 and TABLE 3 respectively. In addition, we can also see that the ensemble approach improves significantly the performance in terms of VOE, but not RVD.

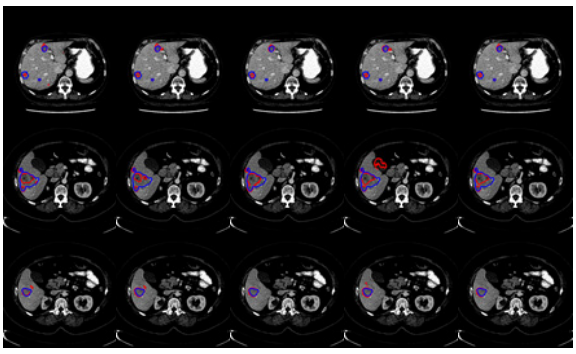
Slices displayed in FIGURE 6 and FIGURE 7 illustrate that the ensemble approach can take a comprehensive consideration from the segmentation results produced by every single model. We can find that every single model has some false positive predictions on the displayed slices, and the ensemble model got the best segmentation results among them without over-segmentation.

The blue line denotes the Ground truth, while the red line denotes the automatic segmentation result of each loss function: (a) DL; (b) GDL; (c) TL; (d) GTL; (e) Ensemble.



(a) (b) (c) (d) (e)

FIGURE 6. Liver segmentation conducted by different Loss functions.



(a) (b) (c) (d) (e)

FIGURE 7. Lesion segmentation conducted by different loss functions. The blue line de-notes the Ground truth, while the red line denotes the automatic segmentation result of each loss function: (a)DL; (b) GDL; (c) TL; (d) GTL; (e) Ensemble.

TABLE 4. Comparing with different deep learning models on liver segmentation.

Methods	VOE	RVD	Dice (%)
U-Net	14.28	21.22	90.82
SegNet	18.31	52.39	89.46
FCN-8s	-	-	89.0
Cascade-2D-FCN	0.128	-3.3	93.1
2D-Densenet	0.15	-0.8	92.3
3D-2D FCN+CRF	-	-	93.5
GTL	0.116	-0.056	93.8
Ensemble	**0.095**	**0.021**	**94.9**

VI. DISCUSSION

A. THE SIMILARITY-BASED LOSS FUNCTION

By jointly analyzing the results in TABLE 2 and TABLE 3, it is obvious that WCE is inferior to the other 4 similarity-based loss functions. Despite WCE introduce weights to eliminate the effect of data imbalance, it still can't compete with the similarity-based methods. In other words, the result shows that adapting the similarity-based measurements as loss function is a better strategy than only adding weights to the loss. Furthermore, we find that the combination of the re-weighting strategy and similarity-based methods doesn't mean better results. As the GTL adds weights to the TL, we observe an apparent decline in the lesion segmentation result. The GDL also get a minor decline on the performance on lesion after introducing the re-weight item. We attribute this results to the formula "N-n/n" which may cause the instability of the network.

TABLE 5. Comparing with different deep learning models on lESION segmentation.

Methods	VOE	RVD	Dice (%)
U-Net	-	-	65.0
ResNet	-	-	67.3
2D-FCNs	0.45	**0.040**	67.0
Cascade-2D-FCN	0.411	19.75	72.5
GDL	0.393	-0.124	73.9
Ensemble	**0.379**	-0.159	**75.2**

B. COMPARISONS WITH EXISTING POPULAR DEEP LEARNING NETWORKS

We further compared our cascade U-ResNets with six popular deep learning structures: U-Net [24], SegNet [3], FCN-8s [2], Cascade-2D-FCN [7], 2D-Densenet [21] and 3D-2D FCN+CRF [25], 2D-FCNs [17] on the liver and lesion segmentation, and our comparison results are shown in TABLE 4 and TABLE 5.

VII. CONCLUSION

In this paper, we first proposed an end-to-end liver and lesion segmentation model, which is composed of two cascaded U-ResNets. In the cascaded structure, besides the intra-network skip connection inherent in U-ResNet, the inter-network skip connections are introduced to our model to ensure the information exchange between two U-ResNets during training. We further investigated the performance of our model using five popular loss functions (i.e., WCE, DL, GDL, TL, and GTL). Our experimental results show that the similarity-based loss functions perform much better than WCE. To take advantage of each loss function, we assembled models trained with different loss functions (i.e., DL, GDL, TL, and GTL) to jointly segment the liver CT volume. Our experimental results showed our ensemble model can achieve much better results than each model.

In this paper, we only used a fix re-weight term $(N\text{-}n_i)/n_i$ to adjust the two popular similarity loss functions (i.e., DL and TL). This is a very simple solution. In the future, we will find re-weighted strategies to integrate the re-weighted strategies to similarity-based loss functions on the liver and lesion segmentation. We also have a great interest in introducing our model to other medical image segmentation tasks.

ACKNOWLEDGMENT

The authors declare that they have no competing interests. Since a publicly available dataset was used in this paper, so the ethical approval is not applicable.

REFERENCES

[1] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing. RCA engineer," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[2] A. Ben-Cohen, I. K. E. Diamant, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 77–85.

[3] 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.

[4] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062

[6] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmad, and B. H. Menze, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2016, pp. 415–423.

[7] P. F. Christ, F. Ettlinger, F. Grün, M. Ezzeldin A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze, "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," 2017, *arXiv:1702.05970*. [Online]. Available: http://arxiv.org/abs/1702.05970

[8] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," 2017, *arXiv:1707.04912*. [Online]. Available: http://arxiv.org/abs/1707.04912

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[10] M. Diwakar and M. Kumar, "A review on CT image noise and its denoising," *Biomed. Signal Process. Control*, vol. 42, pp. 73–88, Apr. 2018.

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

[12] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 171–182, Feb. 2017.

[13] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[14] T. Heimann *et al.*, "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1251–1265, Aug. 2009.

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

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

[17] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," 2017, *arXiv:1704.07239*. [Online]. Available: http://arxiv.org/abs/1704.07239

[18] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFS with Gaussian edge potential," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[20] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[21] K. Chaitanya Kaluva, M. Khened, A. Kori, and G. Krishnamurthi, "2D-densely connected convolution neural networks for automatic liver and tumor segmentation," 2018, *arXiv:1802.02182*. [Online]. Available: http://arxiv.org/abs/1802.02182

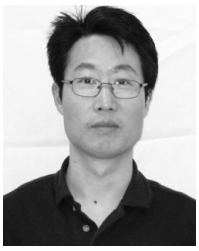
[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
[23] P. Bilic *et al.*, "The liver tumor segmentation benchmark (LiTS)," 2019, *arXiv:1901.04056*. [Online]. Available: https://arxiv.org/abs/1901.04056
[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.
[25] S. Rafiei, E. Nasr-Esfahani, K. Najarian, N. Karimi, S. Samavi, and S. M. R. Soroushmehr, "Liver segmentation in CT images using three dimensional to two dimensional fully convolutional network," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2067–2071.
[26] P. Y. Simard, Y. A. LeCun, J. S. Denker, and B. Victorri, "Transformation invariance in pattern recognition—Tangent distance and tangent propagation," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 1998, pp. 239–274.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[29] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep network," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, Sep. 2017, pp. 379–387.
[30] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 240–248.
[31] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.
[32] A. Tversky, "Features of similarity," *Psycholog. Rev.*, vol. 84, no. 4, p. 327, 1977.
[33] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
[34] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury, "Liver lesion segmentation informed by joint liver segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1332–1335.

XUE-FENG XI was born in Suzhou, Jiangsu, China, in 1978. He received the B.S. degree from Jiangsu University, in 2000, and the M.S. degree from Southeast University, in 2004, and the Ph.D. degree from Soochow University, in 2017.

From 2000 to 2011, he was a Research Assistant from 2000 to 2006 and an Assistant Professor from 2007 to 2011 with the Electronic Information Department, Suzhou University of Science and Technology. Since 2012, he has been an Associate Professor with the Computer Science and Engineering Department, Suzhou University of Science and Technology. He is the author of two books, more than 20 articles. He holds three patents. His research interests include natural language processing, computer vision, human–machine interaction, machine learning, and software engineering.

Dr. Xi was a recipient of the Award for Science and Technology Progress of Jiangsu Province, in 2018, and the Award for Teaching Achievement of Jiangsu Province, in 2017.

LEI WANG was born in Gaoyou, Jiangsu, China, in 1994. He received the B.S. and M.S. degrees from the Yanchen Institute of Technology, Suzhou University of Science and Technology, in 2016 and 2019, respectively. From January 2018 to June 2019, he was a Research Scholar with the University of Central Arkansas, USA. His research interests include data mining, machine learning, computer vision, and related application.



VICTOR S. SHENG (Senior Member, IEEE) received the master's degree in computer science from the University of New Brunswick, Canada, in 2003, and the Ph.D. degree in computer science from Western University, London, ON, Canada, in 2007.

He was an Associate Research Scientist and a NSERC Postdoctoral Fellow of information systems with the Stern Business School, New York University, after he received his Ph.D. He was an Associate Professor of computer science with the University of Central Arkansas, and the Founding Director of the Data Analytics Lab (DAL). Since 2018, he has been a Tenured Associate Professor with the Department of Computer Science, Texas Tech University, Lubbock, TX, USA. His research interests include data mining, machine learning, and related applications.

Dr. Sheng is a lifetime Member of the ACM. He received the best paper award runner-up from KDD '08, and the best paper award from ICDM '11. He is also a PC member for a number of international conferences and a reviewer for several international journals.

ZHIMING CUI was born in Shanghai, China, in 1961. He is currently a Professor with the Computer Science and Engineering Department. He is also a Vice President of the Suzhou University of Science and Technology. He is the author of 13 books, more than 300 articles. He holds 34 patents. His research interests include data mining, machine learning, computer vision, human–machine interaction, and related applications.

BAOCHUAN FU was born in Zhengzhou, Henan, China, in March 1964. He received the B.S. degree in physical science from Henan University, in 1986, and the Ph.D. degree in control theory from Tongji University, in 2008.

Since 2006, he has been a Professor with the Institute of Electronics and Information Engineering, Suzhou University of Science and Technology. He is the author of more than 80 articles. His research interests include machine learning, intelligent control and building energy conservation, and so on.

FUYUAN HU received the Ph.D. degree from Northwestern Polytechnical University.

He was a Postdoctoral Researcher with Vrije Universiteit Brussel, Belgium. He was a Visiting Ph.D. Student with the City University of HongKong. Since 2012, he has been a Professor of computer vision and machine learning with the Suzhou University of Science and Technology. His research interests include machine learning, computer vision, and image processing.