Journal of Physics: Conference Series

**PAPER** • **OPEN ACCESS**

# Auto Whole Heart Segmentation from CT images Using an Improved Unet-GAN

To cite this article: Kening Le *et al* 2021 *J. Phys.: Conf. Ser.* **1769** 012016



# Auto Whole Heart Segmentation from CT images Using an Improved Unet-GAN

**Kening Le, Zeyu Lou, Weiliang Huo, Xiaolin Tian***

Faculty of Information Technology, Macau University of Science and Technology, Macau SAR, China

*Corresponding author's email: xltian@must.edu.mo

**Abstract.** The development of deep learning is rapid, and convolutional neural network especially U-Net plays an important role in the medical image segmentation tasks, which is lack of data. Lots of models and methods are proposed to segment cardiac CT images. In this paper, we proposed a new network architecture. The network architecture is based on a traditional architecture called conditional generative adversarial network (cGAN), where R2U-Net acts as the generative network and FCN as the discriminative network. The performance of this model running on the dataset from MICCAI 2017 Multi-Modality Whole Heart Segmentation Challenge (MM-WHS 2017) is good.

**Keywords.** R2U-Net, U-Net, GAN, learning rate policy, cardiac CT image, whole heart segmentation

## 1. Background

Cardiovascular diseases (CVDs) have the highest morbidity and mortality in the world. It is useful for effective treatments that CVDs are detected early. With the rapid development of medical imaging techniques, computed tomography (CT) scan makes timely diagnosing possible. With scanned cardiac CT slices, the algorithm can generate a 3D heart model which is built according to the patient's heart as a result, which would be used to instruct doctors to do operations.

Cardiac CT image segmentation is the first and important step of the whole process. It is a challenge for computer to segment the regions of heart from the cardiac CT images automatically because of the high anatomical and signal intensity variations. The lack of high-level labeled dataset is always the biggest problem.

Due to the development of deep learning in digital image processing these years, due to the development of deep learning, deep convolutional neural networks undertake the segmentation tasks. As a typical convolutional neural network which needs little data, U-net [1] is used in medical image segmentation area widely. Kinds of neural network models based on U-net were proposed. We can apply them to deal with the problems about medical image segmentation, including the segmentation of cardiac CT images.

Generative adversarial nets (GAN) [2] is a network frame. There are two neural networks, including a generative network and a discriminative network. The core idea is that the generator generates new data which is like the training set to deceive the discriminator, while the discriminator identities and evaluates the data. The two networks contest with each other and adjust parameters constantly until the discriminator cannot identify whether the data provided by generator is true or false.

## 2. Related work

Since the CT slices can be seen as slices of a volume which contains a whole organ which is wanted, 3D U-Net was designed [3] and applied with an additional data augmentation process in [4], which presents a nice accuracy of cardiac CT images segmentation.

Liu used the first U-Net to segment the whole heart from the original cardiac CT images and the second one to segment each part from the results of the first stage [5]. Xu combined the Faster R-CNN and U-Net, whose duties are localizing and detecting [6].

Aimed to the disadvantage that the generator of GAN generates the data randomly instead of images with specific properties, conditional generative adversarial nets (cGAN) [7] was designed by replacing all the probabilities in the original GAN with conditional probabilities. Research in [8] used Unet-GAN to segment medical images and it performed well.

Reducing learning rate during the process of training a neural network always helps a lot. Learning rate schedules can reduce the learning rate according to a pre-defined schedule during the training process. As a common learning rate schedule, step decay is commonly used in either training process or neural architecture search (NAS), it drops the learning rate a certain rate every certain number of epochs [9-11].

## 3. Proposed method

The architecture of the whole segmentation algorithm is based on cGAN, which also contains a generative network for generating fake masks and a discriminative network for discriminating them. As the workflow shown in Figure 1, the generative network generates a fake mask according to the original image, which is the input, then the discriminative network discriminates the generated fake mask according to the real labeled mask. The result of discrimination would be propagated back to then generative and discriminative networks to instruct the parameter modification.

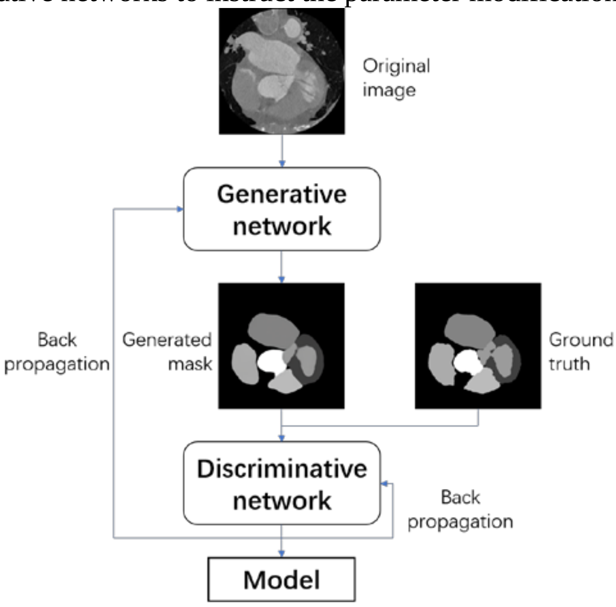



**Figure 1.** The Workflow of our Method (R2Unet-GAN)

### *3.1. Generative network*

Recurrent residual convolutional Neural network (R2U-Net) [12] is chosen as the generative network in this frame. As shown in Figure. 2, the network architecture of R2U-Net is like that of a basic U-Net. Replacing all regular forward convolutional layers by recurrent residual blocks makes the R2U-Net model deeper than original U-net. Skip-connection parts in regular U-Net model are also modified here, where the cropping and copying units are removed. In this model, the features are accumulated with different time-step, which makes the feature representation better and makes sense of low-level feature extraction.

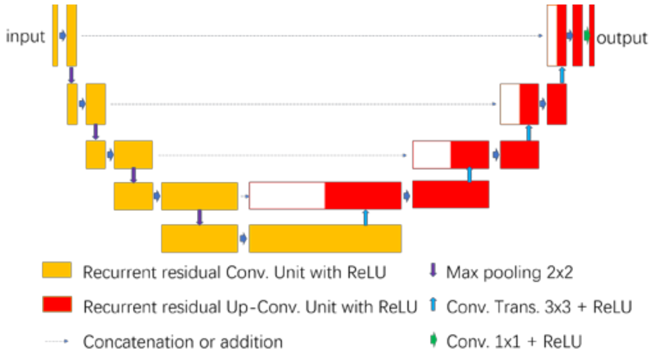



**Figure 2.** The Network Architecture of the Generative Network (R2-Unet)

### *3.2. Discriminative network*

The regular GAN discriminator is not suitable for the area of image segmentation which requires high resolution and more details. PatchGAN in pixel-to-pixel frame, another extending of cGAN, can judge and penalize the structure in each piece of patch with an exact size [13]. In this frame, an FCN [14] which has 3 layers with 1x1 kernel and 1x1 stride replaces the PatchGAN with 1x1 patch. As shown in Figure 3.

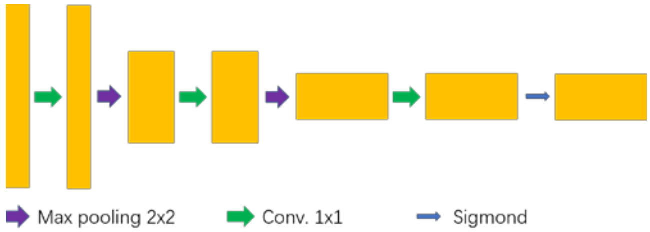



**Figure 3.** The Network Architecture of the Discriminative Network (FCN)

### *3.3. Learning rate policy*

Stochastic Gradient Descent with Warm Restarts (SGDR) shows a more effective schedule called cosine decay, start from a large learning rate, relatively quickly reduce to a minimum, and then quickly increase [9][15]. The cosine decay is like this in Figure 4.

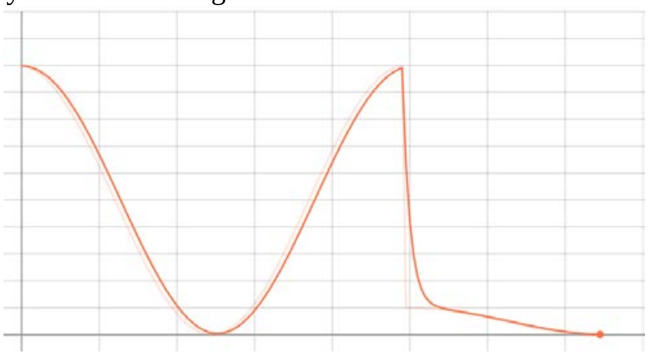


**Figure 4.** Cosine Decay

## 4. Results of the proposed method

### *4.1. Datasets and pre-processing*

The datasets we used are presented by MM-WHS Challenge 2017 [16,17]. There are 20 sets of labeled cardiac CT slices with a pixel resolution of 512 x 512. Among these data, about a number between 40 and 80 of slices of each set have a pure dark label, which means that any part of heart does not appear in these slices. To increase the adaptability of our model in complex environments, we chose to remain these slices.

We extracted the data and divided it into training data and testing data two parts. We set two groups of sets, setting No. 1 to 15 and 1 to 17 sets as the data sets for training and others are regarded as the testing data sets, to valid the effect of the mount of training data.

### *4.2. Performances and comparisons*

All models running in this work are with the PyTorch deep learning architecture on a Nvidia Tesla V100 32 GB GPU. Batch-size equals to 4, the initial learning rate is set as 0.0002. The models ran for 150 epochs.

We trained Unet-GAN with different generative networks (basic U-Net and R2U-Net), different learning rate policies (step and cosine) and different datasets (N0. 1 to 15 and No. 1 to 17).

The evaluation method we used in this paper is the Dice similarity coefficient (DSC) [18].

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Where A presents the segmentation result, and the B presents the labeled mask.

The segmentation results of the number 205 slice from Heart No. 20 are shown in Figure 5. With the high accuracies, each model outputs great results. If we observe carefully, several differences among these results can be pointed. With the same learning rate policy, the segmentation results of R2Unet-GAN are little better than these of Unet-GAN. With the same generative network, models with cosine decay perform better than models with step decay obviously.

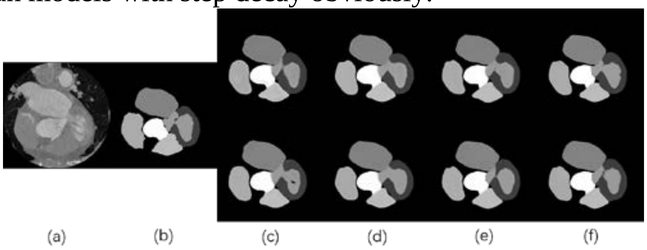
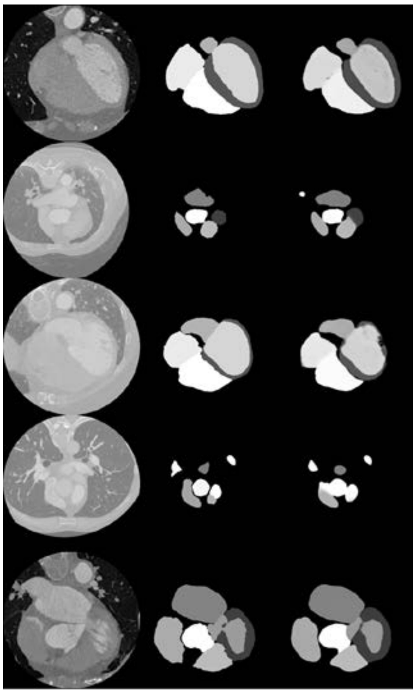



**Figure 5.** (a) Original Cardiac CT Slices (b) Labeled Masks, Segmentation Results of Experiments with different configurations (c) Unet-GAN + step decay, (d) Unet-GAN + cosine decay, (e) R2Unet-GAN + step decay, (F) R2Unet-GAN + cosine decay, upper four training on 15 sets and the below four training on 17 sets

Figure 6 shows the specific segmentation results of R2Unet-GAN, the shown slices are chosen randomly from each set. (slice number 131 from heart 16, slice number 174 from heart 17, slice 86 from heart 18, slice 231 from heart 19 and slice 205 from heart 20).



(a) (b) (c)

**Figure 6.** (a) Original Cardiac CT Slices (b) Labeled Masks (c) Results of R2Unet-GAN

**Table 1**. The DSC scores of R2Unet-GAN + cosine (%)

| Heart No. | 16 | 17 | 18 | 19 | 20 | Avg. |
|---|---|---|---|---|---|---|
| R2Unet-GAN | 80.8 | 88.5 | 90.3 | 91.1 | 93.8 | 88.9 |

The performance of our method (R2Unet-GAN + cosine) running on 15 sets are shown in Table 1. From the table we can see, the model has difference performance while inputting different cardiac CT sets. For example, the DSC score of heart 20 is up to 0.938 but that of heart 16 is only 0.808. Because of the differences among the types, contrasts, or localizations of different cardiac CT sets, the accuracy of segmentation model would change if the list of sets chosen to act as training data changed.

The average DSC scores of all testing data of experiments whose results shown in Figure 5 are shown in Table 2. With especial DSC scores, we can much more intuitively see that R2Unet-GAN performs better than Unet-GAN, cosine decay helps more than step decays and more training slices makes the accuracy up, of course.

Data in Table 3 is the DSC of the whole heart (WH) segmentation results of different methods, including the proposed method and methods related in section 2. And the comparative results showed that R2Unet-GAN performs well.

**Table 2.** The average DSC of models with different generative networks, learning rate policies and datasets (%)

| Datasets | 15 training sets | | 17 training sets | |
|---|---|---|---|---|
| LR policy | step | cosine | step | cosine |
| Unet-GAN | 86.6 | 87.9 | 90.6 | 91.2 |
| R2Unet-GAN | 87.3 | 88.9 | 90.9 | 91.5 |

**Table 3.** The comparison among different methods (%)

| | Dice of WH |
|---|---|
| U-Net | 78.5 |
| 2-stage U-Net | 79.3 |

| | |
|---|---|
| 3D U-Net | 89.0 |
| R2Unet-GAN | 88.9 |

## 5. Conclusion
This paper proposed an improved Unet-GAN model, R2Unet-GAN, where the generator is R2U-Net and the discriminator is FCN, the learning rate policy is cosine decay. The accuracy increased up to 88.9% while training on 15 sets (the max accuracy of single set is 94.0%). Although the performance is not bad, still many challenges need to be deal with. The difference between the performance of different Unet-GAN architectures with different generators and learning rate policies are not obvious enough. Because of the huge difference between different cardiac CT images, the model did not perform well for some of the testing data.

As the future work, we will extend the method to segment each parts of heart from cardiac CT images, aiming to increasing the accuracy. Focusing on slices with long-scale contrast, we will optimize the algorithm with novel pre-processing.

## Acknowledgments
This research is supported by The Science and Technology Development Fund, Macau SAR (File no. 0009/2018/A).

## References
[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. *Springer*, 20.
[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *In Advances in neural information processing systems*, pages 2672–2680, 20.
[3] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D u-net: learning dense volumetric segmentation from sparse annotation. *In International conference on medical image computing and computer-assisted inter-vention*, pages 424–432. Springer, 20.
[4] Marija Habijan, Hrvoje Leventi ́c, Irena Gali ́c, and Danilo Babin. Whole heart segmentation from ct images using 3d u-net architecture. *In 2019 International Conference on Systems, Signals, and Image Processing (IWSSIP)*, pages 121–126. IEEE, 20.
[5] Tao Liu, Yun Tian, Shifeng Zhao, Xiaoying Huang, and Qingjun Wang. Automatic whole heart segmentation using a two-stage u-net framework and an adaptive threshold window. *IEEE Access*, 7: 83628–83636, 201.
[6] Zhanwei Xu, Ziyi Wu, and Jianjiang Feng. Cfun: Combining faster r-cnn and u-net network for efficient whole heart segmentation. *arXiv preprint* arXiv:1812.04914, 2018.
[7] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:*1411.1784, 2014.
[8] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia, and Qian Tao. The domain shift problem of medical image segmentation and vendor-adaptation by unet-gan. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 623–631. *Springer*, 20.
[9] Mengtian Li, Ersin Yumer, and Deva Ramanan. Budgeted training: Rethinking deep neural network training under resource constraints. *arXiv preprint arXiv*:1905.04753, 2019.
[10] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv*:1611.01578, 2016.
[11] Hieu Pham, Melody Y Guan, Barret Zoph, Quoc V Le, and Jeff Dean. Efficient neural architecture search via parameter sharing. *arXiv preprint arXiv:*1802.03268, 2018.

[12] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan KAsari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:*1802.06955, 2018.
[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 20.
[14] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* pages 3431–3440, 20.
[15] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:*1608.03983, 2016.
[16] Xiahai Zhuang. Challenges and methodologies of fully automatic whole heart segmentation: a review. *Journal of healthcare engineering*, 4, 2013.
[17] Xiahai Zhuang, Wenjia Bai, Jingjing Song, Songhua Zhan, Xiaohua Qian, Wenzhe Shi, Yanyun Lian, and Daniel Rueckert. Multiatlas whole heart segmentation of ct data using conditional entropy for atlas ranking and selection. *Medical physics,* 42(7): 3822–3833, 201.
[18] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology,* 26(3): 297–302, 194. Conference Short Name: WOODSTOCK' 18.