# CANet: Compact Attention Network for Automatic Melanoma Segmentation

Yingyan Hou[1(✉)] and Kaichuang Liu[2]

[1] School of Software, Tsinghua University, Beijing 100084, China
hyy20@mails.tsinghua.edu.cn
[2] Department of Automation, Tsinghua University, Beijing 100084, China
liukc20@mails.tsinghua.edu.cn

**Abstract.** In the study of skin cancer, particularly melanoma, automatic and accurate segmentation as a crucial step in Computer-Aided Diagnosis (CAD) provides a reliable basis for clinical diagnosis efficiency and pathology research. However, due to the variability of skin lesions in texture, shape, and complex boundaries, automatic and accurate segmentation is still an unsolved challenge. In this paper, we propose a new automatic segmentation network for melanoma segmentation named Compact Attention Network (CANet). Based on the fully convolutional networks, the CANet removes down-sampling so as not to reduce the spatial accuracy. The CANet expands the receptive field by the designed atrous convolution, which could avoid the gridding issue. In order to refine the feature map information and make the segmentation edge smoother, we add an attention module after every designed atrous convolution. Finally, our model achieves State-of-the-Art (SOTA) performance in the task of melanoma segmentation compared with U-net, SegNet, FrCN, and so on. We conduct ablation experiments to prove the effectiveness of each element of the network. Our results show that the melanoma segmentation of the CANet is 91.7% in Sensitivity and 90.7% in Dice scores for the International Skin Imaging Collaboration (ISIC) test dataset. The CANet outperforms FrCN, U-Net, SegNet, Mask R-CNN and nnU-Net in Dice by 3.6%, 14.5%, 8.6%, 5.4% and 1.7% respectively, exhibiting better performance than these classic networks. The CANet can perform medical image segmentation more accurately and quickly, provide an important reference for medical workers in diagnosing diseases, and improve diagnosis efficiency.

**Keywords:** Melanoma segmentation · Atrous convolution · Attention mechanism

## 1 Introduction

Semantic segmentation of medical images is one of the important steps in artificial intelligence-assisted medical diagnosis and has been widely used in the

Hunan Key Research and Development Program (2019WK2072).


X. Gao et al. (Eds.): MobiHealth 2021, LNICST 440, pp. 329–342, 2022.
https://doi.org/10.1007/978-3-031-06368-8_22

field of medical image analysis. Using medical image segmentation technology to extract human tissues, organs, and lesions, which provides important references for medical workers to diagnose diseases, reduces the misdiagnosis rate, and improves diagnostic efficiency.

Cancer is one of the leading causes of human unnatural death. The World Health Organization (WHO) has published the latest world cancer report for 2020 [1]. The report shows that there were 19.3 million newly diagnosed cancer cases worldwide in 2020, with nearly 10 million deaths. Skin cancer is one of the most common cancers. The majority of skin cancer deaths are caused by melanoma. Melanoma has been reported to grow at an annual rate of 7% at the American Society of Clinical Oncology (ASCO) annual meeting. Early diagnosis and identification of melanoma are increasingly important.

Early symptoms of melanoma are difficult to distinguish from benign skin lesions on the epidermis, leading to the misdiagnosis of melanoma. Melanoma or benign skin lesions have irregular colors and shapes, and manual segmentation is cumbersome and time-consuming. Therefore, it is necessary to study an automatic and accurate method for melanoma segmentation. In recent years, semantically segmented networks have been commonly used for melanoma segmentation. In earlier studies, there have been many traditional image segmentation algorithms to solve this problem. In 2009, Yuan et al. [3] proposed a skin lesion segmentation method based on region fusion and narrowband energy graph partitioning, which can handle topological changes, weak edges, and asymmetric skin lesion areas, and can accurately detect complex outlines of lesion areas. In 2011, Schaefer et al. [4] proposed a technique for extracting skin lesion areas using an iterative measurement of non-lesion pixels and co-operative neural networks. In the same year, Zhou et al. [5] proposed a gradient vector flow algorithm based on the mean drift to extract the contours of skin lesions. Wong et al. [6] proposed an iterative random area merging method to extract lesion areas from conventional macro-skin images. However, traditional image segmentation algorithms can only segment-specific cases, and neither speed nor accuracy is very high.

In recent years, deep learning has made remarkable progress in the field of computer vision. It has been widely applied in the field of image, and its effect is higher than the previous SOTA performance, which provides inspiration for the segmentation of skin lesions. Long et al. [7] proposed a Fully Convolution Network (FCN), which removed the original full connection layer of the convolution neural network and replaces it with transposed convolution layer. FCN has two main points, one is to expand the receptive field through the pooling layer, the other is to expand the size of the image through up-sampling. Ronneberger et al. [8] proposed U-net, which applied the results of the pooling layer to the decoding process and introduces more coding information. The U-net is a very classical medical image segmentation model because it can be trained with a small amount of data. Zhao et al. [9] proposed a pyramid scene analysis network that used a pooling layer, which had a large core to expand the receptive field. He et al. [10] proposed Mask R-CNN based on Fast R-CNN, which can achieve high-quality

semantic segmentation. Peng et al. [11] proposed an encoder-decoder architecture with a large convolution core, which used ResNet [12] structure as the encoder and used the graph convolution network [13] as the decoder. Badrinarayanan et al. [14] proposed an encoder-decoder architecture for semantic pixel-wise segmentation termed SegNet, in which the decoder up-samples its lower resolution input feature maps. Chen et al. [15] proposed Atrous Spatial Pyramid Pooling (ASPP), which could combine information of different sizes. Al-Masni et al. [16] proposed FrCN using the full spatial resolution of the input image to reduce the loss of information. Roy et al. [2] improved the Squeeze & Excitation (SE) module [17] in the field of image classification, and proposed the scSE module that matched semantic segmentation, which greatly improved the accuracy of semantic segmentation. Stringer et al. [18] introduced a generalist, deep learning-based segmentation method called Cellpose, which can precisely segment cells from a wide range of image types and does not require model retraining or parameter adjustments. Fabian et al. [19] proposed the nnU-Net, a deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing for any new task.

Most semantically segmented networks down-sample the feature maps of the middle layer and use the full convolution network of encoder-decoder to expand the receptive field of the captured image context. Currently, the common improvement directions are enlarging the receptive field by pooling layers and recovering the input quality by up-sampling. However, the pooling layers lose precise location information and reduce accuracy. The up-sampling layer cannot be fully recovered, and it also incurs additional computative costs.

In this paper, we propose a compact attention network architecture without down-sampling, different from common semantic segmentation. This structure adds the atrous convolution to expand the receptive field. We design the rate of atrous convolution to avoid the gridding issue and add scSE module [2] behind the convolution layers to improve the feature map. In order to further improve performance, we use pre-processing methods such as data augmentation, erosion, and dilation. In Sect. 3, we design ablation experiments and compare our network with existing models (such as U-net). Compared with existing classic models, the CANet has the best advanced performance in melanoma segmentation. The proposed method is innovative in the field of medical image segmentation. The CANet can perform medical image segmentation more accurately and quickly, provide an important reference for medical workers in diagnosing diseases, and improve diagnosis efficiency.

## 2 Methodology

We propose a method for skin lesion segmentation, which integrates the proposed new compact fully convolutional network with attention module and data processing methods oriented to the features of the ISIC dataset.

### 2.1 On the Elements of the Proposed Network

**Atrous Convolution.** In the pixel-wise semantic segmentation task, as mentioned above, most of these network architectures have encoder-decoder architecture. It lowers spatial resolution and cannot completely restore it. Instead, Li et al. [20] proposed a novel 3D architecture that incorporated high spatial resolution feature maps throughout the layers. They designed a compact network architecture without down-sampling for the segmentation of volumetric images. This architecture used atrous convolution to expand the receptive field instead of the pooling layer.

Part of our network architecture draws inspiration from it. We build a compact convolutional neural network with atrous convolution. Atrous convolution maintains image resolution and computes with a high spatial resolution by inserting "holes" between pixels in convolutional kernels. The apparent advantage of atrous convolution is that it enlarges the size of the receptive field without losing spatial resolution. Chen et al. [15] used atrous convolution with up-sampled kernels for semantic image segmentation. Atrous convolution can be used to produce accurate dense predictions and detailed segmentation along object boundaries. For example, the atrous convolution rate is set to 2, the receptive field of each convolution is $3 \times 3$, and the receptive field of the entire convolution kernel is $7 \times 7$. Furthermore, it has been applied to a broader range of tasks, such as optical flow [21].

**Gridding.** Stacking convolution layers with the same atrous rate would have an effect on the receptive field due to the gridding. The receptive field from the same atrous convolution covers an area with non-zero values. The atrous convolution can be regarded as the standard convolution on different feature maps. If the pixels of each feature map has no further interaction, it will cause discontinuity of pixel connection. Besides, repeated stacking would further aggravate the gridding. In the convolution kernel with a large atrous rate, the receptive field is too sparse to cover any information.

Wang et al. [22] alleviated this issue by using a range of relatively prime atrous rates instead of the same atrous rate (see Fig. 1). They noted that the gridding issue would still exist if a series of atrous rates have a common factor relationship (such as 2, 4, 8, etc.). We would find that there are many pixels in the receptive field that are not used and a lot of holes appear. Proper atrous rates can increase the receptive field effectively.

**Attention - ScSE Module.** Hu et al. [17] proposed an architectural component, SE block, which could be integrated within any convolutional neural network model. Inspired by SE, Roy et al. [2] proposed three block for image segmentation:

– Spatial squeeze and channel excitation block - cSE as shown in Eq. 1.

$$\hat{\mathbf{I}}_{cSE} = F_{cSE}(\mathbf{I}) = [\sigma(\hat{z_1})\,\mathbf{i}_1, \sigma(\hat{z_2})\,\mathbf{i}_2, \cdots, \sigma(\hat{z_C})\,\mathbf{i}_C] \tag{1}$$

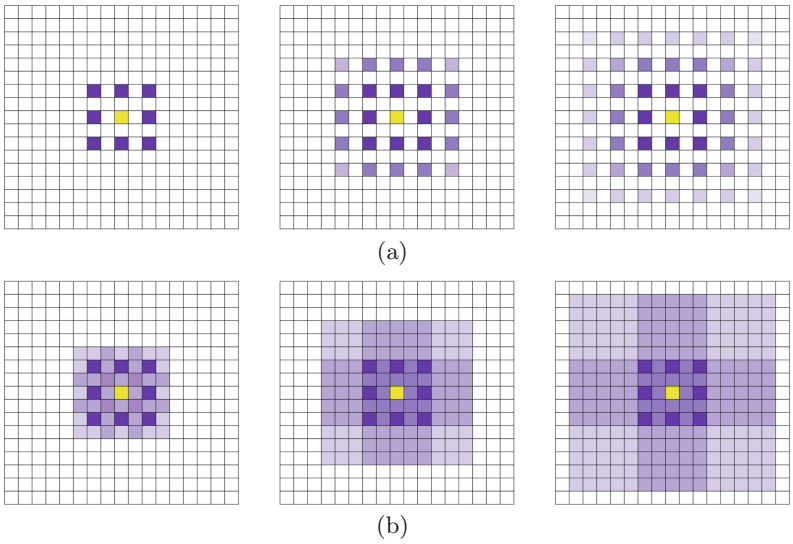



**Fig. 1.** Illustration of the gridding issue: (a) all convolutional layers have a atrous rate of 2; (b) avoid the gridding issue by setting relatively prime atrous rates.

where $I$ is the input feature map $I = [i_1, i_2, ..., i_c]$, $\sigma()$ is a sigmoid layer, $\hat{z_k}$ is performed by a global average pooling layer and weights.

– Channel squeeze and spatial excitation block - sSE as shown in Eq. 2.

$$\hat{\mathbf{T}}_{sSE} = F_{sSE}(\mathbf{T}) = \left[\sigma\left(q_{1,1}\right)\mathbf{t}^{1,1}, \cdots, \sigma\left(q_{i,j}\right)\mathbf{t}^{i,j}, \cdots, \sigma\left(q_{H,W}\right)\mathbf{t}^{H,W}\right] \quad (2)$$

where $T$ is the input tensor $\mathbf{T} = \left[\mathbf{t}^{1,1}, \mathbf{t}^{1,2}, \cdots, \mathbf{t}^{i,j}, \cdots, \mathbf{t}^{H,W}\right]$, $\sigma()$ is a sigmoid layer, $q_{i,j}$ represents all channels linear combination of a spatial location $(i, j)$.

– Patial and channel squeeze & excitation -scSE is element-wise addition of cSE and sSE as shown in Eq. 3:

$$\hat{\mathbf{U}}_{scSE} = \hat{\mathbf{U}}_{cSE} + \hat{\mathbf{U}}_{sSE} \quad (3)$$

We incorporate scSE module within the CANet.

**Loss Function.** We design the loss function due to unbalanced data. Inspired by Support Vector Machine (SVM) soft margin classification and large margin classification [23], the hinge loss function is uesd for the last layer of the network without any activation function.

We record positive and negative samples as $\pm 1$. In the training data, a n-voxel volume $\{x_i\}_{i=1}^{n}$ and its segmentation map $\{y_i\}_{i=1}^{n} (y_i \in \{-1, 1\})$ are calculated in the hinge loss function as shown in Eq. 4:

$$L_{hinge}(\{x_i\},\{y_i\}) = \sum_{i=1}^{n} [1 - y_i\,(w \cdot x_i + b)]_+ + \lambda \|w\|^2 \tag{4}$$

$$[R]_+ = \begin{cases} R, R > 0 \\ 0, R \leq 0 \end{cases}$$

where the first term $\sum_{i=1}^{n} [1 - y_i\,(w \cdot x_i + b)]_+$ denotes the loss, and the second term $\lambda \|w\|^2$ denotes the regularization term.

However, the training data is apparently unbalanced, which is typical of medical image segmentation. Equation 4 leads to a strongly biased estimation towards the majority class, which is the non-melanoma area. Thus the hinge loss function adds weight $\alpha$ as shown in Eq. 5:

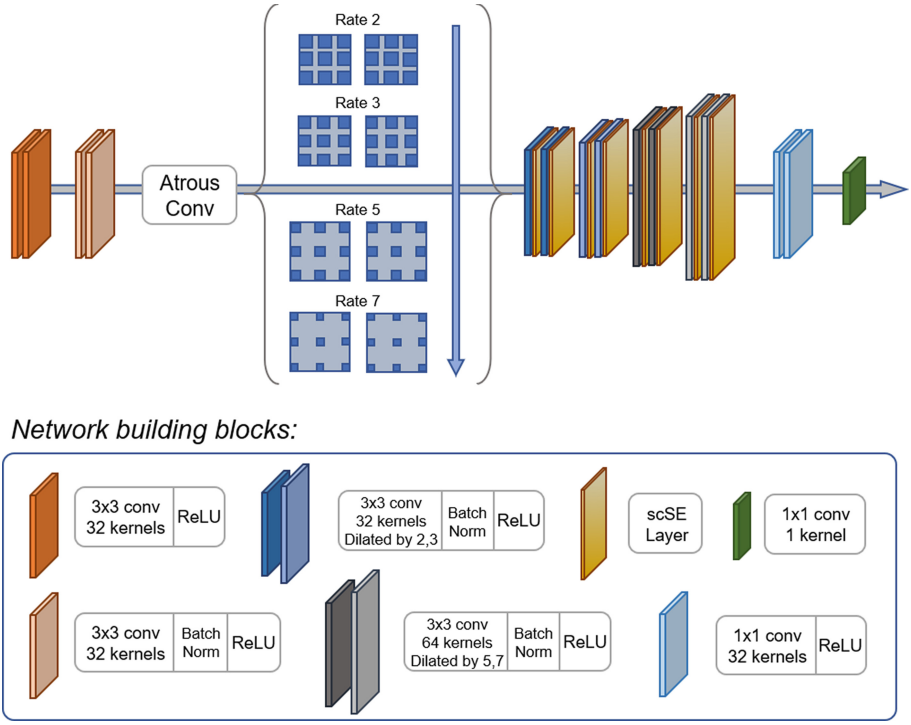



**Fig. 2.** Our network architecture for image segmentation.

$$L^*_{hinge}(\{x_i\},\{y_i\}) = \begin{cases} \sum_{i=1}^{n} \alpha \cdot [1 - y_i\,(w \cdot x_i + b)]_+ + \lambda \|w\|^2, y_i = 1 \\ \sum_{i=1}^{n} [1 - y_i\,(w \cdot x_i + b)]_+ + \lambda \|w\|^2, y_i = -1 \end{cases} \tag{5}$$

By choosing $\alpha > 1$, the network could pay more attention to the segmentation target and make parameters more actively optimized, so that the training network can get a reasonable segmentation model faster. To a certain extent, loss function weight could improve performance. We conduct experiments to discuss the choice of $\alpha$ in Sect. 3.

### 2.2 Network Architecture

The proposed network consists of 13 layers of convolutions (see Fig. 2). The first four convolution layers use $3 \times 3$ pixel convolutions without atrous convolution. Stacking two layers of $3 \times 3$ pixel convolution gives the same receptive field as a convolutional kernel with $5 \times 5$ pixel convolution, but parameters of them are only about half of $5 \times 5$ pixel convolution parameters. The third and fourth layer are associated with the Batch Normalization (BN) [24]. The first four layers are designed to capture low-level features of the input image.

In the subsequent 8 convolutional layers, each convolutional layer is associated with an element-wise Rectified Linear Unit (ReLU) layer, a BN layer and a scSE layer. A scSE layer could get a more accurate calibrated feature map. The kernels are dilated by the designed rates of 2, 3, 5, 7, which not have a common factor relationship. Choosing proper rates can effectively expand the receptive field and avoid the gridding issue [22]. We use atrous convolution instead of the pooling layer to efficiently enlarge the receptive field without reducing resolution. Due to the different sizes of the melanoma, the rates of atrous convolutions are gradually increased to incorporate features at multiple scales when the layer goes deeper. Except for the last layer of our network, every convolutional layer is associated with a ReLU layer. The final layer gives binary classification labels for every pixel.

### 2.3 Data Preprocessing

In the experiment, each network uses the processed training data set for training.

**Erosion and Dilation of Image Noise.** The ISIC dataset has hair noise in some dermoscopy images. Segmenting the melanoma, which is black and has many forms, requires minimizing the effects of hair noise.

We perform closing by erosion and dilation to remove the noise on the ISIC dataset. Erosion and dilation are mathematical morphology transformations [25].

Erosion uses vector subtraction as shown in Eq. 6.

$$M \ominus N = \{x, y \mid (N)_{xy} \subseteq M\} \tag{6}$$

where N denotes a structural element. M is the region to erode whose pixel values are all 1 in the binary image. It should be noted that N needs to define an origin, whose coordinate is (x, y). During the movement process of N, when the pixels of N are completely contained by M, the pixel of M covered by the origin of N is set to 1 otherwise 0.

Dilation uses vector addition as shown in Eq. 7.

$$M \oplus N = \{x, y \mid (N)_{xy} \cap M \neq \emptyset\} \tag{7}$$

where N is as the same definition as Eq. 6, and M denotes the region to dilate. During the movement process of N, when pixels of N have intersections with those of M, the pixel of M covered by the origin of N is set to 1 otherwise 0.

**Data Augmentation.** Training a well-performing network commonly requires a large amount of data, but datasets in the medical field are generally small. On a small dataset, the model would easily overfit. Data augmentation is a conventional method for training models with a small dataset. For each image in the training dataset, we use three augmentation functions from the following list:

- Herizontal Flip.
- Vertical Flip.
- Rotation with angle 180°.

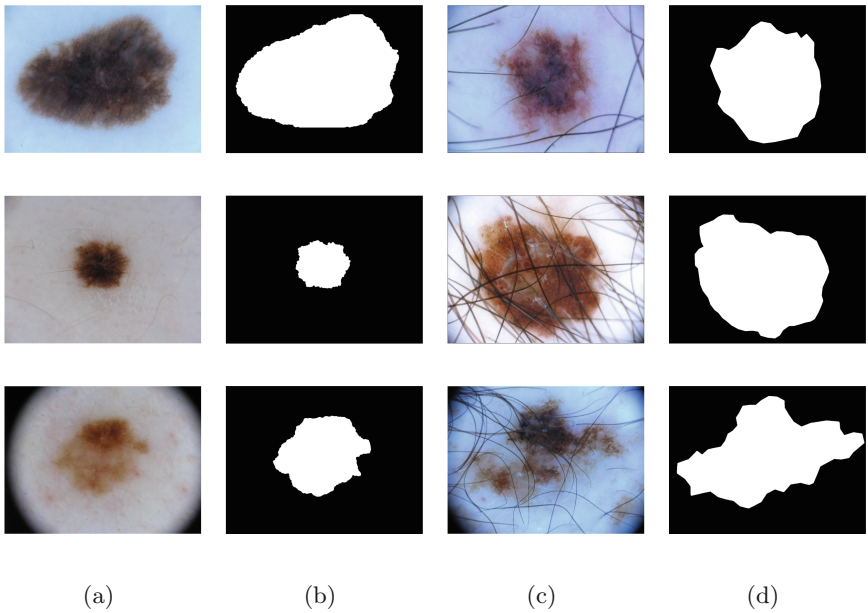
## 3 Experiments

This section introduces the dataset, evaluation metrics, and experiment configurations. We conduct experiments to explore our proposed network, CANet. The experiments are presented in two parts. The first part is ablation experiments to prove the effectiveness of each element of the network. In the second part, we compare our method with the SOTA networks.

In this paper, the experiments are performed and tested with the following configurations: Intel Core i9 processor at 3.5 GHz, NVIDIA GeForce RTX 2080 GPU with Ubuntu 18.04. The learning rate is set to 0.0001 in all experiments.

### 3.1 Dataset

The ISIC is a well-known skin medical organization, which collects a large number of skin images and provides annotation for these images. The organization holds a skin lesion challenge every year to attract researchers in the field of computer vision to participate in, so as to improve the recognition algorithm of skin lesions and make more people realize the harm of skin cancer. In this paper, we used ISIC 2017 challenge dataset. There are 2000 skin images and the corresponding skin lesion area images marked by experts in the training set. The ratio of the training set, validation set, and testing set is 40:3:12, and the resolution of each image varies from 540 × 722 to 4499 × 6748 pixels. The dataset is a collection of real images, some of which have serious noise (see Fig. 3).



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

**Fig. 3.** Various skin lesion images in ISIC. (a) noise-free images. (b) the annotation of noise-free images. (c) noisy images. (d) the annotation of noisy images.

### 3.2 Performance Evaluation Metrics

A quantitative analysis of experiments is carried out based on True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) [26]. In this work, image segmentation methods are evaluated according to the following evaluation metrics:

- **Sensitivity**.

$$Sensitivity = \frac{TP}{TP + FN}$$

- **Specificity**.

$$Specificity = \frac{TN}{FP + TN}$$

- **Accuracy**.

$$Accuracy = \frac{TP + TN}{FP + TN + TP + FN}$$

- **Jaccard Similarity Index (JSI)**.

$$JSI = \frac{TP}{FP + TP + FN}$$

- **Dice Similarity Coefficient (Dice)**.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}$$

### 3.3 Results and Discussions

In this section, we first perform ablation experiments on our network to prove the effectiveness of each element. Then we compare our model with the SOTA models in the industry. We show results and draw box-plots to visually highlight the superior performance of our proposed method.

**Ablation Experiments.** The ablation experiments are as follows:

- The CANet without atrous convolution.
- Atrous rates of 2, 4, 6, 8, exists the gridding issue.
- The CANet without scSE module.
- Loss function without weight $\alpha$.
- Loss function weight $\alpha = 4$.
- Loss function weight $\alpha = 6$.
- The CANet ($\alpha = 2$).

Table 1 lists the mean Sensitivity, Specificity, Accuracy, Jaccard Similarity Index, and Dice scores on the test set of the ablation experiments. Comparing along the rows with the final row, it is proved that the designed atrous convolution rates 2, 3, 5, 7 is effective, which expands the receptive field and reduces the gridding effect. Attention module scSE provides a statistically significant increase in comparison to the CANet. In addition, it is necessary to set the weight of the loss function, and the weight $\alpha = 2$ is the optimal solution for the melanoma data set. Fig 4 visualizes the results of ablation experiments under the Dice scores. The CANet outperforms "No Atrous Convolution", "Rates: 2, 4, 6, 8", "No scSE Module", "No Weight $\alpha$", "Weight $\alpha = 4$" and "Weight $\alpha = 6$" in Dice by 4.8%, 1.5%, 4.0%, 2.3%, 2.7% and 3.4% respectively, showing the effectiveness of each element and great segmentation performance with the melanoma in CANet.

**Table 1.** Ablation experiments.

| Strategy | Evaluation metrics | | | | |
|---|---|---|---|---|---|
| | Sensitivity | Specificity | Accuracy | Jaccard | Dice scores |
| No Atrous convolution | 0.877 | 0.940 | 0.923 | 0.754 | 0.859 |
| Atrous rates: 2, 4, 6, 8 (gridding) | 0.904 | 0.953 | 0.939 | 0.806 | 0.892 |
| No scSE module | 0.879 | 0.938 | 0.920 | 0.765 | 0.867 |
| Loss function no weight $\alpha$ | 0.896 | 0.950 | 0.935 | 0.793 | 0.884 |
| Loss Weight $\alpha = 4$ | 0.892 | 0.949 | 0.933 | 0.786 | 0.880 |
| Loss Weight $\alpha = 6$ | 0.886 | 0.949 | 0.932 | 0.776 | 0.873 |
| CANet ($\alpha = 2$) | **0.917** | **0.955** | **0.944** | **0.830** | **0.907** |

The best performance is highlighted in boldface.

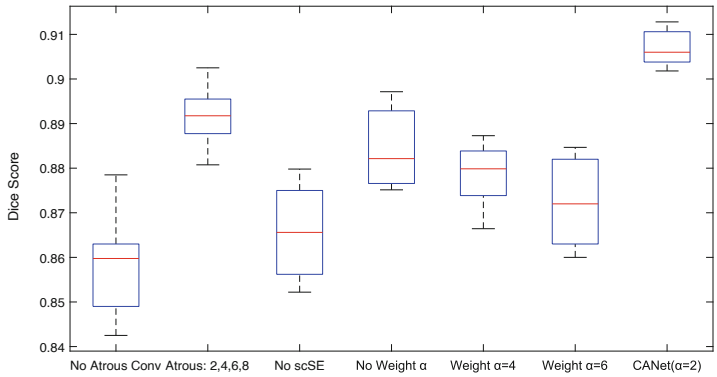



**Fig. 4.** Boxplot of Dice scores for ablation experiments.

**Comparative Experiments.** The CANet compares with the U-net [8], the SegNet [14], the FrCN [16], the Mask R-CNN [10] and the nnU-Net [19]. The required network parameters of these algorithms are provided in their papers.

**Table 2.** Results among CANet, Mask R-CNN, U-net, FrCN, SegNet, nnU-Net.

| Network | Evaluation metrics | | | | |
|---|---|---|---|---|---|
| | Sensitivity | Specificity | Accuracy | Jaccard | Dice scores |
| Mask R-CNN [10] | 0.848 | 0.960 | 0.935 | 0.743 | 0.853 |
| U-net [8] | 0.672 | **0.972** | 0.901 | 0.616 | 0.762 |
| FrCN [16] | 0.854 | 0.967 | 0.940 | 0.771 | 0.871 |
| SegNet [14] | 0.801 | 0.954 | 0.918 | 0.696 | 0.821 |
| nnU-net [19] | 0.899 | 0.958 | 0.943 | 0.802 | 0.890 |
| CANet | **0.917** | 0.955 | **0.944** | **0.830** | **0.907** |

The best performance is highlighted in boldface.

Table 2 and Fig 5 compare the performance on the test set. Table 2 directly shows that our method achieves the best performance under the four metrics of Sensitivity, Accuracy, Jaccard Similarity Index, and Dice scores, and there is a small gap between our method and the best performance under the metric of Specificity. Results show that the melanoma segmentation of the CANet is 91.7% in Sensitivity and 90.7% in Dice scores for the ISIC test dataset. The CANet outperforms FrCN, U-Net, SegNet, Mask R-CNN and nnU-net in Dice by 3.6%, 14.5%, 8.6%, 5.4% and 1.7% respectively, exhibiting better performance than these classic networks. Figure 5 visualizes the results of five networks under the Dice scores. The CANet is much better than other networks in terms of mean value and stability. There is little difference between the boundary line segmented by CAnet and the boundary line marked by experts, which would be

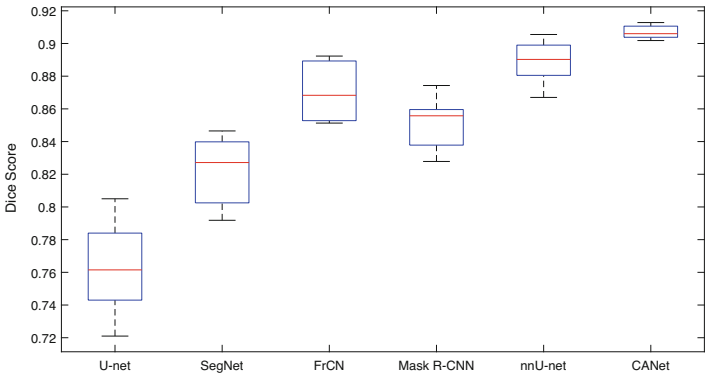



**Fig. 5.** Boxplot of Dice scores for six networks.

of great help to the subsequent doctors' judgment and diagnosis of the lesion area (see Fig. 6).

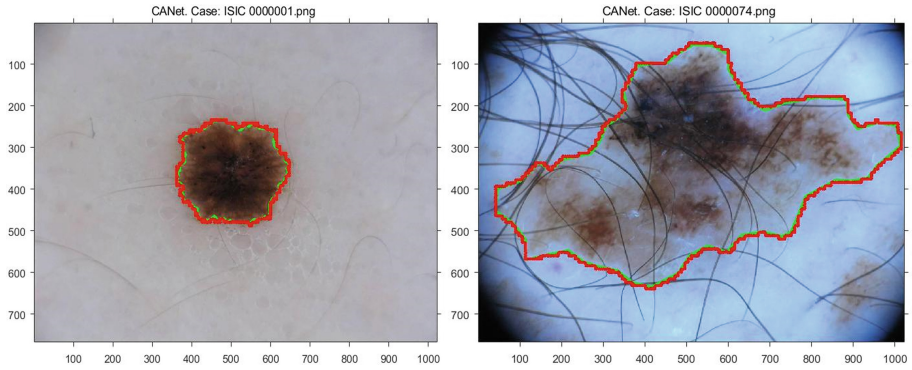



**Fig. 6.** Visual lesion segmentation of CANet. The segmentation results of the ground truth (green) and CANet (red). (Color figure online)

## 4 Conclusions

In this paper, we propose a compact attention network architecture that incorporates attention module scSE and designed atrous convolution for expanding the receptive field. On the segmentation of the melanoma, the CANet architecture performs better than the U-net, SegNet, FrCN, and nnU-net. In particular, the CANet is simpler and more compact than the SOTA segmentation network nnU-net. It is also worth noting that even in the noisy images, melanoma segmentation results are in good agreement with the ground truth. The CANet potentially provides a good point for other segmentation tasks.

In the future, we would extensively test the CANet segmentation ability in more datasets. Furthermore, we note that experts spent too much time annotating in medical images. In the subsequent research, we would try adopting a semi-automatic annotating method. The above issues would be regarded in our future research.

## References


1. World Cancer Report [DB/OL]. https://www.iarc.who.int/cards_page/world-cancer-report/. Accessed 29 Feb 2021
2. Roy, A.G., Navab, N., Wachinger, C.: Concurrent spatial and channel 'Squeeze & excitation' in fully convolutional networks. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 421–429. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_48
3. Yuan, X., Ning, S., Zouridakis, G.: A narrow band graph partitioning method for skin lesion segmentation. Pattern Recogn. **42**(6), 1017–1028 (2009)
4. Schaefer, G., Rajab, M.I., Celebi, M.E., et al.: Colour and contrast enhancement for improved skin lesion segmentation. Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc. **35**(2), 99–104 (2011)
5. Zhou, H., Schaefer, G., Celebi, M.E., et al.: Gradient vector flow with mean shift for skin lesion segmentation. Comput. Med. Imaging Graph. **35**(2), 121–127 (2011)
6. Wong, A., Scharcanski, J., Fieguth, P.: Automatic skin lesion segmentation via iterative stochastic region merging. IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc. **15**(6), 929–36 (2011)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2015)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Zhao, H., Shi, J., Qi, X., et al.: Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239. IEEE Computer Society (2017). https://doi.org/10.1109/CVPR.2017.660
10. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 386–397 (2020). https://doi.org/10.1109/TPAMI.2018.2844175
11. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters - improve semantic segmentation by global convolutional network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1743–1751. IEEE (2017). https://doi.org/10.1109/CVPR.2017.189
12. He, K., Zhang, X., Ren, S., et al. : Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE Computer Society (2016). https://doi.org/10.1109/CVPR.2016.90
13. Kip, F.T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv e-prints arXiv:1609.02907 (2016)
14. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)

15. Chen, L.C., Papandreou, G., Kokkinos, I., et al.: DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017)
16. Al-Masni, M.A., Al-Antari, M.A., Choi, M.T., et al.: Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Comput. Methods Program. Biomed. **162**, 221–231 (2018)
17. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. IEEE (2018). https://doi.org/10.1007/978-3-030-00928-1_48
18. Stringer, C., Wang, T., Michaelos, M., et al.: Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods **18**, 100–106 (2021). https://doi.org/10.1038/s41592-020-01018-x
19. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. methods **18**(2), 203–211 (2021)
20. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 348–360. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_28
21. Sevilla-Lara, L., Sun, D., Jampani, V., et al.: Optical flow with semantic segmentation and localized layers. In: 2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3889–3898. IEEE (2016). https://doi.org/10.1109/cvpr.2016.422
22. Wang, P., Chen, P., Yuan, Y., et al. : Understanding convolution for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 1451–1460. IEEE (2018). https://doi.org/10.1109/wacv.2018.00163
23. Tang, Y.: Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239 (2013)
24. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. Proc. Mach. Learn. Res. **37**, 448–456 (2015)
25. Haralick, R.M., Sternberg, S.R., Zhuang, X.: Image analysis using mathematical morphology. IEEE Trans. Pattern Anal. Mach. Intell. **9**(04), 532–50 (2009)
26. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)