Dense-U-Net: densely connected convolutional network for semantic segmentation with a small number of samples

Yuanyi Zeng, Xiaoyu Chen, Yi Zhang, Lianfa Bai, Jing Han

Yuanyi Zeng, Xiaoyu Chen, Yi Zhang, Lianfa Bai, Jing Han, "Dense-U-Net: densely connected convolutional network for semantic segmentation with a small number of samples," Proc. SPIE 11069, Tenth International Conference on Graphics and Image Processing (ICGIP 2018), 110692B (6 May 2019); doi: 10.1117/12.2524406

Event: Tenth International Conference on Graphic and Image Processing (ICGIP 2018), 2018, Chengdu, China

Dense-U-Net: Densely Connected Convolutional Network for Semantic Segmentation with a Small Number of Samples

Yuanyi Zeng, Xiaoyu Chen, Yi Zhang, Lianfa Bai, Jing Han*
Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China

ABSTRACT

The main contribution of this work is the proposal of a densely connected convolutional network for semantic segmentation, which strengthens utilization of features and improves segmentation results even with limited training samples. To achieve this, we combine the U-Net network and our resulting system is called Dense-U-Net. Compared to traditional convolutional networks such as U-Net, there are additional concatenation layers between each pair of convolutional layers which have the same size of outputs in our Dense-U-Net, each layer can get the feature-maps of all its preceding layers as inputs while its feature-maps can be passed to all subsequent layers, and a higher segmentation quality can be achieved without a need for increasing the volume of datasets finally. We evaluate our proposed architecture by segmentation accuracy, foreground-restricted rand scoring after border thinning V^{Rand} and foreground-restricted information theoretic scoring after border thinning V^{Info} at the same time, and the results are shown on three different segmentation tasks: ISBI challenge 2012 for segmentation of neuronal structures in electron microscopic stacks, ISBI cell tracking challenge 2014(Glioblastoma-astrocytoma U373 cells) and 2015(HeLa cells), our Dense-U-Net achieves better results than U-Net and several other state-of-the-art networks on all tasks.

Keywords: Densely connected convolutional network, semantic segmentation, concatenation layers, ISBI challenge.

1. INTRODUCTION

With the developments of hardware technologies, deep convolutional neural networks have been deployed in various visual recognition tasks and achieved excellent results in the recent years. One of the important applications is semantic segmentation, which is used to get a very detailed detection for shape, size and outline of the object, therefore, it needs a prediction at every pixel[1], and it is more and more important to achieve a good result even with a small number of samples.

This main contribution of this paper is an approach for introducing additional context into state-of-the-art general semantic segmentation, we enhance the information flow between layers in the networks. Under the premise of a limited number of training sets, the result achieves the current highest accuracy for semantic segmentation with a single network, and the comparable speed can also be maintained. We achieve this result combined with an excellent framework called U-Net[2], augment U-Net with additional concatenation layers between each pair of convolutional layers which have the same size of outputs to introduce additional large-scale context in semantic segmentation. Crucially, in contrast to the original U-Net, each layers gets extra inputs from preceding layers and gives its own feature-maps to subsequent layers, hence, a layer has a input that consists of the feature-maps of all preceding convolutional blocks, the whole network has more connections than the original U-Net. Because of the dense connectivity pattern, our architecture is called as Dense-U-Net.

Our proposed Dense-U-Net architecture connects all layers that have matching feature-map sizes, this improves the flow information and gradients throughout the network, each layer can has more comprehensive access to the gradients from the loss function and the original input data, leading to an implicit deep supervision, which helps the process of train to be easier and further improves the accuracy of segmentation[3]. In addition, compared with the original U-Net, our Dense-U-Net has a better effect on reducing over-fitting while the training set sizes are small.

We demonstrate state-of-the-art results on the ISBI challenge for segmentation of neuronal structures and cells[4,5,6]. With very little training data, our Dense-U-Net can reach higher accuracy and lower loss than the original U-Net, the result of segmentation is better significantly.

Tenth International Conference on Graphics and Image Processing (ICGIP 2018), edited by
Chunming Li, Hui Yu, Zhigeng Pan, Yifei Pu, Proc. of SPIE Vol. 11069, 110692B
© 2019 SPIE · CCC code: 0277-786X/19/$18 · doi: 10.1117/12.2524406

2. RELATED WORK

Semantic segmentation algorithms based on deep convolutional neural networks have been improved all the time. In the early day, one of the popular approaches is patch classification[7], each pixel was classified combined with a patch of image around it, it is defective because it requires fixed size input images and it is too slow. Later in 2014, Long et al.[8] proposed an architecture for dense predictions without any fully connected layers called fully convolutional networks (FCN), which replaces the fully connected layer in deep convolutional neural networks with a convolutional layer and outputs the spatial feature map directly, then the map will be upsampled to generate dense outputs pixel by pixel, FCN allows segmentation maps to be generated for image of any size and is much faster compared to the patch classification approach[7], but it is also limited by its fixed-size receptive, coarse label map and overly simple deconvolutional procedure[9].

Hence, there are many improved algorithms have been proposed, and around the issue of pooling layers, which increase the field of view but discard the 'where' information, two different architectures of network have developed. One use dilated convolutions and do away with pooling layers, it usually uses conditional random field[10] (CRF) to improve the segmentation, but it has much cost of computation and memory although it may achieve a good segmentation; The next approach is encoder-decoder architecture, encoder gradually reduces the spatial dimension with pooling layers and decoder gradually recovers the object details and spatial dimension, such as U-Net, which was proposed by Ronneberger et al.[2] in 2015. U-Net has been a very popular and simple approach. The architecture of U-Net is consist of a chain of convolutional layers and a chain of upsampling layers, it is efficient and can work well with very few training images, but it does not yield precise enough segmentation when applied to some complex scenes[11].

3. METHODOLOGY

3.1 Architecture

The Dense-U-Net architecture is illustrated in Figure 1. The whole network is composed of two parts: encoder and decoder (convolution and upsampling). The encoder consists of 5 pairs of convolutional layers, all convolutional layers have a size of 3×3 with no pad, the number of feature channels is same between the convolutional layers in one pair and doubled between every pair in order, simultaneously, each pair of convolutional layers is connected with another by a 2×2 max pooling operation with stride 2 for downsampling. Moreover, the critical part of Dense-U-Net is the dense blocks, which connected the convolutional layers in every pair by using a batch normalization layer and a scale layer followed by a rectified linear unit (ReLU), then the two convolutions are concatenated at last. In the general architecture, the decoder is similar to the encoder, it consists of four expansive steps, each steps composed of an upsampling of the feature map followed by a 2×2 convolutional layer that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the encoder (the crop layer is necessary depending on the size of feature map output by the convolutional layers both in encoder and decoder, it can lead to better performance at edges), and a pair of 3×3 convolutional layers, in which the convolutional layers is also connected by a dense block. At the end of network, a convolutional layer with a size of 1×1 is used to map each component feature vector to the number of classes set up in the beginning. Dense-U-Net has 23 convolutional layers and dropout layers are used to avoid overfitting.

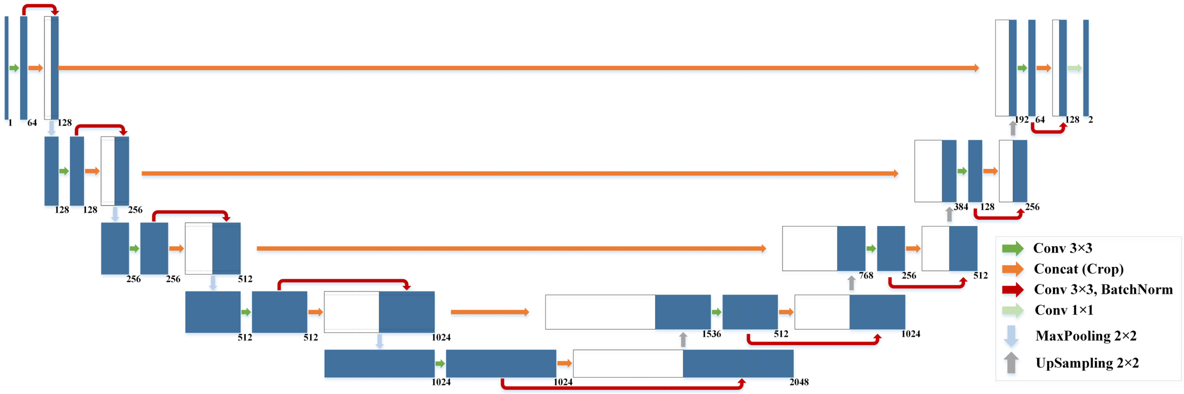


Figure 1. Detailed architecture of Dense-U-Net

3.2 Training

Our model is trained by using Adam optimizer, which is set up with a learning rate of 1e-4 and trained for 500 epochs totally (the training process has been convergent). Simultaneously, in order to make the update in the current optimization step can be determined by the previously seen training samples, we set the momentum to 0.99.

All trainings and inferences are achieve on Nvidia GTX1070 with a memory of 8G. The training parameters can apply to most experiments and only the resizing size of the input images should be chosen because of the limited memory, then the batch size is set to 2. In addition, in order to make the receptive field of network be enough large according to different experiments, it is necessary to set the resolution of the input images reasonably.

Moreover, in order to build a more valid model and avoid to be restricted by input image orientation, data augmentation is necessary when there are only few training data[1], such as rotations at different angles, horizontal flip and vertical flip. Data augmentation is useful to enhance the universality of our model.

3.3 Evaluation Methods

In order to score the segmentation qualitatively in a reasonable and effective way, except the accuracy, we also use foreground-restricted rand scoring after border thinning V^{Rand} and foreground-restricted information theoretic scoring after border thinning V^{Info} [12,13]:

$$V_{\alpha}^{Rand} = \frac{1}{\alpha \frac{1}{V_{split}^{Rand}} + (1\text{-}\alpha)\frac{1}{V_{merge}^{Rand}}}, \tag{1}$$

$$V_{split}^{Rand} = \frac{\sum_{ij} p_{ij}^2}{\sum_k t_k^2} \;,\; V_{merge}^{Rand} = \frac{\sum_{ij} p_{ij}^2}{\sum_k s_k^2} . \tag{2}$$

Where p_{ij} is the probability that a randomly chosen pixel belongs to segment i in S and segment j in T (S is the predicted segmentation and T is the ground truth segmentation, this joint probability distribution satisfies the normalization condition $\sum_{ij} p_{ij} = 1$), the marginal distribution $s_i = \sum_j p_{ij}$ is the probability that a randomly chosen pixel belongs to segment i in S, and $t_j = \sum_i p_{ij}$ is defined similarly. V_{split}^{Rand} is the probability that two randomly chosen voxels belong to the same segment in S, given that they belong to the same segment in T, V_{merge}^{Rand} is the probability that two randomly chosen voxels belong to the same segment in T, given that they belong to the same segment in S. We set $\alpha = 0.5$ to weight split and merge errors equally.

$$V_{\alpha}^{Info} = \frac{1}{(1\text{-}\alpha)\frac{1}{V_{split}^{Info}} + \alpha \frac{1}{V_{merge}^{Info}}}, \tag{3}$$

$$V_{split}^{Info} = \frac{I(S;T)}{H(S)} \;,\; V_{merge}^{Info} = \frac{I(S;T)}{H(T)} . \tag{4}$$

Where $I(S;T) = \sum_{ij} p_{ij} \log p_{ij} - \sum_i s_i \log s_i - \sum_j t_j \log t_j$ is the mutual information, V_{split}^{Info} is the information theoretic split score and V_{merge}^{Info} is the information theoretic merge score. Here we also set $\alpha = 0.5$.

4. EXPERIMENT

4.1 Datasets

For the purpose investigate the performance of Dense-U-Net and demonstrate its generality, we apply Dense-U-Net to three different tasks related to semantic segmentation. Three tasks consist of the segmentation of neuronal structures in electron microscopic recording (EM segmentation challenge started at ISBI 2012[4], the training data is a set of 30 images with a size of 512×512), the segmentation of Glioblastoma-astrocytoma U373 cells on a polyacrylimide substrate recorded by phase contrast microscopy (the training data is a set of 34 images with a size of 696×520) and the segmentation of HeLa cells on a flat glass recorded by differential interference contrast (DIC) microscopy (the training data is a set of 18 images with a size of 512×512), the latter two are the cell segmentation task in light microscopic image that are parts of the ISBI cell tracking challenge 2014 and 2015[5,6].

4.2 Results

On the three datasets mentioned above, we also use some other excellent algorithms to do the segmentation tasks and make a comparison (V^{Rand}, V^{Info} and Accuracy) with our Dense-U-Net, with this, the performance of Dense-U-Net can be showed obviously.

1) The visualization of the segmentation output of all algorithms on neuronal structures are showed in Figure 2 and their specific performance indicators are showed in Table 1.

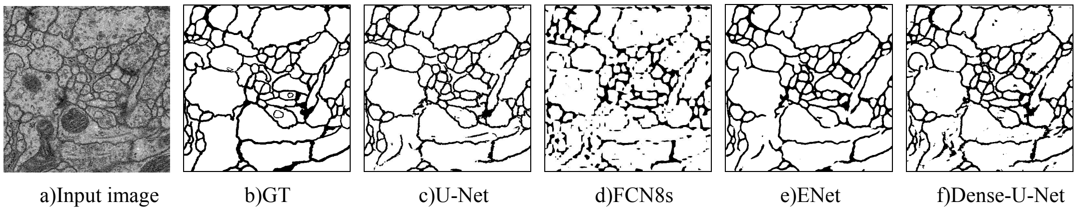


Figure 2. Segmentation results on neuronal structures

Table 1. Quantitative results on neuronal structures

Method	V^{Rand}	V^{Info}	Accuracy
U-Net	0.7334	0.9125	94.93%
FCN8s	0.1003	0.1354	87.05%
ENet	0.5863	0.8725	95.69%
Dense-U-Net	0.7412	0.9206	96.62%

2) The visualization of the segmentation output of all algorithms on Glioblastoma-astrocytoma U373 cells are showed in Figure 3 and their specific performance indicators are showed in Table 2.

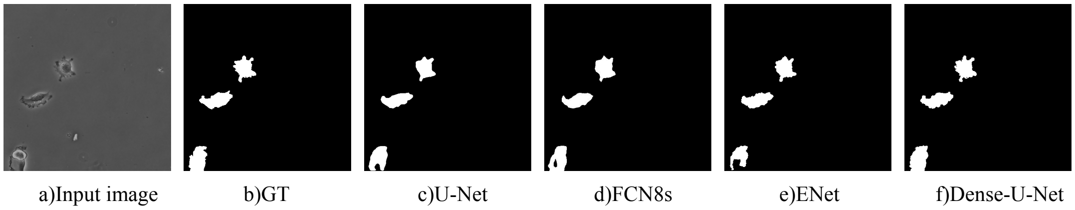
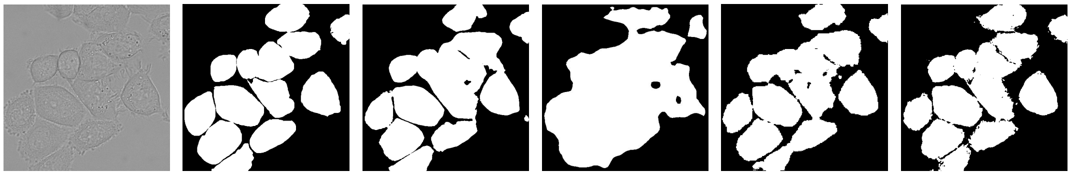


Figure 3. Segmentation results on Glioblastoma-astrocytoma U373 cells

Table 2. Quantitative results on Glioblastoma-astrocytoma U373 cells

Method	V^{Rand}	V^{Info}	Accuracy
U-Net	0.9932	0.9810	98.86%
FCN8s	0.9941	0.9829	91.23%
ENet	0.9949	0.9840	92.32%
Dense-U-Net	0.9950	0.9874	99.54%

3) The visualization of the segmentation output of all algorithms on HeLa cells are showed in Figure 4 and their specific performance indicators are showed in Table 3.



a)Input image b)GT c)U-Net d)FCN8s e)ENet f)Dense-U-Net

Figure 4. Segmentation results on HeLa cells

Table 3. Quantitative results on HeLa cells

Method	V^{Rand}	V^{Info}	**Accuracy**
U-Net	0.5676	0.4883	96.98%
FCN8s	0.5122	0.2397	86.56%
ENet	0.5940	0.5120	97.37%
Dense-U-Net	0.6017	0.5735	98.28%

5. CONCLUSION

We propose a dense connection method to minimize the loss of features which achieves very good performance on very different semantic segmentation applications with very few training images. Our new Dense-U-Net framework which is based on U-Net gets the state-of-the-art segmentation results respectively on ISBI challenge 2012 for segmentation of neuronal structures in electron microscopic stacks(segmentation accuracy reaches 96.62%, foreground-restricted rand scoring after border thinning V^{Rand} reaches 0.7412, foreground-restricted information theoretic scoring after border thinning V^{Info} reaches 0.9206), ISBI cell tracking challenge 2014 for Glioblastoma-astrocytoma U373 cells(accuracy reaches 99.54%, V^{Rand} reaches 0.9950, V^{Info} reaches 0.9874) and ISBI cell tracking challenge 2015 for HeLa cells(accuracy reaches 98.28%, V^{Rand} reaches 0.6017, V^{Info} reaches 0.5735), which are obvious better than original U-Net and several other advanced frameworks. In view of the excellent performance while the training set sizes are small, we are sure that our Dense-U-Net can be competent for more other tasks, such as PASCAL VOC and MS COCO. In addition, while we only apply our dense connection model to the U-Net framework, some other structures that enhance the information flow between layers in the networks, such as residual structure, can be also used on U-Net. Moreover, these approaches can be applied to more other semantic segmentation networks and there will be considerable performance improvement.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (61501235).

REFERENCES

[1] Khalel A. and Elsaban M., “Automatic Pixelwise Object Labeling for Aerial Imagery Using Stacked U-Nets”, (2018).

[2] Ronneberger O., Fischer P. and Brox T., “U-Net: Convolutional Networks for Biomedical Image Segmentation”, MICCAI. 9351, 234-241 (2015).

[3] Huang G., Liu Z. and Maaten L. V. D., “Densely Connected Convolutional Networks”, IEEE Trans. Computer Society. 243, 2261-2269 (2017).

[4] WWW: Web page of the em segmentation challenge, http://brainiac2.mit.edu/isbi_challenge/.

[5] Maška M., Ulman V. and Svoboda D., “A benchmark for comparison of cell tracking algorithms”, Bioinformatics. 30(11), 1609-1617, (2014).

[6] WWW: Web page of the cell tracking challenge, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html.

[7] Sasank Chilamkurthy, "A 2017 Guide to Semantic Segmentation with Deep Learning", (2017).

[8] Long J., Shelhamer E. and Darrell T., "Fully convolutional networks for semantic segmentation", IEEE Trans. Computer Society. 39, 3431-3440 (2015).

[9] Noh H., Hong S. and Han B., "Learning Deconvolution Network for Semantic Segmentation", IEEE Trans. Computer Society. 178, 1520-1528 (2015).

[10] Koltun V., "Efficient inference in fully connected CRFs with Gaussian edge potentials", ICONIP. 24, 109-117 (2011).

[11] Shah S., Ghosh P. and Davis L. S., "Stacked U-Nets: A No-Frills Approach to Natural Image Segmentation", (2018).

[12] Iglovikov V., Mushinskiy S. and Osin V., "Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition", (2017).

[13] Argandacarreras I., Turaga S. C. and Berger D. R., "Crowdsourcing the creation of image segmentation algorithms for connectomics", Frontiers in Neuroanatomy. Nov 5, 9-142 (2015).