A Dense-Gated U-Net for Brain Lesion Segmentation

Zhongyi Ji[1], Xiao Han[1], Tong Lin[*2], and Wenmin Wang[3]
[1]School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School
[2]The Key Laboratory of Machine Perception (MOE), School of EECS, Peking University; and Peng Cheng Laboratory
[3]International Institute of Next Generation Internet, Macau University of Science and Technology
jizhongyi@pku.edu.cn, hanxiao18@pku.edu.cn, lintong@pku.edu.cn, wmwang@must.edu.mo

***Abstract*—Brain lesion segmentation plays a crucial role in diagnosis and monitoring of disease progression. DenseNets have been widely used for medical image segmentation, but much redundancy arises in dense-connected feature maps and the training process becomes harder. In this paper, we address the brain lesion segmentation task by proposing a Dense-Gated U-Net (DGNet), which is a hybrid of Dense-gated blocks and U-Net. The main contribution lies in the dense-gated blocks that explicitly model dependencies among concatenated layers and alleviate redundancy. Based on dense-gated blocks, DGNet can achieve weighted concatenation and suppress useless features. Extensive experiments on MICCAI BraTS 2018 challenge and our collected intracranial hemorrhage dataset demonstrate that our approach outperforms a powerful backbone model and other state-of-the-art methods.**

***Index Terms*—Brain lesion segmentation, U-Net, dense connections, dense gates**

I. Introduction

The huge amounts of medical images have necessitated the use of computers and artificial intelligence to process and analyze. Medical image segmentation algorithms play a vital role in various biomedical-imaging applications, such as the quantification of diagnosis, location of lesion, analysis of anatomical structure and treatment planning. Additionally, manual annotation of lesion regions is quite time-consuming and tedious for doctors. Therefore automatic algorithms for medical image segmentation are essential.

In medical image segmentation tasks, semantic information from shallow layers and deep layers are usually combined for more accurate results. Residual skips [1] and dense connections [2] emerged as popular approaches to strengthen feature propagations among different layers. Residual skips are to employ residual connections between the encoder and decoder, like U-Net [3]. Dense connections are to densely connect each layer with its all preceding layers, like DenseNet. Specifically, DenseNet has been widely used and achieved state-of-the-art results in various medical segmentation challenges, such as volumetric brain segmentation [4].

However, much redundancy caused by dense connections from preceding layers remains a challenging task. We argue that inherent inefficacy exists in the DenseNet architecture design. Specifically, in one dense block, each layer is connected to every other layer in a feed-forward manner, which results in the redundancy of features and a waste of computation time. For example, when a deep network is applied to segment an image, color, shape and texture information are all processed together and equally. But we feel that shapes are more important to obtain detailed object boundaries, while color and texture might contain irrelevant information for segmentation. Another aspect is that it is well known that with skip connections, the mainstream module only produces residual additive outputs. But for many dense connections from previous layers, it is unclear what a role the mainstream module plays.

In order to address these issues of DenseNet, we introduce a gating mechanism to model the layer relationships in densely connected blocks. This idea is inspired by SENet [5]. Dense-gated blocks can learn optimal importance weights of varying dense-connected layers, and the learned weights can help enhance the feature maps from informative layers and suppress redundant contents. Instead of simple concatenations of dense connections used in DenseNet, our gating mechanism alleviates redundancy and makes the training process easier.

Similar gating mechanisms have been proposed for semantic segmentation tasks, such as Attention Gates [6] and CBAM [7]. These prior works employ adaptive feature pooling to focus on local image regions, or model the interdependence between feature channels. Different from previous studies, our method pays more attention to the weighted concatenation of layers, instead of spatial or channel attention on feature maps.

Our contributions are summarized as follows:

1) We propose a novel dense-gated block, which leverages the gating mechanism to reweight each dense connections of previous layers. Gates are learned for enhancing useful connections and suppressing useless others. We then apply dense-gated blocks to U-Net and construct the DGNet for the task of brain lesion segmentation.
2) Extensive experiments are conducted on two datasets — MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS) 2018 and our collected intracranial hemorrhage CT images dataset. Advantages and shortcomings of different competing methods are discussed, and we testify the effectiveness of our proposed method over state-of-the-art methods.

*The corresponding author.

978-1-7281-8068-7/20/$31.00 ©2020 IEEE

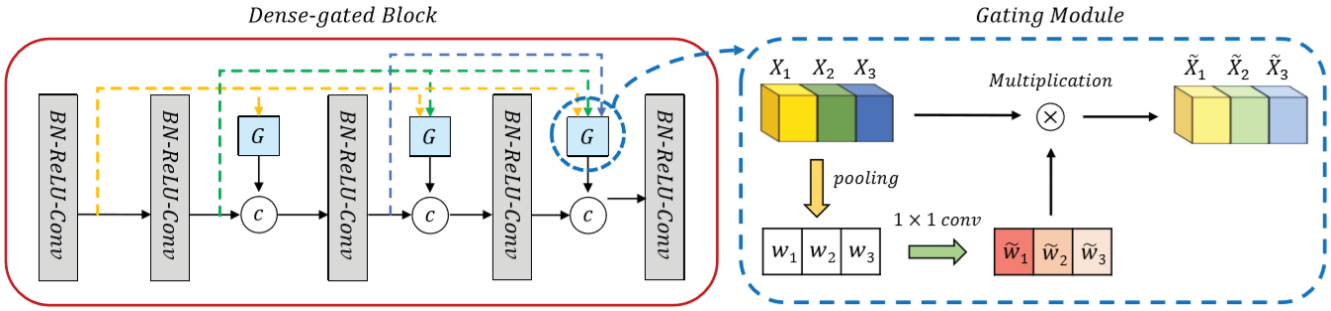

Fig. 1. Dense-gated Blocks. The c above indicates a concatenation and G indicates a gating module, with its details depicted in the right blue box. The gating module applies global average pooling to generate a layer descriptor. Then, two 1×1 convolutional layers are used to model layer relationships and estimate weights of each layer. Finally, the original feature map is rescaled through multiplication.

II. Method

Gated Fusion. Gating is a mature mechanism to measure the usefulness of each feature vector in a feature map and aggregates information accordingly. In this paper, a dense-gated block is designed based on the simple concatenation-based fusion by controlling information flow with gates. Besides that useful information can be regulated to the right place through gates, useless information can also be effectively suppressed on both the sender and receiver sides. Thus information redundancy can be alleviated because the information is only received when concatenation of the current layer has useful features.

As illustrated in Fig.1, we change the original dense connections by reweighting each feature maps before concatenation and design a gating module to make the network focus on more informative feature maps. More precisely, we perform feature compression and turn each of feature maps concatenated into a layer descriptor. This layer descriptor has a global receptive field in certain degrees, and the output dimension matches the number of input characteristic concatenation. It represents the global distribution of responses on the characteristic concatenation. Assuming the $(l+1)$-th layer receives the feature maps of all preceding layers, $x_0, x_1, \ldots, x_l$, as input and $[x_0, x_1, \ldots, x_l]$ represents the concatenation operation. For the $(l+1)$-th layer, a statistic $\mathrm{z} \in R^l$ is generated by squeezing the feature maps X and z_c as the c-th element of z can be expressed as:

$$z_c = \frac{1}{D_c \times W_c \times H_c} \sum_{i=1}^{D_c} \sum_{j=1}^{W_c} \sum_{k=1}^{H_c} x_c(i,j,k), c \in (0,1,\ldots,l) \tag{1}$$

Where D_c, W_c and H_c denote the number of channels, the width and height of x_c respectively.

After applying global average pooling above, we then use 1×1 convolutional layers to explicitly model the correlation between different layers. Parameters are used to generated rescaling weights for each characteristic concatenation part. To meet these criteria, the sigmoid activation is used as a gating mechanism:

$$s = \sigma\left(W_b \delta\left(W_a z\right)\right), \tag{2}$$

Where $W_a \in R^{\frac{l}{r} \times l}$ and $W_b \in R^{l \times \frac{l}{r}}$ represent the parameters of two 1×1 convolutional layers. δ refers to the ReLU function, and σ is the sigmoid. r denotes the reduction ratio. The final output is a reweight operation. It is obtained by rescaling the transformation output X with the activations:

$$\widetilde{x_c} = s_c \odot x_c, \tag{3}$$

Where $\tilde{X} = [\widetilde{x_0}, \widetilde{x_1}, \ldots, \widetilde{x_c}]$, and $\odot$ refers to the element-wise product between the feature map $x_c \in R^{D \times W \times H}$ and the scalar s_c. In this way, feature maps X are converted into new feature maps $\tilde{X}$, which contain more valid information and less redundancy.

Dense-Gated Blocks. Since dense connections can strengthen feature propagation, we densely connect the feature maps in a left-right manner within a block. Each dense-gated block has 5 convolutional layers, which is formulated as:

$$h_l = [x_0, x_1, \ldots, x_l], \tag{4}$$

$$x_{l+1} = \varphi\left([x_l, g_l\left(h_l\right)]\right), \tag{5}$$

$$\varphi\left(x_i\right) = \mathrm{W} * \delta\left(\mathrm{B}\left(x_i\right)\right), i \in \{0, \ldots, l\}. \tag{6}$$

Where W is the weight matrix, g denotes the gating mechanism, $*$ denotes convolution operation, $B(\cdot)$ represents batch normalization, $\delta(\cdot)$ denotes ReLU for activation.

Network Architecture. The overall framework of our network is shown in Fig. 2. We employ 3D U-Net with skip layers to fuse rich scale feature maps as our baseline. Similar to traditional U-Net, a Dense-gated U-Net has two paths: a downward path (left-down) and an upward path (right-up). The network consists of 4 level encoders in the downward path, 4 level decoders in the upward path and a base level. In the encoder path, each encoder level has a dense-gated block (DGB) which aims at semantic feature extraction. Each layer in the dense block can use the feature maps of all preceding layers as inputs, and use its own feature maps as input into all subsequent gates.

III. Experiments

A. BraTS 2018 Dataset

Dataset. MICCAI BraTS 2018 training set and validation set [8], [9] consist of ample multi-institutional clinically-acquired and multi-modal MRI scans of glioblastoma (HGG)

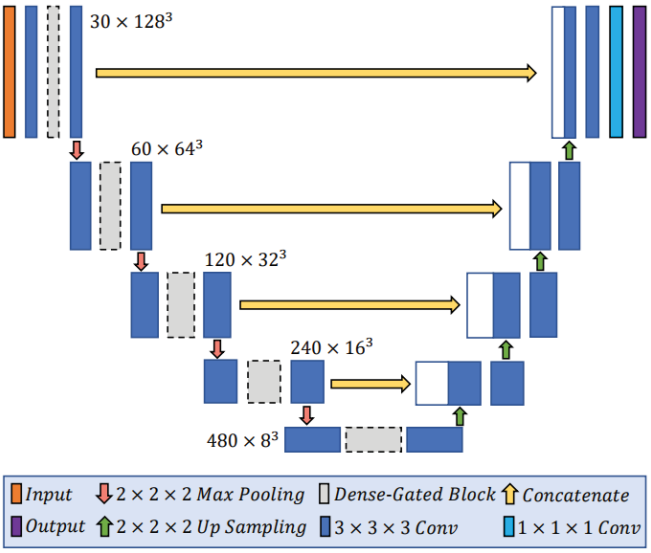

Fig. 2. The framework of our proposed dense-gated U-Net.

and lower-grade glioma (LGG). The training set includes totally 210 HGG patients and 75 LGG patients. Annotations include the GD-enhancing tumor core (ET-label 4), the peritumoral edema (ED-label 2) and the necrotic and non-enhancing tumor core (NCR/NET-label 1). Other pixels except these labels (1,2,4) are labeled as 0. For the sake of convenient description, WT, ET, TC in the following sections refer to the Whole Tumor, the Enhancing Tumor, and the Tumor Core, respectively. As for metrics, we use the Dice and Hausdorff Distance to evaluate the performance of semantic segmentations.

Implementation Details. We train our model with Pytorch framework on NVIDIA Tesla V100 GPUs. Inspired by state-of-the-art semantic segmentation networks, we adopt the multiclass dice loss as the loss function. Our model takes the randomly sampled patches of size $128 \times 128 \times 128$ voxels as inputs and batch size is set as 2. An epoch refers to an iteration over 250 batches, and the model is trained for at most 500 epochs. Early stop strategy is active when the exponential moving average of the validation loss has not been improved within the recent 60 epochs. We used weight decay (L2 weighting factor = 0.00001) for regularization and set the initial learning rate as 10^{-4}.

Qualitative Evaluation. We randomly select a MRI sequence from the training dataset and the segmentation results are shown in Fig. 4(a). We can see that our model has a promising performance for MRI slices and can accurately separate lesion subregions out. The segmentation results of tumor core and enhancing tumor are most impressive. Besides, there are several vessels in the left part of the brain showing similar features with tumors, but our model correctly segments them as background.

Quantitative Evaluation. Quantitative evaluation results are shown in Table I. To further verify the performance, we

TABLE I
QUANTITATIVE RESULTS ON BRATS'18. TOP: ON VALIDATION DATASET. MIDDLE: ON TRAINING DATASET. BOTTOM: ABLATION STUDY RESULTS ON VALIDATION DATASET.

Method	Dice			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Isensee [10]	0.810	0.908	0.854	**2.540**	4.970	7.040
Myro [11]	0.817	0.907	0.860	3.824	4.412	6.841
McKi [12]	0.796	0.903	0.847	3.550	4.170	4.930
Zhou [13]	0.814	0.909	0.853	2.716	4.172	6.545
Our	**0.818**	**0.912**	**0.862**	2.703	**3.901**	**4.595**
Chen [14]	0.740	0.888	0.844	4.631	5.888	5.661
Isensee [10]	0.772	**0.901**	0.843	3.680	5.610	6.000
Our	**0.830**	0.899	**0.899**	**2.167**	**3.772**	**3.791**
Base	0.771	0.898	0.822	3.223	6.306	9.244
Base+DB	0.798	0.892	0.825	4.894	6.225	8.037
Base+DGB	**0.818**	**0.912**	**0.862**	**2.703**	**3.901**	**4.595**

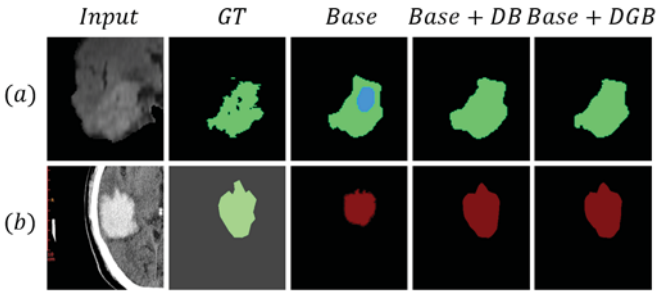

Fig. 3. Visualization results of ablation study on two datasets.

compare the results of our model with state-of-the-art methods. Note that not all work provide results on the training set, so the comparing methods are different. Our method clearly outperforms competing methods on the training dataset and performs on par with the best results on the validation dataset. Such comparative results indicate the promising accuracy of our segmentation method and the effectiveness of our proposed architecture.

Ablation Study. We separately compare the effectiveness of the densely connected block and our dense-gated block. For the non-dense structure (Base), we use the original 3D U-Net framework without any dense blocks. For non-gated dense Structure (Base+DB), we remove all gates in dense-gated blocks but keep dense connection for feature reuse.

The bottom part of Table I shows the contributions of each components on validation set. We can see improvements in dice scores and Hausdorff distances across each tumor regions when we add the proposed strategies gradually. Specifically, the dense-gated block offers large gains compared with its non-gated counterpart. Meanwhile, we also visualize the ablation study in Fig.3(a). We can see that the original 3D U-Net cannot segment the tumor core at all and produces many false regions on tumor core and edema. Compared with Base+DB, our proposed method can produce more accurate boundaries.

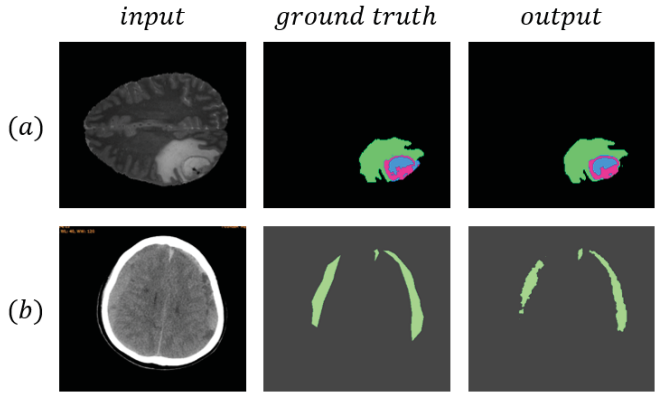

Fig. 4. Segmentation results. (a) BraTS'18: WT in green, ET in purple and TC in blue. (b) Hemorrhage Dataset.

TABLE II
IOU RESULTS ON OUR COLLECTED HEMORRHAGE DATASET. TOP: ABLATION STUDY. BOTTOM: COMPARISON WITH THE STATE-OF-THE-ART RESULTS.

Methods	Lesion	Background	mIoU
Base	0.673	0.995	0.834
Base+DB	0.682	0.996	0.839
Base+DGB	**0.692**	0.996	**0.844**
DeepMedic	0.683	**0.997**	0.840
UNet++	0.677	0.995	0.836
DenseASPP	0.670	0.996	0.833
DeepLabv3	0.639	0.995	0.817

B. Our Collected Hemorrhage Dataset

Dataset and Preprocessing. It is made up of intracranial hemorrhage CT images consisting of 500 collected patients from hospitals. The Annotations include intracranial hemorrhage area (lesion, labeled as 1), while other pixels are all labeled as 0. We divide the total 5000 slices into the training set (4000 slices) and test set (1000 slices). At each epoch, 3200 slices are randomly selected and fed into our model for training and the left 800 slices are for validation.

Results. Qualitative and quantitative results are displayed in Fig.3(b) and Table II. From the top part of Table II, we can see that the proposed Base + DGB version performs the best according to the IoU measure and it can be justified that the removal of each block impairs the overall performance to some extent. For further comparison, we compare our proposed framework with the state-of-the-art methods in semantic segmentation in the bottom part of Table II. It is noted that our method achieves better performance than other comparing methods. To our surprise, we find that DeepLabv3 produces worse segmentation results even than our plain baseline method, though DeepLabv3 achieves superb results on natural image segmentation. This may reveal that there exists a big gap between natural image segmentation and medical image segmentation.

IV. CONCLUSION

In this paper, we introduced a dense-gated block and applied it to U-Net for brain lesion segmentation. We proposed a new gating mechanism to connect the intermediate layers in original dense blocks by reweighting their concatenations. Compared to U-Net, our DGNet is able to gather features from previous dense connections according to the amount of information they contain. Our experiments show that dense-gated blocks can lead to a highly effective architecture that produces more accurate segmentation. Our method significantly improves over the strong baseline algorithm and performs on par with the state-of-the-art methods on the challenging BraTS2018 dataset and our collected intracranial hemorrhage dataset.

ACKNOWLEDGMENT

This research was supported by National Key R&D Program of China (2018AAA0100300) and the Seeding Grant for Medicine and Information Sciences of Peking University.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[2] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017, pp. 4700–4708.
[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
[4] L. Chen, Y. Wu, A. M DSouza, A. Z Abidin, A. W uller, and C. Xu, "Mri tumor segmentation with densely connected 3d cnn," in *Medical Imaging 2018: Image Processing. International Society for Optics and Photonics*, 2018, vol. 10574, p. 105741F.
[5] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018, pp. 7132–7141.
[6] J. Schlemper, O. Oktay, and M. Schaap, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019.
[7] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018, pp. 3–19.
[8] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S Kirby, J. B Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, pp. 170117, 2017.
[9] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, et al., "Identifying the best machinelearning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*,2018.
[10] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H Maier-Hein, "No new-net," in *International MICCAI Brain lesion Workshop*, Springer, 2018, pp. 234–244.
[11] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 311–320.
[12] R. McKinley, R. Meier, and R. Wiest, "Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 456–465.
[13] C. Zhou, S. Chen, C. Ding, and D. Tao, "Learning contextual and attentive information for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 497–507.
[14] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3d-unet: Separable 3d u-net for brain tumor segmentation," in *International MICCAI Brainlesion Workshop. Springer*, 2018, pp. 358–368.