# Learning High-Resolution and Efficient Non-local Features for Brain Glioma Segmentation in MR Images

Haozhe Jia[1,2,4], Yong Xia[1,2(✉)], Weidong Cai[3], and Heng Huang[4,5]

[1] Research and Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
yxia@nwpu.edu.cn
[2] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[3] School of Computer Science, University of Sydney, Sydney, NSW 2006, Australia
[4] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
[5] JD Finance America Corporation, Mountain View, CA 94043, USA

**Abstract.** Brain glioma segmentation using multi-parametric magnetic resonance (MR) imaging has significant clinical value. Although 3D convolutional neural networks (CNNs) have become increasingly prevalent in delivering this segmentation task, these models still suffer from an insufficient ability to high-resolution feature representation for small and irregular regions, limited local receptive fields, and poor long-range dependencies. In this paper, we propose a 3D High-resolution and Non-local Feature Network (HNF-Net) for brain glioma segmentation using multi-parametric MR imaging. We construct HNF-Net based mainly on the parallel multi-scale fusion (PMF) module, which helps produce strong high-resolution feature representation and aggregate multi-scale contextual information. We also introduce the expectation-maximization attention (EMA) module to HNF-Net, aiming to capture the long-range dependent contextual information and reduce the feature redundancy in a lightweight fashion. We evaluated our HNF-Net on the BraTS 2019 Challenge dataset against eight top-ranking methods listed on the challenge leaderboard. Our results suggest that the proposed HNF-Net achieves improved overall performance over these methods, and our ablation study demonstrates the effectiveness of the PMF module and EMA module.

**Keywords:** Brain glioma segmentation · High-resolution feature representation · Lightweight non-local module

## 1 Introduction

As the most common primary brain malignancy, gliomas generally contain heterogeneous histological sub-regions, i.e. edema/invasion, active tumor structures,


A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12264, pp. 480–490, 2020.
https://doi.org/10.1007/978-3-030-59719-1_47

cystic/necrotic components, and non-enhancing gross abnormality. Accurate and automated segmentation of these intrinsic sub-regions using multi-parametric magnetic resonance (MR) imaging is critical for the potential diagnosis and treatment of this disease. This task, however, remains challenging, due to the heterogeneity of gliomas in shape, appearance, and location [2].

With deep learning being widely applied to computer vision applications, convolutional neural networks (CNNs) have been designed for this segmentation task and have shown convincing performance. In [8], Havaei *et al.* proposed a 2D CNN to exploit both local features and global contextual features simultaneously, which outperforms the previous methods while being over 30 times faster. Dong *et al.* [6] utilized a 2D U-Net [16] to segment brain gliomas in a fully convolutional fashion and introduced a soft Dice loss [13] to further improve the accuracy. However, the performance of these models remains limited, since they only focus on 2D intra-slice features and ignore 3D global contextual information.

Kamnitsas *et al.* [10] constructed a 3D dual pathway CNN, namely DeepMedic, which simultaneously processes the input image at multiple scales with a dual pathway architecture so as to exploit both local and global contextual information. DeepMedic also uses a 3D fully connected conditional random field to remove false positives. This model, however, tends to lose the global structural information since it uses small local image patches as its input, and the inference is inefficient for the same reason. To jointly use the 3D contextual information and global structural information, a recent popular practice is to construct segmentation models based on 3D U-Net [4]. In [9], Isensee *et al.* achieved outstanding segmentation performance using a 3D U-Net with instance normalization and leaky ReLU activation, in conjunction with a combination loss function and a region-based training strategy. Chen *et al.* [3] proposed a highly efficient 3D U-Net, which leverages the 3D multi-fiber unit to reduce the computational cost and uses 3D dilated convolutions to build multi-scale feature representation, and achieved satisfactory accuracy in the real-time segmentation of brain gliomas. In [14], Myronenko *et al.* incorporated a variational auto-encoder (VAE) based reconstruction decoder into a 3D U-Net to regularize the shared encoder, and achieved the state-of-the-art segmentation performance at that time.

Despite their performance gains over previous approaches, these 3D U-Net based methods still have two major limitations. First, the encoders of these models have limited convolutional layers and parameters in high-resolution stages to avoid high GPU memory usage, and the decoder may lose some detailed contextual information when low-resolution features are upsampled to form high-resolution features. Consequently, these models usually have insufficient high-resolution feature representation ability for small and irregular regions such as the enhancing tumor and tumor core. Second, these models can hardly overcome the difficulties caused by the highly variable locations, morphology, and sub-structures of gliomas, since they usually extract features with fixed geometric structures which result in limited local receptive fields and poor long-range dependencies. Although Non-local self-attention modules [7,20,21] have been

proposed to build long-range dependencies of spatial contextual information for computer vision tasks, the high computation complexity and GPU memory occupation make it hard to be applied to volumetric medical image segmentation.

In this paper, we propose a 3D High-resolution and Non-local Feature Network (HNF-Net) for brain glioma segmentation using multi-parametric MR imaging. We construct HNF-Net based mainly on the parallel multi-scale fusion (PMF) module, which maintains strong high-resolution feature representation and aggregates multi-scale contextual information. We further reformulate the Non-local self-attention mechanism in a light-weight expectation-maximization iteration manner, and thus introduce the expectation-maximization attention (EMA) module [11] to HNF-Net to enhance the long-range dependent spatial contextual information at the cost of acceptable computational complexity. We evaluated the proposed HNF-Net on the BraTS 2019 dataset against eight top-tier solutions. Our results demonstrate the effectiveness of the proposed PMF and EMA modules and the superior performance of our HNF-Net.

The contributions of this work are two-fold: (1) we construct HNF-Net with the PMF module to learn high-resolution feature representation and aggregate multi-scale contextual information; and (2) we introduce the EMA module to capture Non-local long-range dependent spatial contextual information in a lightweight fashion.

## 2 Dataset

We used the BraTS 2019 Challenge database [1,2,12] for this research, which contains 335 training and 125 validation multi-parametric brain MR studies. Each study has four MR images, including T1-weighted (T1), post-contrast T1-weighted (T1ce), T2-weighted (T2), and fluid attenuated inversion recovery (Flair) sequences. All MR images have the same size of $240 \times 240 \times 155$ and the same voxel spacing of $1 \times 1 \times 1\,\mathrm{mm}^3$. For each study, the enhancing tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NCR/NET) were annotated on a voxel-by-voxel basis by experts. The annotations for training studies are publicly available, and the annotations for validation studies are withheld for online evaluation.

## 3 Method

The proposed HNF-Net has an encoder-decoder structure. For each study, four multi-parametric brain MR sequences are first concatenated to form a four-channel input and then processed at five scales i.e., $r$, $1/2r$, ... $1/16r$, highlighted in green, yellow, blue, pink, and orange in Fig. 1, respectively. At the original scale $r$, there are four convolutional blocks, two for encoding and the other two for decoding. The connection from the encoder to the decoder skips the processing at other scales so as to maintain the high resolution and spatial information in a long-range residual fashion. At other four scales, four PMF modules are jointly used as a high-resolution and multi-scale aggregated feature extractor.

At the end of the last PMF module, the output feature maps at four scales are first recovered to the $1/2r$ scale and then concatenated as mixed features. Next, the EMA module is used to efficiently capture long-range dependent contextual information and reduce the redundancy of the obtained mixed features. Finally, the output of the EMA module is recovered to original scale $r$ and 32 channels via $1 \times 1 \times 1$ convolutions and upsampling and then added to the full-resolution feature map produced by the encoder for the dense prediction of voxel labels. We adopt a combination of generalized Dice loss [18] and cross-entropy loss as the loss function. We now delve into the details of the PMF module and EMA module.

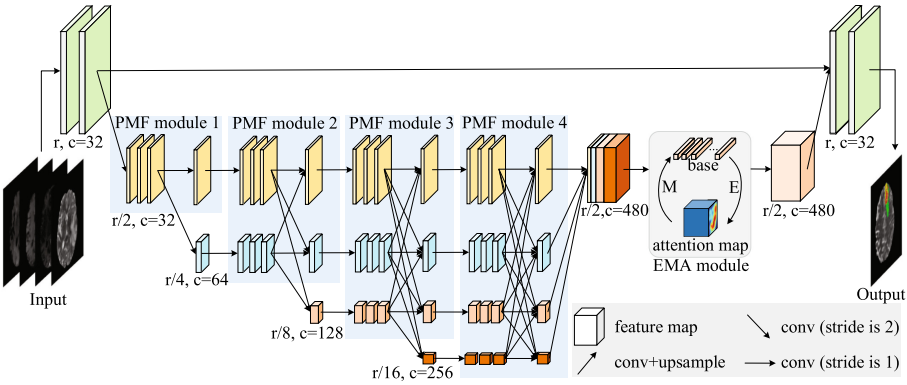


**Fig. 1.** Diagram of the proposed HNF-Net. r denotes the original resolution and c denotes the channel number of feature maps.

### 3.1 PMF Module for High-Resolution Feature Representation

It has been observed that the high-resolution low-level features produced at the early stages of CNNs reveal fine spatial information but have weak semantic guidance, whereas the low-resolution high-level features have strong semantic consistency but give coarse spatial predictions. Given that learning strong high-resolution representation is essential for small object segmentation, high-resolution features are maintained throughout the segmentation process in recent solutions [15,17,19], and it has been shown that this strategy contributes to convincing performance.

For this study, we construct the PMF module as the main component of our HNF-Net. The PMF module has two parts, the parallel multi-scale convolutional block and fully connected fusion block, as shown in Fig. 2(a) and Fig. 2(b), respectively. The former has a set of parallel branches similar to the group convolution, and each branch is built with repeated residual convolutional blocks at a specific scale. The latter fuses all output features of the parallel multi-scale convolutional

block in a parallel but fully connected fashion, where each branch is the summation of the output features of all resolution branches. Thus, in each PMF Module, the parallel multi-scale convolutional block can fully exploit multi-resolution features but maintain high-resolution feature representation, and the fully connected fusion block can aggregate rich multi-scale contextual information.

Moreover, we cascade multiple PMF modules, in which the number of branches increases progressively with depth, as shown in Fig. 2(c). As a result, from the perspective of the highest resolution stage, its high-resolution feature representation is boosted with repeated fusion of multi-scale low-resolution representations. Meanwhile, the cascade network can be regarded as an ensemble of several U-shape sub-networks with different depths and widths, which tends to further reduce the semantic gap of the features at different depths.

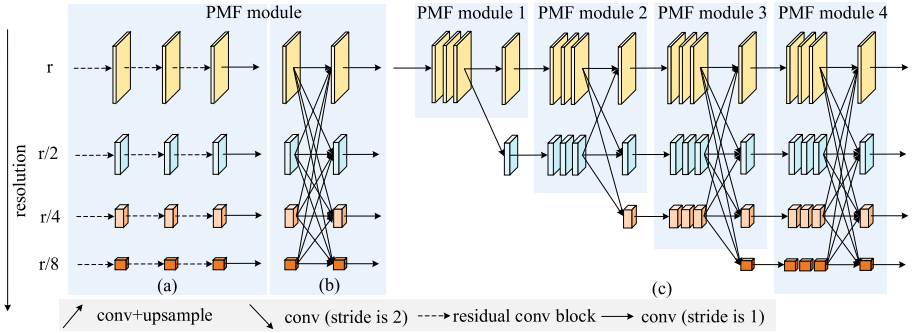


**Fig. 2.** Architecture of the PMF module. The PMF module has two parts, the parallel multi-scale convolutional block and fully connected fusion block, as shown in (a) and (b), respectively. In our proposed HNF-Net, we cascade multiple PMF modules together with the structure shown in (c). All downsample operations are achieved with 2-stride convolutions, and all upsample operations are achieved with joint $1 \times 1 \times 1$ convolutions and trilinear interpolation. It is noted that, since it is inconvenient to show 4D feature maps ($C \times D \times H \times W$) in the figure, we show all feature maps without depth information, and the thickness of each feature map reveals its channel number.

### 3.2 EMA Module for Non-local Features

The Non-local self-attention mechanism can help to aggregate contextual information from all spatial positions and capture long-range dependencies, and hence has been widely used for semantic image segmentation [7,20,21]. However, when incorporating this mechanism into a CNN with 3D convolutions, the spatial and computational complexity of the model increases dramatically, since it requires the calculation of point-wise spatial attention and generation of large attention maps, which are extremely time and memory consuming. Hence, we introduce the EMA module [11] to our glioma segmentation model, aiming to incorporate a lightweight Non-local attention mechanism into our model.

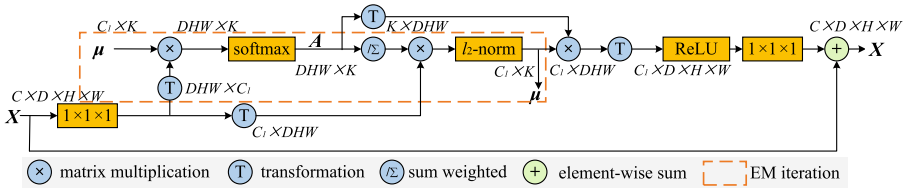


**Fig. 3.** The structure of EMA module. The shape of the input feature maps $X$ are $C \times D \times H \times W$, where $C$ is the channel number. The dotted box denotes the EM algorithm operation, where the Non-local spatial attention maps $A$ and the bases $\mu$ are generated alternately as the E step and M step, respectively. After convergence, we can use the obtained $A$ and $\mu$ to generate reconstructed feature maps $\widetilde{X}$. At both the beginning and ending of the EM algorithm operation, the $1 \times 1 \times 1$ convolutions are adopted to change the channel number. In addition, to avoid overfitting, we further sum $\widetilde{X}$ with $X$ in a residual fashion.

The main concept of the EMA module is operating the Non-local attention on a set of feature reconstruction bases rather than directly achieving this on the high-resolution feature maps. Since the reconstruction bases have much less elements than the original feature maps, the computation cost of the Non-local attention can be significantly reduced. The pipeline of the EMA module is illustrated in Fig. 3. Given the input feature maps $\boldsymbol{X} \in R^{C \times D \times H \times W}$ and the reconstruction bases $\boldsymbol{\mu} \in R^{C \times K}$, where $C$ is the number of channels and $K$ is the number of bases, we regard the reconstruction bases $\boldsymbol{\mu}$ and Non-local self-attention $\boldsymbol{A} \in R^{D \times H \times W \times K}$ as the learned parameters and latent variables, respectively. The expectation-maximization (EM) algorithm [5] aims to find the optimal learnable parameters, which maximize the complete data likelihood. Specifically, the expectation (E) step estimates the expectation of $\boldsymbol{A}$ and the maximization (M) step updates $\boldsymbol{\mu}$ by maximizing the likelihood, shown as follows:

$$\boldsymbol{A}^t = softmax(\boldsymbol{X}^T(\boldsymbol{\mu}^{t-1})) \tag{1}$$

$$\boldsymbol{\mu}_k^t = \frac{\sum_{n=1}^{N}(\boldsymbol{A}_{nk}^t \boldsymbol{X}_n)}{\sum_{m=1}^{N} \boldsymbol{A}_{mk}^t} \tag{2}$$

where $t$ denotes the $t$-th iteration, $N = D \times H \times W$, $\boldsymbol{\mu}^k$ is the $k$-th base, and $\boldsymbol{A}_{nk}$ is the attention vector in the $k$-th channel at location $n$. As suggested by [11], we alternatively execute the E step and M step for three iterations. Then, we use the normalized attention $\boldsymbol{A}$ and bases $\boldsymbol{\mu}$ to obtain the reconstructed feature maps $\widetilde{\boldsymbol{X}}$, which have long-range dependencies and can be formulated as:

$$\widetilde{\boldsymbol{X}} = \boldsymbol{\mu}\boldsymbol{A}^T \tag{3}$$

Since we set $K \ll D \times H \times W$, the EMA module can significantly reduce the computational complexity from $O((D \times H \times W)^2)$ to $O(D \times H \times W \times K)$, as compared to the traditional point-wise Non-local mechanism. Moreover, as $\widetilde{\boldsymbol{X}}$ lies in a subspace of $\boldsymbol{X}$, the EMA module can further reduce the redundancy and noise of feature maps.

**Table 1.** Segmentation performances of the proposed HNF-Net and eight top-ranking methods on the BraTS 2019 validation set. All results were directly adopted from the BraTS 2019 Challenge leaderboard.

| Method | Dice (%) | | | 95%HD (mm) | | | Mean-rank |
|---|---|---|---|---|---|---|---|
| | ET | WT | TC | ET | WT | TC | |
| HNF-Net (**ours**) | $81.16_{(3)}$ | $91.12_{(2)}$ | $84.52_{(6)}$ | $3.49_{(2)}$ | $4.13_{(2)}$ | $\mathbf{5.25}_{(1)}$ | **2.67** |
| Questionmarks | $80.21_{(5)}$ | $90.94_{(4)}$ | $\mathbf{86.47}_{(1)}$ | $\mathbf{3.15}_{(1)}$ | $4.26_{(3)}$ | $5.44_{(3)}$ | 2.83 |
| NVDLMED [14] | $\mathbf{82.28}_{(1)}$ | $91.01_{(3)}$ | $86.22_{(2)}$ | $3.62_{(3)}$ | $4.42_{(5)}$ | $5.46_{(4)}$ | 3.00 |
| SVIG1 | $81.33_{(2)}$ | $\mathbf{91.16}_{(1)}$ | $85.79_{(3)}$ | $4.21_{(6)}$ | $\mathbf{4.10}_{(1)}$ | $5.92_{(6)}$ | 3.17 |
| lfn [9] | $80.24_{(4)}$ | $90.91_{(5)}$ | $85.49_{(4)}$ | $3.88_{(5)}$ | $4.36_{(4)}$ | $5.32_{(2)}$ | 4.00 |
| NWPU_ASGO | $73.54_{(9)}$ | $90.71_{(7)}$ | $84.92_{(5)}$ | $3.63_{(4)}$ | $5.27_{(6)}$ | $5.81_{(5)}$ | 6.00 |
| GHcheng | $78.69_{(8)}$ | $90.80_{(6)}$ | $83.31_{(8)}$ | $4.33_{(7)}$ | $5.36_{(7)}$ | $6.97_{(8)}$ | 7.33 |
| ANSIR | $80.06_{(6)}$ | $90.09_{(8)}$ | $84.38_{(7)}$ | $4.52_{(8)}$ | $6.94_{(9)}$ | $6.00_{(7)}$ | 7.50 |
| deecamp | $79.22_{(7)}$ | $89.40_{(9)}$ | $80.27_{(9)}$ | $6.08_{(9)}$ | $5.61_{(8)}$ | $8.68_{(9)}$ | 8.50 |

## 4 Experiments and Results

**Implementation Details:** Suggested by some previous works [3,9], we pre-processed each brain MR sequence, including brain stripping, clipping voxel intensity with a window of [0.5%–99.5%], and normalizing voxel values into zero mean and unit variance. To reduce the potential overfitting, we further employed several online data augmentation techniques, including random flipping (on all three planes independently), random rotation ($\pm 10°$ on all three planes independently), and random intensity shift of [$\pm 0.1$]. In the training phase, we concatenated four MR sequences along the channel dimension and randomly cropped the input image into a fixed size of $128 \times 128 \times 128 \times 4$. We adopted the Adam optimizer with an initial learning rate of 0.001 and a batch size of 4 with synchronized batch normalization. All experiments were performed using the Pytorch on a workstation with 4 NVIDIA Geforce GTX 2080Ti GPUs.

**Comparison with State-of-the-Art Methods:** We trained our HNF-Net on the BraTS 2019 training set and submitted the segmentation results of the validation set for the online evaluation. In the BraTS 2019 challenge, the segmentation performance of three sub-regions, namely enhancing tumor (ET), tumor core (TC), and whole tumor (WT), are evaluated by the Dice score and 95% Hausdorff distance (%95HD). Since there are a total of 140 valid entries on the

validation leaderboard, we present the segmentation performance of the HNF-Net and the performance of eight top-ranking methods in Table 1. It is shown that the proposed HNF-Net achieves a Dice score of 81.16% for ET, 91.12% for WT, and 84.52% for TC, which ranks the 3rd, 2nd, and 6th among nine competing methods, respectively, and achieves a %95HD of 3.49 for ET, 4.13 for WT, and 5.25 for TC, which ranks the 2nd, 2nd, and 1st among nine methods, respectively. Since no method achieves the best performance in terms of all metrics, we use the average of the six ranks (mean-rank) as the overall performance indicator. A lower mean-rank value means better segmentation accuracy. Thus, the mean-rank of our HNF-Net is 2.67, which is the lowest among nine competing methods. Therefore, we believe that our HNF-Net is very competitive with respective to those eight top-ranking methods, particularly with less outliers even in those small and irregular ET and TC regions as indicated by low %95HD values.

**Table 2.** Segmentation performances of 3D U-Net (baseline1), 3D U-Net++ (baseline2), HNF-Net but without & with original Non-local modules [20], and our proposed HNF-Net on the BraTS 2019 training set. All results were obtained using the same five-fold cross-validation.

| Method | Params (M) | FLOPs (G) | Dice (%) | | |
|---|---|---|---|---|---|
| | | | ET | WT | TC |
| 3D U-Net++ [22] | **27.47** | **2019.31** | 74.80 | 89.53 | 82.91 |
| 3D U-Net [4] | 15.80 | 1240.27 | 74.82 | 88.35 | 80.74 |
| HNF-Net (w/o Non-local) | 16.79 | 420.36 | 79.93 | 90.30 | 84.25 |
| HNF-Net (w Non-local [20]) | 17.31 | 558.01 | 80.43 | 90.64 | 85.77 |
| HNF-Net (ours) | 16.82 | 420.49 | **80.96** | **91.12** | **86.40** |

**Ablation Study:** To evaluate the effectiveness of the PMF module and EMA module in HNF-Net, we conducted an ablation study on the BraTS 2019 training set using five-fold cross-validation. We chose 3D U-Net [4] and 3D U-Net++ [22] as two baseline models, which uses neither the PMF module nor the EMA module. Besides the propsoed HNF-Net, we also tested its two variants where the first one has no Non-local module and the second one uses the original Non-local module [20]. The experiments were performed under the same settings and the results were given in Table 2. It shows that (1) compared to 3D U-Net [4], using the PMF module not only dramatically improved the Dice score by 5.11% for ET, 1.95% for WT, and 3.51% for TC but also reduced the computations significantly with 1/3 of the FLOPs, though having moderately increased parameters; (2) incorporating the EMA module into the segmentation model further improves the Dice score by 1.03% for ET, 0.82% for WT, and 2.15% for TC but only increases the parameters and FLOPs slightly; (3) compared to 3D U-Net++ [22], another U-shape baseline model also with denser layer-forwarding policy,

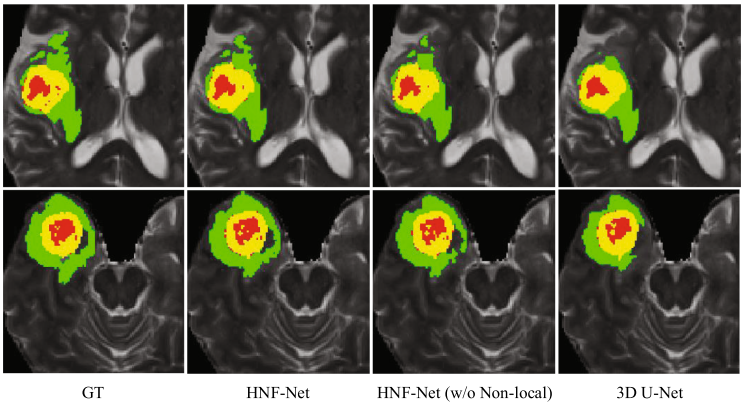


**Fig. 4.** The visual comparison of the segmentation results of HNF-Net, HNF-Net (w/o Non-local), and 3D U-Net. GT represents ground truth. The NCR/NET, ED, and ET regions are highlighted in red, green, and yellow, respectively. (Color figure online)

the proposed HNF-Net also significantly improves the Dice score by 6.16% for ET, 1.59% for WT, and 3.49% for TC but only with around 60% parameters and 20% FLOPs; (4) Using EMA module to replace the original Non-local module [20] can obtain a slight performance gain but has obviously smaller computation cost, especially for FLOPs. The results suggest that the high-resolution feature representation produced by the PMF module is beneficial to the glioma segmentation and that the EMA module is effective in improving segmentation accuracy via capturing long-range dependencies. In addition, the segmentation results of 3D U-Net [4], HNF-Net (w/o Non-local), and HNF-Net were visualized in Fig. 4. It is shown that both HNF-Net and HNF-Net (w/o Non-local) obtained obviously better segmentation results than 3D U-Net, and HNF-Net achieved best segmentation results. The conclusion drawn from this table is consistent with the conclusions drawn from quantitative evaluation.

## 5 Conclusion

In this paper, we have proposed HNF-Net for brain glioma segmentation using multi-parametric MR imaging, which uses the PMF module to exploit multi-resolution feature representation and employs the EMA module to capture long-range dependent contextual information in a lightweight fashion. Our results on the BraTS 2019 dataset show that the proposed HNF-Net achieves the state-of-the-art performance and also suggest the effectiveness of the proposed PMF and EMA modules. One of our potential future works is to extend the proposed HNF-Net to the segmentation of other small and irregular lesions and organs such as kidney tumors or hippocampus.

**Acknowledgement.** Haozhe Jia and Yong Xia were partially supported by the Science, Technology and Innovation Commission of Shenzhen Municipality, China under

Grant JCYJ20180306171334997, the National Natural Science Foundation of China under Grant 61771397, and the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX202042.

## References

1. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**, 170117 (2017)
2. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629 (2018)
3. Chen, C., Liu, X., Ding, M., Zheng, J., Li, J.: 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 184–192. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_21
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc.: Ser. B (Methodol.) **39**(1), 1–22 (1977)
6. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 506–517. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_44
7. Fu, J., et al.: Dual attention network for scene segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3146–3154 (2019)
8. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
9. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No new-net. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11384, pp. 234–244. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_21
10. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
11. Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., Liu, H.: Expectation-maximization attention networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 9167–9176 (2019)
12. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2014)
13. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
14. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11384, pp. 311–320. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_28

15. Pohlen, T., Hermans, A., Mathias, M., Leibe, B.: Full-resolution residual networks for semantic segmentation in street scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4151–4160 (2017)
16. 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
17. Saxena, S., Verbeek, J.: Convolutional neural fabrics. In: Advances in Neural Information Processing Systems, pp. 4053–4061 (2016)
18. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
19. Sun, K., et al.: High-resolution representations for labeling pixels and regions. arXiv preprint arXiv:1904.04514 (2019)
20. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803 (2018)
21. Zhao, H., et al.: PsaNet: Point-wise spatial attention network for scene parsing. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 267–283 (2018)
22. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1