Supplementary material
RFNet: Region-aware Fusion Network
for Incomplete Multi-modal Brain Tumor Segmentation

Yuhang Ding[1,2] Xin Yu[2] Yi Yang[2]
[1] Baidu Research   [2] ReLER, University of Technology Sydney
dyh.ustc.uts@gmail.com, {xin.yu, yi.yang}@uts.edu.au

This document supplements the main paper. *Firstly*, comparisons with the state-of-the-art methods under three other criteria on BRATS2020 are reported in Table 1. *Secondly*, more comparisons with the state-of-the-art methods (*i.e.*, HeMIS [3], U-HVED [2] and RobustSeg [1]) on BRATS2015 and BRATS2018 are reported in Table 2 and Table 3, respectively. *Thirdly*, more visual results of RFNet are illustrated in Fig. 1. Furthermore, more qualitative comparisons with the state-of-the-art are illustrated in Fig. 2. *Fourthly*, illustrations of probability maps of the combinations of different modalities are shown in Fig. 3. *Fifthly*, more examples of generated attention weights are illustrated in Fig. 4 and Fig. 5. *Finally*, Figure 6 illustrates the detailed architecture of the encoder and decoder in RFNet.

Comparisons under other criteria on BRATS2020: Table 1 report the specificity, sensitivity and Hausdorff distance (95%) on BRATS2020. As shown in Table 1, our RFNet also outperforms the state-of-the-art in terms of the three criteria on BRATS2020.

Comparisons on BRATS2015 and BRATS2018: Table 2 and Table 3 report the segmentation accuracy from all modal combinations on BRATS2015 and BRATS2018, respectively. As visible in Table 2 and Table 3, our RFNet is able to improve the accuracy for all the modal combinations significantly on the two datasets. This demonstrates the superiority of our method.

Visualization of the predicted segmentation masks: As seen in Fig. 1, our RFNet is able to segment brain tumors in various combinations of modalities. Furthermore, we also compare our RFNet with three state-of-the-art methods qualitatively, as shown in Fig. 2. By comparing the predictions of our RFNet and other methods, we can see that our method predicts more accurate segmentation maps.

Visualization of the estimated probability maps: As shown in Fig. 3, our RFNet is able to generate accurate probability maps from various combinations of image modalities. Thus, the probability maps significantly facilitate our region-aware fusion and improve the accuracy of

Table 1. Comparisons under three testing criteria on BRATS2020. 'Com', 'Cor' and 'Enh' denotes the whole tumor, the tumor core and the enhancing tumor.

Methods	specificity (%) ↑			sensitivity (%) ↑			Hausdorff95 (mm) ↓		
	Com	Cor	Enh	Com	Cor	Enh	Com	Cor	Enh
HeMIS [14]	82.72	66.88	62.40	99.55	99.79	99.77	12.73	13.30	32.84
U-HVED [10]	81.73	68.16	64.09	**99.60**	99.78	99.78	13.13	16.01	32.69
RobustSeg [3]	88.52	75.70	72.22	99.48	99.78	99.77	18.37	21.34	33.98
Ours	**89.42**	**80.91**	**73.65**	**99.60**	**99.81**	**99.81**	**7.88**	**8.80**	**25.85**

our segmentation results.

Visualization of the generated attention weights: Figure 4 and Figure 5 illustrate the generated attention weights of our RFM. Particularly, the attention weights for all the combinations of four image modalities in the fourth stage are reported in Fig. 4, while Fig. 5 illustrates the generated attention weights from the full set (without missing modalities) in all four stages. As shown in Fig. 5 and Fig. 4, our RFNet is able to pay more attention to more sensitive modalities in each region, thus achieving superior segmentation performance.

Architecture of RFNet: RFNet employs the same encoder and decoder architectures as in [1], as visible in Fig. 6. Therefore, our performance gain mainly comes from our proposed RFM and regularization. As shown in Fig. 6, the four-stage/level encoder and decoder are employed. In each stage, each encoder employs a residual convolutional block (including two $3 \times 3 \times 3$ convolutional layers followed by the corresponding instance normalization layers and Leaky ReLU activation) to extract features. Furthermore, convolutional layers with a kernel size of $3 \times 3 \times 3$ and a stride of 2 are employed to reduce the feature resolutions and increase the channel numbers accordingly. The initial channel number is set to 16. In our decoder, an up-sampling layer followed by convolutional layers is used to increase the resolutions of feature maps. For RFM, we leverage $3 \times 3 \times 3$ convolutional layers to extract features and $1 \times 1 \times 1$ convolutional layers for the probability map generation.

Table 2. Comparisons with the state-of-the-art methods, including HeMIS [3], U-HVED [2] and RobustSeg [1], on BRATS2015. Complete, Core and Enhancing denote the Dice scores of the whole tumor, the tumor core and the enhancing tumor, respectively. The results of U-HVED [2] are reproduced based on the authors' codes.

Modalities				Dice scores (%)											
				Complete				Core				Enhancing			
F	T1	T1c	T2	HeMIS	U-HVED	RobustSeg	Ours	HeMIS	U-HVED	RobustSeg	Ours	HeMIS	U-HVED	RobustSeg	Ours
○	○	○	●	58.48	81.19	85.49	**86.89**	40.18	53.40	58.66	**63.81**	20.31	29.05	37.66	**40.07**
○	○	●	○	33.46	67.48	71.86	**74.95**	44.55	68.24	72.87	**72.64**	49.93	71.54	70.22	**81.40**
○	●	○	○	33.22	53.58	68.40	**74.20**	17.42	41.14	50.00	**61.27**	4.67	19.16	22.67	**29.44**
●	○	○	○	71.26	83.82	83.02	**86.91**	37.45	51.37	46.67	**58.71**	5.57	22.18	28.30	**35.23**
○	○	●	●	67.59	84.77	87.53	**88.39**	63.39	73.18	78.46	**77.50**	65.38	83.54	76.82	**86.97**
○	●	●	○	45.93	69.65	74.59	**78.13**	55.06	68.85	76.40	**74.06**	62.40	76.96	73.95	**82.48**
●	●	○	○	80.28	85.82	87.66	**88.51**	49.52	58.39	60.17	**66.88**	22.26	26.65	35.28	**40.95**
○	●	○	●	69.56	82.17	87.87	**88.25**	47.26	57.58	64.88	**67.24**	23.56	33.94	41.05	**40.58**
●	○	○	●	82.10	87.74	89.08	**89.62**	53.42	59.13	63.51	**68.74**	23.19	30.31	39.72	**44.64**
●	○	●	○	79.80	87.48	88.01	**88.45**	66.12	74.27	78.09	**79.30**	67.12	84.30	76.62	**86.15**
●	●	●	○	80.88	87.91	87.73	**88.75**	69.26	75.82	80.68	**80.46**	71.30	84.33	78.81	**87.30**
●	●	○	●	83.87	87.59	89.07	**89.93**	57.76	62.43	65.99	**69.75**	28.46	33.21	43.04	**44.21**
●	○	●	●	82.78	89.85	89.06	**90.07**	70.62	75.10	79.47	**79.29**	70.52	86.03	78.07	**87.34**
○	●	●	●	70.98	84.72	88.26	**88.41**	66.60	74.85	80.84	**79.18**	67.84	84.03	78.56	**87.47**
●	●	●	●	83.15	89.79	89.07	**90.49**	72.50	76.48	81.19	**80.16**	75.37	86.12	79.13	**87.68**
Average				68.22	81.57	84.45	**86.13**	54.07	64.68	69.19	**71.93**	43.86	56.76	57.33	**64.13**

Table 3. Comparisons with the state-of-the-art methods, including HeMIS [3], U-HVED [2] and RobustSeg [1], on BRATS2018. Complete, Core and Enhancing denote the Dice scores of the whole tumor, the tumor core and the enhancing tumor, respectively. The results of RobustSeg [1] are provided by the authors.

Modalities				Dice scores (%)											
				Complete				Core				Enhancing			
F	T1	T1c	T2	HeMIS	U-HVED	RobustSeg	Ours	HeMIS	U-HVED	RobustSeg	Ours	HeMIS	U-HVED	RobustSeg	Ours
○	○	○	●	79.20	80.90	82.24	**84.30**	40.18	54.10	57.49	**67.62**	20.31	30.80	28.97	**40.71**
○	○	●	○	58.50	62.40	73.31	**74.93**	44.55	66.70	76.83	**80.99**	49.93	65.50	67.07	**69.43**
○	●	○	○	54.30	52.40	70.11	**74.68**	17.42	37.20	47.90	**64.42**	4.67	13.70	17.29	**34.43**
●	○	○	○	79.90	82.10	85.69	**86.46**	37.45	50.40	53.57	**64.89**	5.57	24.80	25.69	**33.92**
○	○	●	●	81.00	82.70	85.19	**86.39**	63.39	73.70	80.20	**83.27**	65.38	70.20	69.71	**73.01**
○	●	●	○	63.80	66.80	77.18	**78.59**	55.06	69.70	78.72	**82.22**	62.40	67.00	69.06	**70.73**
●	●	○	○	83.90	84.30	88.24	**88.78**	49.52	55.30	60.68	**71.59**	22.26	24.20	32.13	**39.68**
○	●	○	●	80.80	82.20	84.78	**86.15**	47.26	57.20	62.19	**70.89**	23.56	30.70	32.01	**41.42**
●	○	○	●	86.00	87.50	88.28	**89.12**	53.42	59.70	61.16	**70.82**	23.19	34.60	33.84	**43.77**
●	○	●	○	83.30	85.50	88.51	**89.17**	66.12	72.90	80.62	**82.94**	67.12	70.30	70.30	**72.84**
●	●	●	○	85.10	86.20	88.73	**89.71**	69.26	74.20	81.06	**83.77**	71.30	71.10	70.78	**73.17**
●	●	○	●	87.00	88.00	88.81	**89.68**	57.76	61.50	64.38	**73.09**	28.46	34.10	36.41	**44.79**
●	○	●	●	87.00	88.60	89.27	**90.06**	70.62	75.60	80.72	**83.54**	70.52	71.20	70.88	**73.13**
○	●	●	●	82.10	83.30	86.01	**86.78**	66.60	75.30	80.33	**83.97**	67.84	71.10	70.10	**72.56**
●	●	●	●	87.60	88.80	89.45	**90.26**	72.50	76.40	80.86	**84.02**	75.37	71.70	71.13	**73.21**
Average				78.60	80.10	84.39	**85.67**	54.07	64.00	69.78	**76.53**	43.86	50.00	51.02	**57.12**

References

[1] Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 447–456, 2019.

[2] Reuben Dorent, Samuel Joutard, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. Hetero-modal variational encoder-decoder for joint modality completion and segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 74–82, 2019.

[3] Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. Hemis: Hetero-modal image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 469–477, 2016.

Input modalities

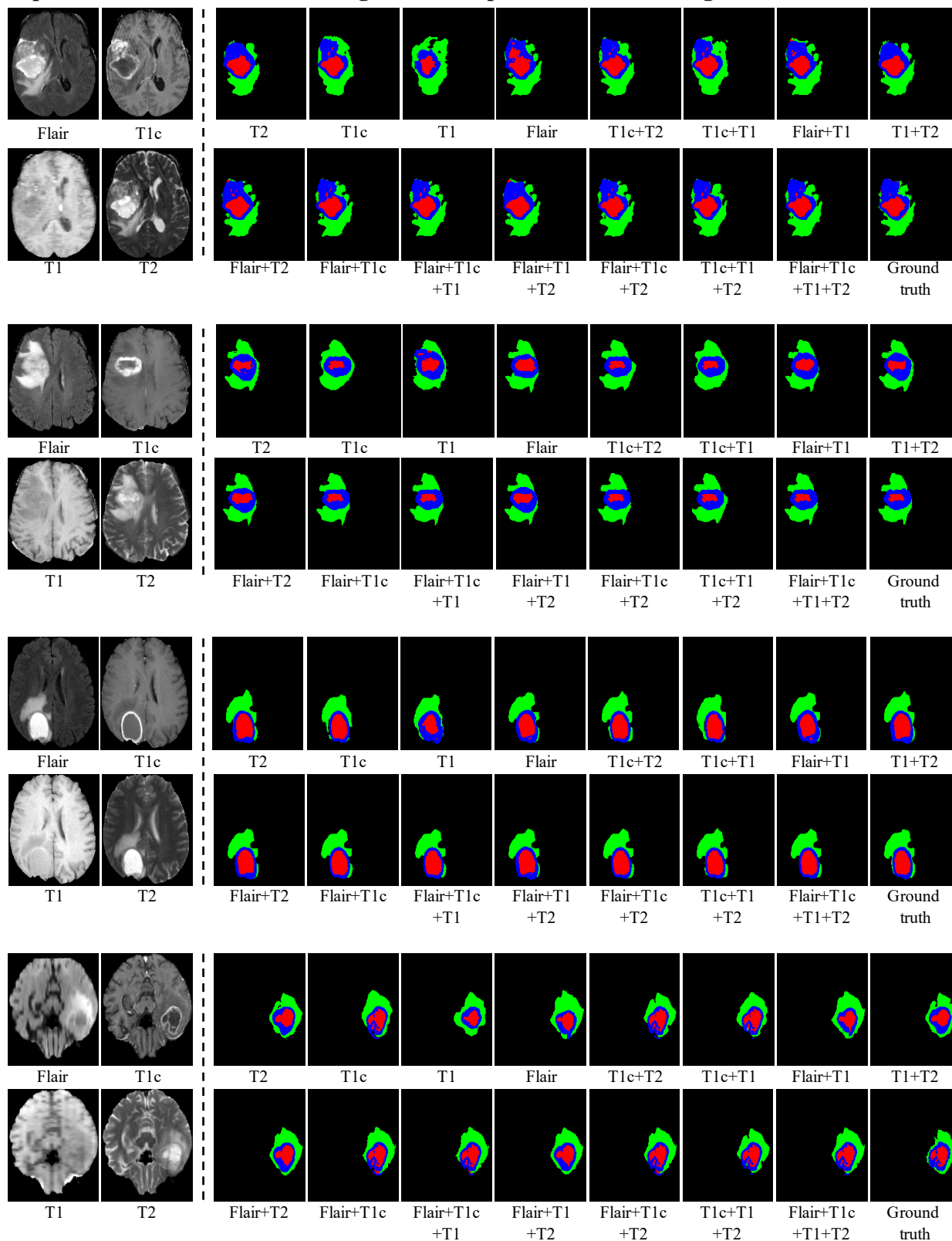**Segmentation predictions and the ground truth**

Figure 1. Visualization of the predicted segmentation maps. Left: four image modalities. Right: segmentation maps predicted by our RFNet from all fifteen combinations of image modalities and the corresponding ground-truth.
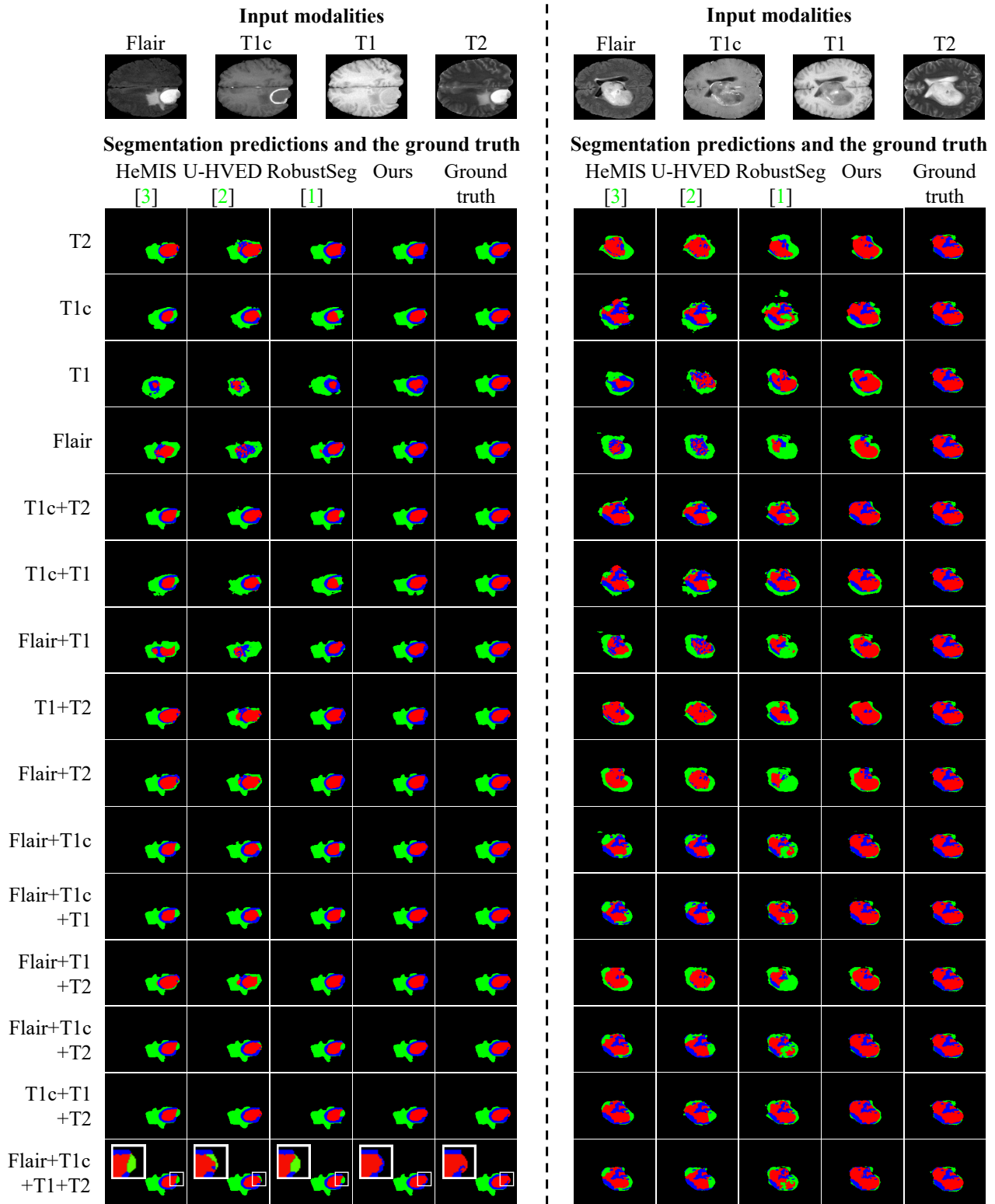
Figure 2. Qualitative comparisons with the state-of-the-art methods. Two tumor cases are illustrated in left and right columns. Top: four image modalities are shown; Bottom: segmentation maps predicted by four methods (*i.e.*, HeMIS [3], U-HVED [2], RobustSeg [1] and our RFNet) and the corresponding ground-truth are shown.
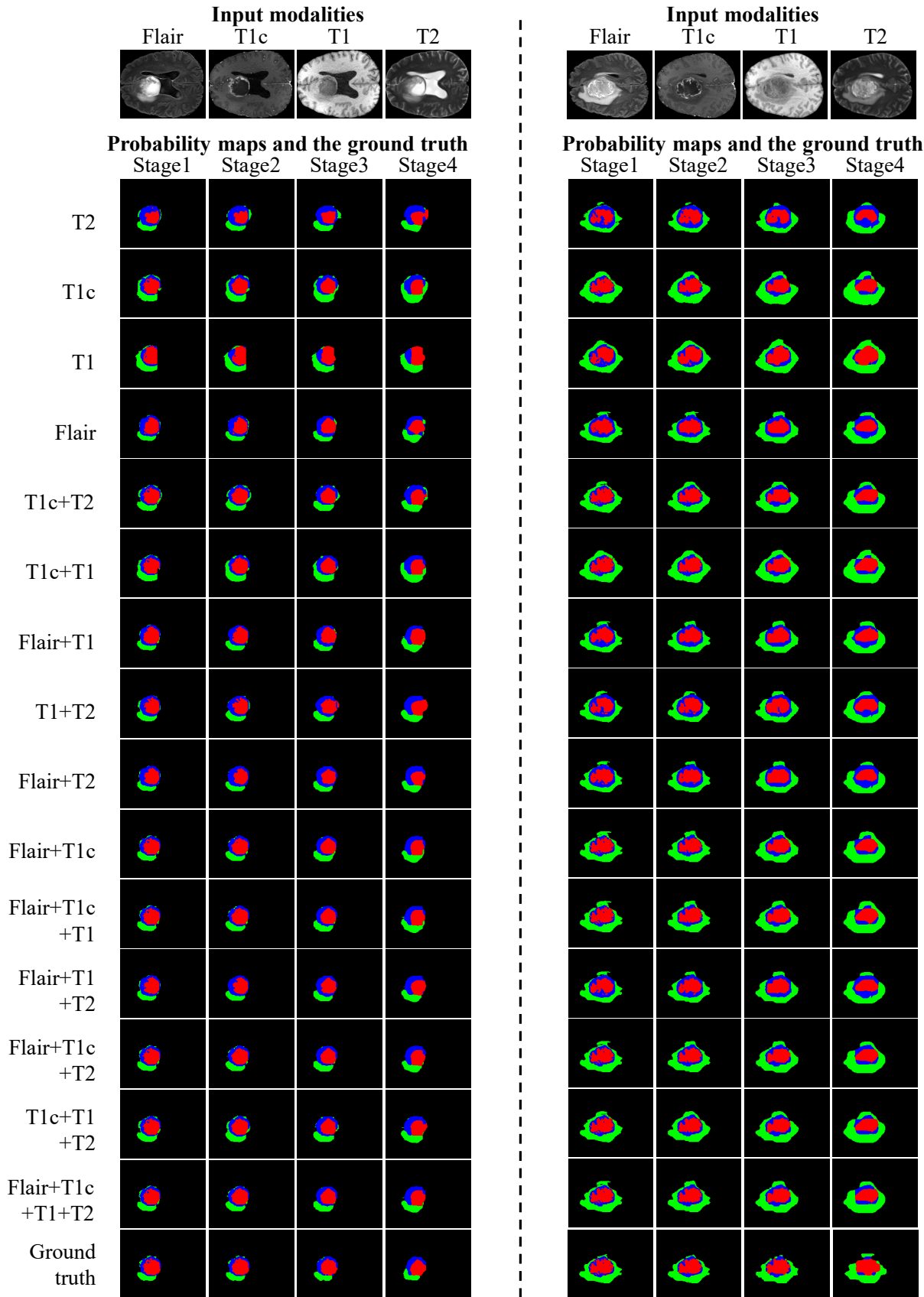
Figure 3. Visualization of the probability maps. Two tumor cases are illustrated in left and right columns. Top: four image modalities are shown. Bottom: our estimated probability maps from all the combinations of image modalities at different stages and their corresponding ground-truth are illustrated.
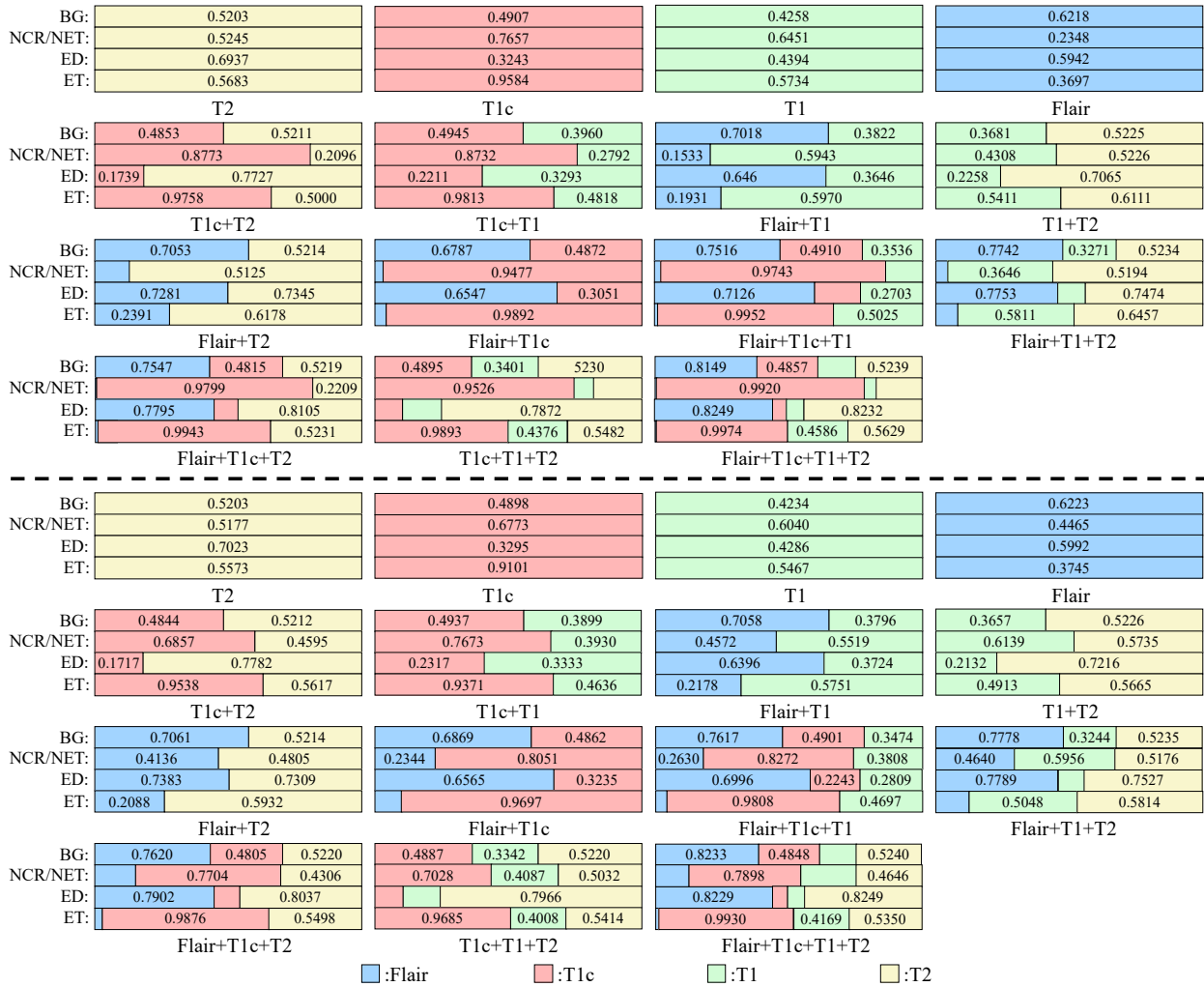
:Flair   :T1c   :T1   :T2

Figure 4. Visualization of the generated attention weights by our RFM at the fourth stage. Two examples are shown in the top and bottom parts, respectively. For each example, fifteen panels illustrate all combinations of four image modalities. In each panel, attention weights (in numbers) are used to aggregate available modalities (in colors) adaptively in diverse regions (in rows). Larger colored bars denote larger attention weights for the corresponding modality.
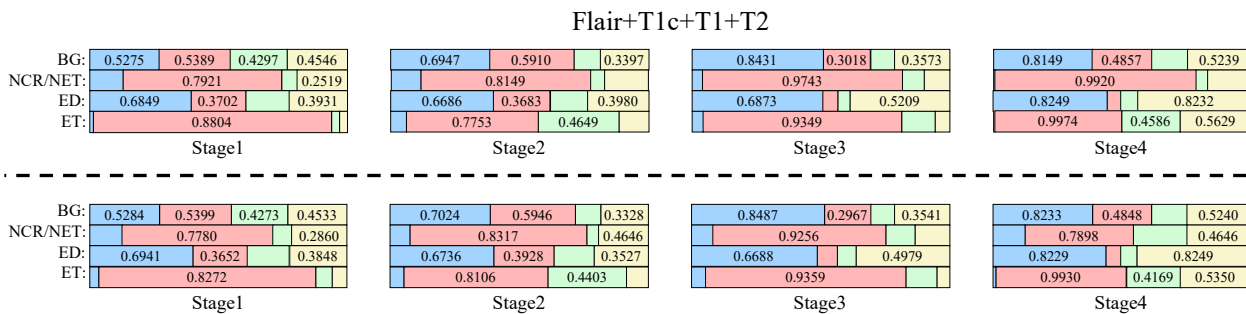
Flair+T1c+T1+T2

Figure 5. Visualization of the generated attention weights from the full set of four image modalities in different stages. Two examples are shown in the top and bottom parts, respectively. In each example, four panels indicate the attention weights at different stages in our network.
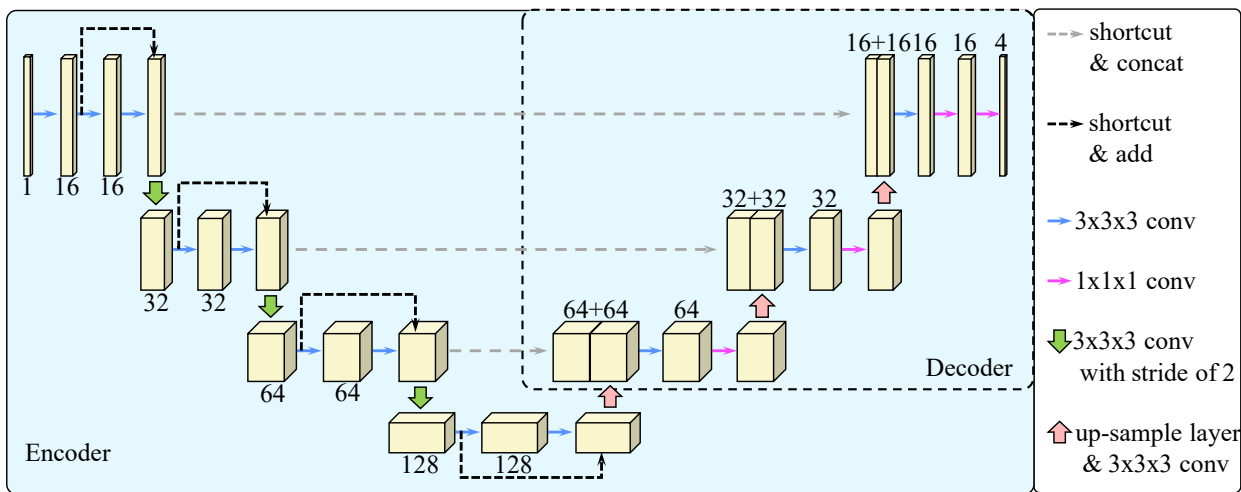
Figure 6. Visualization of the encoder and decoder in RFNet. Each convolutional layer, except for the last segmentation layer, is followed by an instance normalization layer and Leaky ReLU activation.