Appendix: High-Quality Fusion and Visualization for MR-PET Brain Tumor Images via Multi-Dimensional Features

Jinyu Wen, Asad Khan, Amei Chen, Weilong Peng, Meie Fang, C. L. Philip Chen, *Fellow, IEEE*, and Ping Li, *Member, IEEE*

I. Quality Metric

The specific calculation equation of quality evaluation metric (rSFe [1]) is as:

$$\mathrm{rSFe} = \frac{\sqrt{\sum_{k\in\Omega_1} SF_k^2} - \sqrt{\sum_{l\in\Omega_2} Grad_l I_R}}{\sqrt{\sum_{l\in\Omega_2} Grad_l I_R}}, \tag{1}$$

where $\Omega_1 = \{\mathrm{RF, CF, MDF, SDF}\}, \Omega_2 = \{\mathrm{H, V, MD, SD}\}$, I_R is the fused image, I_A and I_B are two source images. And SF_k are calculated according to the following formulas:

$$SF_1 = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[I_R(i,j) - I_R(i,j-1)]^2}, \tag{2}$$

$$SF_2 = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[I_R(i,j) - I_R(i-1,j)]^2}, \tag{3}$$

$$SF_3 = \sqrt{w_\mathrm{d}\cdot\frac{1}{MN}\sum_{i=2}^{M}\sum_{j=2}^{N}[I_R(i,j) - I_R(i-1,j-1)]^2}, \tag{4}$$

$$SF_4 = \sqrt{w_\mathrm{d}\cdot\frac{1}{MN}\sum_{j=1}^{N-1}\sum_{i=2}^{M}[I_R(i,j) - I_R(i-1,j+1)]^2}, \tag{5}$$

Manuscript received 25 September 2023; revised 02 April 2024; accepted 22 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62072126 and Grant 61772164, in part by the Fundamental Research Projects Jointly Funded by Guangzhou Council and Municipal Universities under Grant 202102010439, in part by the PolyU Research Institute for Sports Science and Technology under Grant P0044571, and in part by The Hong Kong Polytechnic University under Grant P0048387, Grant P0042740, Grant P0044520, Grant P0043906, Grant P0049586, and Grant P0050657. *(Corresponding author: Meie Fang.)*

Jinyu Wen, Asad Khan, Weilong Peng, and Meie Fang are with the Metaverse Research Institute, School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 511400, China (e-mail: wjy1361120721@163.com; {asad, wlpeng, fme}@gzhu.edu.cn).

Amei Chen is with the Department of Radiology, Second Affiliated Hospital of South China University of Technology, Guangzhou 510641, China (e-mail: 17690189@qq.com).

C. L. Philip Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China, and also with the Pazhou Lab, Guangzhou 510335, China (e-mail: philip.chen@ieee.org).

Ping Li is with the Department of Computing, the School of Design, and the Research Institute for Sports Science and Technology, The Hong Kong Polytechnic University, Hong Kong (e-mail: p.li@polyu.edu.hk).

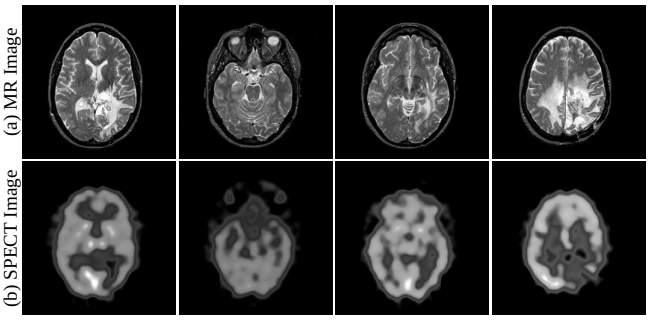

Fig. 1. Four pairs of MR-PET sources images.

where the image size is $M \times N$, w_d represents the distance weight, we take its value as $1/\sqrt{2}$.

Further, $Grad_l I_R$ can be calculated as:

$$Grad_l I_\mathrm{R} = \max_{l\in\Omega_2}\{\mathrm{abs}(Grad_l I_A), \mathrm{abs}(Grad_l I_B)\}. \tag{6}$$

II. Loss Function

The total loss function consists of structural similarity L_S and pixel loss L_M. I_f represents the result of fusion, $I_i, i = 1, 2$ represents the input image, the source image. $SSIM$ represents the structural similarity of two images [11]. For the two types of images, the structural similarity is calculated and combined. Assuming that the structural similarity of the two images is F_{SSIM_i}, $F_{SSIM_i} = 1 - SSIM(I_f, I_i), i = 1, 2$, and the total structural similarity function is calculated as follows:

$$L_S = \sum_{i=1}^{2}\left(1 - \frac{\left(2\mu_{I_f}\mu_{I_i} + c_1\right)\left(2\sigma_{I_f}\sigma_{I_i} + c_2\right)}{\left(\mu_{I_f}^2 + \mu_{I_i}^2 + c_1\right)\left(\sigma_{I_f}^2 + \sigma_{I_i}^2 + c_2\right)}\right), \tag{7}$$

where, μ_{I_f} is the mean of I_f, μ_{I_1} and μ_{I_2} are mean values of I_1 and I_2 respectively. $\mu_{I_f}^2$ is the variance of I_f, $\mu_{I_1}^2$ and $\mu_{I_2}^2$ are variance of I_1 and I_2 respectively. $\sigma_{I_f I_1}$ and $\sigma_{I_f I_2}$ are the covariance between original image and fused image, c_i are constants. The pixel loss is the l_2-norm. Assuming that the pixel losses of the two images are $F_{M_i}, i = 1, 2$, $F_{M_i} = \|I_f - I_i\|_2$, and the total pixel loss function is calculated as follows:

$$L_M = \sum_{j=1}^{n}\left(\left(I_{f_j} - I_{1_j}\right)^2 + \left(I_{f_j} - I_{2_j}\right)^2\right). \tag{8}$$

TABLE I
COMPARE THE INFORMATION ENTROPY OF THE FUSED IMAGE WITH THAT OF THE TWO SOURCE IMAGES AND ANALYZE STATISTICAL SIGNIFICANCE USING T-TESTS.

	MR-CT		**MR-PET**		**MR-SPECT**	
EN	vs. MR	vs. CT	vs. MR	vs. PET	vs. MR	vs. SPECT
P-value	3.89e-14	4.52e-18	1.85e-30	1.49e-19	3.03e-60	5.17E-27
T-statistic	18.6975	29.8863	31.6935	16.3363	54.4362	17.0669

TABLE II
COMPARISON WITH SOTA METHODS IN QUANTITATIVE RESULTS OF MEDICAL IMAGE FUSION OF 21 MR-CT PAIRS RESPECTIVELY.

MR-CT	**EN**	**MI**	**SD**	**VIF**	**SCD**	**SSIM**	**CC**	**MSE**	**rSFe**
FunFuseAn [2]	5.0245	3.1844	58.5545	0.6295	0.9037	1.5134	0.8263	0.0940	-0.4218
FusionDN [3]	4.8268	**3.2174**	69.0312	0.3832	0.9328	0.4696	0.7698	0.1007	-0.4246
U2Fusion [4]	4.9148	2.1846	55.1028	0.3055	**1.6707**	0.4371	**0.8347**	0.0460	-0.5763
DeFusion [5]	4.6002	2.1969	66.1716	0.4649	1.1152	1.2976	0.8153	3.1185	-0.4319
CDDFuse [6]	4.8309	2.2431	**79.4820**	0.6100	1.4100	1.3094	0.8285	0.0753	**-0.0255**
MdAFuse	**5.1195**	3.0399	56.9286	**0.6790**	0.9250	**1.5666**	0.8343	**0.0241**	-0.4661

TABLE III
COMPARISON WITH SOTA METHODS IN VARIANCE OF MEDICAL IMAGE FUSION OF 74 MR-SPECT PAIRS AND 42 MR-PET PAIRS RESPECTIVELY.

MR-SPECT	**EN**	**MI**	**SD**	**VIF**	**SCD**	**SSIM**	**CC**	**MSE**	**rSFe**
FunFuseAn [2]	0.4269	0.2738	8.3265	0.1237	0.3010	0.0990	**0.0912**	0.0192	0.0533
FusionDN [3]	0.4303	0.2393	**5.7993**	0.0825	0.4063	**0.0521**	0.1044	0.0264	0.0772
U2Fusion [4]	**0.4112**	0.5997	8.5349	**0.0291**	0.2147	0.1715	0.1075	0.0092	**0.0220**
DeFusion [5]	0.4352	**0.2019**	7.6711	0.0552	0.2872	0.1305	0.1161	2.9011	0.0294
CDDFuse [6]	0.4610	0.2809	10.6004	0.0417	**0.1856**	0.1428	0.1097	0.0265	0.0299
MdAFuse	0.4585	0.2827	7.4872	0.1826	0.2173	0.0912	0.1014	**0.0079**	0.0598
MR-PET	**EN**	**MI**	**SD**	**VIF**	**SCD**	**SSIM**	**CC**	**MSE**	**rSFe**
FunFuseAn [2]	0.7291	0.3290	7.8834	0.0705	0.2697	0.0829	0.0511	0.0351	0.0629
FusionDN [3]	**0.7024**	0.3151	**5.3841**	0.0711	0.2097	0.0950	0.0579	0.0261	0.0782
U2Fusion [4]	0.7078	0.3118	8.2173	**0.0584**	**0.0993**	**0.0569**	**0.0499**	0.0128	**0.0390**
DeFusion [5]	0.7282	**0.2291**	7.1749	0.1042	0.2535	0.1139	0.0622	0.0202	0.1050
CDDFuse [6]	0.7164	0.2763	8.4459	0.1356	0.1787	0.0973	0.0616	0.0121	1.6354
MdAFuse	0.7630	0.3050	6.9127	0.0878	0.2658	0.0903	0.0519	**0.0071**	0.0743
MR-CT	**EN**	**MI**	**SD**	**VIF**	**SCD**	**SSIM**	**CC**	**MSE**	**rSFe**
FunFuseAn [2]	0.2048	0.1222	3.2153	0.0995	0.0916	0.0429	0.0107	0.0178	0.0284
FusionDN [3]	0.2075	0.1632	3.8607	0.0435	0.1608	0.0328	0.0182	0.0196	0.0156
U2Fusion [4]	0.1824	**0.0994**	**2.2788**	0.0345	**0.0269**	**0.0314**	0.0135	0.0067	0.0171
DeFusion [5]	0.1884	0.1029	4.5483	**0.0175**	0.0890	0.0558	0.0147	0.0103	0.0243
CDDFuse [6]	0.2129	0.2623	4.4536	0.0495	0.0838	0.0384	0.0135	0.0119	**0.0049**
MdAFuse	**0.1808**	0.1413	2.8292	0.0897	0.0903	0.0433	**0.0103**	**0.0035**	0.0233

III. STATISTICAL SIGNIFICANCE ANALYSIS

In order to validate the statistical significance of our image fusion algorithm, we employed T-tests. As there is no ground truth available in image fusion, we separately computed the information entropy of two original images and the fused image. We then conducted T-tests to compare the information entropy of the fused image with that of the original images. The results of the T-tests are presented in Table I, it is evident that in the fusion across different modalities, the P-values are significantly smaller than the significance level (0.05). This suggests a significant difference between our fusion results and the original images, which is of great statistical significance. Furthermore, the T-statistic values in the table are notably large. T-statistic is a metric used to gauge the difference in means between two sets of data. Our results indicate a significant difference between the image fusion results and the original conditions. Our image fusion algorithm demonstrates distinct effects on the original images, reflecting its effectiveness in aspects such as information preservation and contrast enhancement.

IV. QUANTITATIVE ANALYSIS

In addition to conducting fusion experiments on MR-PET (See Fig. 1) and MR-SPECT, we also performed experiments on MR-CT. We tested 21 pairs of MR-CT images and computed the average values for nine evaluation metrics, as shown

TABLE IV
TIMES SELECTED OF FUSION IMAGES FROM EACH METHOD IN THE QUESTIONNAIRE SURVEY.

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	$\sum$
LatLrr [7]	6	4	4	7	5	10	1	6	10	7	9	17	5	10	5	106
atsIF [8]	9	3	10	3	9	7	8	6	3	11	8	7	12	4	12	112
DeepVTF [9]	5	9	7	6	4	6	8	13	4	3	5	3	3	5	6	87
FusionDN [3]	4	2	3	8	6	2	6	4	8	2	5	3	3	6	4	66
VIFNet [10]	5	15	3	8	5	7	4	3	1	10	4	1	7	3	3	79
Proposed	9	3	10	6	8	7	12	6	13	7	7	7	7	9	10	121

in Table II. From the data in the Table II, it can be observed that our method maintains an advantage in most metrics. For instance, on the SSIM metric, our method improves by 5.32% compared to the second-best result. On MSE, our method achieves a 2.19% reduction compared to the second-best result. Despite CDDFuse being the latest state-of-the-art method, our approach outperforms it on six of the evaluated metrics. Additionally, we calculated the variance after fusion across different modalities, as presented in the Table III.

V. QUESTIONNAIRE SURVEY

To validate the clinical effectiveness of our algorithm, we conducted an online questionnaire survey on the fusion results. We distributed the survey questionnaire to 31 doctors from different neurology and medical imaging departments in various hospitals. The doctors have clinical experience of over 10 years (9 doctors), 5-10 years (1 doctor), 3-5 years (6 doctors), and less than 3 years (15 doctors). To enable medical professionals to efficiently complete the questionnaire within a limited time, we selected the fusion results of 6 methods for investigation. In this questionnaire, we designed 15 questions based on 15 fusion experiments. For each question, the order of fused images from different methods was randomly arranged. Each respondent was allowed to select one or two options for the best fusion results for each question. The doctors did not need to determine how much original image information was retained in the fused image. They only needed to judge which fusion result was more beneficial to their observations and clinical diagnoses based on their clinical experience. In the end, we received valid answers from 29 participants.

The statistical results are shown in Table IV, in which the times selected for the fusion images from each method are recorded. In each column marked with $T_i (i = 1, \cdots, 15)$, the following numbers correspond to the times selected by doctors for the fusion images of the six methods. In 15 groups of experiments, the fusion images produced by proposed approach are most frequently selected as the best fusion images in 8 groups, and second frequently selected as the best fusion images in 4 groups. From the view of clinical doctors, the fusion effect of proposed approach far exceeds that of any other considered method. The last column marked with $\sum$ lists the total times for each kind of fusion method whose fusion images were selected. The calculation shows that our method is selected 9 more times than the suboptimal method.

REFERENCES

[1] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Information Fusion*, vol. 8, no. 2, pp. 177–192, 2007.

[2] N. Kumar, N. Hoffmann, M. Oelschlägel, E. Koch, M. Kirsch, and S. Gumhold, "Structural similarity based anatomical and functional brain imaging fusion," in *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*, 2019, vol. 11846, pp. 121–129.

[3] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A unified densely connected network for image fusion," in *AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 484–12 491.

[4] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 502–518, 2022.

[5] P. Liang, J. Jiang, X. Liu, and J. Ma, "Fusion from decomposition: A self-supervised decomposition approach for image fusion," in *European Conference on Computer Vision*, vol. 13678, 2022, pp. 719–735.

[6] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool, "CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5906–5916.

[7] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," *arXiv preprint arXiv:1804.08992*, pp. 1–6, 2018.

[8] J. Du, M. Fang, Y. Yu, and G. Lu, "An adaptive two-scale biomedical image fusion method with statistical comparisons," *Computer Methods and Programs in Biomedicine*, vol. 196, pp. 105 603:1–105 603:13, 2020.

[9] I. Shopovska, L. Jovanov, and W. Philips, "Deep visible and thermal image fusion for enhanced pedestrian visibility," *Sensors*, vol. 19, no. 17, pp. 3727:1–3727:21, 2019.

[10] R. Hou, D. Zhou, R. Nie, D. Liu, L. Xiong, Y. Guo, and C. Yu, "VIF-Net: An unsupervised framework for infrared and visible image fusion," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 640–651, 2020.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.