











Information fusion for fully automated segmentation of head and neck tumors from PET and CT images

Isaac Shiri¹  | Mehdi Amini¹  | Fereshteh Yousefirizi² | Alireza Vafaei Sadr^{3,4}  |
 Ghasem Hajianfar¹  | Yazdan Salimi¹  | Zahra Mansouri¹  | Elnaz Jenabi⁵ |
 Mehdi Maghsudi⁶ | Ismini Mainta¹  | Minerva Becker⁷  | Arman Rahmim^{2,8}  |
 Habib Zaidi^{1,9,10,11} 

¹Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland

²Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, British Columbia, Canada

³Institute of Pathology, RWTH Aachen University Hospital, Aachen, Germany

⁴Department of Public Health Sciences, College of Medicine, The Pennsylvania State University, Hershey, USA

⁵Research Center for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

⁶Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran

⁷Service of Radiology, Geneva University Hospital, Geneva, Switzerland

⁸Department of Radiology and Physics, University of British Columbia, Vancouver, Canada

⁹Geneva University Neurocenter, Geneva University, Geneva, Switzerland

¹⁰Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

¹¹Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

Correspondence

Habib Zaidi, Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland.
 Email: habib.zaidi@hcuge.ch

Abstract

Background: PET/CT images combining anatomic and metabolic data provide complementary information that can improve clinical task performance. PET image segmentation algorithms exploiting the multi-modal information available are still lacking.

Purpose: Our study aimed to assess the performance of PET and CT image fusion for gross tumor volume (GTV) segmentations of head and neck cancers (HNCs) utilizing conventional, deep learning (DL), and output-level voting-based fusions.

Methods: The current study is based on a total of 328 histologically confirmed HNCs from six different centers. The images were automatically cropped to a 200 × 200 head and neck region box, and CT and PET images were normalized for further processing. Eighteen conventional image-level fusions were implemented. In addition, a modified U2-Net architecture as DL fusion model baseline was used. Three different input, layer, and decision-level information fusions were used. Simultaneous truth and performance level estimation (STAPLE) and majority voting to merge different segmentation outputs (from PET and image-level and network-level fusions), that is, output-level information fusion (voting-based fusions) were employed. Different networks were trained in a 2D manner with a batch size of 64. Twenty percent of the dataset with stratification concerning the centers (20% in each center) were used for final result reporting. Different standard segmentation metrics and conventional PET metrics, such as SUV, were calculated.

Results: In single modalities, PET had a reasonable performance with a Dice score of 0.77 ± 0.09 , while CT did not perform acceptably and reached a Dice score of only 0.38 ± 0.22 . Conventional fusion algorithms obtained a Dice score range of [0.76–0.81] with guided-filter-based context enhancement (GFCE) at the low-end, and anisotropic diffusion and Karhunen–Loeve transform fusion (ADF), multi-resolution singular value decomposition (MSVD), and multi-level image decomposition based on latent low-rank representation (MDLatLRR) at the high-end. All DL fusion models achieved Dice scores of 0.80. Output-level voting-based models outperformed all other models, achieving superior results with a Dice score of 0.84 for Majority_ImgFus, Majority_All, and Majority_Fast. A

Isaac Shiri and Mehdi Amini contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

Funding information

Swiss National Science Foundation,
Grant/Award Numbers: SNSF
320030_176052, 320030_173091/1; Natural
Sciences and Engineering Research Council
of Canada (NSERC), Grant/Award Number:
RGPIN-2019-06467

mean error of almost zero was achieved for all fusions using SUV_{peak} , SUV_{mean} and SUV_{median} .

Conclusion: PET/CT information fusion adds significant value to segmentation tasks, considerably outperforming PET-only and CT-only methods. In addition, both conventional image-level and DL fusions achieve competitive results. Meanwhile, output-level voting-based fusion using majority voting of several algorithms results in statistically significant improvements in the segmentation of HNC.

KEYWORDS

deep learning, fusion, head and neck cancer, PET/CT, segmentation

1 | INTRODUCTION

Radiation therapy is a standard treatment approach for head and neck cancer (HNC); therefore, it is crucial to use non-invasive, specialized techniques for improved diagnosis and focused therapies based on tumor phenotype.¹ The evaluation of HNC has greatly profited from combined [¹⁸F]fluoro-2-deoxyglucose positron-emission tomography ([¹⁸F]-FDG PET) and computed tomography (CT) imaging. Due to the combined assessment of anatomy and metabolism, PET/CT holds a prominent role in the detection and characterization of HNCs locally, at the nodal level and for detecting distant metastases.^{2,3} [¹⁸F]FDG uptake in HNC, as a biomarker of glucose metabolism, is typically higher at disease sites than in normal tissues. However, due to partial volume effects, inherent tumor heterogeneity and inferior PET's spatial resolution compared to CT and MR imaging, assessment of deep submucosal tumor spread, which is essential for tumor staging and treatment planning remains imprecise. With the help of CT, abnormal [¹⁸F]FDG uptake is identified more accurately⁴ not only for primary tumor staging and follow-up of HNC but also for the development and evaluation of diagnostic and prognostic models based on HNC segmentation. Medical image segmentation aims to extract regions with distinct anatomical and/or functional features and categorizing pixels (voxels) according to their texture, grayscale, and other factors.⁵ On one hand, segmentation can help radiation oncologists plan treatments more effectively by reducing tumor delineation time and by enhancing observer reproducibility.⁶ On the other hand, the usefulness of radiomics for prognosis, diagnosis, and treatment assessment is undisputable, and segmentation is an essential first step.^{7–11} Furthermore, radiomics analysis relies on engineered feature extraction within specified lesions or volumes of interest (VOIs) for robust and reproducible modeling.¹² Therefore, efficient solutions for the automatic segmentation of Gross Tumor Volume of primary tumors (GTVt) are of major interest.

Since manual segmentation is challenging, time-consuming, and prone to error, deep learning (DL)-based automatic segmentation appears as an attractive alter-

native for automated tumor segmentation.^{12,13} However, the segmentation task remains challenging due to the complexity of PET/CT imaging and the high cost of processing 3D data.¹⁴ Convolutional neural networks (CNNs)-based automated tumor segmentation methods have been presented as a possible solution.^{15,16} Although missing small lesions, some approaches have demonstrated excellent results despite requiring the development of significant computational resources.^{15,16} However, before employing CNN results for additional analysis, verification and correction by an imaging expert is required.^{15,16} This necessitates a comprehensive visual examination of each 3D [¹⁸F]-FDG PET/CT scan and identifying lesions the algorithm may have missed.⁵ As a result, creating a pipeline that would facilitate this process of checking and adjusting is highly desirable in clinical practice.⁵

Although CT images offer detailed anatomical information, it can be challenging to determine the precise tumor extent given the relatively low lesion-to-background ratio in infiltrating tumors and tumors with large areas of necrosis, which are the hallmarks of HNC.^{17,18} Algorithms using complementary data from multi-modality acquisitions, including quantitative assessments based on PET, in particular, standardized uptake values (SUVs) have been proposed to increase the accuracy of tumor segmentation.¹⁹ Integrating the information from different modalities can be performed at several levels.²⁰ The first level includes fusing scans from different modalities before entering the networks by using conventional image fusion techniques. The next level comprises feeding the network with different modalities simultaneously and combining the information from these modalities at different stages of the learning process, that is, input layers, network levels, and output layers of the network architecture. The final strategy includes combining the decision of different models using staple or voting methods. When considering both functional and anatomical disease extent, any algorithms that analysis imaging modality independently are inherently limited. In contrast, techniques that combine data from two modalities often 'prioritize' data from one of the two modalities for various tasks using a priori knowledge of different modality features (i.e., PET

and CT). Alternatively, data may be combined using a fusion model incorporating the two modalities.

Image fusion aims to merge information from different scenes or perspectives into a single data while ideally preserving important information, reducing information loss, and preventing emerging artifacts and misinformation on fused images.²¹ Furthermore, medical image fusion often leads to generating new patterns and textures, which are otherwise not visible to experts.²² Hence, higher performance for the detection of abnormalities, their classification, segmentation, and estimation of improved prognosis can be achieved.²³ Conventional PET/CT segmentation fusion techniques include hybrid techniques as suggested by Bagci et al.⁴ simultaneously delineating ROIs in PET, PET/CT, PET/MR imaging, and fused MRI/PET/CT segmentation using a random walk segmentation with an automated similarity-based seed selection process. In order to classify solitary pulmonary nodules using CT textural features and PET metabolic features, Zhao et al.²⁴ recommended combining dynamic thresholding, watershed segmentation, and support vector machine. Ju et al.,²⁵ Yang et al.,²⁶ and Song et al.²⁷ used PET and CT regions to characterize tumors with spatial and visual consistency. Han et al.²⁸ segmented tumors from PET/CT scans using a Markov Random Field formulation with specific energy terms for both multimodalities.

Different scenarios were proposed for using the complementary information of multi-modality images based on fusion at different levels of DL-based (input-level, feature-level, and decision-level) segmentation in multi-modality imaging.^{29,30} Zhong et al.,³¹ used a graph-cut approach to combine U-Net training results of PET and CT. Li et al.³² described a variational model to segment lung cancers using PET image and a CNN-derived CT probability map. Kumar et al.³³ suggested that the appropriate fusion of multi-modality images should be learned from the underlying visual properties of a single modality, while DL techniques provided the most advanced feature learning, selection, and extraction techniques. By learning to combine complementary information from PET/CT images, Kumar et al.³³ proposed the creation of a fusion by CNN that explicitly quantifies fusion feature weight in each modality.

The purpose of the current study was: (i) to evaluate several information fusion approaches, including conventional, DL and output level fusion, and (ii) to investigate the efficacy of fusion at various levels for HNC GTV segmentation on PET/CT images.

2 | MATERIALS AND METHODS

2.1 | PET/CT data acquisition and description

The data used in this study included 328 histologically confirmed head and neck (HN) cancer patients,

undergone PET/CT scans in six different clinics (dataset #1–6) with 23, 32, 34, 59, 81, and 99 patients, respectively. Patient demographics and imaging characteristics (image acquisition protocols, scanner specifications, image reconstruction, and correction algorithms) of the datasets are presented in Table 1. More in-depth details of the dataset can be found in.^{2,16,34–41} For training the models, the whole dataset was divided into segments of 70%, 10%, and 20% (by preserving the ratios of centers) for training, validation, and testing, respectively.

2.2 | Manual image segmentation and pre-processing

The primary tumor was delineated on PET/CT images for radiation therapy purposes.^{16,34–41} However, before enrolling in this study, an experienced nuclear medicine physician edited/verified segmentations for errors, such as missing slices or airway lumens. The revised tumor segmentations were used as the ground truth. Metal artifacts in CT images (Hounsfield Unit (HU) maps) were corrected using Iterative Metal Artifact Reduction (iMAR) algorithm.⁴² PET images were converted to Standardized Uptake Value (SUV) maps. All the scans were resized to $1 \times 1 \times 1 \text{ mm}^3$ voxels, then automatically cropped to 200×200 grids containing HN regions. This automatic procedure was performed by identifying the starting slice of the HN region using a DL-based lung segmentation (helped to identify the slice above the lung apex) and body contouring using PET intensity thresholding. For a better contrast in CT images, HU maps were clipped to $[-1024, 1200]$. Then, both CT and PET images were normalized. All pre-processing steps were conducted automatically to enable fully automated HN tumor segmentations on PET/CT images.

2.3 | Conventional image fusion

A variety of image fusion methods have been developed in the previous decades. Image fusion methods can be categorized based on the level of fusion (pixel-level, feature-level, decision-level), fusion domain (spatial domain or transform domain), application (multi-modal, multi-focus, multi-exposure, etc.), and algorithm type. Regarding the backbone algorithm, image fusion methods can be classified into seven groups: DL approach, multi-scale transform-based, sparse representation-based, subspace-based, saliency-based, hybrid models, and other miscellaneous algorithms.

To cover all groups, we fused PET and CT images using 18 different methods, including most recent state-of-the-art methods. We investigated the following methods: anisotropic diffusion and Karhunen–Loeve transform fusion (ADF),⁴³ cross bilateral filter (CBF),⁴⁴ guided filtering fusion (GFF),⁴⁵ guided-filter-based

TABLE 1 Image acquisition and reconstruction parameters from the different centers.

	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6
Number	23	32	34	59	81	99
Demographics	Weight (Kg) 81.3 ± 11 Age 53.2 ± 12 PET Siemens Biograph Truepoint	84.04 ± 18.34 55.78 ± 7.64 GE-Discovery	74.57 ± 22.53 66.60 ± 9.19 GE-Discovery	76.85 ± 12.10 63.46 ± 9.19 GE-Discovery ST, Phillips Guardian Body	77.15 ± 17.18 61.09 ± 10.95 GE-Discovery ST	75.88 ± 19.01 63.89 ± 10.05 Phillips Guardian Body
CT acquisition and reconstruction parameters	kVp 120 Average tube current 223.7 ± 32.1	120 248.2 ± 82.1	120,140 197.1 ± 118.5	120 179.2 ± 87.6	120,140 208.4 ± 103.9	120, 140 374.1 ± 35.7
PET acquisition and reconstruction parameters	Matrix size 512 × 512 Injected activity (MBq) 402.23 ± 83.12 Time to scan (min) 64.31 ± 21.04 Time per bed (min) 2.6 ± 0.13	512 × 512 588.73 ± 115.83	512 × 512 472.02 ± 164.15	512 × 512 374.35 ± 352.62	512 × 512 564.77 ± 62.03	512 × 512 323.32 ± 75.23
Scatter correction	Model-based	Model-based	Model-based	Convolution subtraction, SS-SIMUL	Convolution subtraction	SS-SIMUL
Reconstruction	OSEM	OSEM	OSEM	OSEM	OSEM	LOR-RAMLA
Matrix size	168 × 168	128 × 128	128 × 128	128 × 128, 144 × 144	128 × 128	144 × 144
Slice thickness	3	3.27	3.27	3.75 ± 0.32	3.27	4

Abbreviations: CT, computed tomography; kVp, kilovoltage peak; LOR-RAMLA, line-of-response row-action maximum likelihood algorithm; OSEM, ordered subsets expectation maximization; SS-SIMUL, single-scatter simulation.

context enhancement (GFCE),⁴⁶ hybrid multi-scale decomposition (Hybrid-MSD),⁴⁷ multi-scale guided filtered-based fusion (MGFF),⁴⁸ hybrid multi-scale decomposition fusion method using guided image filtering (HMSD-GF),⁴⁶ multi-resolution singular value decomposition (MSVD),⁴⁹ and wavelet fusion,⁵⁰ one subspace-based group method, fourth order partial differential equations (FPDE),⁵¹ two saliency-based group methods (a multi-level image decomposition method based on latent low-rank representation (MDLatLRR),⁵² and two-scale image fusion (TIF),⁵³ four hybrid methods (ratio of low-pass pyramid with sparse representation (RP-SR),⁵⁴ visual saliency map and weighted least square optimization (VSM-WLS),⁵⁵ multi-scale transform with sparse representation (MST-SR),⁵⁴ non-subsampled contourlet transform with sparse representation (NSCT-SR)⁵⁴), and finally two methods from other categories gradient transfer and total variation minimization fusion (GTF),⁵⁶ and infrared feature extraction and visual information preservation (IFEVIP).⁵⁷ All fusion algorithms were applied on 2D slices, except the wavelet method. All methods were adopted from the Visible and Infrared Image Fusion Benchmark (VIFB) package.⁵⁸

2.4 | DL model

We implemented the modified U2-Net architecture as the baseline of DL model.⁵⁹ Deep supervision and residual U-blocks were integrated into the U2-Net architecture.⁵⁹ The U-structure of the U-Net³¹ was kept, but inside each block, another U-Net with symmetric encoder-decoder architecture was implemented. This U2-Net architecture helps to prevent resolution degradation and feature extraction with a mixture of receptive fields and keeps the computational and memory footprint low.⁵⁹ Figure 1 illustrates the network architecture and fusion implemented in this study.

2.5 | DL fusion

In the input-level fusion scenario, images from PET and CT modalities are fused at the first channels, inputting as multi-channel to one network for the segmentation task. As such, PET and CT are regarded as two input channels in which, in this scenario, different crucial information (anatomical and metabolic) for this task could be fully exploited at different feature levels. In the decision-level fusion scenario, two different models were trained separately for PET and CT, each with a single input, and at the end, the segmentation results are integrated. In our implementation, averaging of the confidence of each network was used. Finally, information fusion is performed in layers by connecting different layers in the network-level fusion scenario. In our implementation, PET and CT go through various

networks, and the weights of the different layers are connected during training. All layers are connected in Fusion_NetLev fusion.

2.6 | STAPLE and majority voting

Simultaneous truth, performance level estimation (STAPLE)⁶⁰ and majority voting were implemented to merge different segmentation outputs as output-level voting-based information fusion. We implemented output-level voting-based fusion of different scenarios, including output-level voting-based fusion of all image segmentation methods (conventional and DL-based), conventional image fusion output, DL-based output, and fast (excluding time-consuming methods) NSCT, MDLatLRR, and CBF methods output fusion.

2.7 | Training the networks

The networks were trained in a 2D fashion with a batch size of 64, an Adam optimizer, a learning rate of 0.001, a Dice loss, and a weight decay of 0.0001. Twenty percent of the dataset with stratification of centers was used for the final result reporting, where 80% of the data set was used for training and validation (70/10%, respectively), while the rest (20% of the data set) was untouched for the hold-out test set.

2.8 | Quantitative evaluation metabolic activity intensity and shape analysis

Standard segmentation metrics, including Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), false-negative rate, false-positive rate, volume similarity, and mean and standard deviation (SD) of surface distance with manual segmentation were calculated to evaluate and compare models' performance. Moreover, the performance of all models was assessed from a clinical perspective using image-derived PET metrics, including SUV_{peak} , SUV_{mean} , SUV_{median} , SUV_{max} , Metabolic Tumor Volume (MTV) and Total Lesion Glycolysis (TLG). In addition to the abovementioned conventional PET metrics, morphological features, including sphericity, asphericity, elongation and flatness, were mined using the SERA radiomics package.^{60–62} The mean error (ME) of the abovementioned metrics was reported for each model with respect to manual segmentations. All quantitative metrics were extracted from the 3D segmented volumes.

2.9 | Statistical analysis

All metrics were described statistically with mean \pm SD and 95% confidence interval (CI). First, conventional

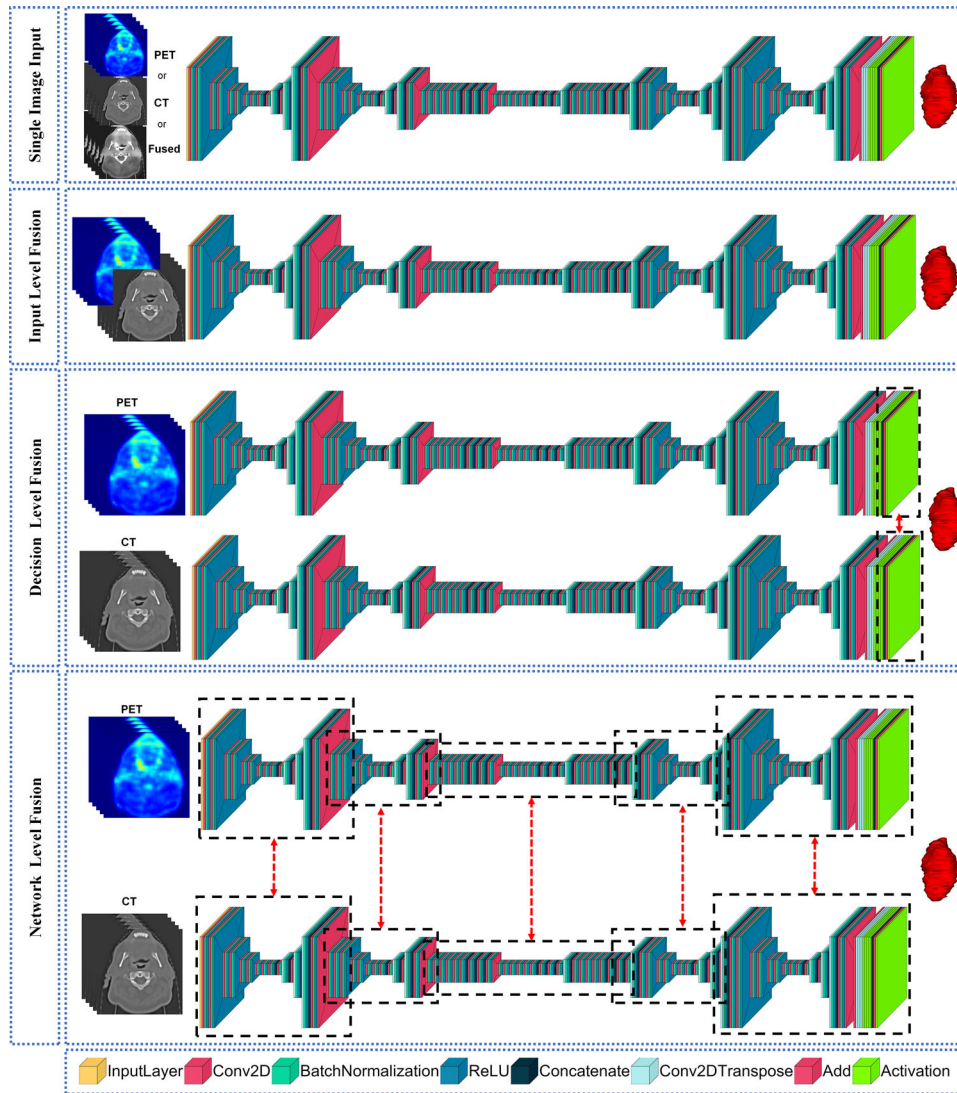


FIGURE 1 Network architecture in different scenarios, including single image input (PET, CT, or fused image), input-level network fusion (PET and CT input simultaneously), decision-level network fusion (PET and CT go through different networks, and fusion performed on the final layer), and network-level fusion (weights are connected at different level of networks). CT, computed tomography.

PET metrics derived from each model segmentation mask were compared with their peers derived from the manual segmentation mask, using the ME. Next, the Kolmogorov-Smirnov test was utilized to evaluate normal distributions. Then, the Wilcoxon Signed Rank Test was used to compare the paired distributions of different segmentation metrics. A p -value <0.05 was set as the significance threshold. The algorithms were implemented on a desktop PC with Intel(R) Core(TM) i9-10900KF CPU 3.70 GHz, “32 GB RAM, and NVIDIA GEFORCE RTX 2080 Ti GPU.”

3 | RESULTS

The mean and standard deviation of performance metrics for image segmentations are shown in Table 2 for

CT, PET, and the different fusion models. Their 95% confidence interval is presented in Table S1. The corresponding boxplots are illustrated in Figure 2. Since the loss function was based on the Dice score, we reported and compared models using the Dice score metric. Nevertheless, other metrics are presented for a more thorough comparison. Regarding single modalities, PET had a reasonable performance with a Dice score of 0.77 ± 0.09 , while CT did not perform well, reaching a Dice score of only 0.38 ± 0.22 . Conventional algorithms obtained a Dice score range of [0.76, 0.81] with GFCE (0.76 ± 0.1) at the low-end and ADF (0.81 ± 0.08), MSVD (0.81 ± 0.09), and MDLatLRR (0.81 ± 0.07) at the high-end, respectively. Network fusion models achieved Dice scores of 0.80 ± 0.09 , 0.80 ± 0.09 , and 0.80 ± 0.1 for InpLev, DesLev, and NetLev, respectively. Output fusion models (voting

TABLE 2 Summary of quantitative image segmentation performance metrics (CI 95%) for the different image fusion and network fusion models.

Fusion name	Dice	False negative	False positive	Hausdorff distance	Jaccard	Max surface distance	Mean surface distance	Std surface distance	Avg. Hausdorff Distance
CT	0.38 ± 0.22	0.68 ± 0.21	0.39 ± 0.29	21.13 ± 9.11	0.26 ± 0.17	21.13 ± 9.1	1.9 ± 2.06	3.52 ± 2.72	3.66 ± 4.62
PET	0.77 ± 0.09	0.2 ± 0.12	0.23 ± 0.16	16.38 ± 14.63	0.63 ± 0.12	16.83 ± 14.33	0.46 ± 0.4	1.39 ± 1.16	0.57 ± 0.72
ADF	0.81 ± 0.08	0.17 ± 0.1	0.18 ± 0.13	13.58 ± 9.77	0.69 ± 0.11	13.61 ± 9.75	0.4 ± 0.39	1.27 ± 1.29	0.49 ± 0.66
CBF	0.78 ± 0.14	0.22 ± 0.16	0.19 ± 0.14	21.21 ± 17.4	0.65 ± 0.15	21.24 ± 17.37	0.52 ± 0.5	1.69 ± 1.33	0.61 ± 0.77
GFCE	0.76 ± 0.1	0.21 ± 0.14	0.23 ± 0.16	19.31 ± 16.8	0.63 ± 0.13	19.45 ± 16.7	0.61 ± 0.85	1.75 ± 1.94	0.73 ± 1.13
GFF	0.79 ± 0.12	0.19 ± 0.15	0.19 ± 0.12	21.57 ± 18.7	0.67 ± 0.14	21.58 ± 18.69	0.5 ± 0.5	1.66 ± 1.58	0.58 ± 0.66
HMSD_GF	0.79 ± 0.08	0.14 ± 0.09	0.23 ± 0.15	23.15 ± 20.51	0.67 ± 0.11	23.21 ± 20.47	0.47 ± 0.44	1.64 ± 1.49	0.53 ± 0.7
Hybrid_MSD	0.8 ± 0.08	0.16 ± 0.11	0.22 ± 0.14	19.19 ± 14.82	0.67 ± 0.11	19.23 ± 14.78	0.43 ± 0.4	1.46 ± 1.49	0.5 ± 0.69
MGFF	0.8 ± 0.08	0.19 ± 0.11	0.19 ± 0.13	17.06 ± 15.92	0.67 ± 0.1	17.2 ± 15.82	0.39 ± 0.34	1.24 ± 1	0.45 ± 0.62
MSVD	0.81 ± 0.09	0.16 ± 0.11	0.2 ± 0.15	15.49 ± 19.36	0.68 ± 0.11	15.55 ± 19.32	0.4 ± 0.47	1.19 ± 1.45	0.47 ± 0.79
Wavelet	0.8 ± 0.09	0.15 ± 0.1	0.22 ± 0.15	21 ± 20.08	0.67 ± 0.11	21.11 ± 19.99	0.39 ± 0.35	1.31 ± 1.11	0.46 ± 0.66
FPDE	0.77 ± 0.12	0.2 ± 0.15	0.22 ± 0.14	26.7 ± 24.43	0.65 ± 0.14	26.81 ± 24.34	0.49 ± 0.41	1.57 ± 1.18	0.6 ± 0.76
MDLatLRR	0.81 ± 0.07	0.15 ± 0.09	0.2 ± 0.13	17.6 ± 11.82	0.69 ± 0.1	17.63 ± 11.79	0.39 ± 0.34	1.32 ± 1.19	0.45 ± 0.57
TIF	0.80 ± 0.08	0.16 ± 0.11	0.21 ± 0.14	19.63 ± 18.41	0.68 ± 0.11	19.76 ± 18.3	0.4 ± 0.36	1.29 ± 1.06	0.43 ± 0.46
MST_SR	0.78 ± 0.1	0.18 ± 0.13	0.21 ± 0.15	15.67 ± 11.75	0.65 ± 0.12	15.76 ± 11.69	0.47 ± 0.5	1.42 ± 1.52	0.53 ± 0.67
NSCT_SR	0.8 ± 0.09	0.18 ± 0.12	0.2 ± 0.13	18.54 ± 13.32	0.67 ± 0.11	18.61 ± 13.25	0.47 ± 0.43	1.59 ± 1.41	0.54 ± 0.69
RP_SR	0.78 ± 0.09	0.18 ± 0.11	0.22 ± 0.15	17.52 ± 12.41	0.65 ± 0.12	17.67 ± 12.26	0.5 ± 0.46	1.65 ± 1.47	0.59 ± 0.74
VSM_WLS	0.8 ± 0.09	0.17 ± 0.11	0.21 ± 0.14	16.99 ± 12.49	0.68 ± 0.11	17.18 ± 12.31	0.41 ± 0.42	1.34 ± 1.3	0.47 ± 0.72
GTF	0.79 ± 0.09	0.18 ± 0.12	0.22 ± 0.14	18.31 ± 17.43	0.66 ± 0.12	18.41 ± 17.36	0.47 ± 0.44	1.46 ± 1.36	0.54 ± 0.68
IFEVIP	0.79 ± 0.08	0.19 ± 0.1	0.21 ± 0.13	20.55 ± 20.67	0.66 ± 0.1	20.64 ± 20.59	0.43 ± 0.45	1.43 ± 1.38	0.51 ± 0.81
Fusion_1npLev	0.80 ± 0.09	0.16 ± 0.11	0.2 ± 0.16	16.26 ± 17.83	0.68 ± 0.12	16.31 ± 17.8	0.43 ± 0.43	1.45 ± 1.6	0.52 ± 0.69
Fusion_DesLev	0.80 ± 0.09	0.16 ± 0.12	0.21 ± 0.15	19.78 ± 16.21	0.67 ± 0.11	19.83 ± 16.17	0.41 ± 0.39	1.34 ± 1.18	0.48 ± 0.68
Fusion_NetLev	0.80 ± 0.1	0.16 ± 0.12	0.21 ± 0.16	13.44 ± 10.95	0.68 ± 0.13	13.64 ± 10.82	0.37 ± 0.34	1.08 ± 1.03	0.44 ± 0.6
Staple_1mgFus	0.81 ± 0.1	0.07 ± 0.06	0.27 ± 0.16	11.56 ± 9.03	0.69 ± 0.14	11.91 ± 8.85	0.37 ± 0.38	1.11 ± 1.16	0.43 ± 0.7
Staple_NetFus	0.82 ± 0.09	0.15 ± 0.11	0.18 ± 0.15	9.02 ± 8.19	0.7 ± 0.12	9.07 ± 8.16	0.33 ± 0.32	0.95 ± 1.01	0.38 ± 0.6
Staple_All	0.80 ± 0.11	0.06 ± 0.06	0.28 ± 0.17	11.66 ± 9.08	0.68 ± 0.14	12.04 ± 8.88	0.37 ± 0.38	1.12 ± 1.16	0.44 ± 0.7
Staple_Fast	0.80 ± 0.11	0.06 ± 0.06	0.27 ± 0.17	11.2 ± 8.92	0.68 ± 0.14	11.54 ± 8.74	0.37 ± 0.39	1.1 ± 1.17	0.44 ± 0.71
Majority_1mgFus	0.84 ± 0.07	0.15 ± 0.1	0.15 ± 0.12	6.52 ± 5.66	0.73 ± 0.1	6.8 ± 5.63	0.26 ± 0.28	0.74 ± 0.81	0.31 ± 0.6
Majority_NetFus	0.82 ± 0.09	0.15 ± 0.11	0.18 ± 0.15	9.02 ± 8.19	0.7 ± 0.12	9.07 ± 8.16	0.33 ± 0.32	0.95 ± 1.01	0.38 ± 0.6
Majority_All	0.84 ± 0.07	0.14 ± 0.1	0.16 ± 0.13	6.36 ± 5.61	0.73 ± 0.1	6.59 ± 5.59	0.26 ± 0.28	0.75 ± 0.82	0.31 ± 0.6
Majority_Fast	0.84 ± 0.07	0.14 ± 0.1	0.16 ± 0.13	6.45 ± 5.55	0.73 ± 0.1	6.67 ± 5.53	0.26 ± 0.28	0.74 ± 0.81	0.31 ± 0.61

Abbreviations: ADF, anisotropic diffusion and Karhunen–Loeve transform fusion; CT, computed tomography; CBF, cross bilateral filter; FPDE, fourth order partial differential equation; GFCE, guided-filter-based context enhancement; GFF, guided filtering fusion; GTF, gradient transfer and total variation minimization fusion; Hybrid-MSD, hybrid multi-scale decomposition; IFEVIP, infrared feature extraction and visual information preservation; MDLatLRR, multi-level image decomposition based on latent low-rank representation; MGFF, multi-scale guided filtered-based fusion; MSVD, multi-resolution singular value decomposition; TIF, two-scale image fusion.

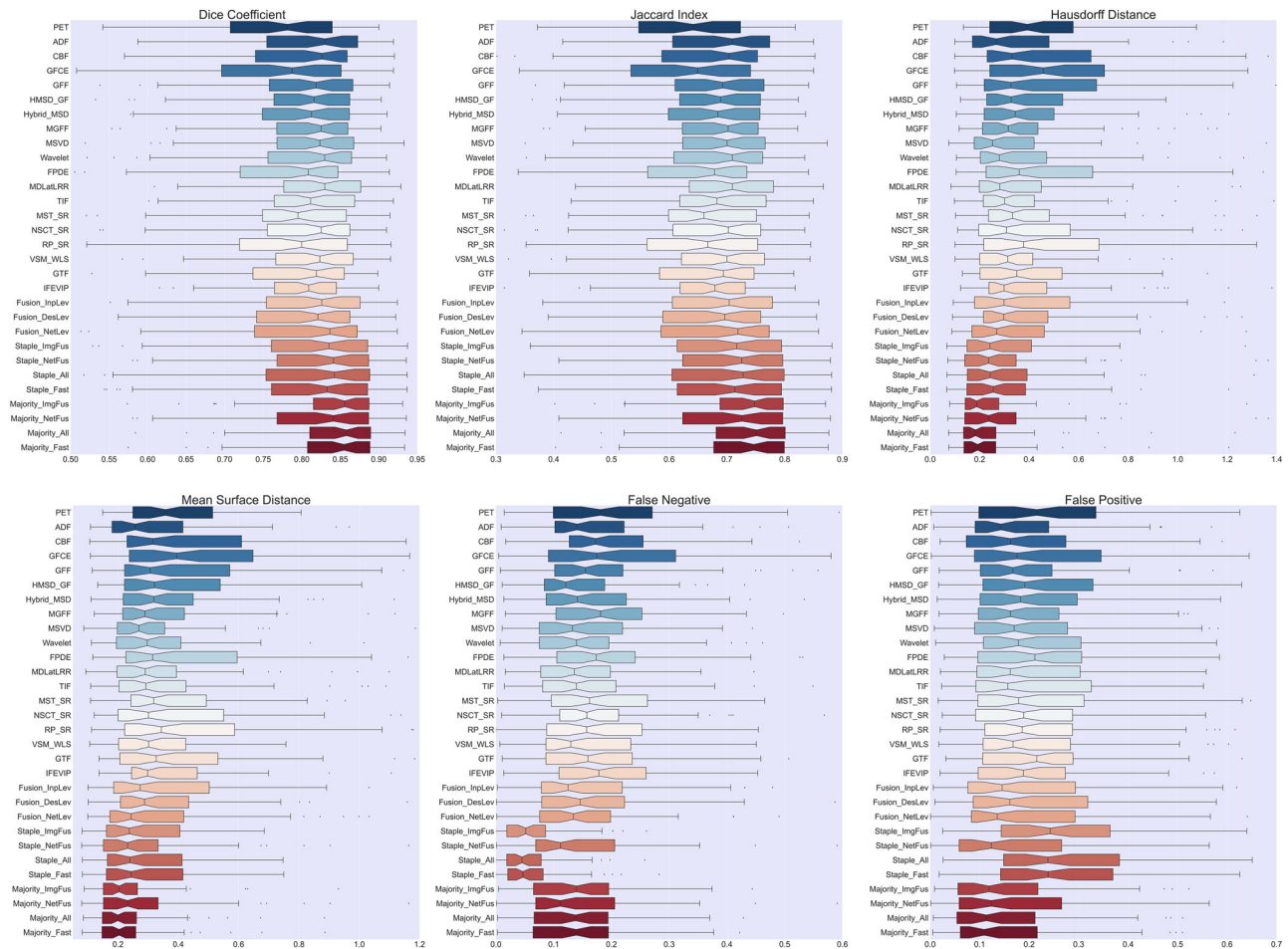


FIGURE 2 Box plots of different segmentation metrics for different segmentation models, including DSC, JSC, false-negative rate, false-positive rate, volume similarity, and mean and SD of surface distance. DSC, Dice similarity coefficient; JSC, Jaccard similarity coefficient; SD, standard deviation.

models) outperformed all other models, while Majority voting models achieved superior results with a Dice score of 0.84 ± 0.07 for Majority_ImgFus, Majority_All, and Majority_Fast.

Figure 3 shows the segmentations delineated by the best models for 2D slices of four patients. Figure 4 illustrates the 3D segmentation of HNCs by the best models for different cases with different sizes of tumors. Moreover, the same result for all models is presented in Figures S1–S3.

Table 3 lists the mean \pm SD of the ME of quantitative PET metrics and morphological radiomic features extracted from manual segmentation for all models. Their confidence interval is listed in Table S2. PET metrics, including SUV_{peak} , SUV_{mean} , SUV_{median} , and SUV_{max} , were perfectly matched between manual segmentation and segmentations from PET and all fusion models. In contrast, the CT model did not show conformity in any metric.

Figure 5 shows the results of model performance comparison using paired distributions Wilcoxon Signed

Rank Test and t -values on the corresponding Dice scores. CT had significantly lower results compared to all other models. Compared to PET, among image fusion methods, ADF, MGFF, MSVD, and Wavelet from the Multi-scale group, MDLatLRR, and TIF from the Saliency-based, and VSM-WLS from hybrid methods outperformed PET. Among network fusion models, only Fusion-InpLev model outperformed PET. All models from the output level fusion strategy (Staple and Majority Voting models) showed superiority over PET alone. Finally, the best model was Majority-All which significantly outperformed all models, except Staple-NetFus, Majority-Fast, Majority-NetFus, and Majority-ImFus. In addition to the Dice score, other metrics were also compared and the results presented in Figures S4 and S5.

Table 4 lists the timing results of all models. The first column shows the time spent by each model on average slice (averaged on 150 different slices from different regions of the HN and different tumors). The second column shows the time spent for an average

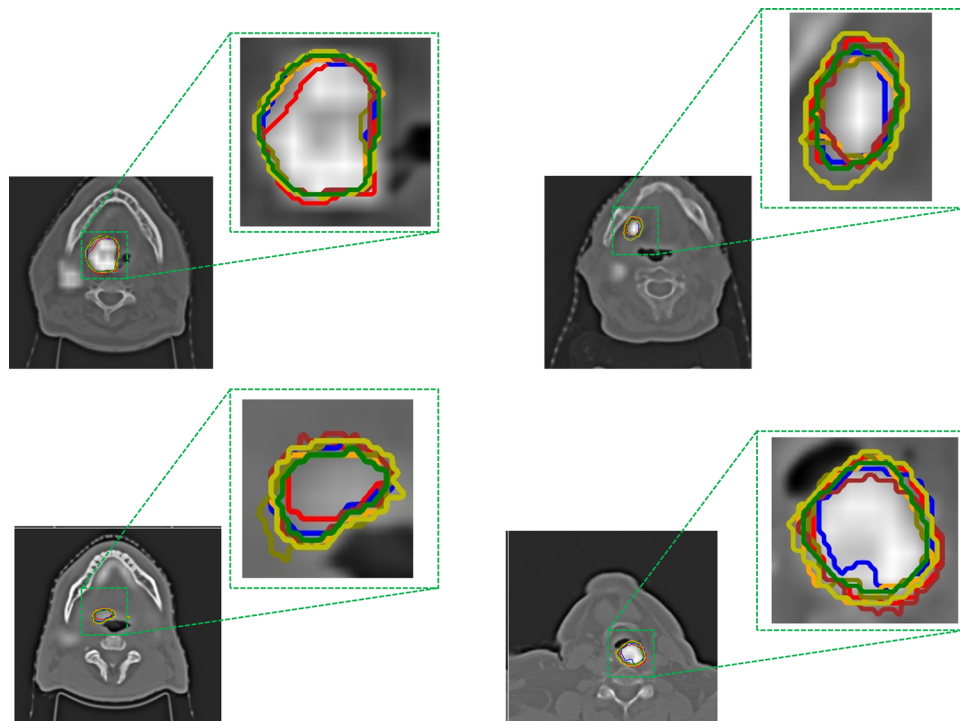


FIGURE 3 Visualization of segmentation output of representative clinical studies on transaxial slices with their corresponding zoomed version. Manual: Red; PET: Green, MDLatLRRR: Blue, Fusion_Ils: orange, Fusion_DLs: Olive, FusionNLs: Brown, Staple All: Yellow, Majority All: Cyan.

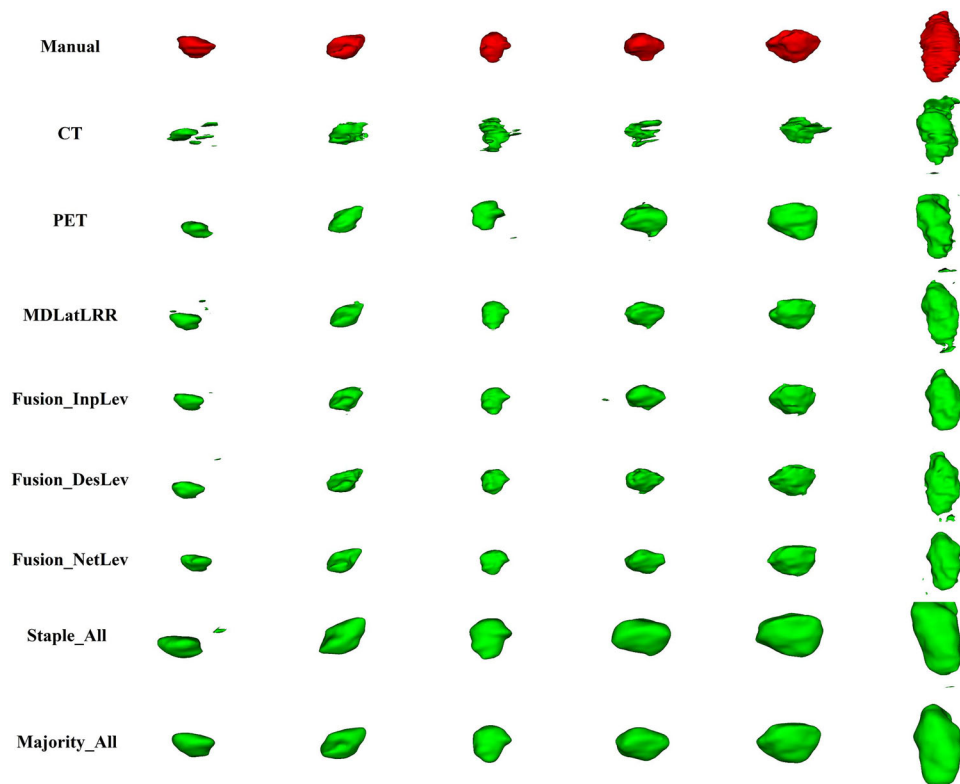


FIGURE 4 Representative 3D view of the different tumors with different sizes from different centers for manual, CT, PET, MDLatLRRR, Fusion_Ils, Fusion_DLs, FusionNLs, Staple, and Majority All. CT, computed tomography.

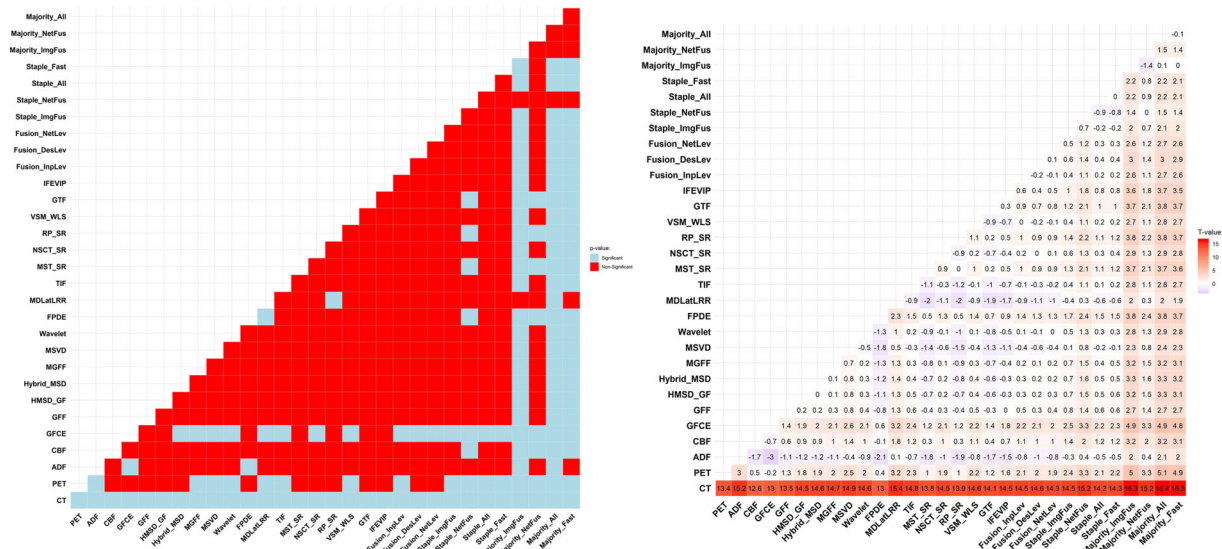


FIGURE 5 Heat map of p -values (left) and t -values (right) for comparing different segmentation methods pair-wise. The row is the reference for t -values.

patient, assuming the patient has 50 slices. Regarding image fusion methods, CBF, Hybrid-MSD, MDLAtLRR, and NSCT-SR, achieving 1.081, 0.225, 6.528, and 0.686 s, respectively, were the slowest models.

All other image fusion models were used for fast Staple and fast Majority Voting models. In the majority voting models, majority-All achieved the lowest computational time (9.295 s) for each slice. While having statistically comparable performance with the Majority-All model, the Majority-Fast model took only 0.772 s on average per slice and 38.5 s on average per patient.

4 | DISCUSSION

Gross tumor volume (GTV) estimation using reliable and robust segmentations of HNC is essential to correctly and successfully plan patient treatment and predict outcomes. However, manual GTV segmentation on PET/CT images is time-consuming, often requiring to switch between PET and CT images. In addition, it is subject to intra- and inter-observer variability. Therefore, accurate and fast automated PET/CT image segmentation is essential. Moreover, PET/CT may offer complementary information regarding improved analysis of submucosal deep tumor invasion, monitoring of treatment, and more precise interpretation of disease prognosis.

HNC tumor segmentation on CT images is challenging owing to low soft-tissue contrast since the distinction between the tumor and peri-tumoral inflammation is hardly possible despite CT providing anatomical information and good contrast between soft tissues, bone, and air. In our study, CT-only images achieved a Dice score of only 0.38 ± 0.22 in the test set. PET-only images, however, achieved a Dice score of

0.77 ± 0.09 in the test set. Nevertheless, by combining PET and CT information integrated at different levels, our study clearly showed that high Dice scores of 0.81, 0.80, and 0.84 for image level, DL network level, and output-level voting-based information fusion could be achieved.

Until now, most published studies focused on extracting feature representations from a single modality. Fusion techniques have been suggested to utilize the complimentary data from multi-modality images. The accuracy of tumor detection, correct staging, and segmentation can be increased rationally by using complementary information from multiple imaging modalities.^{2,3,32,63} Numerous PET and CT fusion techniques have been investigated, including in the image domain (conventional algorithms), DL-based (input, layer, and decision level), and output level (voting from the output of multiple models). In the current study, we employed different PET and CT information fusions for HNC segmentation. We showed that information fusion of PET and CT using conventional algorithms in the image domain and DL-based fusion could achieve competitive performance. Furthermore, image segmentation information fusion using majority voting from different algorithms outperformed other algorithms.

Andrearczyk et al.⁶⁴ developed 2D and 3D fully convolutional V-Net models for tumor and nodal metastasis segmentation in the HN region of 202 patients. While the achieved Dice scores on single CT and PET modalities were 0.48 and 0.58, respectively, the PET/CT multi-modal model (integrated on a late fusion approach) significantly improved the Dice score (0.60). In addition, their 2D model (Dice = 0.60) slightly outperformed the 3D model (Dice = 0.59). Moe et al.⁶⁵ developed U-Net CNN models for completely automated segmentation

TABLE 3 Summary of mean errors of calculated quantitative PET metrics (Mean ± SD) for the different image fusion and network fusion models.

Fusion	MTV	Sphericity	Asphericity	Elongation	flatness	SUV _{peak}	SUV _{mean}	SUV _{median}	SUV _{max}	TLG
CT	-5.93 ± 6.61	-0.22 ± 0.11	0.77 ± 0.51	-0.14 ± 0.19	-0.1 ± 0.18	-0.36 ± 1.44	-1.13 ± 2.31	-0.95 ± 2.93	-0.99 ± 2.16	-63.87 ± 71.12
PET	0.044 ± 3.54	-0.02 ± 0.1	0.04 ± 0.26	-0.04 ± 0.11	0 ± 0.10	0.04 ± 0.34	-0.11 ± 1.26	-0.08 ± 1.59	0.02 ± 0.16	0.21 ± 20.59
ADF	-0.39 ± 2.88	0.01 ± 0.09	-0.02 ± 0.22	-0.04 ± 0.11	-0.01 ± 0.11	0 ± 0	-0.02 ± 0.98	0.01 ± 1.21	0 ± 0	-0.71 ± 16.10
CBF	-0.51 ± 2.54	-0.05 ± 0.1	0.12 ± 0.27	-0.04 ± 0.14	-0.03 ± 0.12	0 ± 0	-0.05 ± 1.04	0 ± 1.31	-0.02 ± 0.15	-2.20 ± 14.87
GFCE	-0.56 ± 3.71	-0.07 ± 0.1	0.19 ± 0.29	-0.03 ± 0.12	0 ± 0.12	0 ± 0	-0.07 ± 1.15	0.06 ± 1.4	0 ± 0.03	-5.89 ± 23.34
GFF	-0.33 ± 2.68	-0.05 ± 0.09	0.12 ± 0.24	-0.03 ± 0.11	-0.01 ± 0.11	0 ± 0.02	-0.11 ± 0.9	-0.09 ± 1.12	-0.02 ± 0.09	-1.28 ± 15.34
HMSD_GF	0.63 ± 2.85	-0.03 ± 0.09	0.06 ± 0.24	-0.03 ± 0.12	0 ± 0.10	0 ± 0	-0.4 ± 1.06	-0.43 ± 1.32	0 ± 0	3.86 ± 16.11
Hybrid_MSD	0.07 ± 2.92	-0.01 ± 0.09	0.02 ± 0.22	-0.02 ± 0.11	-0.01 ± 0.10	0 ± 0	-0.15 ± 0.98	-0.14 ± 1.21	0 ± 0	1.26 ± 16.31
MGFF	-0.40 ± 2.77	0 ± 0.09	-0.01 ± 0.22	-0.03 ± 0.08	0 ± 0.09	0.01 ± 0.05	0.01 ± 1.01	0.06 ± 1.21	0 ± 0	-1.42 ± 15.97
MSVD	-0.01 ± 2.9	0 ± 0.08	-0.01 ± 0.21	-0.02 ± 0.11	-0.01 ± 0.10	-0.09 ± 0.77	-0.1 ± 1.03	-0.1 ± 1.26	0.03 ± 0.21	1.43 ± 17.04
Wavelet	0.32 ± 2.92	-0.01 ± 0.08	0.01 ± 0.21	-0.02 ± 0.09	0.01 ± 0.09	0 ± 0	-0.27 ± 1.08	-0.3 ± 1.36	0 ± 0	2.91 ± 17.48
FPDE	0.18 ± 2.74	-0.04 ± 0.09	0.1 ± 0.24	-0.03 ± 0.12	-0.01 ± 0.10	0.01 ± 0.06	-0.28 ± 1.07	-0.29 ± 1.3	-0.01 ± 0.04	1.64 ± 15.06
MDLatLRR	0.16 ± 2.72	0 ± 0.10	0.01 ± 0.23	-0.05 ± 0.09	0 ± 0.09	0 ± 0	-0.14 ± 0.87	-0.13 ± 1.08	0 ± 0	2.40 ± 15.16
TIF	0.18 ± 2.59	0 ± 0.10	0.02 ± 0.26	-0.02 ± 0.10	0.02 ± 0.10	0 ± 0	-0.2 ± 0.95	-0.21 ± 1.18	0 ± 0	2.17 ± 15.43
MST_SR	-0.02 ± 3.13	-0.03 ± 0.10	0.09 ± 0.29	-0.03 ± 0.10	0 ± 0.09	0 ± 0	-0.22 ± 1.1	-0.23 ± 1.39	0 ± 0	1.17 ± 18.70
NSCT_SR	-0.06 ± 2.55	-0.03 ± 0.10	0.10 ± 0.26	-0.03 ± 0.10	0 ± 0.11	0.03 ± 0.24	-0.18 ± 0.97	-0.18 ± 1.21	0.06 ± 0.46	1.52 ± 15.47
RP_SR	0.27 ± 2.82	-0.05 ± 0.10	0.13 ± 0.26	-0.05 ± 0.10	-0.02 ± 0.10	0.04 ± 0.31	-0.27 ± 1.05	-0.28 ± 1.34	0 ± 0.07	2.06 ± 16.61
VSM_WLS	0.06 ± 2.65	-0.02 ± 0.09	0.05 ± 0.23	-0.04 ± 0.11	-0.01 ± 0.10	0 ± 0.02	-0.15 ± 1.03	-0.14 ± 1.28	-0.01 ± 0.07	1.59 ± 14.55
GTF	0.24 ± 2.89	-0.04 ± 0.1	0.12 ± 0.27	-0.02 ± 0.11	0 ± 0.10	0.01 ± 0.05	-0.24 ± 1.06	-0.25 ± 1.3	0 ± 0.04	2.64 ± 16.09
IFEVIP	-0.05 ± 2.88	-0.01 ± 0.10	0.04 ± 0.26	-0.03 ± 0.08	-0.01 ± 0.10	0 ± 0	-0.18 ± 1.04	-0.17 ± 1.3	0 ± 0	0.89 ± 17.81
Fusion_InpLev	-0.36 ± 3.11	0.01 ± 0.09	-0.02 ± 0.22	-0.03 ± 0.12	-0.01 ± 0.11	-0.08 ± 0.68	0.07 ± 1.14	0.11 ± 1.4	0.04 ± 0.35	-1.86 ± 17.72
Fusion_DesLev	-0.17 ± 2.83	-0.04 ± 0.09	0.09 ± 0.24	-0.01 ± 0.09	0 ± 0.08	0 ± 0	0.06 ± 1.12	0.12 ± 1.39	0 ± 0	-1.10 ± 16.74
Fusion_NetLev	-0.19 ± 3.01	0.03 ± 0.09	-0.06 ± 0.2	-0.01 ± 0.09	-0.01 ± 0.09	0 ± 0	0.07 ± 1.3	0.09 ± 1.57	0 ± 0	-0.74 ± 17.28
Staple_ImgFus	2.0438 ± 2.73	0.07 ± 0.09	-0.14 ± 0.21	-0.02 ± 0.10	0.01 ± 0.09	0 ± 0	-0.69 ± 1.02	-0.89 ± 1.29	0 ± 0	13.32 ± 16.95
Staple_NetFus	-0.41 ± 2.89	0.04 ± 0.08	-0.08 ± 0.19	-0.01 ± 0.08	0 ± 0.08	0 ± 0	0.21 ± 1.25	0.22 ± 1.51	0 ± 0	-1.36 ± 16.55
Staple_All	2.19 ± 2.69	0.07 ± 0.09	-0.15 ± 0.21	-0.02 ± 0.10	0.01 ± 0.09	0 ± 0	-0.71 ± 1.0	-0.93 ± 1.26	0 ± 0	13.96 ± 16.66
Staple_Fast	2.01 ± 2.68	0.07 ± 0.09	-0.14 ± 0.22	-0.01 ± 0.10	0.01 ± 0.09	0 ± 0	-0.69 ± 1.0	-0.89 ± 1.26	0 ± 0	13.33 ± 16.44
Majority_ImgFus	-0.36 ± 2.61	0.10 ± 0.080	-0.20 ± 0.18	-0.01 ± 0.07	0 ± 0.07	0 ± 0	0.13 ± 0.93	0.1 ± 1.15	0 ± 0	1.08 ± 15.08
Majority_NetFus	-0.41 ± 2.89	0.04 ± 0.08	-0.08 ± 0.19	-0.01 ± 0.08	0 ± 0.08	0 ± 0	0.21 ± 1.25	0.22 ± 1.51	0 ± 0	-1.36 ± 16.56
Majority_All	-0.35 ± 2.59	0.04 ± 0.08	-0.09 ± 0.18	-0.02 ± 0.07	0 ± 0.07	0 ± 0	0.14 ± 0.94	0.12 ± 1.16	0 ± 0	3.08 ± 15.01
Majority_Fast	-0.38 ± 2.64	0.09 ± 0.08	-0.19 ± 0.18	-0.01 ± 0.07	0 ± 0.07	0 ± 0	0.14 ± 0.95	0.13 ± 1.17	0 ± 0	0.72 ± 15.02

Abbreviations: ADF, anisotropic diffusion and Karhunen–Loeve transform fusion; CBF, cross bilateral filter; CT, computed tomography; FPDE, fourth order partial differential equations; GFCE, guided-filter-based context enhancement; GFF, guided filtering fusion; GTF, gradient transfer and total variation minimization fusion; Hybrid-MSD, hybrid multi-scale decomposition; IFEVIP, infrared feature extraction and visual information preservation; MDLatLRR, multi-level image decomposition based on latent low-rank representation; MGFF, multi-scale guided filtered-based fusion; MSVD, multi-resolution singular value decomposition; MTV, Metabolic Tumor Volume; PET, positron emission tomography; SUV, standardized uptake values; TIF, two-scale image fusion; TLG, Total Lesion Glycolysis.

TABLE 4 Inference time of the different algorithms for information fusion for the different strategies.

Strategy	Model	Average slice	Average patient
Image fusion	ADF	0.055	2.739
	CBF	1.081	54.030
	GFCE	0.091	4.564
	GFF	0.048	2.400
	HMSD_GF	0.088	4.385
	Hybrid_MSD	0.225	11.243
	MGFF	0.081	4.070
	MSVD	0.038	1.897
	Wavelet	0.012	0.578
	FPDE	0.081	4.040
	MDLatLRR	6.528	326.413
	TIF	0.004	0.203
	MST_SR	0.010	0.500
	NSCT_SR	0.686	34.323
	RP_SR	0.010	0.513
	VSM_WLS	0.114	5.688
	GTF	0.080	4.005
IFEVIP	0.036	1.776	
Network fusion	Fusion_InpLev	0.003	0.155
	Fusion_DesLev	0.005	0.227
	Fusion_NetLev	0.005	0.227
STAPLE	Staple_ImgFus	9.303	465.160
	Staple_NetFus	0.019	0.930
	Staple_All	9.368	468.416
	Staple_Fast	0.796	39.787
Majority voting	Majority_ImgFus	9.280	463.985
	Majority_NetFus	0.015	0.726
	Majority_All	9.295	464.736
	Majority_Fast	0.772	38.593

Abbreviations: ADF, anisotropic diffusion and Karhunen–Loeve transform fusion; CBF, cross bilateral filter; FPDE, fourth order partial differential equations; GFCE, guided-filter-based context enhancement; GFF, guided filtering fusion; GTF, gradient transfer and total variation minimization fusion; Hybrid-MSD, hybrid multi-scale decomposition; IFEVIP, infrared feature extraction and visual information preservation; MDLatLRR, multi-level image decomposition based on latent low-rank representation; MGFF, multi-scale guided filtered-based fusion; MSVD, multi-resolution singular value decomposition; STAPLE, Simultaneous truth and performance level estimation; TIF, two-scale image fusion.

of HNC GTV using CT, PET, and PET/CT images. The model trained, validated, and tested on 142, 15, and 40 patients, respectively, achieved a Dice score of 0.55, 0.69, and 0.71 for CT, PET, and PET/CT models, respectively. In addition, they reported the importance of PET images in the DL model development for HNC GTV segmentation. More importantly, the winner of the 2021 MICCAI HNC tumor segmentation challenge (HECKTOR 2022) developed a SegResNet semantic segmentation model for 3D segmentation of tumors

and lymph nodes using PET/CT images and achieved a Dice score of 0.788.⁶⁶ Their study included 883 cases, with 524 cases used for training and the rest for model testing.

Guo et al.²⁹ proposed three different neural network architectures for integrating multi-modality medical images in the network's feature-learning, classifier, and decision-making levels. The authors used these structures to develop automated image segmentation models for segmenting soft-tissue sarcomas using MR, CT, and PET images.²⁹ Superior performances for multi-modal models compared to single modalities were reported.²⁹ The authors concluded that fusing within the network (convolutional or fully connected layers) generally resulted in superior performance compared to models fused at the output (decision level). We examined various weight flow configurations during the fusion of networks in the Fusion_NetLev setting (low-level to high-level feature fusion). However, our experiments did not reveal significant differences in PET image segmentation performance across these settings. As indicated in the results, network fusion attained consistent performance metrics for Fusion_InpLev (0.80 ± 0.09), Fusion_DesLev (0.80 ± 0.09), and Fusion_NetLev (0.80 ± 0.1). This suggests that a different way of fusion in Fusion_NetLev (low or high level) could converge to the performance observed in Fusion_InpLev and Fusion_DesLev. We did not observe significant differences between the fusion results at various levels. Therefore, we chose to report only one setting to maintain coherence and simplify the presentation of our findings.

Using multi-modality images is challenging because CT and PET resolution and overall imaging characteristics of HNC on these modalities are different. To achieve similar resolution on both imaging modalities prior to fusion in the pre-processing step, one imaging modality must be down- or up-sampled. Image distortion usually occurs during the up or down sampling procedure, thus limiting the structural features that may be seen on PET compared to CT.⁶⁷ Our study inherently bears some limitations, including the size of the dataset, not involving lymph node region segmentation, and the lack of an external validation set. Future studies should be performed on larger datasets and tested on external test sets.⁶⁸ However, we used a standard split of data sets in our study. In addition, our study focused on the tumor GTV; yet, lymph node segmentation is equally required for radiation therapy. Future studies should include them for segmentation purposes.

5 | CONCLUSION

In the current study, we considered different fusion techniques applied to PET/CT images, employing

conventional and DL methods, as well as output-level voting fusions, towards improved HNC segmentation. We conclude that image fusion adds significant value to segmentation tasks, considerably outperforming PET-only and CT-only methods. Meanwhile, output-level voting-based fusion using majority voting on outputs of different algorithms considerably improved HNC segmentation performance, while both conventional and DL algorithms (without voting) produced competitive yet inferior results. Nevertheless, they significantly outperformed single-modality-based segmentations.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under grants SNSF 320030_176052 and 320030_173091/1, and the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2019-06467.

Open access funding provided by Universite de Geneve.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

Isaac Shiri  <https://orcid.org/0000-0002-5735-0736>
 Mehdi Amini  <https://orcid.org/0000-0002-4370-680X>
 Alireza Vafaei Sadr  <https://orcid.org/0000-0002-5733-6678>
 Ghasem Hajianfar  <https://orcid.org/0000-0001-5359-2407>
 Yazdan Salimi  <https://orcid.org/0000-0002-1233-9576>
 Zahra Mansouri  <https://orcid.org/0000-0003-2087-9721>
 Ismini Mainta  <https://orcid.org/0000-0001-5296-0116>
 Minerva Becker  <https://orcid.org/0000-0001-9644-9053>
 Arman Rahmim  <https://orcid.org/0000-0002-9980-2403>
 Habib Zaidi  <https://orcid.org/0000-0001-7559-5297>

REFERENCES

1. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50(suppl 1):122s-150s.
2. Katirtzidou E, Rager O, Varoquaux AD, et al. Detection of distant metastases and distant second primary cancers in head and neck squamous cell carcinoma: comparison of [¹⁸F]FDG PET/MRI and [¹⁸F]FDG PET/CT. *Insights Imaging*. 2022;13(1):121.
3. Becker M, Varoquaux AD, Combescure C, et al. Local recurrence of squamous cell carcinoma of the head and neck after radio(chemo)therapy: diagnostic performance of FDG-PET/MRI with diffusion-weighted sequences. *Eur Radiol*. 2018;28(2):651-663.
4. Bagci U, Udupa JK, Mendhiratta N, et al. Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. *Med Image Anal*. 2013;17(8):929-945.
5. Yousefirizi F, Jha AK, Brosch-Lenz J, Saboury B, Rahmim A. Toward high-throughput artificial intelligence-based segmentation in oncological PET imaging. *PET clinics*. 2021;16(4):577-596.
6. Hatt M, Lee J, Schmidlein CR, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: report of AAPM task group no. 211. *Med Phys*. 2017;44(6):e1-e42.
7. Yousefirizi F, Pierre D, Amyar A, Ruan S, Saboury B, Rahmim A. AI-Based detection, classification and prediction/prognosis in medical imaging:: towards radiophenomics. *PET Clinics*. 2022;17(1):183-212.
8. Amini M, Nazari M, Shiri I, et al. Multi-level multi-modality (PET and CT) fusion radiomics: prognostic modeling for non-small cell lung carcinoma. *Phys Med Biol*. 2021;66(20).
9. Khodabakhshi Z, Amini M, Mostafaei S, et al. Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information. *J Digit Imaging*. 2021;34(5):1086-1098.
10. Khodabakhshi Z, Mostafaei S, Arabi H, Oveisi M, Shiri I, Zaidi H. Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. *Comput Biol Med*. 2021;136:104752.
11. Amini M, Hajianfar G, Hadadi Avval A, et al. Overall survival prognostic modelling of non-small cell lung cancer patients using positron emission tomography/computed tomography harmonised radiomics features: the quest for the optimal machine learning algorithm. *Clin Oncol (R Coll Radiol)*. 2022;34(2):114-127.
12. Zaidi H, El Naqa I. Quantitative molecular positron emission tomography imaging using advanced deep learning techniques. *Annu Rev Biomed Eng*. 2021;23:249-276.
13. Zaidi H, Abdoli M, Fuentes C, El Naqa I. Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging*. 2012;39(5):881-891.
14. Andrearczyk V, Oreiller V, Abobakr M, et al. Overview of the HECKTOR challenge at MICCAI 2022: automatic head and neck tumor segmentation and outcome prediction in PET/CT. 2022;1-30. doi:10.1007/978-3-031-27420-6_1
15. Yousefirizi F, Shiri I, Bloise I, et al. Semi-supervised learning towards automated segmentation of PET images with limited annotations: application to lymphoma patients. *arXiv preprint arXiv:221209908*. 2022.
16. Shiri I, Arabi H, Sanaat A, Jenabi E, Becker M, Zaidi H. Fully automated gross tumor volume delineation from PET in head and neck cancer using deep learning algorithms. *Clin Nucl Med*. 2021;46(11):872-883.
17. Oreiller V, Andrearczyk V, Jreige M, et al. Head and neck tumor segmentation in PET/CT: the HECKTOR challenge. *Med Image Anal*. 2022;77:102336.
18. Becker M, Monnier Y, de Vito C. MR imaging of laryngeal and hypopharyngeal cancer. *Magn Reson Imaging Clin N Am*. 2022;30(1):53-72.
19. Zhu Z, Chai Y, Yin H, Li Y, Liu Z. A novel dictionary learning approach for multi-modality medical image fusion. *Neurocomputing*. 2016;214:471-482.
20. Markel D, Zaidi H, El Naqa I. Novel multimodality segmentation using level sets and Jensen-Rényi divergence. *Med Phys*. 2013;40(12):121908-121915.
21. Hermessi H, Mourali O, Zagrouba E. Multi-modal medical image fusion review: theoretical background and recent advances. *Signal Process*. 2021;183:108036.

22. James AP, Dasarathy BV. Medical image fusion: a survey of the state of the art. *Inf Fusion*. 2014;19:4-19.
23. Liu Y, Chen X, Cheng J, Peng H, Wang Z, Infrared and visible image fusion with convolutional neural networks. *Int J Wavelets Multiresolution Inf Process*. 2018;16(03):1850018.
24. Zhao J, Ji G, Qiang Y, Han X, Pei B, Shi Z. A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm. *PLoS one*. 2015;10(4):e0123694.
25. Ju W, Xiang D, Zhang B, Wang L, Kopriva I, Chen X. Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. *IEEE Trans Image Process*. 2015;24(12):5854-5867.
26. Yang S, Weidong C, Heng H, et al. Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies. *IEEE Trans Med Imaging*. 2014;33(2):408-421.
27. Song Q, Bai J, Han D, et al. Optimal co-segmentation of tumor in PET-CT images with context information. *IEEE Trans Med Imaging*. 2013;32(9):1685-1697.
28. Han D, Bayouth J, Song Q, et al. Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. *Inf Process Med Imaging Proc Conf*. 2011;22:245-256.
29. Guo Z, Li X, Huang H, Guo N, Li Q. Deep learning-based image segmentation on multi-modal medical imaging. *IEEE Trans Radiat Plasma Med Sci*. 2019;3(2):162-169.
30. Bradshaw T, Perk T, Chen S, et al. Deep learning for classification of benign and malignant bone lesions in [F-18] NaF PET/CT images. *J Nucl Med*. 2018;59(supplemental):327.
31. Zhong Z, Kim Y, Zhou L, et al. 3D fully convolutional networks for co-segmentation of tumors on pet-ct images. *Proc IEEE Int Symp Biomed Imaging*. 2018;228-231. doi:10.1109/ISBI.2018.8363561
32. Li L, Zhao X, Lu W, Tan S. Deep learning for variational multimodality tumor segmentation in PET/CT. *Neurocomputing*. 2020;392:277-295.
33. Kumar A, Fulham M, Feng D, Kim J. Co-learning feature fusion maps from PET-CT images of lung cancer. *IEEE Trans Med Imaging*. 2019. doi:10.1109/tmi.2019.2923601
34. Shiri I, Vafaei Sadr A, Amini M, et al. Decentralized Distributed multi-institutional PET image segmentation using a federated deep learning framework. *Clin Nucl Med*. 2022;47(7):606-617. gr
35. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by non-invasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
36. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Sci Data*. 2017;4:170077.
37. Grossberg AJ, Mohamed ASR, Elhalawani H, et al. Author correction: imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Sci Data*. 2018;5(1):1.
38. Grossberg A, Elhalawani H, Mohamed A, et al. MD Anderson Cancer Center head and neck quantitative imaging working group. *HNSCC The Cancer Imaging Archive*. 2020.
39. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045-1057.
40. Vallières M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep*. 2017;7(1):10117.
41. Vallières M, Kay-Rivest E, Perrin LJ, et al. Data from head-neck-PET-CT. The Cancer Imaging Archive. Published; 2017.
42. Mehranian A, Ay MR, Rahmim A, Zaidi H. 3D prior image constrained projection completion for X-ray CT metal artifact reduction. *IEEE Trans Nucl Sci*. 2013;60(5):3318-3332.
43. Bavisetti DP, Dhuli R. Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform. *IEEE Sens J*. 2015;16(1):203-209.
44. Shreyamsha Kumar B. Image fusion based on pixel significance using cross bilateral filter. *Signal Image and Video Process*. 2015;9(5):1193-1204.
45. Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process*. 2013;22(7):2864-2875.
46. Zhou Z, Dong M, Xie X, Gao Z. Fusion of infrared and visible images for night-vision context enhancement. *Appl Opt*. 2016;55(23):6480-6490.
47. Zhou Z, Wang B, Li S, Dong M. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters. *Inf Fusion*. 2016;30:15-26.
48. Bavisetti DP, Xiao G, Zhao J, Dhuli R, Liu G. Multi-scale guided image and video fusion: a fast and efficient approach. *Circuits Syst Signal Process*. 2019;38(12):5576-5605.
49. Naidu V. Image fusion technique using multi-resolution singular value decomposition. *Def Sci J*. 2011;61(5):479.
50. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*. 2015;60(14):5471.
51. Bavisetti DP, Xiao G, Liu G. Multi-sensor image fusion based on fourth order partial differential equations. In 2017 20th International conference on information fusion (Fusion). Vallières 2017. (pp. 1-9).
52. Li H, Wu X-J, Kittler J. MDLatLRR: a novel decomposition method for infrared and visible image fusion. *IEEE Trans Image Process*. 2020;29:4733-4746.
53. Bavisetti DP, Dhuli R. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Phys Technol*. 2016;76:52-64.
54. Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf Fusion*. 2015;24:147-164.
55. Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys Technol*. 2017;82:8-17.
56. Ma J, Chen C, Li C, Huang J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf Fusion*. 2016;31:100-109.
57. Zhang Y, Zhang L, Bai X, Zhang L. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Phys Technol*. 2017;83:227-237.
58. Zhang X, Ye P, Xiao G. VIFB: a visible and infrared image fusion benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2020. (pp. 104-105).
59. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U2-Net: going deeper with nested U-structure for salient object detection. *Pattern Recognit*. 2020;106:107404.
60. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004;23(7):903-921. doi:10.1109/TMI.2004.828354
61. Ashrafinia S. *Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics*. The Johns Hopkins University; 2019.
62. Depeursinge A, Andrearczyk V, Whybra P, et al. Standardised convolutional filtering for radiomics. *arXiv preprint arXiv:200605470*. 2020.
63. Varoquaux A, Rager O, Dulguerov P, Burkhardt K, Ailianou A, Becker M. Diffusion-weighted and PET/MR imaging after radiation therapy of malignant head and neck tumors. *Radiographics*. 2015;35(5):1502-1527.
64. Andrearczyk V, Oreiller V, Vallières M, et al. Automatic segmentation of head and neck tumors and nodal metastases in PET-CT

- scans. In *Medical imaging with deep learning*. 2020. (pp. 33-43). PMLR.
65. Moe YM, Groendahl AR, Tomic O, Dale E, Malinen E, Futsaether CM. Deep learning-based auto-delineation of gross tumour volumes and involved nodes in PET/CT images of head and neck cancer patients. *Eur J Nucl Med Mol Imaging*. 2021;48(9):2782-2792.
 66. Myronenko A, Siddiquee MMR, Yang D, He Y, Xu D. Automated head and neck tumor segmentation from 3D PET/CT. *arXiv preprint arXiv:220910809*. 2022.
 67. Xiao N, Yang W, Qiang Y, et al. PET and CT image fusion of lung cancer with siamese pyramid fusion network. *Front Med*. 2022;9:792390.
 68. Shiri I, Vafaei Sadr A, Akhavan A, et al. Decentralized collaborative multi-institutional PET attenuation and scatter correction using federated deep learning. *Eur J Nucl Med Mol Imaging*. 2023;50(4):1034-1050.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shiri I, Amini M, Yousefirizi F, et al. Information fusion for fully automated segmentation of head and neck tumors from PET and CT images. *Med Phys*. 2024;51:319–333.
<https://doi.org/10.1002/mp.16615>