Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning Where to Fuse: A Lightweight Pyramid Weighting Network for Multimodal Medical Image Fusion

Qiang Xie

xieqiang@chzc.edu.cn

Chuzhou Polytechnic

Kai He

Chuzhou Polytechnic

Shun Liu

Chuzhou Polytechnic

Article

Keywords: Multimodal medical image fusion, Laplacian pyramid, Weight map prediction, Structural re-parameterization, Lightweight network, Fusion evaluation

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11580238/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learning Where to Fuse: A Lightweight Pyramid Weighting Network for Multimodal Medical Image Fusion

Xie Qiang[1,*], He Kai[1], Liu Shun[1]

[1]School of Mechanical and Automotive Engineering, Chuzhou Polytechnic, Chuzhou, China

*Correspondence: xieqiang@chzc.edu.cn

Abstract

Multimodal medical image fusion combines anatomical and functional images into one view. Most recent deep fusion methods synthesize every fused intensity with a large generator, although for registered images the main task is to decide where to trust which modality. We propose LPW-Net, a lightweight pyramid weighting network that decomposes both sources with a fixed Laplacian pyramid, predicts one weight map per level with a single coarse-to-fine network, and reconstructs the fused image from convex combinations of the source coefficients. Edge-aware training branches are merged into plain convolutions for inference, leaving 7,841 parameters. We show that such weight predictors tend to collapse to the MRI through saturated logits, remove this with a logit penalty, and use the resulting logits for an inference-time bias that shifts the fusion between the sources without retraining. Against twelve deep methods, ten of them from 2024 and 2025, and eight classical methods on CT, PET and SPECT fusion with MRI, LPW-Net ranked first on the three structure-based fusion metrics in every task of the public test split, ran in 1.98 ms per image pair on a GPU, and dominated nearly all compared methods in a per-source edge-preservation trade-off. The evaluation is complemented by an analysis that measures separately how much edge information each method keeps from each source, and the learned weight maps of LPW-Net localize bone and functional uptake as intended.

Keywords: Multimodal medical image fusion; Laplacian pyramid; Weight map prediction; Structural re-parameterization; Lightweight network; Fusion evaluation

1. Introduction

Anatomical and functional imaging answer different clinical questions. CT shows dense structures such as bone, MRI resolves soft tissue, and PET and SPECT map metabolism or perfusion at a much coarser resolution. A single image that carries the relevant content of two registered modalities helps a reader localize a functional finding on the anatomy and supports diagnosis and treatment planning, which has made multimodal medical image fusion a long-standing research topic[1]. Classical fusion methods decompose the sources with a multi-scale transform or a sparse dictionary, measure the activity of each coefficient, and combine the coefficients with a hand-designed rule[2,3]. Their output is a selection or blend of source content, but the activity measures and rules are designed by hand, and their parameters usually have to be tuned for each modality pair.

Deep learning moved most of this design into data. An early medical method used a convolutional network to produce a weight map that was then applied inside a pyramid, but the field soon adopted encoder–decoder networks that synthesize the fused image directly[4,5]. Recent methods continue this line with Transformers, state-space models, diffusion models and general frameworks that serve several fusion tasks with one model[6,7]. These networks can produce visually rich results, but every output intensity is synthesized. In medical images intensity often carries meaning, for instance the density of bone or the uptake of a tracer, and a generator may change it in ways that are hard to predict. The networks are also growing, which works against deployment on the modest hardware found next to a scanner or in a clinic. Lightweight designs exist[8], yet they too synthesize the fused intensities.

We start from a different view of the problem. Given two registered images, fusion mainly has to decide, at every position and every scale, how much to trust each modality; the content itself is already present in the sources, and such a decision needs far less capacity than synthesis. Following this view, we propose LPW-Net, a lightweight pyramid weighting network. LPW-Net decomposes both sources with a fixed Laplacian pyramid[9], predicts one soft weight map per level with a single small network shared by all levels, from coarse to fine, and reconstructs the fused image from the convex combination of the source coefficients, so every fused coefficient lies between the corresponding source coefficients. During training each layer of the predictor is an edge-aware multi-branch block, which is merged into one 3×3 convolution for inference[10]; the deployed network has 7,841 parameters.

Making this formulation work required one ingredient that is easy to overlook. Without it, training frequently collapses: the weights saturate, the gradients vanish, and the network outputs the MRI unchanged. We trace this to unbounded growth of the weight logits and remove it with a small logit penalty. The same property gives LPW-Net a control that generators lack: adding a scalar bias to all logits at inference moves the result continuously from one source to the other without retraining. The collapse also exposed a weakness of the usual evaluation. Fusion papers rank methods by averaging metrics such as $Q^{AB/F}$, Q_Y, VIF and MI over test images[11,12]. Several of these metrics reward an output that stays close to one source, and a high average alone is therefore not sufficient evidence that information from the second modality was transferred.

We therefore complement the standard comparison with an analysis that measures separately how much edge information each method keeps from each source, traces the whole trade-off of LPW-Net with the inference bias, and uses a linear blend of the two sources as a null reference. LPW-Net is compared with twelve deep methods, ten of them published in 2024 or 2025, and with eight classical methods on CT–MRI, PET–MRI and SPECT–MRI pairs from the Harvard Whole Brain Atlas, using official code and weights for every compared method. The contributions of this paper are as follows.

- We formulate medical image fusion as per-level weight prediction over a parameter-free Laplacian pyramid. The convex combination keeps every fused coefficient within the

range of the source coefficients, and a scale-shared, coarse-to-fine predictor built from edge-aware re-parameterizable blocks needs 7,841 parameters at inference.

- We identify logit saturation as the cause of the collapse of weight-predicting fusion networks to one source, show that a logit penalty removes it, and use the resulting well-conditioned logits for an inference-time bias that adjusts the fusion between the two sources without retraining.
- We evaluate LPW-Net against 20 methods on two test sets with twelve metrics, paired significance tests and a per-source trade-off analysis with a null reference. On the public test split LPW-Net ranked first on $Q^{AB/F}$, Q_Y and Q_{CB} for all three tasks, and its trade-off curve passed above nearly all compared methods in the edge-preservation plane.

Section 2 reviews related work, Section 3 describes LPW-Net, Section 4 reports the experiments, Section 5 analyses the results and the limitations, and Section 6 concludes the paper.

2. Related Works

2.1. Rule-Based Multi-Scale and Sparse Fusion

Conventional medical image fusion combines a fixed signal representation with hand-designed fusion rules; a review is given in[1]. Multi-scale transform methods decompose each source into low-pass and band-pass components, merge corresponding components with an activity measure and a selection or averaging rule, and invert the transform. The Laplacian pyramid[9] is the simplest such decomposition, and later medical methods favoured shift-invariant directional transforms. NSCT-PCLL[2] uses phase congruency and local Laplacian energy as activity measures in the nonsubsampled contourlet domain, and NSST-PAPCNN[13] fuses nonsubsampled shearlet subbands with a pulse-coupled neural network whose parameters are set adaptively from the input. Sparse-representation methods code the sources over a dictionary and fuse the sparse coefficients; CSR[3] uses convolutional sparse representation of the whole image, and CSMCA[14] separates cartoon and texture components that are fused separately. Structure-aware fusion[15] works in the spatial domain.

These methods need no training data and are easy to interpret, because the fused image is a selection or blend of source content. Their quality, however, depends on activity measures and rules that are fixed for all images and scales, and on parameters such as the number of decomposition levels, the dictionary size or the filter radii, which usually need tuning for each modality pair. LPW-Net keeps the fixed decomposition and the blend of source coefficients, and replaces the hand-designed rule with a learned, level-specific weight.

2.2. Deep Fusion Networks

Early deep methods kept the classical pipeline and used a network only to measure activity. Liu et al.[16] trained a Siamese network as a patch classifier, used its output as a weight map, and

applied it inside an image pyramid with a hand-crafted local-similarity rule. This is the closest ancestor of weight prediction for medical fusion, but the network was not optimized for the fused result, a single weight map served all levels, and the per-level combination remained rule-based. Later networks learn the mapping from the sources to the fused image directly. IFCNN[4] merges convolutional features with an element-wise rule and decodes them; U2Fusion[17] trains one unsupervised network for several fusion tasks; SwinFusion[18] and MATR[19] use Transformer blocks to model long-range dependencies; and CDDFuse[5] decomposes features into a Transformer base branch and an invertible detail branch before decoding them.

Methods published since 2024 extend this encoder–decoder design in several directions. Diffusion models are used either as samplers, as in DDFM[20], or to train the feature extractor of a unified network, as in DM-FNet[6]. General frameworks serve several fusion tasks with one model: EMMA[21] trains under an equivariance prior, GIFNet[7] relies on interaction between low-level tasks, ReFusion[22] learns its loss by meta-learning, and TTD[23] adapts a pre-trained network such as CDDFuse at test time. Among medical-specific networks, GeSeNet[24] uses semantic guidance with a couple mask ensemble, and BSAFusion[25] aligns unregistered pairs and fuses them jointly. Table 1 lists the deep methods compared in this paper. Almost all of them produce the fused intensities with a decoder or generator; the output is therefore not constrained to the range spanned by the sources, and part of the model capacity is spent on image synthesis rather than on the fusion decision.

Table 1. Deep methods compared in this paper. Parameters were counted by us on the official models as the parameters of the layers executed at inference.

Method	Venue	Year	Released weights used	Parameters
MATR[19]	IEEE TIP	2022	medical	13,230
CDDFuse[5]	CVPR	2023	medical	1,188,272
EMMA[21]	CVPR	2024	infrared–visible	1,517,657
GeSeNet[24]	IEEE TNNLS	2024	medical	240,585
LRFNet[26]	Comput. Biol. Med.	2024	PET/SPECT–MRI	55,185
MMDRFuse[8]	ACM MM	2024	infrared–visible	113
TTD[23]	NeurIPS	2024	medical (CDDFuse)	1,188,272
BSAFusion[25]	AAAI	2025	medical, per task	9,760,776
DM-FNet[6]	IEEE TMM	2025	medical	23,696,721
GIFNet[7]	CVPR	2025	unified	743,032
MACTFusion[27]	IEEE JBHI	2025	CT–MRI	571,035
ReFusion[22]	IJCV	2025	medical	59,856
LPW-Net (ours)	–	–	medical	7,841

2.3. Lightweight Fusion and Structural Re-parameterization

Lightweight fusion networks have mainly been obtained by distillation or by compact architectures. MMDRFuse[8] distills a mini-model with 113 parameters from a larger teacher with

a dynamic refresh strategy, and LUT-Fuse[28] distills an infrared–visible fusion network into learnable look-up tables. LRFNet[26] targets real-time fusion of MRI with functional images, and MACTFusion[27] keeps its cross Transformer small. These designs still synthesize the fused intensities.

Structural re-parameterization offers another route, which decouples the network used during training from the one deployed. RepVGG[10] trains blocks with parallel 3×3, 1×1 and identity branches and, because all branches are linear, merges them into a single 3×3 convolution for inference. The Diverse Branch Block[29] adds further branches that still convert into one convolution. ECBSR[30] brought this idea to real-time super-resolution and added branches in which a 1×1 convolution is followed by fixed Sobel and Laplacian filters with learnable scaling, which inject an edge prior during training at no inference cost. These edge branches are the direct model for the block of LPW-Net.

LPW-Net differs from the nearest work in three respects. Unlike the encoder–decoder networks above, it outputs only one weight map per pyramid level, and each fused coefficient is a convex combination of the co-located source coefficients. Unlike Liu et al.[16] and the rule-based transform methods[2,13], the weights are neither derived from a fixed activity measure nor produced by a classifier trained for another objective; they are learned end to end with an unsupervised fusion loss, separately for every level and from coarse to fine. Unlike distillation-based lightweight fusion, its small size comes from what it predicts and from re-parameterization; MMDRFuse is still smaller, so the distinction lies in the formulation rather than in the parameter count. Finally, whereas controllable diffusion methods steer the output through conditions during iterative sampling[31,32], LPW-Net moves the balance between the sources with one scalar bias at inference.

3. Methodology

3.1. Notation and Problem Formulation

Let A denote the anatomical image (MRI) and B the second source, which is either a CT image or the luminance of a PET or SPECT image (Sect. 3.8). Both images are registered, have size $H \times W$ and are scaled to $[0,1]$. A fusion method maps the pair (A, B) to a fused image F that should keep the clinically relevant content of both sources. No reference fused image exists, so the mapping has to be learned from unpaired training data with an unsupervised objective. Table 2 lists the symbols used in this section.

Table 2. Notation used in the description of LPW-Net.

Symbol	Meaning
A, B, F	MRI source, CT or functional luminance source, fused image
L, l	number of pyramid levels, level index ($l = 0$ is the finest)
G_X^l, P_X^l	Gaussian and Laplacian coefficients of image X at level l

Symbol	Meaning
D, U, $\mathcal{R}$	downsampling, expansion and reconstruction operators
f_θ	weight predictor shared by all levels, with parameters θ
$\boldsymbol{x}^l$	input of the predictor at level l
z^l, $\tilde{z}^l$	logit map of level l, upsampled logit map of level $l+1$
W^l	weight map of level l (the weight given to A)
σ	logistic sigmoid, $\sigma(z) = 1/(1+e^{-z})$
b	scalar fusion bias applied at inference (zero during training)
C	number of channels of the hidden layers of f_θ
α, β, γ	weights of the gradient, structural and logit terms of the loss
λ	local-variance weight inside the structural loss
ϵ	small constant that avoids division by zero

LPW-Net never generates pixels. It decomposes A and B with the same fixed Laplacian pyramid, predicts one weight map per pyramid level with the small network f_θ, mixes the two sets of coefficients with these weights, and reconstructs F from the mixed pyramid. The only learned part is the mapping from local pyramid content to a weight in (0,1), that is, the decision of *where* to trust which modality at which scale. Figure 1 gives an overview of the pipeline and of the training objective.

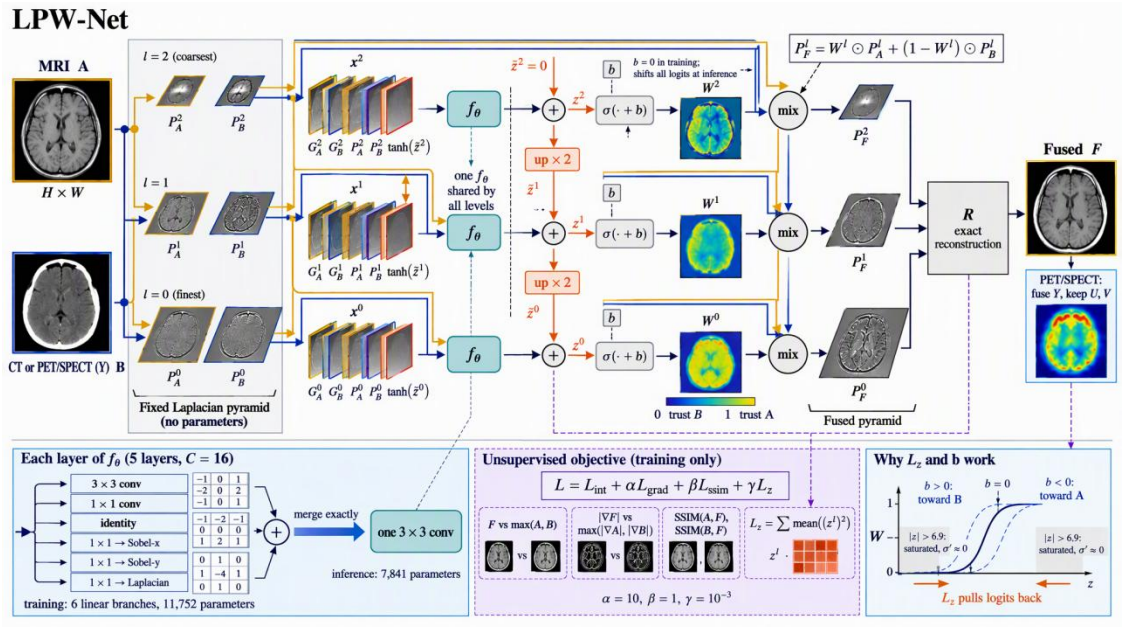


Figure 1. Overview of LPW-Net. Top: both sources are decomposed by the same fixed Laplacian pyramid. One predictor f_θ, shared by all levels, estimates a logit map per level from the coarsest to the finest level; the upsampled logit of a coarser level is the starting point of the next one. The sigmoid weights W^l (yellow trusts the MRI, dark blue the second source) mix the source coefficients convexly, and the fused image is the exact reconstruction $\mathcal{R}$ of the mixed pyramid. Bottom, from left to right: during training each layer of f_θ sums six linear branches, which are merged into one 3×3 convolution for inference; the unsupervised training objective; the saturation of the sigmoid

for large $|z|$, which the logit penalty $\mathcal{L}_z$ counteracts. The images are schematic; the weight maps of a real pair are shown in Fig. 9.

3.2. Laplacian Pyramid Decomposition

We use the Burt–Adelson pyramid[9] with the separable five-tap kernel $k = \frac{1}{16}[1,4,6,4,1]$. For an image $X \in \{A, B\}$ the Gaussian pyramid is $G_X^0 = X$ and $G_X^{l+1} = \mathrm{D}(G_X^l)$, where D blurs with $k \otimes k$ (reflect padding) and keeps every second row and column. With the expansion operator U, which inserts zeros, blurs with $4\,k \otimes k$ and crops to the target size, the Laplacian pyramid with L levels is

$$P_X^l = G_X^l - \mathrm{U}\big(G_X^{l+1}\big), l = 0, \dots, L-2, \qquad P_X^{L-1} = G_X^{L-1}. \qquad (1)$$

The reconstruction $\mathcal{R}$ sets $X^{L-1} = P^{L-1}$ and $X^l = P^l + \mathrm{U}(X^{l+1})$ down to $l = 0$, and recovers X exactly from $\{P_X^l\}$.

The predictor receives both kinds of coefficient for a reason: the Gaussian coefficients tell it the local brightness of each source, for instance whether a region is bone, soft tissue or background, and the Laplacian coefficients tell it where each source has structure. The fixed pyramid itself is chosen for three reasons. It has no parameters, so the network does not spend capacity on learning a transform, and every level has a known meaning: the top level carries local brightness, the other levels carry band-pass contrast. It is exactly invertible, so a weight map equal to one everywhere reproduces A without any loss, and any difference between F and the sources is caused by the weights alone. Finally, it is cheap: the decomposition and the reconstruction together cost a few separable five-tap filters per level. We use $L = 3$ levels, selected on the validation set (Sect. 4.1).

3.3. Scale-Shared, Coarse-to-Fine Weight Prediction

The weights are predicted from the coarsest level to the finest. At level l the predictor receives the Gaussian and Laplacian coefficients of both sources together with the decision already made at the coarser level,

$$\boldsymbol{x}^l = \big[\, G_A^l,\ G_B^l,\ P_A^l,\ P_B^l,\ \tanh(\tilde{z}^l) \,\big] \in \mathbb{R}^{5 \times H_l \times W_l}, \qquad (2)$$

where $H_l \times W_l$ is the size of level l, $\tilde{z}^{L-1} = 0$, and $\tilde{z}^l$ for $l < L-1$ is the logit map of level $l+1$ upsampled bilinearly to the size of level l. The predictor outputs a residual update of the logit,

$$z^l = \tilde{z}^l + f_\theta(\boldsymbol{x}^l), \qquad W^l = \sigma(z^l + b), \qquad (3)$$

where b is zero during training (Sect. 3.7). The fused coefficients and the fused image are

$$P_F^l = W^l \odot P_A^l + (1 - W^l) \odot P_B^l, \qquad F = \mathcal{R}\big(\{P_F^l\}_{l=0}^{L-1}\big), \qquad (4)$$

where $\odot$ is the element-wise product.

Three design choices follow from treating fusion as a decision problem. First, Eq. (4) is a convex combination, so every fused coefficient lies between the two corresponding source coefficients. The network can select and blend structures that exist in the sources, but it cannot add a coefficient outside their range. This matters for medical images, where intensity often carries meaning, such as the density of bone in CT or the tracer uptake in PET. Second, the same f_θ is applied at every level. The input of Eq. (2) is expressed in pyramid coefficients, whose meaning (local brightness, local contrast) is the same at every scale, so sharing the predictor costs little accuracy and keeps the number of parameters independent of L; Sect. 4.3 compares it with one predictor per level. Third, the coarse-to-fine residual form lets a coarse level settle the large-scale assignment, for instance which region follows the functional image, and leaves the finer levels to refine it near edges. The bounded input $\tanh(\tilde{z}^l)$ tells the predictor what was decided above without letting a large logit dominate its other inputs.

The predictor f_θ is a plain stack of five 3×3 convolutional layers with widths $5 \to C \to C \to C \to C \to 1$ and $C = 16$, with LeakyReLU (slope 0.1) between layers and no activation after the last layer. Each layer is the re-parameterizable block of Sect. 3.5 during training and a single convolution at inference. Five 3×3 layers give a receptive field of 11×11 coefficients at each level, which corresponds to 44×44 pixels at the coarsest level when $L = 3$; together with the coarse-to-fine path this lets the weights depend on a neighbourhood much larger than one edge. To start training from an unbiased mixture, the last block is initialized so that its output is zero, which gives $W^l = 0.5$ everywhere at the first iteration.

3.4. Properties of the Formulation

Writing the fusion as Eqs. (3) and (4) has consequences that a synthesis network does not share. The first concerns the training signal. By the chain rule, the gradient of the loss with respect to the logits of level l is

$$\frac{\partial \mathcal{L}}{\partial z^l} = \sigma'(z^l) \odot \left(P_A^l - P_B^l\right) \odot \frac{\partial \mathcal{L}}{\partial P_F^l}, \qquad (5)$$

where $\sigma'(z) = \sigma(z)(1 - \sigma(z))$ and the last factor is the gradient that reaches the fused coefficients through the reconstruction. The weights therefore learn only where the two sources differ. Where both sources agree, for instance in the background, the choice does not matter and the weight receives no gradient, which is why the weight maps in such regions stay near their initial value of 0.5. Equation (5) also shows why saturated logits are harmful: when $\sigma'(z^l)$ is close to zero, no difference between the sources can move the weight, which is the collapse addressed in Sect. 3.6.

The second consequence is consistency. If the two sources are identical, every weight map yields $F = A$; if the weights are one everywhere, the output is exactly A, and if they are zero, it is exactly B. Between these extremes every fused coefficient lies between the source coefficients.

The convexity holds for the pyramid coefficients and not for individual pixels, because the reconstruction adds bands that may come from different sources, so the fused image is clipped to $[0,1]$ after reconstruction. The formulation also cannot invert the contrast of a source, a failure that a pixel-regression network of the same size exhibited in our experiments (Sect. 5).

The third consequence is cost. Because the same predictor runs on every level and level l has a quarter of the pixels of level $l-1$, the multiply–accumulate operations of the predictor amount to the per-pixel cost of the five merged convolutions, multiplied by $HW\sum_l 4^{-l}$. For $L=3$ and a 256×256 pair this gives 0.67 GMACs, about one third more than running the predictor once at full resolution, and the pyramid operations add only a few separable filters per level.

3.5. Edge-Aware Re-parameterizable Block

Weight maps have to change sharply at tissue boundaries, while a five-layer network with 16 channels has little capacity to discover edge detectors on its own. During training, each layer therefore sums several linear branches,

$$\boldsymbol{y} = \mathrm{Conv}_{3\times3}(\boldsymbol{x}) + \mathrm{Conv}_{1\times1}(\boldsymbol{x}) + 1[c_{\mathrm{in}} = c_{\mathrm{out}}]\,\boldsymbol{x} + \sum_{e\in\{s_x,s_y,\Delta\}} (\boldsymbol{s}_e \odot (M_e * \mathrm{Conv}^{e}_{1\times1}(\boldsymbol{x})) + \boldsymbol{b}_e)\,, \qquad (6)$$

where c_{in} and c_{out} are the numbers of input and output channels, M_{s_x} and M_{s_y} are the horizontal and vertical Sobel masks, and M_Δ is the four-neighbour Laplacian mask. Each edge branch applies a learnable 1×1 convolution without bias, followed by the fixed mask applied depth-wise; $\boldsymbol{s}_e, \boldsymbol{b}_e \in \mathbb{R}^{c_{\mathrm{out}}}$ are a learnable per-channel scale and bias. The fixed masks give the block edge responses from the first iteration, in the spirit of the edge-oriented branches of ECBSR[30], and the 1×1 and identity branches follow RepVGG and DBB[10,29], which use such branches to ease the optimization of plain networks.

All branches are linear and use zero padding on the same 3×3 support, so after training they are merged into one 3×3 convolution with kernel and bias

$$K = K_3 + \mathrm{pad}(K_1) + I_3 + \sum_e (\boldsymbol{s}_e \odot K_1^e) \otimes M_e, \qquad \boldsymbol{b} = \boldsymbol{b}_3 + \boldsymbol{b}_1 + \sum_e \boldsymbol{b}_e\,, \qquad (7)$$

where $K_3, \boldsymbol{b}_3$ and $K_1, \boldsymbol{b}_1$ are the kernels and biases of the 3×3 and 1×1 branches, $\mathrm{pad}(K_1)$ places the 1×1 kernel at the centre of a 3×3 kernel, I_3 is the identity kernel (present only when $c_{\mathrm{in}} = c_{\mathrm{out}}$), and $(\boldsymbol{s}_e \odot K_1^e) \otimes M_e$ is the $c_{\mathrm{out}}\times c_{\mathrm{in}}\times3\times3$ kernel whose (o,i) slice is $s_{e,o}\,K^e_{1,oi}\,M_e$. The merge is exact up to floating-point rounding. The deployed network is therefore a plain five-layer CNN with 7,841 parameters, while the training-time network has 11,752. The extra branches change only how the same function class is parameterized during optimization, not what the deployed network can represent, which is why their effect is measured empirically in Sect. 4.3.

3.6. Unsupervised Objective and Logit Penalty

No fused ground truth exists, so the network is trained with three unsupervised terms of the kind used by SwinFusion and related methods[18] and with one regularizer that is specific to the weighting formulation:

$$\mathcal{L} = \mathcal{L}_{\text{int}} + \alpha\,\mathcal{L}_{\text{grad}} + \beta\,\mathcal{L}_{\text{ssim}} + \gamma\,\mathcal{L}_{z}. \qquad (8)$$

The intensity term $\mathcal{L}_{\text{int}} = \frac{1}{HW}\|F - \max(A, B)\|_1$ favours the bright, salient content of either source: dense bone in CT, high uptake in PET and SPECT, and bright soft tissue in MRI. The gradient term

$$\mathcal{L}_{\text{grad}} = \frac{1}{HW}\,\|\;|\nabla F| - \max(|\nabla A|, |\nabla B|)\;\|_1, \qquad |\nabla X| = |S_x * X| + |S_y * X|, \qquad (9)$$

with the Sobel operators S_x and S_y, asks every edge of the fused image to be as strong as the stronger of the two source edges at that position. The structural term

$$\mathcal{L}_{\text{ssim}} = 1 - \text{mean}(\lambda \odot \text{SSIM}(A, F) + (1 - \lambda) \odot \text{SSIM}(B, F)), \qquad \lambda = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2 + \epsilon}, \qquad (10)$$

uses local SSIM maps[33] and the local variances σ_A^2 and σ_B^2, all computed in an 11×11 Gaussian window with standard deviation 1.5, so that each region is compared mainly with the source that carries more structure there.

The fourth term, $\mathcal{L}_z = \sum_l \text{mean}\,((z^l)^2)$, penalizes large logits. It addresses a failure mode of weight prediction that we met repeatedly while developing the method: training ends with $W^l \equiv 1$, and by the exactness of $\mathcal{R}$ the network then reproduces the MRI pixel for pixel. The weighting formulation makes this state hard to leave. The gradient of the loss with respect to a logit is the gradient with respect to W^l multiplied by $\sigma'(z^l) = \sigma(z^l)(1 - \sigma(z^l))$, which decays exponentially in $|z^l|$. Once the logits have grown large, the network receives almost no signal that would move the weights back, however much the loss would gain from it. Starting from $W^l = 0.5$ does not prevent this, and neither does a soft bound on the magnitude of the logits that still allows the saturated regime (Sect. 5). The quadratic penalty acts on every logit at every iteration and pulls it back toward the range in which the sigmoid still passes gradient. We use $\alpha = 10$, $\beta = 1$ and $\gamma = 10^{-3}$, chosen on the validation set.

3.7. Inference-Time Fusion Bias

The scalar b in Eq. (3) is zero during training and in all main results. Because it shifts every logit by the same amount, a positive b moves each fused coefficient toward the MRI coefficient and a negative b toward the CT or functional coefficient, continuously and without retraining; $b \rightarrow$

$+\infty$ returns A and $b \to -\infty$ returns B. A network that synthesizes the fused image has no comparable control. We use b in Sect. 5.2 to trace the whole trade-off between the information kept from each source, and a reader can use it to let the functional or the anatomical content dominate. The logit penalty is what makes this control usable: with saturated logits a moderate b hardly changes the output, whereas with the penalty the logits stay in a range where b has a graded effect.

3.8. Colour Images

PET and SPECT images are pseudo-coloured. We convert them to YUV with the ITU-R BT.601 coefficients[34], fuse the luminance Y with the MRI as image B, keep the chrominance U and V of the functional image, and convert (F, U, V) back to RGB. CT–MRI fusion is performed on the grey images directly. This keeps the colour coding of the functional image, which encodes uptake, unchanged, while the anatomical detail enters through the luminance. Fusing the three colour channels separately would weight them differently at the same position and could shift the hue, which a reader would interpret as a different uptake. The price of this choice is that the functional image contributes to the luminance only where the weights select it, a point we return to in Sect. 5.2.

3.9. Training and Inference

Algorithm 1 summarizes training and inference. LPW-Net is trained with Adam (learning rate 10^{-3} with cosine decay to 10^{-5}) for 20,000 iterations on batches of 32 random 128×128 crops, augmented with horizontal and vertical flips and rotations by multiples of $90°$. CT–MRI, PET–MRI and SPECT–MRI pairs are mixed in one training set, so a single model serves all three tasks. The task is not given to the network as an input. The MRI is always the source A, and the predictor has to infer from the coefficients themselves whether the second source shows bone, metabolism or perfusion, which is possible because the three kinds of image differ clearly in their brightness and frequency content. Mixing the tasks also gives the small predictor more examples of tissue boundaries than any single task would. The 128×128 crops keep a 32×32 coarsest level for $L = 3$, large enough for the receptive field of the predictor, while allowing a batch of 32 crops to be processed quickly. After training every block is merged with Eq. (7), and inference is a single forward pass of the merged network through the pyramid, with no iterative refinement.

Algorithm 1. Training and inference of LPW-Net.

Input: training pairs $\{(A, B)\}$, levels L, loss weights α, β, γ, iterations T.

1. **for** $t = 1, \dots, T$ **do**
2. Sample a batch of registered crops (A, B) and augment them.
3. Compute $\{G_A^l, P_A^l\}$ and $\{G_B^l, P_B^l\}$ with Eq. (1).

4. Set $\tilde{z}^{L-1} \leftarrow 0$.
5. **for** $l = L-1, \ldots, 0$ **do**
6. $z^l \leftarrow \tilde{z}^l + f_\theta(\boldsymbol{x}^l)$ with $\boldsymbol{x}^l$ from Eq. (2); $W^l \leftarrow \sigma(z^l)$.
7. $P_F^l \leftarrow W^l \odot P_A^l + (1 - W^l) \odot P_B^l$.
8. **if** $l > 0$ **then** $\tilde{z}^{l-1} \leftarrow$ bilinear upsampling of z^l.
9. **end for**
10. $F \leftarrow \mathcal{R}(\{P_F^l\})$; compute $\mathcal{L}$ with Eq. (8); update θ with Adam.
11. **end for**
12. Merge every block of f_θ into one 3×3 convolution with Eq. (7).
13. **Inference:** run the loop body once with the merged network and $W^l \leftarrow \sigma(z^l + b)$; for PET/SPECT fuse the luminance and restore the chrominance (Sect. 3.8).

4. Experimental Results

4.1. Experimental Setup

Data. All images come from the Harvard Whole Brain Atlas[35]. We use the registered 256×256 CT–MRI, PET–MRI and SPECT–MRI pairs and the public train/test split released with ASFE-Fusion[36]. Test set B is the public test split with 24 pairs per task (72 pairs). Test set A consists of nine pairs, three per task, for which the outputs of eight classical methods were available to us [TO BE COMPLETED BY AUTHORS: origin of test set A]. Five of these nine pairs are identical or nearly identical to atlas slices, so we removed from the training data the scans that contribute them and, as a precaution, one further scan, six scans in total. One file pair of test set A had its two modalities labelled the wrong way round; we corrected it before any experiment reported here. After the exclusion, 648 training pairs remained (112 CT–MRI, 245 PET–MRI, 291 SPECT–MRI). A fixed validation set of ten pairs per task, drawn from these pairs, was used for every design decision; the test sets were not used for any choice of architecture, hyperparameter or checkpoint.

Compared methods. We compare LPW-Net with the twelve deep methods of Table 1 and, on test set A, with eight classical methods: NSCT-PCLL[2], NSST-PAPCNN[13], CSR[3], LRD[37], SA[15], CSMCA[14], CBF[38] and ASR[39]. Every deep method was run with its official code and pretrained weights, without retraining; only file paths and input/output handling were changed. EMMA and MMDRFuse release infrared–visible weights only, which we applied directly, as the EMMA authors do for medical images. LRFNet releases weights for PET/SPECT–MRI and MACTFusion for CT–MRI only, so each is reported for those tasks. For PET and SPECT, methods that fuse grey images received the colour handling of Sect. 3.8, while ReFusion, GIFNet, DM-FNet and BSAFusion kept their own. Three deviations from the official code were needed.

The MATR output was clipped to [0,1] before conversion to 8 bits, because the official script lets 101 of 5,308,416 pixels wrap around. The sampling noise of DM-FNet was seeded per image; two seeds gave outputs that agree to 47.9 dB PSNR on average. Two library imports of BSAFusion were replaced by equivalent ones, after which our pipeline reproduced the released outputs to above 53.4 dB PSNR. BSAFusion includes a registration module that slightly deforms the functional image even for registered input; we kept it as part of the method.

Metrics. We report twelve common metrics[12,40]: entropy (EN)[41], standard deviation (SD)[40], spatial frequency (SF)[42], average gradient (AG)[43], mutual information (MI)[44], normalized mutual information (NMI)[45], the gradient-based $Q^{AB/F}$[11], the SSIM-based Q_Y[46], the perception-based Q_{CB}[47], visual information fidelity (VIF)[48], SSIM[33] and the sum of the correlations of differences (SCD)[49]. MI, VIF and SSIM are summed over the two sources, higher is better for all twelve, and for PET and SPECT all metrics are computed on the luminance. EN, SD, SF and AG describe the fused image alone, whereas the others compare it with the sources. $Q^{AB/F}$ is a gradient-weighted average of two edge-preservation maps; we also report its two parts $Q^{A/F}$ and $Q^{B/F}$, each weighted by the gradient strength of its own source, because they show from which source a method keeps edges.

Protocol. The hyperparameters (L, C, depth, α, β, γ) were selected by validation $Q^{AB/F}$. The logit penalty was insensitive over $\gamma \in [10^{-4}, 3 \times 10^{-3}]$, with validation $Q^{AB/F}$ between 0.758 and 0.760, and validation $Q^{AB/F}$ dropped to 0.753 at $\gamma = 10^{-2}$. The final models were trained on all 648 pairs, which include the validation pairs, so we kept their last iteration instead of selecting a checkpoint. We trained five final models with different seeds and report their mean; the figures show the model with seed 0, fixed in advance. Training one model took 13.2 minutes on one NVIDIA RTX 4090D.

4.2. Comparison with Existing Methods

Table 3. Results on test set B (24 pairs per task). LPW-Net: mean over five seeds; the standard deviation over seeds is given in the s.d. row. Best in bold, second underlined. LRFNet and MACTFusion release weights for one task family only.

Method	EN	SD	SF	AG	MI	NMI	$Q^{AB/F}$	Q_Y	Q_{CB}	VIF	SSIM	SCD
CT-MRI												
MATR	4.852	50.548	18.386	6.558	**4.051**	**0.981**	0.529	0.520	0.354	**0.763**	0.591	-0.170
CDDFuse	4.726	80.980	36.617	10.661	3.105	0.777	0.615	0.918	0.665	0.526	**1.480**	1.464
EMMA	**5.611**	79.452	24.909	8.754	2.943	0.662	0.509	0.848	0.455	0.466	1.365	1.518
GeSeNet	5.206	81.938	35.729	10.648	3.001	0.711	<u>0.629</u>	0.732	0.367	0.508	1.088	1.467
MMDRFuse	4.771	69.272	19.765	6.474	2.978	0.741	0.398	0.399	0.198	0.448	0.556	1.354
TTD	4.560	<u>84.273</u>	<u>38.305</u>	10.617	3.055	0.781	0.583	0.922	0.666	0.527	<u>1.474</u>	1.451
BSAFusion	4.664	82.378	28.892	8.922	3.131	0.787	0.536	0.916	0.653	0.510	1.448	1.410
DM-FNet	4.580	**85.928**	36.555	10.181	3.024	0.771	0.587	<u>0.922</u>	<u>0.667</u>	0.525	1.468	<u>1.520</u>
GIFNet	5.023	71.124	**41.788**	**12.283**	2.602	0.629	0.441	0.487	0.355	0.323	0.729	1.166

Method	EN	SD	SF	AG	MI	NMI	$Q^{AB/F}$	Q_Y	Q_{CB}	VIF	SSIM	SCD
MACTFusion	4.668	82.145	32.513	11.041	2.656	0.667	0.565	0.872	0.657	0.447	1.383	**1.614**
ReFusion	5.465	77.976	32.855	10.528	2.745	0.630	0.617	0.532	0.346	0.458	0.744	1.466
LPW-Net (ours)	4.629	81.319	35.288	9.929	3.384	0.854	**0.649**	**0.938**	**0.707**	0.649	1.456	1.378
s.d.	0.015	0.138	0.086	0.020	0.030	0.007	0.002	0.000	0.001	0.003	0.001	0.004
PET-MRI												
MATR	5.679	67.612	27.176	10.687	3.598	0.740	0.710	0.607	0.349	0.682	0.715	0.186
CDDFuse	5.743	87.999	34.639	13.605	3.668	0.757	0.723	0.951	0.611	0.749	**1.407**	1.482
EMMA	6.332	88.128	31.276	12.464	3.062	0.600	0.670	0.870	0.454	0.594	1.285	1.486
GeSeNet	6.045	85.114	33.970	13.541	3.148	0.635	0.732	0.765	0.407	0.665	1.032	1.277
LRFNet	5.095	66.640	24.609	9.607	3.210	0.716	0.645	0.846	0.587	0.635	1.313	0.762
MMDRFuse	5.688	83.625	29.894	11.552	3.249	0.680	0.644	0.534	0.273	0.613	0.684	1.354
TTD	5.728	88.929	**39.254**	**15.140**	3.339	0.693	0.710	0.941	0.592	0.649	1.376	1.368
BSAFusion	5.264	**91.301**	37.719	14.320	3.500	0.761	0.724	0.946	0.611	0.679	1.383	**1.552**
DM-FNet	5.547	89.754	36.379	14.052	3.145	0.670	0.713	0.943	0.613	0.639	1.395	1.454
GIFNet	5.654	83.897	35.381	11.646	2.597	0.554	0.436	0.501	0.376	0.398	0.704	1.321
ReFusion	**6.390**	85.692	32.568	13.448	3.172	0.615	0.722	0.660	0.388	0.650	0.864	1.418
LPW-Net (ours)	5.271	81.289	35.189	14.006	**4.087**	**0.884**	**0.784**	**0.976**	**0.682**	**0.880**	1.385	0.757
s.d.	0.025	0.041	0.017	0.011	0.025	0.008	0.000	0.000	0.002	0.003	0.001	0.009
SPECT-MRI												
MATR	5.053	50.769	18.392	6.938	4.030	0.916	0.709	0.562	0.349	0.748	0.667	0.011
CDDFuse	5.127	59.994	21.799	8.192	4.412	1.000	0.752	0.979	0.688	0.909	1.425	1.055
EMMA	5.735	64.600	21.528	8.321	3.224	0.684	0.660	0.886	0.459	0.595	1.330	1.542
GeSeNet	5.462	62.170	21.878	8.394	3.332	0.728	0.721	0.757	0.401	0.666	1.042	1.447
LRFNet	4.912	**68.405**	22.285	8.237	3.307	0.772	0.679	0.857	0.571	0.630	1.286	**1.777**
MMDRFuse	5.339	67.649	20.722	7.887	3.294	0.733	0.636	0.496	0.260	0.597	0.638	1.728
TTD	5.209	60.130	22.334	8.325	3.723	0.840	0.725	0.972	0.672	0.778	1.419	0.805
BSAFusion	4.759	62.891	**24.449**	**8.916**	3.603	0.853	0.732	0.971	0.669	0.815	1.428	1.270
DM-FNet	5.068	63.458	23.433	8.621	3.175	0.726	0.681	0.941	0.604	0.621	1.433	1.481
GIFNet	5.057	56.783	23.128	7.634	2.816	0.650	0.481	0.565	0.347	0.412	0.805	1.068
ReFusion	**5.857**	62.766	21.454	8.568	3.324	0.697	0.712	0.631	0.384	0.696	0.823	1.521
LPW-Net (ours)	4.723	58.953	22.114	8.439	**4.510**	**1.066**	**0.768**	**0.990**	**0.737**	**0.976**	**1.444**	0.853
s.d.	0.019	0.037	0.005	0.007	0.022	0.007	0.000	0.001	0.002	0.002	0.001	0.013

Table 3 gives the results on test set B. LPW-Net ranked first on $Q^{AB/F}$, Q_Y and Q_{CB} in all three tasks. Its $Q^{AB/F}$ was 0.649, 0.784 and 0.768 for CT–MRI, PET–MRI and SPECT–MRI, against 0.629 (GeSeNet), 0.732 (GeSeNet) and 0.752 (CDDFuse) for the strongest compared method. Over the twelve metrics it was first on 3, 6 and 7 metrics for the three tasks; for PET and SPECT the first places also include MI, NMI and VIF, and for SPECT the SSIM. The standard deviation over seeds was at most 0.002 for $Q^{AB/F}$, so the ranking does not depend on the seed. On EN, SD

and SCD, which reward a large intensity spread and large differences from the sources, LPW-Net stays close to the sources by design (Sect. 5).

Paired two-sided Wilcoxon signed-rank tests over the 24 pairs of each task compared LPW-Net with every compared method on every metric, which gives 132 comparisons per task. At an uncorrected level of $p < 0.05$, LPW-Net was significantly higher than the compared method in 89, 93 and 87 of the 132 comparisons for CT–MRI, PET–MRI and SPECT–MRI, and not significantly different in 6, 5 and 13.

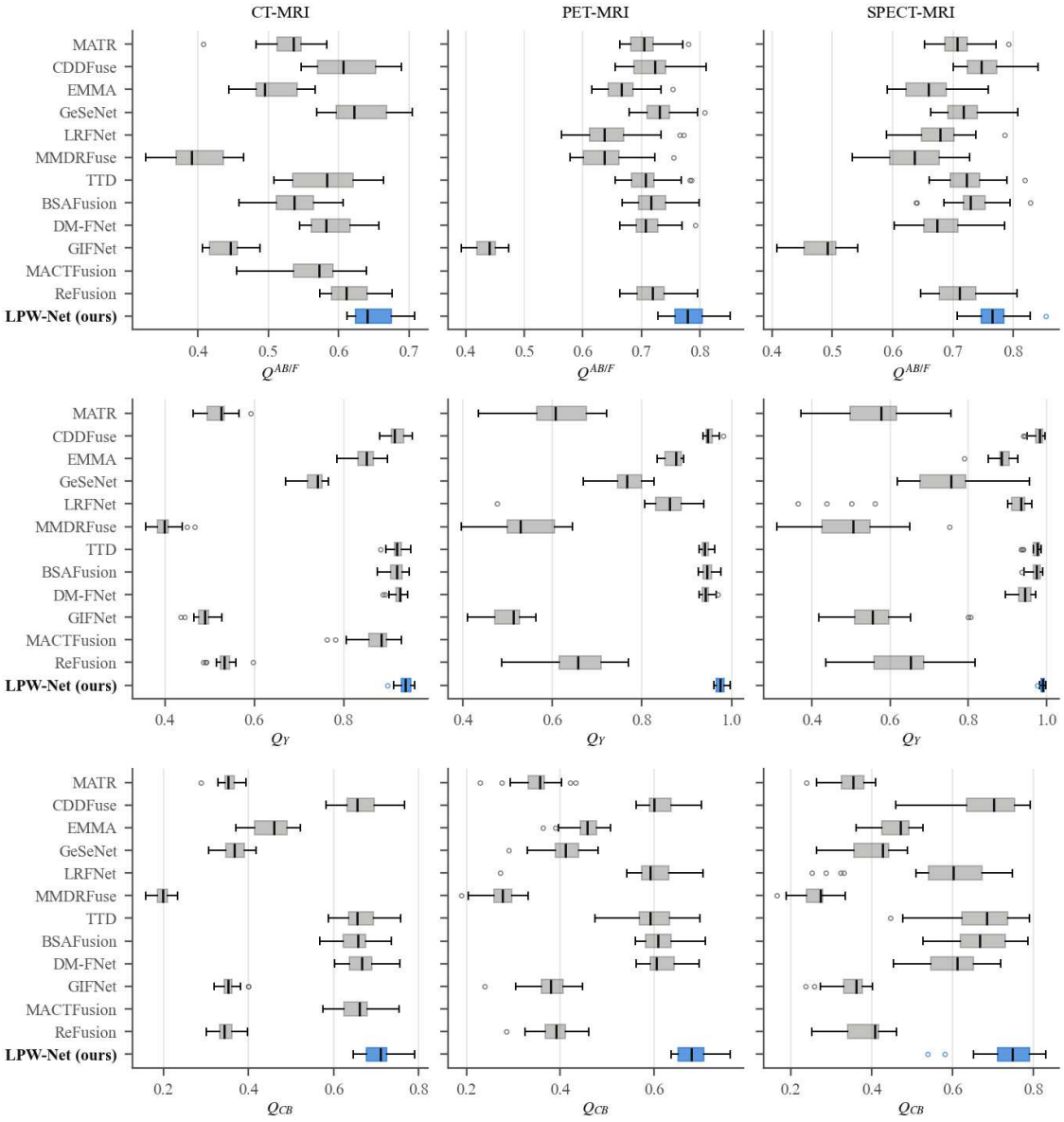


Figure 2. Per-image $Q^{AB/F}$, Q_Y and Q_{CB} on test set B (24 pairs per task). LPW-Net: per-image mean over five seeds. Boxes span the interquartile range with the median marked, whiskers reach the most extreme values within one and a half interquartile ranges, and circles are outliers. LRFNet and MACTFusion appear only in the tasks for which weights are released.

Figure 2 shows the per-image values behind these means. LPW-Net had the highest $Q^{AB/F}$ on 21, 24 and 24 of the 24 pairs of CT–MRI, PET–MRI and SPECT–MRI, the highest Q_Y on 21, 24 and 21, and the highest Q_{CB} on all 24 pairs of every task. On PET–MRI its lower quartile lay above the upper quartile of every compared method for all three metrics. On CT–MRI and SPECT–MRI the boxes of CDDFuse and of some other methods overlapped with those of LPW-Net,.

Table 4. Results on test set A (three pairs per task), including the eight classical methods. LPW-Net: mean over five seeds. Best in bold, second underlined.

Method	EN	SD	SF	AG	MI	NMI	$Q^{AB/F}$	Q_Y	Q_{CB}	VIF	SSIM	SCD
CT-MRI												
NSCT-PCLL	5.579	75.540	26.842	8.596	2.969	0.635	0.581	0.899	0.539	0.625	1.246	1.515
NSST-PAPCNN	5.636	76.194	26.306	8.367	2.935	0.623	0.569	0.896	0.514	0.573	1.260	1.595
CSR	5.321	74.340	30.697	8.793	3.739	0.822	0.614	0.918	0.571	0.704	1.234	1.276
LRD	5.545	78.134	24.862	7.993	2.994	0.645	0.507	0.875	0.493	0.468	1.239	1.607
SA	5.187	68.050	25.556	8.238	3.821	0.846	0.655	0.947	0.593	0.711	1.278	1.214
CSMCA	5.165	66.534	25.346	7.415	2.663	0.610	0.611	0.886	0.570	0.640	1.246	1.424
CBF	5.184	63.054	27.880	9.225	3.417	0.738	0.604	0.891	0.551	0.489	1.256	1.148
ASR	5.018	57.017	23.730	7.104	2.626	0.604	0.577	0.844	0.559	0.495	1.245	1.224
MATR	5.391	51.651	16.089	6.041	**4.368**	**0.930**	0.552	0.655	0.316	**0.749**	0.689	0.135
CDDFuse	5.129	71.969	29.213	9.166	3.023	0.684	0.617	0.854	0.545	0.536	1.262	1.554
EMMA	**5.969**	74.882	21.754	8.066	3.301	0.662	0.568	0.856	0.392	0.560	1.220	**1.667**
GeSeNet	5.727	75.807	29.191	9.620	3.475	0.714	0.675	0.797	0.350	0.597	1.032	1.618
MMDRFuse	5.445	68.165	18.232	6.545	3.514	0.740	0.503	0.576	0.241	0.550	0.683	1.565
TTD	5.096	75.071	30.992	9.429	2.987	0.680	0.599	0.875	0.546	0.530	1.271	1.546
BSAFusion	5.322	77.041	26.778	8.812	2.834	0.639	0.495	0.829	0.533	0.438	1.164	1.216
DM-FNet	4.987	**81.390**	31.038	9.702	3.037	0.696	0.613	0.912	**0.599**	0.566	1.280	1.660
GIFNet	5.145	64.213	**33.939**	**10.340**	2.665	0.595	0.414	0.543	0.332	0.359	0.716	1.305
MACTFusion	4.841	74.604	26.889	9.394	2.658	0.618	0.605	0.868	0.597	0.502	1.229	1.662
ReFusion	5.932	71.953	27.613	9.566	2.984	0.607	0.663	0.661	0.331	0.553	0.795	1.596
LPW-Net (ours)	5.237	75.249	28.402	8.988	3.935	0.853	**0.686**	**0.948**	0.588	0.725	**1.294**	1.551
s.d.	0.026	0.126	0.059	0.021	0.034	0.008	0.001	0.001	0.003	0.002	0.001	0.004
PET-MRI												
NSCT-PCLL	4.829	76.995	30.406	10.571	2.906	0.648	0.578	0.870	0.581	0.488	1.373	1.574
NSST-PAPCNN	4.826	80.012	30.417	10.613	2.977	0.664	0.576	0.876	0.600	0.491	1.391	**1.690**
CSR	4.500	57.881	31.933	10.935	**4.424**	**1.018**	**0.702**	**0.990**	**0.799**	**0.949**	**1.472**	0.458
LRD	4.770	**81.751**	28.803	10.350	3.082	0.692	0.508	0.872	0.614	0.425	1.401	1.678
SA	4.460	60.915	32.085	11.146	4.158	0.964	0.683	0.946	0.762	0.719	1.436	0.825
CSMCA	4.338	61.886	30.851	10.449	2.955	0.697	0.644	0.908	0.660	0.578	1.433	1.334
CBF	4.447	57.615	31.474	11.249	3.719	0.864	0.634	0.920	0.733	0.583	1.434	0.944
ASR	4.052	59.048	30.148	10.313	2.873	0.701	0.609	0.883	0.673	0.490	1.423	1.271
MATR	4.878	47.251	24.914	8.386	3.262	0.721	0.620	0.452	0.288	0.616	0.515	0.213

Method	EN	SD	SF	AG	MI	NMI	$Q^{AB/F}$	Q_Y	Q_{CB}	VIF	SSIM	SCD
CDDFuse	4.766	72.244	30.831	10.395	3.420	0.768	0.627	0.929	0.679	0.536	1.466	1.554
EMMA	5.408	77.543	25.583	9.054	3.234	0.678	0.535	0.864	0.399	0.455	1.370	1.659
GeSeNet	4.913	79.163	29.999	10.245	3.385	0.749	0.611	0.656	0.285	0.500	0.926	1.595
LRFNet	4.997	68.036	30.132	10.309	3.384	0.740	0.603	0.490	0.314	0.560	0.658	1.565
MMDRFuse	4.700	73.285	23.970	8.121	3.226	0.730	0.472	0.390	0.188	0.480	0.491	1.595
TTD	4.736	71.319	34.044	11.244	3.333	0.751	0.611	0.920	0.680	0.499	1.447	1.478
BSAFusion	**6.083**	68.857	**34.683**	**11.436**	2.973	0.581	0.572	0.529	0.360	0.473	0.697	1.316
DM-FNet	5.272	80.460	29.749	10.234	3.170	0.673	0.580	0.875	0.424	0.485	1.364	1.668
GIFNet	4.625	66.235	32.653	10.183	3.095	0.707	0.390	0.431	0.287	0.324	0.630	1.347
ReFusion	5.372	72.593	28.654	10.379	3.165	0.666	0.619	0.521	0.279	0.458	0.694	1.514
LPW-Net (ours)	4.520	74.297	31.655	11.179	3.603	0.832	0.650	0.934	0.716	0.629	1.459	1.423
s.d.	0.004	0.302	0.065	0.056	0.032	0.007	0.003	0.003	0.002	0.008	0.003	0.022
SPECT-MRI												
NSCT-PCLL	4.687	58.955	20.217	7.097	3.402	0.789	0.749	0.962	0.687	0.835	1.486	1.415
NSST-PAPCNN	4.730	59.573	20.199	7.122	3.377	0.781	0.745	0.960	0.677	0.809	1.492	1.570
CSR	4.631	57.108	20.800	6.874	4.008	0.934	0.787	0.981	0.721	0.956	1.477	1.040
LRD	4.961	60.415	19.366	6.849	3.266	0.738	0.688	0.931	0.629	0.660	1.447	1.516
SA	4.476	56.774	20.071	6.884	4.128	0.972	0.786	0.981	0.739	0.944	1.494	1.242
CSMCA	4.260	46.442	18.968	6.311	2.450	0.601	0.730	0.925	0.664	0.712	1.449	1.026
CBF	4.546	54.847	20.206	7.083	3.823	0.897	0.760	0.963	0.722	0.828	1.483	1.050
ASR	4.178	43.056	17.673	5.975	2.480	0.619	0.682	0.913	0.684	0.604	1.471	0.948
MATR	4.879	48.564	17.010	6.034	3.770	0.857	0.750	0.514	0.301	0.793	0.602	0.361
CDDFuse	4.762	57.949	19.914	6.847	4.194	0.970	0.788	0.989	0.712	0.952	1.490	1.346
EMMA	5.305	60.708	19.622	7.036	3.172	0.695	0.712	0.899	0.409	0.688	1.382	1.643
GeSeNet	4.991	59.061	20.166	7.088	3.242	0.735	0.768	0.714	0.341	0.756	1.004	1.587
LRFNet	4.810	**67.623**	21.334	7.456	3.276	0.764	0.726	0.854	0.550	0.761	1.307	**1.848**
MMDRFuse	4.848	63.376	19.721	6.945	3.270	0.755	0.705	0.455	0.236	0.710	0.568	1.773
TTD	4.789	58.909	20.805	7.059	3.683	0.855	0.767	0.982	0.702	0.855	1.485	0.942
BSAFusion	4.474	60.035	22.062	7.407	3.275	0.786	0.760	0.977	0.652	0.848	1.487	1.448
DM-FNet	4.818	60.291	**22.914**	**7.861**	3.022	0.701	0.723	0.952	0.625	0.685	1.485	1.551
GIFNet	4.529	47.207	19.378	5.887	2.676	0.647	0.482	0.477	0.305	0.472	0.689	1.127
ReFusion	**5.341**	59.589	19.555	7.156	3.188	0.694	0.750	0.568	0.320	0.752	0.742	1.635
LPW-Net (ours)	4.485	57.713	20.216	7.033	**4.274**	**1.008**	**0.799**	**0.991**	**0.745**	**0.994**	**1.500**	1.334
s.d.	0.012	0.063	0.008	0.006	0.030	0.007	0.000	0.001	0.002	0.003	0.001	0.015

Test set A (Table 4) contains only three pairs per task, but it adds the classical methods. LPW-Net was first on 3 metrics for CT–MRI ($Q^{AB/F}$, Q_Y, SSIM) and on 7 for SPECT–MRI. On the PET–MRI pairs, CSR led on $Q^{AB/F}$, Q_Y, Q_{CB}, MI, VIF and SSIM; LPW-Net ranked third on $Q^{AB/F}$, Q_Y and VIF, ahead of all deep methods.

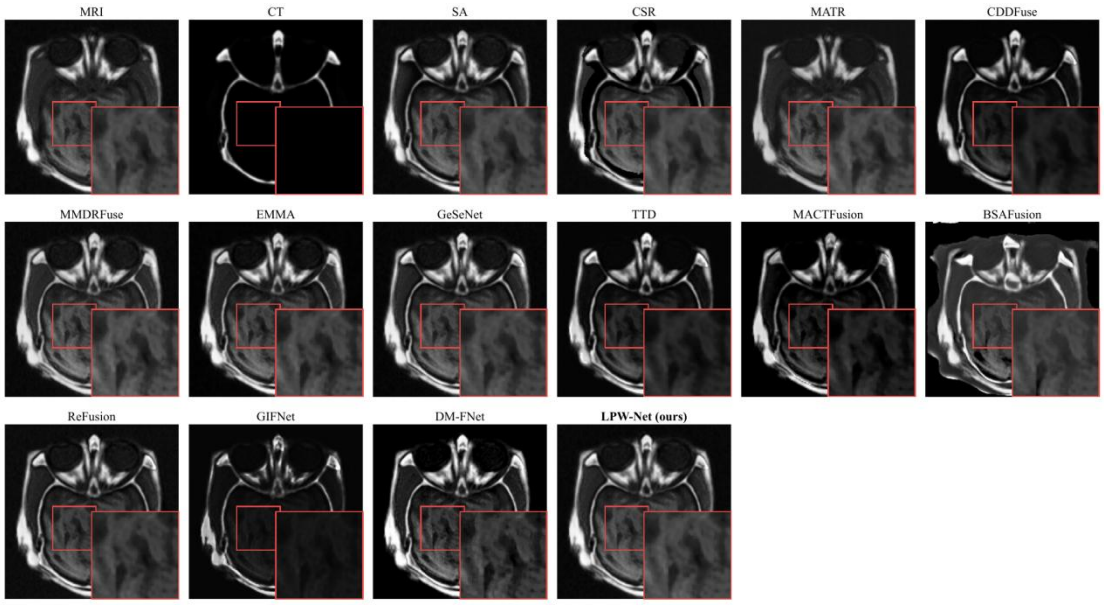


Figure 3. CT–MRI pair of test set A (fixed in advance). The insets enlarge the red box. SA and CSR represent the classical methods.

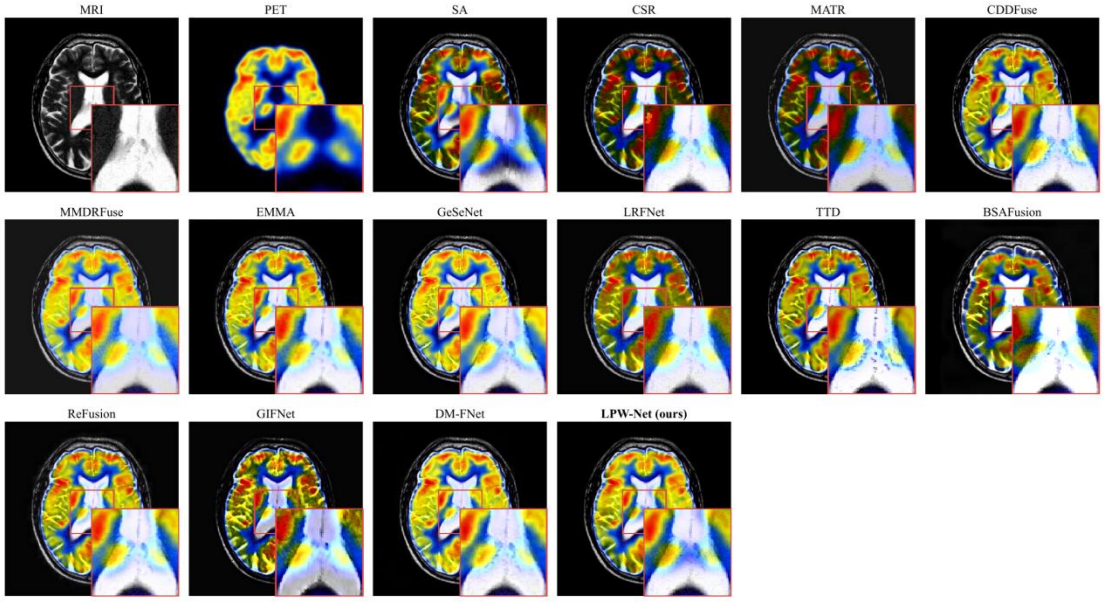


Figure 4. PET–MRI pair of test set A (fixed in advance). The insets enlarge the red box.

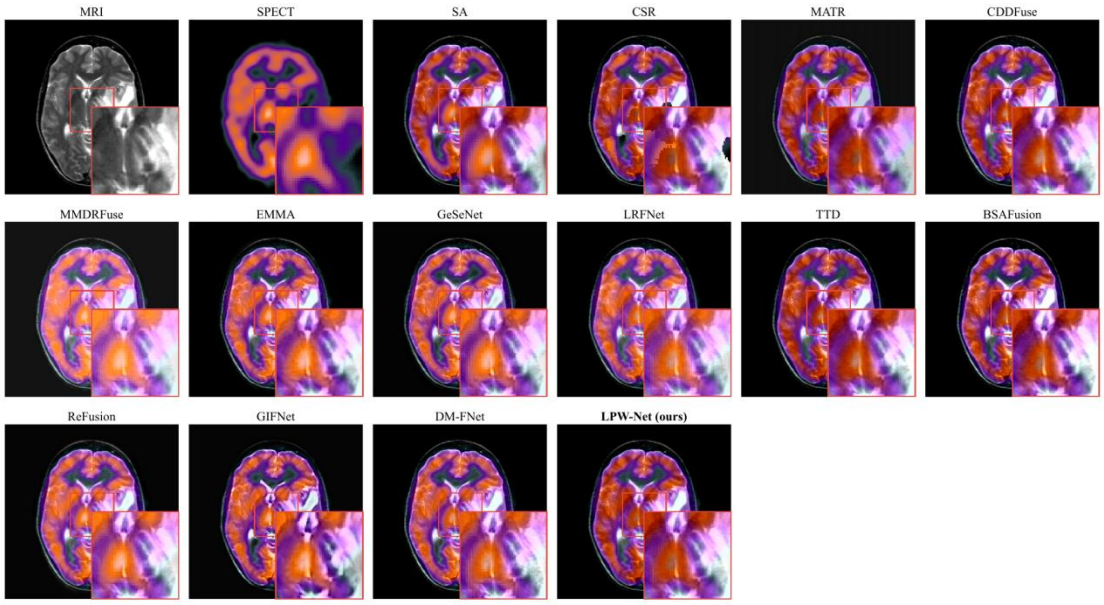


Figure 5. SPECT–MRI pair of test set A (fixed in advance). The insets enlarge the red box.

Figures 3–5 show one pair per task from test set A, chosen before any result was seen. In the CT–MRI pair, LPW-Net kept the bone of the CT together with the soft-tissue texture of the MRI. CDDFuse, TTD and MACTFusion darkened the soft tissue, GIFNet lost most of its texture, DM-FNet added a grid-like pattern, CSR produced dark halos around the skull, and the registration module of BSAFusion deformed the head contour. In the PET–MRI and SPECT–MRI pairs, all methods kept the functional colours. LPW-Net preserved the ventricle boundaries of the MRI without visible artefacts, whereas CSR showed blocky artefacts, TTD left speckles inside the ventricles, and MATR and MMDRFuse raised the background to grey.

4.3. Ablation Study

Table 5. Ablation on test set B (72 pairs), mean over three seeds per variant. "Collapsed" counts seeds whose selected checkpoint reproduced the MRI. All variants use the training split and checkpoint selection on the validation set. Params: parameters of the deployed network.

Variant	Params	Collapsed	Val. $Q^{AB/F}$	$Q^{AB/F}$	$Q^{A/F}$	$Q^{B/F}$	Q_Y	VIF	SSIM	SCD	EN
LPW-Net (full)	7,841	0/3	0.758	0.734	0.817	0.403	0.968	0.837	1.428	0.991	4.879
w/o logit penalty	7,841	2/3	0.732	0.717	0.928	0.187	0.995	1.010	1.424	0.881	4.058
w/o edge branches	7,841	0/3	0.757	0.732	0.817	0.395	0.968	0.840	1.427	0.990	4.847
plain conv (no re-param.)	7,841	0/3	0.759	0.733	0.833	0.357	0.970	0.859	1.424	0.980	4.851
w/o coarse-to-fine	7,841	0/3	0.754	0.728	0.820	0.375	0.965	0.822	1.430	1.046	4.888
per-level predictors	23,523	0/3	0.759	0.734	0.817	0.406	0.965	0.817	1.429	1.037	4.885
$L = 2$	7,841	0/3	0.759	0.737	0.819	0.413	0.965	0.813	1.415	0.959	4.795
$L = 4$	7,841	0/3	0.756	0.732	0.813	0.405	0.966	0.828	1.430	1.060	4.973
w/o $\mathcal{L}_{\text{int}}$	7,841	0/3	0.761	0.735	0.824	0.394	0.968	0.845	1.427	0.924	4.889

Variant	Params	Collapsed	Val. $Q^{AB/F}$	$Q^{AB/F}$	$Q^{A/F}$	$Q^{B/F}$	Q_Y	VIF	SSIM	SCD	EN
w/o $\mathcal{L}_{\mathrm{grad}}$	7,841	0/3	0.704	0.679	0.738	0.409	0.952	0.704	1.462	1.379	5.125
w/o $\mathcal{L}_{\mathrm{ssim}}$	7,841	0/3	0.760	0.735	0.821	0.392	0.969	0.846	1.425	0.976	4.855
pixel regression	7,409	0/3	0.734	0.710	0.829	0.331	0.702	0.595	0.914	0.497	5.250

Table 5 lists the ablations. Each variant was trained with three seeds on the training split, its checkpoint was selected on the validation set, and it was evaluated on test set B. The logit penalty was essential: without it, two of three seeds collapsed to the MRI, and the variant kept edges almost only from the MRI ($Q^{A/F}$ of 0.928 against $Q^{B/F}$ of 0.187). The gradient term was the most important loss term; without it $Q^{AB/F}$ fell from 0.734 to 0.679. The coarse-to-fine path helped (0.728 without it). One predictor per level gave the same $Q^{AB/F}$ of 0.734 with 23,523 instead of 7,841 parameters, so sharing the predictor costs nothing measurable. The remaining components had small effects on these metrics: without the edge branches or without re-parameterization $Q^{AB/F}$ was 0.732 and 0.733, and without the intensity or structural term it was 0.735. With $L = 2$ levels $Q^{AB/F}$ on test set B was 0.737 against 0.734 for $L = 3$; $L = 3$ had been chosen on the validation set, where the two were tied (0.759 and 0.758). Finally, a network of the same size that regresses the fused pixels directly reached 0.710 and was unstable across seeds, which we analyse in Sect. 5.

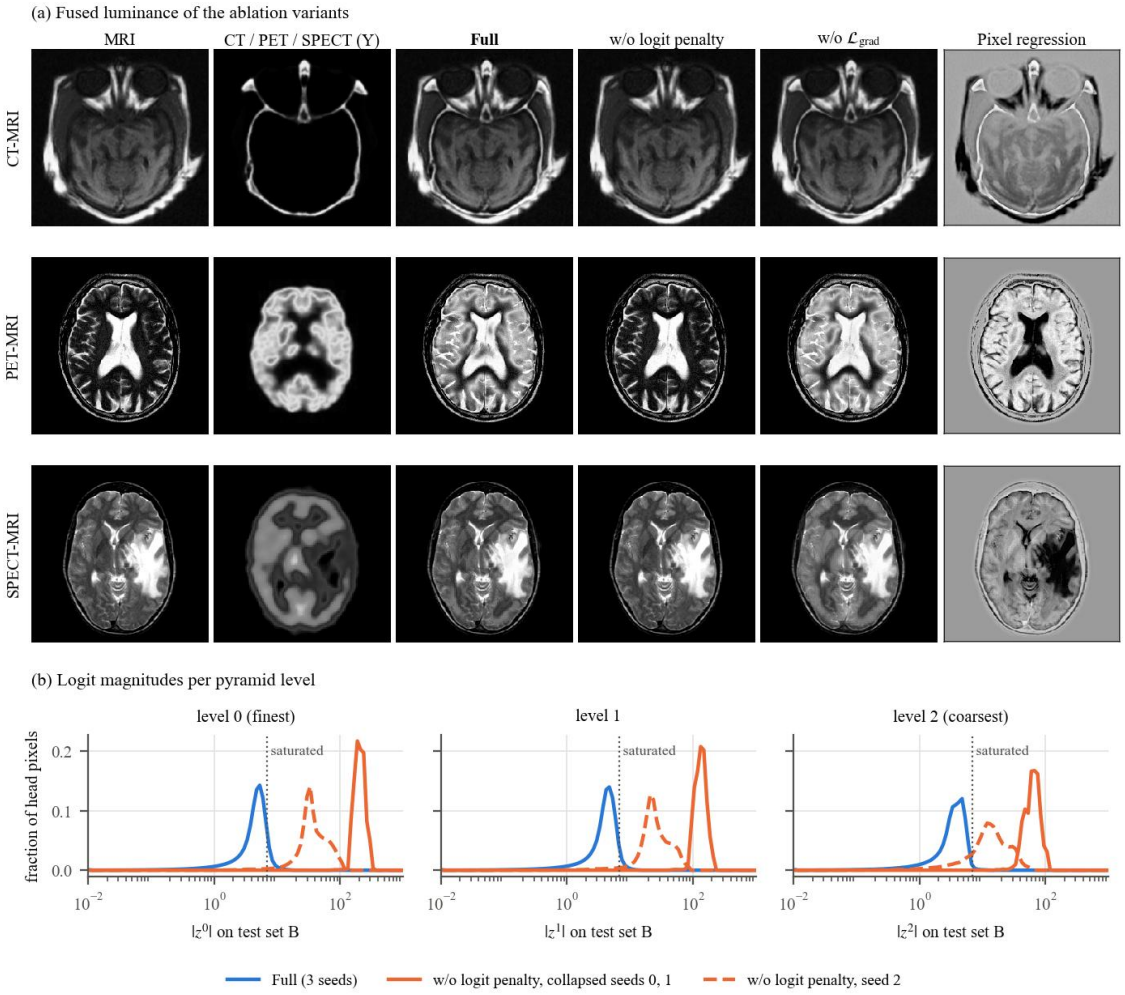


Figure 6. Ablation variants. (a) Fused luminance on the three pairs of Figs. 3–5, seed 0 of each variant except pixel regression, for which seed 2 is shown because its output is inverted; seed 0 without the logit penalty is one of the two collapsed seeds. (b) Distribution of the logit magnitudes $|z^l|$ over the head pixels of test set B, i.e. the pixels where either source exceeds 0.02 at that level. Right of the dotted line ($|z| > 6.9$) the sigmoid is saturated: the weight is above 0.999 or below 0.001.

Figure 6 makes the two failure modes visible. Seed 0 of the variant without the logit penalty is one of its collapsed seeds, and its output is the MRI itself (Fig. 6a). Its logits on the head pixels of test set B had a median magnitude of 228.5 at the finest level and 70.7 at the coarsest, against 4.7 and 3.5 for the full model, and at all three levels 100.0% of them gave the MRI a weight above 0.999 (Fig. 6b). At such magnitudes the factor $\sigma'(z) \approx e^{-|z|}$ of Eq. (5) is negligible, so no gradient can pull the weights back. The seed that did not collapse was saturated as well: 98.1% of its finest-level logits lay beyond the same threshold, against 8.6% for the full model, so its weights acted as a hard selection that a bias of a few units cannot move. Without $\mathcal{L}_{\text{grad}}$ the fused images had visibly weaker edges, most clearly on SPECT–MRI, where more of the smooth functional luminance entered the image. The pixel-regression network of seed 2 inverted the contrast of the MRI, a failure we return to in Sect. 5.1.

5. Discussion and Analysis

5.1. What the Averaged Metrics Miss

Several of the standard metrics reward an output that stays close to one source, and surveys of fusion metrics have noted that different metrics favour different properties and that no single metric is sufficient[12,40]; an averaged metric alone therefore does not show that a method transferred information from the second modality. This motivates the per-source analysis below, and a related blind spot appeared in the ablation. One of the three seeds of the pixel-regression network learned to output an inverted MRI, with a mean correlation of -0.681 between its output and the MRI, while the other two seeds had positive correlations (0.914 and 0.905). $Q^{AB/F}$ measures gradient strength and orientation modulo $180°$[11], so it cannot detect a sign inversion of the contrast. LPW-Net cannot produce such an output, because each fused coefficient lies between the source coefficients. These observations motivate the per-source analysis below.

5.2. Trade-off Between the Two Sources

Every fused image trades similarity to the MRI against similarity to the second source. We measured this trade-off in two planes: the edge plane, spanned by $Q^{A/F}$ and $Q^{B/F}$, and the SSIM plane, spanned by the SSIM of the fused image with each source. For LPW-Net we traced a curve by sweeping the inference bias b from -12 to 12 (mean of the five final models), and as a null reference we traced the linear blend $tA + (1 - t)B$ for t from 0 to 1. A compared method is dominated by a curve if the curve passes above it at the same similarity to the MRI, that is, if the curve keeps more of the second source without giving up anatomical similarity. Figure 7 shows the curves and Table 6 counts the dominated methods.

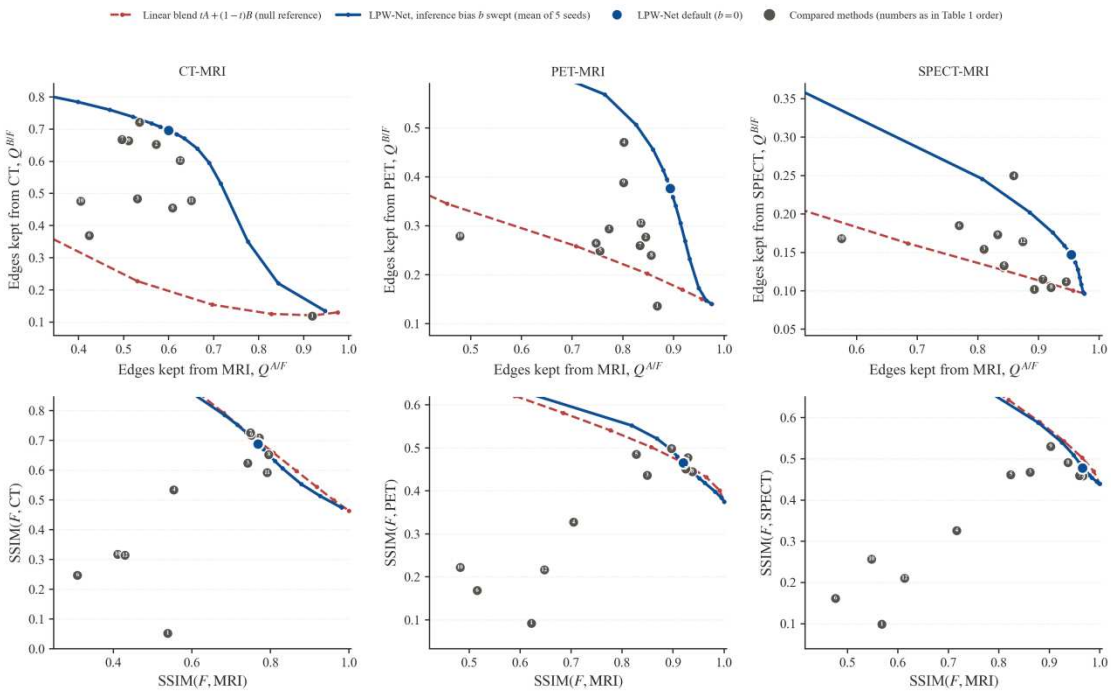


Figure 7. Trade-off between the information kept from the MRI (x-axis) and from the second source (y-axis) on test set B. Top: edge preservation; bottom: SSIM. The LPW-Net curve is traced with the inference bias b; the large dot is the default $b = 0$. The dashed line is the linear blend of the two sources. Grey dots are the compared methods, numbered in the order of Table 1 (1 MATR, 2 CDDFuse, 3 EMMA, 4 GeSeNet, 5 LRFNet, 6 MMDRFuse, 7 TTD, 8 BSAFusion, 9 DM-FNet, 10 GIFNet, 11 MACTFusion, 12 ReFusion).

Table 6. Table 6. Trade-off analysis on test set B. Each entry counts the compared methods that lie below the curve, i.e. for which the curve keeps more information from the second source at the same similarity to the MRI. The LPW-Net curve is traced with the inference bias b (mean of five seeds); the linear blend $tA + (1 - t)B$ is a null reference. Methods outside the x-range of a curve are not counted.

	Edge plane ($Q^{A/F}$, $Q^{B/F}$)		SSIM plane	
Task	LPW-Net curve	linear blend	LPW-Net curve	linear blend
CT-MRI	11/11	1/11	4/8	6/8
PET-MRI	11/11	2/11	8/11	9/11
SPECT-MRI	10/11	3/11	11/11	11/11

In the edge plane, the LPW-Net curve passed above 11 of 11 compared methods on CT–MRI and PET–MRI and above 10 of 11 on SPECT–MRI, whereas the linear blend passed above only 1, 2 and 3 of them. The edge advantage of LPW-Net is therefore not a property of mixing the sources as such, but of where it mixes them. In the SSIM plane the LPW-Net curve passed above 4 of 8, 8 of 11 and 11 of 11 methods, close to the linear blend; the gain of LPW-Net lies in the preservation of edges, not of local luminance statistics. The curves also locate the default operating point. On SPECT–MRI it lay close to the MRI end ($Q^{A/F}$ of 0.954, $Q^{B/F}$ of 0.147), so the luminance followed the MRI in most regions and the functional information was carried mainly by the restored chrominance.

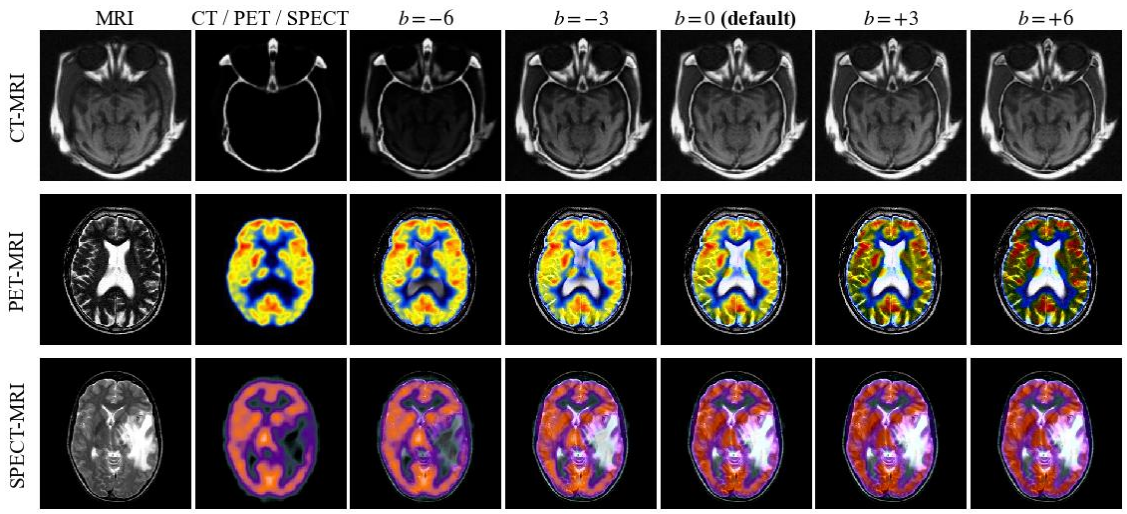


Figure 8. Effect of the inference bias b on LPW-Net (seed 0) for the three pairs of Figs. 3–5. Negative values lean toward the CT or functional image, positive values toward the MRI. PET and SPECT results show the fused luminance with the chrominance of the functional image.

Figure 8 shows what a move along the curve does to the images. A negative bias let the CT bone or the functional luminance dominate, a positive bias returned to the texture of the MRI, and the colour coding of the functional image was kept at every setting. A reader who prefers more functional detail can therefore obtain it without retraining.

5.3. Learned Weight Maps

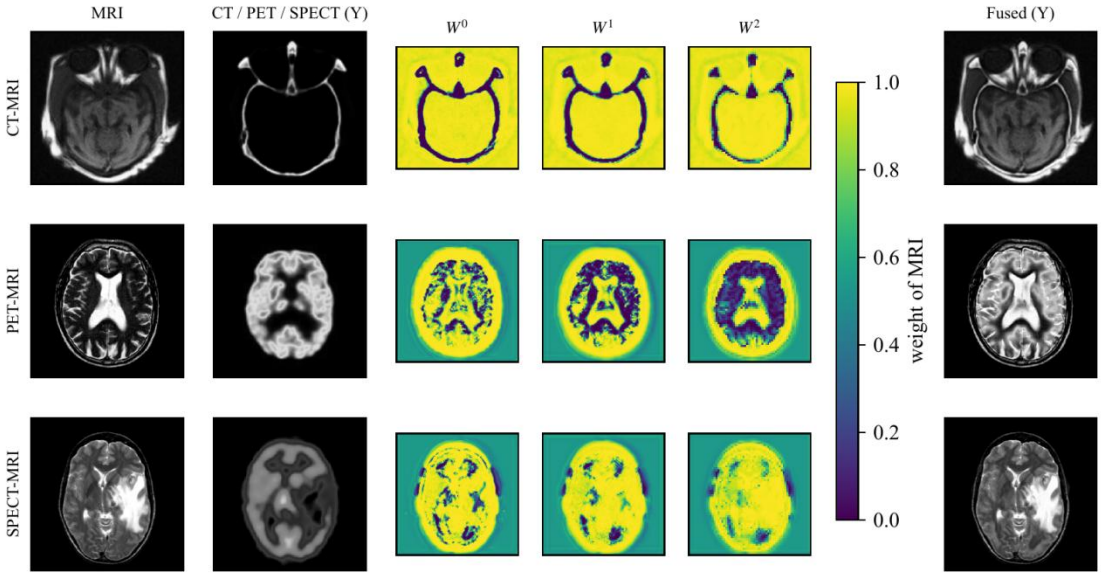


Figure 9. Weight maps W^0 (finest) to W^2 (coarsest) of LPW-Net (seed 0) for the three pairs of Figs. 3–5. Yellow trusts the MRI, dark blue the CT or functional image. In the background both sources are zero, so the weight there has no effect.

Because LPW-Net outputs weights rather than pixels, its decisions can be inspected directly (Fig. 9). For CT–MRI the weights trusted the MRI almost everywhere and switched to the CT exactly on the bone, at every level. For PET–MRI the coarsest level handed most of the brain interior to the PET, and the finer levels returned to the MRI along tissue boundaries, which is the coarse-to-fine division of labour the design intends. For SPECT–MRI the weights favoured the MRI except in a few regions of high uptake, consistent with the operating point discussed above. No supervision on the weights was given; this structure emerged from the unsupervised objective.

5.4. Efficiency

Table 7. Table 7. Inference cost for one 256×256 *pair on one NVIDIA RTX 4090D (batch 1, 10 warm-up runs, mean of the timed runs, model forward only). Parameters as in Table 1.*

Method	Parameters	GPU time (ms)
MMDRFuse	113	0.09
LPW-Net (ours)	7,841	1.98
LRFNet	55,185	1.99
GeSeNet	240,585	3.32
ReFusion	59,856	4.66
GIFNet	743,032	12.67
EMMA	1,517,657	15.79
MATR	13,230	18.04
CDDFuse	1,188,272	24.14
BSAFusion	9,760,776	28.78
MACTFusion	571,035	31.61
TTD	1,188,272	39.28
DM-FNet	23,696,721	102.81

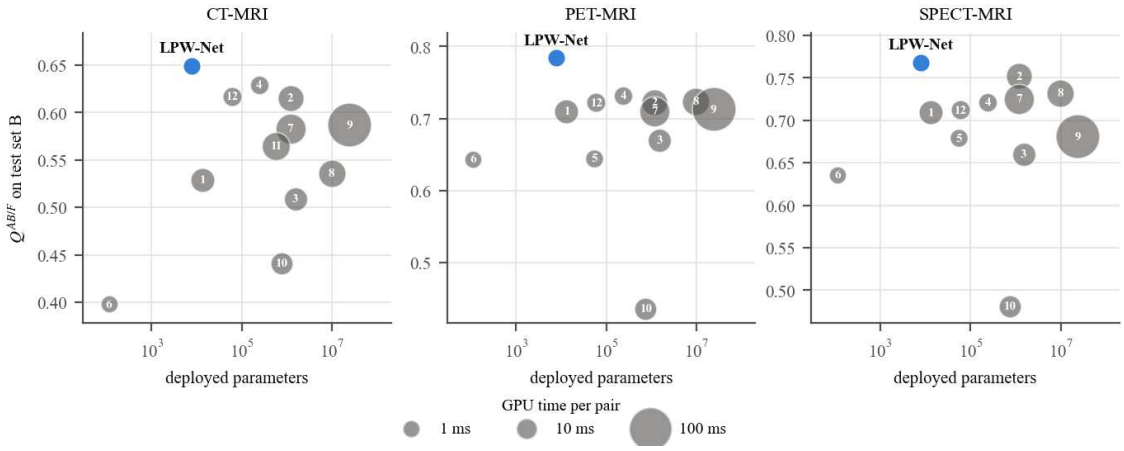


Figure 10. $Q^{AB/F}$ *on test set B against the number of deployed parameters (logarithmic axis). The marker area grows linearly with the GPU time per pair of Table 7. Numbers identify the compared methods in the order of Table 1; TTD (7) runs the network of CDDFuse (2) and therefore has the same parameter count.*

Table 7 reports the inference cost, and Fig. 10 sets it against $Q^{AB/F}$: in each task LPW-Net reached the highest $Q^{AB/F}$, and only MMDRFuse used fewer parameters. With 7,841 parameters and 0.67 GMACs per 256×256 pair, LPW-Net took 1.98 ms on the GPU, the second shortest time after MMDRFuse and on par with LRFNet, and 38.5 ms on a single CPU thread. Most compared networks were an order of magnitude slower or more, for example CDDFuse[5] with 24.14 ms and the diffusion-trained DM-FNet[6] with 102.81 ms. MMDRFuse[8] is smaller and faster, but it ranked eleventh or twelfth of twelve methods on $Q^{AB/F}$, Q_Y, Q_{CB} and SSIM in every task of Table 3, so for LPW-Net the size is a by-product of the formulation rather than the result of compressing a larger model.

5.5. Failure Modes and Limitations

The main failure mode of the weighting formulation, the collapse to the MRI, is removed by the logit penalty: with it, none of the five final models, none of the three ablation seeds and none of the twelve runs of the γ sweep collapsed.

A convex combination of source coefficients does not create contrast beyond the sources; whether a reader prefers higher contrast or higher fidelity to the sources is a clinical question that metrics cannot settle. All experiments used one atlas of registered brain images, PET and SPECT were fused on the luminance only, and the evaluation relied on objective metrics without a reader study by radiologists. Misregistered pairs, other anatomies and clinical acquisition protocols remain to be tested.

6. Conclusion

This paper treated multimodal medical image fusion as a decision about where to trust which modality rather than as image synthesis. LPW-Net decomposes both sources with a fixed Laplacian pyramid, predicts one weight map per level with a single coarse-to-fine network of 7,841 parameters, and reconstructs the fused image from convex combinations of the source coefficients, so it cannot create coefficients outside the range of the sources. We found that such weight predictors tend to collapse to the MRI through saturated logits, removed the collapse with a logit penalty, and used the well-conditioned logits for an inference-time bias that moves the result between the two sources without retraining. On the public test split of the Harvard atlas, LPW-Net ranked first on the structure-based metrics in all three tasks at a cost of 1.98 ms per pair, and in the edge-preservation plane its trade-off curve passed above nearly all of the eleven deep methods compared in each task, whereas a linear blend of the sources did not. Future work should test the method on other anatomies and on clinical data with registration errors, report per-source measures next to averaged metrics, and assess the fused images in a reader study with radiologists.

CRediT authorship contribution statement

Xie Qiang: Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **He Kai**: Writing – review & editing, Supervision, Resources, Conceptualization. **Liu Shun**:Resources, Methodology, Investigation, Formal analysis.

Ethical approval

Not applicable.

Data and code availability statement

The data generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the High-level (High-skilled) Talent Research Startup Fund (QDJZ-2026-6).

Declaration of competing interest

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.

References

1. Hermessi H, Mourali O, Zagrouba E. Multimodal medical image fusion review: theoretical background and recent advances. Signal Process. 2021;183:108036. https://doi.org/10.1016/j.sigpro.2021.108036.

2. Zhu Z, Zheng M, Qi G, Wang D, Xiang Y. A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain. IEEE Access. 2019;7:20811–20824. https://doi.org/10.1109/ACCESS.2019.2898111.

3. Liu Y, Chen X, Ward RK, Wang ZJ. Image fusion with convolutional sparse representation. IEEE Signal Process Lett. 2016;23:1882–1886. https://doi.org/10.1109/LSP.2016.2618776.

4. Zhang Y, et al. IFCNN: a general image fusion framework based on convolutional neural network. Inf Fusion. 2020;54:99–118. https://doi.org/10.1016/j.inffus.2019.07.011.

5. Zhao Z, et al. CDDFuse: correlation-driven dual-branch feature decomposition for multi-modality image fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023:5906–5916. https://doi.org/10.1109/CVPR52729.2023.00572.

6. He D, Li W, Wang G, Huang Y, Liu S. DM-FNet: unified multimodal medical image fusion via diffusion process-trained encoder-decoder. IEEE Trans Multimed. 2025;27:9415–9428. https://doi.org/10.1109/TMM.2025.3613156.

7. Cheng C, et al. One model for ALL: low-level task interaction is a key to task-agnostic image fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025:28102–28112. https://doi.org/10.1109/CVPR52734.2025.02617.

8. Deng Y, Xu T, Cheng C, Wu XJ, Kittler J. MMDRFuse: distilled mini-model with dynamic refresh for multi-modality image fusion. In: Proceedings of the 32nd ACM International Conference on Multimedia (MM '24). 2024:7326–7335. https://doi.org/10.1145/3664647.3681085.

9. Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. IEEE Trans Commun. 1983;31:532–540. https://doi.org/10.1109/TCOM.1983.1095851.

10. Ding X, et al. RepVGG: making VGG-style ConvNets great again. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021:13728–13737. https://doi.org/10.1109/CVPR46437.2021.01352.

11. Xydeas CS, Petrović V. Objective image fusion performance measure. Electron Lett. 2000;36:308–309. https://doi.org/10.1049/el:20000267.

12. Liu Z, et al. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Trans Pattern Anal Mach Intell. 2012;34:94–109. https://doi.org/10.1109/TPAMI.2011.109.

13. Yin M, Liu X, Liu Y, Chen X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. IEEE Trans Instrum Meas. 2019;68:49–64. https://doi.org/10.1109/TIM.2018.2838778.

14. Liu Y, Chen X, Ward RK, Wang ZJ. Medical image fusion via convolutional sparsity based morphological component analysis. IEEE Signal Process Lett. 2019;26:485–489. https://doi.org/10.1109/LSP.2019.2895749.

15. Li W, Xie Y, Zhou H, Han Y, Zhan K. Structure-aware image fusion. Optik. 2018;172:1–11. https://doi.org/10.1016/j.ijleo.2018.06.123.

16. Liu Y, Chen X, Cheng J, Peng H. A medical image fusion method based on convolutional neural networks. In: Proceedings of the 20th International Conference on Information Fusion (FUSION). 2017:1–7. https://doi.org/10.23919/ICIF.2017.8009769.

17. Xu H, Ma J, Jiang J, Guo X, Ling H. U2Fusion: a unified unsupervised image fusion network. IEEE Trans Pattern Anal Mach Intell. 2022;44:502–518. https://doi.org/10.1109/TPAMI.2020.3012548.

18. Ma J, et al. SwinFusion: cross-domain long-range learning for general image fusion via Swin transformer. IEEE/CAA J Autom Sin. 2022;9:1200–1217. https://doi.org/10.1109/JAS.2022.105686.

19. Tang W, He F, Liu Y, Duan Y. MATR: multimodal medical image fusion via multiscale adaptive transformer. IEEE Trans Image Process. 2022;31:5134–5149. https://doi.org/10.1109/TIP.2022.3193288.

20. Zhao Z, et al. DDFM: denoising diffusion model for multi-modality image fusion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023:8048–8059. https://doi.org/10.1109/ICCV51070.2023.00742.

21. Zhao Z, et al. Equivariant multi-modality image fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024:25912–25921. https://doi.org/10.1109/CVPR52733.2024.02448.

22. Bai H, et al. ReFusion: learning image fusion from reconstruction with learnable loss via meta-learning. Int J Comput Vis. 2025;133:2547–2567. https://doi.org/10.1007/s11263-024-02256-8.

23. Cao B, Xia Y, Ding Y, Zhang C, Hu Q. Test-time dynamic image fusion. In: Advances in Neural Information Processing Systems. Vol 37. 2024:2080–2105. https://doi.org/10.52202/079017-0067.

24. Li J, Liu J, Zhou S, Zhang Q, Kasabov NK. GeSeNet: a general semantic-guided network with couple mask ensemble for medical image fusion. IEEE Trans Neural Netw Learn Syst. 2024;35:16248–16261. https://doi.org/10.1109/TNNLS.2023.3293274.

25. Li H, Su D, Cai Q, Zhang Y. BSAFusion: a bidirectional stepwise feature alignment network for unaligned medical image fusion. Proc AAAI Conf Artif Intell. 2025;39:4725–4733. https://doi.org/10.1609/aaai.v39i5.32499.

26. He D, Li W, Wang G, Huang Y, Liu S. LRFNet: a real-time medical image fusion method guided by detail information. Comput Biol Med. 2024;173:108381. https://doi.org/10.1016/j.compbiomed.2024.108381.

27. Xie X, et al. MACTFusion: lightweight cross transformer for adaptive multimodal medical image fusion. IEEE J Biomed Health Inform. 2025;29:3317–3328. https://doi.org/10.1109/JBHI.2024.3391620.

28. Yi X, et al. LUT-Fuse: towards extremely fast infrared and visible image fusion via distillation to learnable look-up tables. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2025:14559–14568. https://doi.org/10.1109/ICCV51701.2025.01351.

29. Ding X, Zhang X, Han J, Ding G. Diverse branch block: building a convolution as an inception-like unit. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021:10881–10890. https://doi.org/10.1109/CVPR46437.2021.01074.

30. Zhang X, Zeng H, Zhang L. Edge-oriented convolution block for real-time super resolution on mobile devices. In: Proceedings of the 29th ACM International Conference on Multimedia. 2021:4034–4043. https://doi.org/10.1145/3474085.3475291.

31. Zhang H, Cao L, Ma J. Text-DiFuse: an interactive multi-modal image fusion framework based on text-modulated diffusion model. In: Advances in Neural Information Processing Systems. Vol 37. 2024:39552–39572. https://doi.org/10.52202/079017-1250.

32. Cao B, Xu X, Zhu P, Wang Q, Hu Q. Conditional controllable image fusion. In: Advances in Neural Information Processing Systems. Vol 37. 2024:120311–120335. https://doi.org/10.52202/079017-3823.

33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13:600–612. https://doi.org/10.1109/TIP.2003.819861.

34. ITU-R. Recommendation ITU-R BT.601-7: studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. ITU; 2011. https://www.itu.int/rec/R-REC-BT.601.

35. Johnson KA, Becker JA. The whole brain atlas. Harvard Medical School; 1999. http://www.med.harvard.edu/AANLIB/.

36. Gu X, et al. Adaptive spatial and frequency experts fusion network for medical image fusion. Biomed Signal Process Control. 2024;96:106478. https://doi.org/10.1016/j.bspc.2024.106478.

37. Li X, et al. Laplacian redecomposition for multimodal medical image fusion. IEEE Trans Instrum Meas. 2020;69:6880–6890. https://doi.org/10.1109/TIM.2020.2975405.

38. Shreyamsha Kumar BK. Image fusion based on pixel significance using cross bilateral filter. Signal Image Video Process. 2015;9:1193–1204. https://doi.org/10.1007/s11760-013-0556-9.

39. Liu Y, Wang Z. Simultaneous image fusion and denoising with adaptive sparse representation. IET Image Process. 2015;9:347–357. https://doi.org/10.1049/iet-ipr.2014.0311.

40. 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 Workshops (CVPRW). 2020:468–478. https://doi.org/10.1109/CVPRW50498.2020.00060.

41. Roberts JW, van Aardt JA, Ahmed FB. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. J Appl Remote Sens. 2008;2:023522. https://doi.org/10.1117/1.2945910.

42. Eskicioglu AM, Fisher PS. Image quality measures and their performance. IEEE Trans Commun. 1995;43:2959–2965. https://doi.org/10.1109/26.477498.

43. Cui G, Feng H, Xu Z, Li Q, Chen Y. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. Opt Commun. 2015;341:199–209. https://doi.org/10.1016/j.optcom.2014.12.032.

44. Qu G, Zhang D, Yan P. Information measure for performance of image fusion. Electron Lett. 2002;38:313–315. https://doi.org/10.1049/el:20020212.

45. Hossny M, Nahavandi S, Creighton D. Comments on ‘information measure for performance of image fusion’. Electron Lett. 2008;44:1066–1067. https://doi.org/10.1049/el:20081754.

46. Yang C, Zhang JQ, Wang XR, Liu X. A novel similarity based quality metric for image fusion. Inf Fusion. 2008;9:156–160. https://doi.org/10.1016/j.inffus.2006.09.001.

47. Chen Y, Blum RS. A new automated quality assessment algorithm for image fusion. Image Vis Comput. 2009;27:1421–1432. https://doi.org/10.1016/j.imavis.2007.12.002.

48. Han Y, Cai Y, Cao Y, Xu X. A new image fusion performance metric based on visual information fidelity. Inf Fusion. 2013;14:127–135. https://doi.org/10.1016/j.inffus.2011.08.002.

49. Aslantas V, Bendes E. A new image quality metric for image fusion: the sum of the correlations of differences. AEU Int J Electron Commun. 2015;69:1890–1896. https://doi.org/10.1016/j.aeue.2015.09.004.