Medical Image Fusion (MRI/PET/CT) with Deep Learning: A Step-by-Step Guide

Medical image fusion combines information from multiple imaging modalities (e.g. CT, MRI, PET) into a single image, capturing complementary details. For example, CT provides high-resolution **structural** detail (bones), MRI adds **soft-tissue** contrast, and PET shows **functional/metabolic** activity. Fusing them can produce an image “with better diagnostic value” than any single source [1] [2]. The fused image retains both anatomical structures and functional signals, supporting more accurate diagnosis.

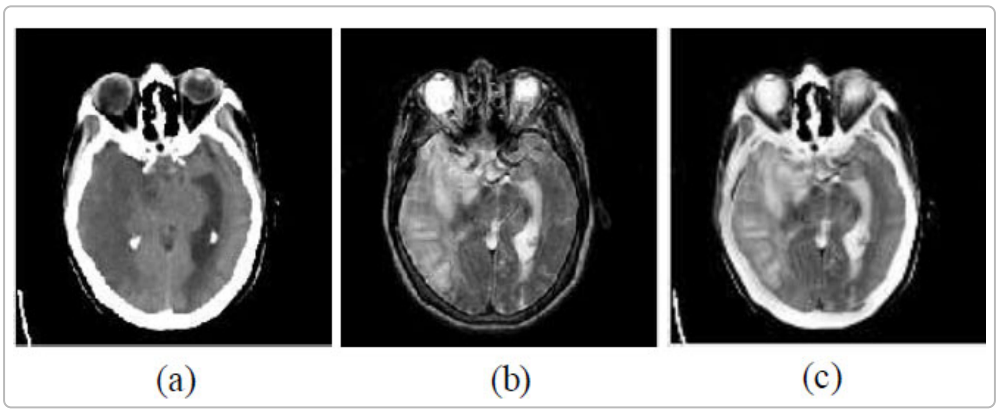

Figure: Example of CT+MRI fusion. (a) Original CT slice, (b) original MRI slice, (c) fused output combining bone detail and soft tissue contrast [1] [3].

Key Deep-Learning Approaches

Modern fusion methods use deep neural networks to **learn** how to merge modalities. Common strategies include:

- **CNN/Autoencoder-based Fusion:** Many models use encoder–decoder CNNs to extract features and reconstruct a fused image. Early works like *DenseFuse* and *IFCNN* used pre-trained autoencoders to encode each modality and merge their features [4]. Fusion architectures range from simple U-Nets to more complex densely-connected nets. CNN-based fusion can operate at the pixel, feature, or decision level [5] [6]. For example, FusionDN (Xu *et al.*, 2020) is an unsupervised densely-connected CNN that fuses images for multiple tasks [6].
- **Generative/Adversarial Models:** GANs treat fusion as a game between a **generator** (creates fused image) and **discriminator**. This can help preserve texture and detail. For instance, Fu *et al.* (2021) used a **dual-stream adversarial GAN** (DSAGAN) to fuse CT/MRI while retaining edges. GAN-based fusion enforces realism but requires careful training (no direct ground truth fused image).
- **Transformer/Attention-based Fusion:** Recent methods incorporate self-attention or Vision Transformers to model long-range dependencies between modalities. These can adaptively weight

features from each input (e.g. Tang *et al.*, Xie *et al.*). Transformers capture global context, at the cost of higher compute. They are an active research area in fusion [6].

- **Diffusion Models:** Very recent work uses *denoising diffusion* processes. For example, He *et al.* (2024) propose **DM-FNet**, a two-stage diffusion-based fusion network [7]. Stage I trains a UNet via progressive denoising (creating rich feature representations); Stage II inputs noisy images at multiple diffusion steps into a fusion network. This "diffusion-trained" encoder captures multiscale details. DM-FNet reported *excellent* objective metrics and high-quality fused images, with code available online [7] [8].
- **Hybrid/Ensemble Models:** Some recent papers combine architectures. Allapakam & Karuna (2024) built an **ensemble** of a pre-trained VGG-19 and a Siamese CNN [9]. Their stacked model leverages VGG for general features and a CNN branch for modality-specific detail, yielding fused MRI/PET images with improved contrast and resolution [9] [10]. Such hybrid designs often outperform classical wavelet or PCA-based fusion methods.
- **Task-Specific Pipelines:** In some studies (e.g. Song *et al.*, 2021), fusion is done via preprocessing rather than a single network. Song *et al.* segmented MRI gray matter and **mask-fused** it with PET to emphasize regions critical for Alzheimer's diagnosis. They created a "GM-PET" composite and fed it to 3D CNN classifiers [11]. This pipeline (registration → segmentation → fusion → CNN) improved AD detection on ADNI data [11]. It shows that careful preprocessing (skull stripping, registration, segmentation) can aid fusion tasks.

Each approach has trade-offs (data needs, compute, interpretability) [12] [7]. As a beginner, focus on understanding one method at a time (e.g., start with a simple autoencoder or CNN fusion, then explore GANs or diffusion).

Available Brain Imaging Datasets

To train and test fusion models, you need **paired images** of the same subjects. Important multimodal brain datasets include:

- **ADNI (Alzheimer's Disease Neuroimaging Initiative):** A large public dataset with longitudinal MRI and FDG-PET scans (plus other biomarkers). Widely used for research. (Access via ADNI website after registration.) Song *et al.* used ADNI for MRI+PET fusion [13].
- **OASIS-3:** Over 1300 subjects, including healthy controls and Alzheimer's patients. Contains thousands of MRI sessions (T1, T2, FLAIR, etc.) and PET sessions (amyloid tracers PIB/AV45 and FDG), plus CT scans [14]. Data access is by request. This is an excellent source of paired MRI+PET brain scans (memory tasks). For example, OASIS-3 has 2842 MRI scans and 2157 PET scans [14].
- **CERMEP-IDB-MRXFDG (EJNMMI Research 2021):** A *multimodal brain database* of 37 healthy adults. Each subject has a T1-weighted MRI, FLAIR MRI, FDG-PET, and CT (head) scan [15]. The images are co-registered and BIDS-formatted. Researchers can request access. This is ideal for testing MRI+PET/CT fusion methods; Song *et al.* note the PET/CT component explicitly [15].
- **PET/MR Brain (NYU CAI2R):** An open dataset of **simultaneously acquired** brain FDG-PET and MRI (MPRAGE) data, including coil sensitivity and attenuation maps [16]. It supports joint PET/MR reconstruction (Knoll *et al.* 2017). You must request it via the NYU CAI2R site, but it provides true aligned PET/MR scans. (Helpful URL in Resources.)
- **BrainWeb (Simulated Data):** A simulated brain MRI database with ground truth. It offers high-quality 3D volumes of normal and MS-affected brains with T1, T2, and PD MRI sequences [17]. While

BrainWeb has no PET/CT, you can use its MRI for practicing fusion pipelines (e.g. fusing simulated MR contrasts).

- **Kaggle Datasets:** Various *brain MRI* collections exist on Kaggle (e.g., Harvard brain tumor MRIs, LGG tumor segmentation). While not fused modalities, you could simulate fusion by pairing slices. Kaggle also hosts "Harvard Medical Image Fusion" collections (CT-MRI, PET-MRI folders) referenced by Allapakam *et al.* [10]. Kaggle's LGG MRI segmentation dataset is often cited for brain MRI (though single modality).

When gathering data, ensure **multimodal alignment**: all images for a subject must be in the same coordinate space. Datasets like OASIS-3 and CERMEP provide coregistered scans. Otherwise, you will need to register images yourself (see next steps).

Step-by-Step Project Plan

1. **Literature Review:** Study foundational papers to understand fusion techniques. Key surveys and examples include: Wang *et al.* (2023) "Multimodal fusion: deep learning perspective" [2] [6], Allapakam & Karuna (2024) for a Siamese+VGG ensemble [9], and He *et al.* (2024) for the diffusion-based DM-FNet [7]. Review CNN vs GAN vs Transformer fusion methods [6]. For hands-on work, identify papers with code or pseudocode (DM-FNet's GitHub is public [18]).

2. **Data Acquisition:** Choose one or more suitable datasets. For a brain fusion project, focus on MRI+PET or MRI+CT pairs: e.g. request ADNI or OASIS-3 data (MRI+PET), or use the CERMEP dataset (healthy brain MR, PET, CT) [15] [14]. Download scans and any available labels. If needed, supplement with BrainWeb simulated MR or Kaggle MRI for additional examples. Keep data organized per subject and modality.

3. **Preprocessing:** Register and normalize the images so that modalities align voxel-by-voxel. Typical steps:

4. **Skull-stripping & Denoising:** Remove non-brain tissue (tools: FSL BET, FreeSurfer, or HD-BET).
5. **Spatial Registration:** Use rigid/affine registration (e.g. ANTs, Elastix, SimpleITK) to align all modalities to a common space (e.g. MNI). Song *et al.* first registered MRI to MNI, then mapped PET to the MRI space [19].
6. **Resampling:** Match voxel dimensions and image size (often to 1mm isotropic).
7. **Intensity Normalization:** Scale intensities (e.g. 0–1 or z-score) for each modality.
8. **Masking (optional):** If focusing on brain tissue, apply a brain mask. For Song *et al.*, they segmented MRI into gray matter and fused only that region with PET [19]. At minimum, ensure only overlapping region is fused.

These steps ensure the input images are well-aligned and comparable. The research pipeline in [19] illustrates registration of MRI to MNI space, segmentation of MRI tissue, and mapping PET into that space for fusion.

1. **Model Selection & Implementation:** Decide on a fusion model to implement (in PyTorch). Possibilities include:

2. **Simple CNN Autoencoder:** Build an encoder-decoder network where you concatenate (or add) features from MRI and PET in the bottleneck. You can modify existing fusion code (e.g. adjust a DenseFuse architecture).
3. **UNet-based Fusion:** Use a UNet that takes both modalities as input channels and outputs a fused image.
4. **GAN:** Implement a generator network and discriminator to train via adversarial loss (e.g. a Pix2Pix-style setup where inputs are two channels, output one channel).
5. **Diffusion Model:** Try DM-FNet's approach: train a UNet with diffusion noise schedule (Stage I), then use noisy inputs in a second network (Stage II) [7] . The authors released code for reference [18] .
6. **Ensemble/Hybrid:** Combine pre-trained models (e.g. a frozen VGG to extract features + a small trainable CNN to fuse). You could replicate Allapakam's Siamese+VGG ensemble [9] .

Use PyTorch (with GPU) for all models. If code is available (e.g. DM-FNet GitHub [18] or related projects on GitHub), study and adapt it. Otherwise, implement from scratch: define data loaders that output (MRI, PET) pairs and the target fused image (if supervised) or just two-channel input (if unsupervised).

1. **Training the Model:** Train on your dataset using a suitable loss. Common losses for fusion include: pixel-wise L1/L2 loss between fused and (pseudo)ground-truth, structural similarity (SSIM) loss, gradient preservation loss, and adversarial loss (if using GAN). If no true fused ground-truth exists, you can train in a self-supervised way (e.g. minimize differences with each source, or use a discriminator to enforce realism). Use metrics like SSIM, PSNR, mutual information (Q_MI) as monitoring signals. Allapakam *et al.* used SSIM, PSNR, MSE to evaluate fusion quality [20] . Train until convergence; use early stopping or validation to prevent overfitting.

2. **Evaluation:** After training, produce fused outputs on test images. Evaluate both qualitatively (visual inspection: check that fused images show details from each modality) and quantitatively. Common quantitative metrics are **SSIM**, **PSNR**, **mutual information** (Q_MI), and task-specific metrics (if fusion is followed by classification). Compare your fused images to standard fusion baselines (simple averaging, PCA, wavelet fusion) as in many papers [10] . For a clinical task (e.g. tumor detection or AD classification), you can feed the fused images into a classifier and measure accuracy/AUC to assess usefulness.

3. **Iteration and Refinement:** Based on results, tweak your approach. Possible refinements: adjust network depth, change loss weights, use attention modules, or incorporate domain knowledge (e.g. focus loss on regions of interest). Pre-training on related tasks (transfer learning) can help if data is scarce. Ensure your fused output preserves clinically relevant features (consult clinicians if possible). Throughout, cite inspiration from the literature (e.g. *DenseFuse*, *U2Fusion*, *CoCoNet*, etc.) to guide improvements.

Useful Resources

- **Code and Libraries:** PyTorch for deep models; NiBabel/SimpleITK for medical image I/O and preprocessing; MONAI or TorchIO for medical imaging pipelines. Many fusion papers publish code (e.g. DM-FNet on GitHub [18] , U2Fusion code on GitHub).
- **Datasets:** ADNI and OASIS require registration but are large. The CERMEP dataset and NYU PET/MR data have request forms but are accessible to researchers. Kaggle and BrainWeb are open. Use these to get enough examples.

- **Papers to Replicate/Reference:** Key papers include: Song *et al.* (2021) for a CNN pipeline on MRI+PET [11], Allapakam & Karuna (2024) for an ensemble fusion model [9], and He *et al.* (2024) for diffusion-based fusion [7]. Also see reviews like Wang *et al.* (2023) for broad context [6].

By following these steps – studying the literature, obtaining aligned multimodal data, preprocessing carefully, implementing a deep fusion model in PyTorch, and iterating – you can build a complete medical image fusion project. Each cited paper above provides implementation details and results you can aim to replicate or improve upon [11] [9] [7].

Sources: We draw on recent reviews and research in multimodal image fusion [1] [6] and exemplar fusion methods (Song *et al.* for MRI/PET [11], Allapakam *et al.* for deep fusion [9], He *et al.* for diffusion models [7]). Datasets are described in ADNI/OASIS documentation and publications [15] [14].

---

[1] [3] [5] [12] A Review of Deep Learning-based Multi-modal Medical Image Fusion
https://openbioinformaticsjournal.com/VOLUME/18/ELOCATOR/e18750362370697/FULLTEXT/

[2] [6] A dual-stream feature decomposition network with weight transformation for multi-modality image fusion | Scientific Reports
https://www.nature.com/articles/s41598-025-92054-0?error=cookies_not_supported&code=2f278c88-1a38-403a-baa6-811b19976da7

[4] [7] [8] [18] DM-FNet: Unified multimodal medical image fusion via diffusion process-trained encoder-decoder
https://arxiv.org/html/2506.15218v1

[9] [10] [20] An ensemble deep learning model for medical image fusion with Siamese neural networks and VGG-19 | PLOS One
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0309651

[11] [13] [19] An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer's Disease Diagnosis - PubMed
https://pubmed.ncbi.nlm.nih.gov/34713109/

[14] OASIS Brains
https://sites.wustl.edu/oasisbrains/home/oasis-3/

[15] CERMEP-IDB-MRXFDG: a database of 37 normal adult human brain [18F]FDG PET, T1 and FLAIR MRI, and CT images available for research | EJNMMI Research | Full Text
https://ejnmmires.springeropen.com/articles/10.1186/s13550-021-00830-6

[16] PET-MR Dataset • Center for Advanced Imaging Innovation and Research
https://cai2r.net/resources/pet-mr-dataset/

[17] BrainWeb: Simulated Brain Database
https://www.bic.mni.mcgill.ca/brainweb/