Displays 80 (2023) 102536

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

SOFNet: Optical-flow based large-scale slice augmentation of brain MRI ☆

Bingxin Wei [a,d], Yang Wen [c,a], Xiaohong Liu [d,a], Xing Qi [b,a], Bin Sheng [a,*]

[a] Department of Computer Scinece and Engineering, Shanghai Jiao Tong University, Shanghai, China
[b] Otolaryngology, HuadongSanatorium, Wuxi, China
[c] Shenzhen University, Shenzhen, China
[d] John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Data augmentation
Optical-flow
Interpolation

ABSTRACT

Brain magnetic resonance imaging (MRI) data from multiple centers often exhibit variations in imaging conditions, such as the types of nuclear magnetic resonance instruments used and the presence of random noise. Additionally, discrepancies in the gap between MRI slices further complicate the usability of the data for advanced artificial intelligence (AI) analysis. Deep learning-based methods have emerged as practical solutions to address the challenge. However, existing research has largely overlooked the augmentation of brain MRI data, particularly when confronted with significant slice gaps, such as around 6 mm observed in our clinical brain MRI slices. In response to this research gap, we aim to develop novel approaches for augmenting brain MRI data, focusing on more significant slice gaps. To achieve this, we propose SOFNet, which utilizes the optical flow-based and encoder–decoder backbone. The primary objective of our model is to interpolate MRI slices while preserving feature consistency. Leveraging the optical flow method, which has exhibited exceptional performance compared to other super-resolution algorithms, our proposed approach has been evaluated on three distinct brain MRI datasets, explicitly addressing the gap between 4.2 mm and 6.0 mm. The experimental results highlight the significant enhancement in super-resolution quality achieved by SOFNet in generating adapted brain MRI data, surpassing other single-image super-resolution (SISR) methods in feature completion. To ensure the credibility of the interpolated brain MRI slices, we conducted experiments on three axes of MRI based on metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). These experiments demonstrate the effectiveness of our approach in transforming low-resolution MRI data into clear and reliable brain MRIs, thereby enabling improved analysis using AI techniques.

1. Introduction

Accurate image analysis and quantitative measurement in medical imaging rely heavily on high spatial resolution, providing crucial structural details and reducing random noise. However, achieving high spatial resolution in MRI scans often has drawbacks such as longer scan times, limited spatial coverage, and lower signal-to-noise ratios (SNR) [1]. To overcome these limitations, a promising approach is reconstructing high-resolution (HR) images from low-resolution (LR) inputs. This reconstruction process can potentially enhance spatial coverage, improve SNR, and achieve better spatial resolution within a shorter scan time.

While deep learning-based methods have shown promising results, they face challenges when dealing with significant resolution gaps between slices, i.e., high resolution in axial view while low resolution in sagittal and coronal view(as shown in Fig. 1, brain MRI slices contain three axes, axial, sagittal and coronal respectively). This discrepancy in resolution across dimensions can lead to suboptimal performance, as the algorithm fails to utilize the available high-resolution information effectively. It is crucial to develop methods that leverage the three-dimensional nature of medical images and consider the relationships between adjacent slices to overcome these limitations. Those approaches can lead to more robust and accurate super-resolution results in brain MRI.

In our research, we have achieved improved results by designing an optical-flow-based encoder–decoder coarse-to-fine network. To enhance the extraction of optical flow and ensure the quality of generated intermediate slices, we utilize a cost volume module (CVM) to generate a representation of the cost volume. This cost volume representation, combined with intermediate feature estimation, enhances the accuracy and fidelity of the optical flow estimation process in our approach. Our

☆ This paper was recommended for publication by Prof Guangtao Zhai.

* Corresponding author.
E-mail addresses: wei-bx@sjtu.edu.cn (B. Wei), wen_yang@szu.edu.cn (Y. Wen), xiaohongliu@sjtu.edu.cn (X. Liu), 15106170926@163.com (X. Qi), shengbin@sjtu.edu.cn (B. Sheng).

https://doi.org/10.1016/j.displa.2023.102536
Received 31 July 2023; Received in revised form 25 August 2023; Accepted 28 August 2023
Available online 3 September 2023

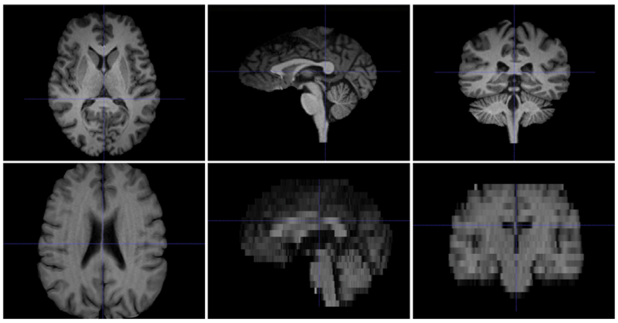

Fig. 1. Three-dimensional views of brain MRI images from the NFBS dataset [2]. The top row shows the original NFBS dataset images, while the bottom row displays the 6× downsampled data with a larger inter-slice gap of approximately 6 mm on axial axis.

work introduces an Optical-Flow generator specifically designed for augmenting brain MRI data, enabling the transfer of low-resolution 3D MRIs to high-resolution 3D MRIs. The main contributions are threefold:

(1) Our approach effectively generates coherent and informative slices, even with a substantial 6 mm gap in most clinical datasets. It showcases robustness in handling challenging scenarios in MRI super-resolution.
(2) We introduced the Cost Volume Module (CVM) for cost volume computation and flow estimation. CVM demonstrates great potential and effectiveness in our approach through the Windows-based computation method, significantly enhancing performance and enabling more accurate flow estimation.
(3) Our proposed method outperforms state-of-the-art approaches in MRI super-resolution, as demonstrated through extensive experiments conducted on both public datasets. These results highlight our approach's superior performance and efficacy compared to existing methods in the field.

2. Related work

We summarize two relevant areas: MRI super-resolution and optical flow frame interpolation, which are closely related to our work.

2.1. MRI super-resolution methods

Initially, interpolation methods such as bilinear, bicubic, and nearest neighbor interpolations were commonly employed as the most straightforward approaches to magnify a low-resolution (LR) slice, resulting in a smoothed version of the image. These methods aimed to enhance the resolution of the LR slice by estimating the missing information between pixels based on neighboring values. While these interpolation methods provided a basic level of upscaling, they often yielded images that needed more delicate details and sharpness. Indeed, various traditional methods, including patch-based super-resolution [3,4] have shown promising results in MRI super-resolution. These methods effectively enhance the resolution of LR images. However, one limitation they often need to improve is over-smoothing, which can result in the loss of intricate anatomical details.

Originating in the realm of natural imagery, deep learning-based single image super-resolution (SISR) methodologies have achieved prominence. Notably, convolutional neural network (CNN)-centered architectures, such as those posited in Refs. [5–7], have garnered substantial attention. Particularly noteworthy is the work by [7], rooted in the UNet architecture, while [5,6] rely on attention mechanisms. These instances underscore the utility of UNet and attention mechanisms in achieving proficient super-resolution.

In parallel, specific SISR techniques have exhibited efficacy in heightening the quality of medical images [8–10]. These methodologies, underpinning the medical imaging arena, provide substantial evidence of their impact. Compared to traditional algorithms, these SISR methods demonstrate superior performance, allowing for seamless transformation between low-resolution (LR) and high-resolution (HR) medical images. A common strategy involves combining super-resolution with CNNs, enabling an end-to-end mapping process. Zeng et al. [8] proposed a model that simultaneously addressed single and multi-contrast SR reconstruction, leveraging CNN's power to enhance image quality. To capture the cubic spatial features of MRI, Du et al. [9] introduced 3D dilated convolutions as an encoder, allowing for extracting high-frequency features and achieving impressive performance in SR. Building upon this model, Pham et al. [10] further developed an SRCNN algorithm tailored explicitly for brain MRI SR. By employing 3D convolutions, this network exhibited excellent performance in enhancing brain MRI images. The integration of CNNs with SR techniques has demonstrated promising results in MRI super-resolution, showcasing the potential of deep learning for enhancing image quality and fine details in this domain.

Moreover, generative adversarial network (GAN) architectures have been utilized for MRI super-resolution. For example, Chen et al. [11] applied a 3D dense GAN for MRI super-resolution. Wang et al. [12] introduced a novel memory-efficient residual-dense block generator for brain MRI super-resolution. You et al. [13] proposed the Fine Perceptive GAN (FP-GAN) to generate HR MRIs from corresponding LR counterparts. It is worth noting that the approaches mentioned above primarily focus on reconstructing images using a single contrast of MRIs. One notable concern in the utilization of generative adversarial network (GAN) architectures for medical image processing, particularly in the context of brain MRI super-resolution, is the potential for geometric distortion. Geometric inconsistencies or distortions pose significant challenges as they can compromise the accuracy and reliability of medical image analysis, making them highly undesirable.

2.2. Optical-flow frame interpolation

Traditionally, optical flow estimation was approached as an optimization problem that aimed to maximize the visual similarity between pairs of images while incorporating regularization terms [14–17]. Significant advancements in this field emerged through improved designs of similarity and regularization terms. However, the introduction of deep neural networks revolutionized optical flow estimation.

The pioneering work of FlowNet [18], an end-to-end convolutional network, marked a major milestone in optical flow estimation. Its successor, FlowNet2.0 [19], introduced a stacked architecture with a warping operation, achieving performance comparable to state-of-the-art methods. Subsequent models, such as SpyNet [20], PWC-Net [21,22], LiteFlowNet [23,24], and VCN [25], adopted coarse-to-fine and iterative estimation methodologies. However, these models often encountered difficulties accurately capturing small, fast-moving objects during the coarse stage. To address this limitation, Teed and Deng proposed RAFT [26], which performs optical flow estimation coarse and finely with a multi-scale search window in each iteration—building upon the RAFT architecture, subsequent works [27–30] aimed to either reduce computational costs or improve flow accuracy.

More recently, optical flow estimation has been extended to more challenging scenarios, such as low-light conditions [31], foggy environments [32], and variations in lighting [33]. These explorations typically involve computing visual similarity by correlating high-dimensional features encoded by convolutional neural networks. The cost volume, which contains pixel pairs' visual similarity, is a core component supporting optical flow estimation. However, the effectiveness of their cost information utilization has been identified as an area for improvement. IFRNet [34] is a recent advancement in optical flow estimation. It builds upon the iterative estimation methodology, aiming to refine

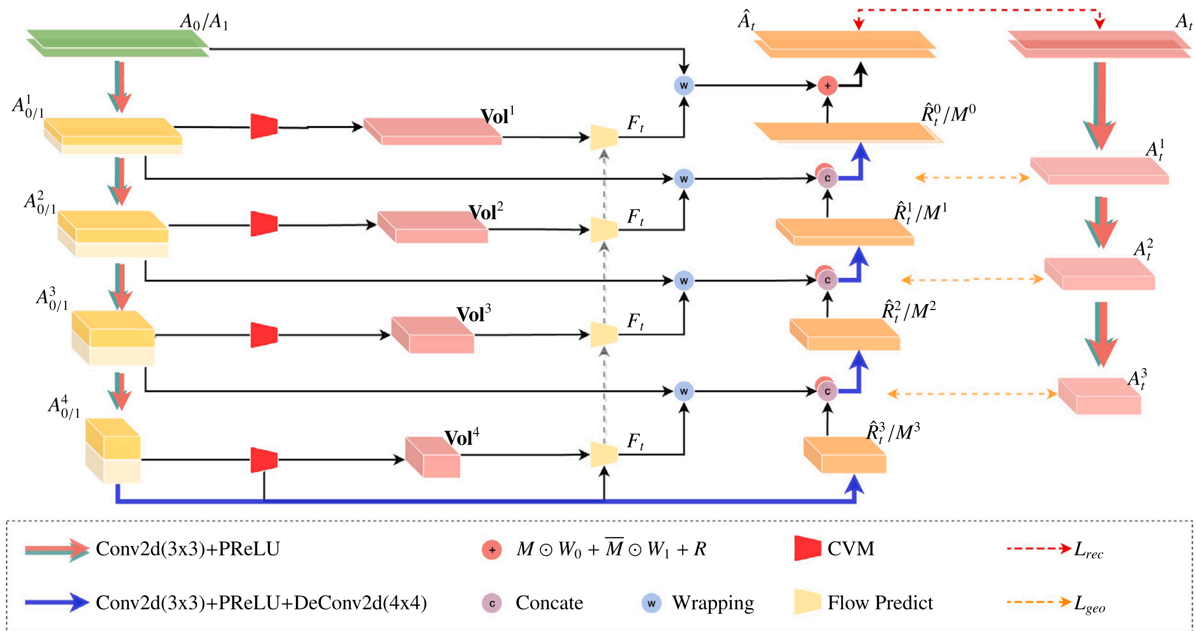

Fig. 2. Architecture overview and loss functions of the generator in SOFNet. The model adopts an efficient encoder–decoder design, extracting pyramid context features from input frames $A_{0/1}$ using a shared encoder. Coarse-to-fine decoders refine bilateral intermediate flow fields $F_{t\to 0}$ and $F_{t\to 1}$, along with reconstructed residual feature $\hat{R}_t^k$, to generate the final output. The generator is guided by image reconstruction loss (L_{rec}), feature space geometry consistency loss (L_{geo}).

flow estimation through successive iterations. By iteratively refining the estimated flow field, IFRNet achieves enhanced accuracy and captures fine details of correlation, making it a notable contribution to the field of optical flow estimation. EMA-VFI [35] is a method proposed to improve the accuracy of variational optical flow estimation. By leveraging exponential moving average (EMA) techniques, EMA-VFI enhances the robustness of flow inference by incorporating temporal information across consecutive frames. FlowFormer [36] is a recent innovation in optical flow estimation that utilizes a transformer-based architecture. Inspired by the success of transformers in various computer vision tasks, it adopts a self-attention mechanism to capture long-range dependencies and spatial relationships within image sequences. This attention mechanism allows FlowFormer to model complex correlation patterns accurately. FlowFormer achieves state-of-the-art performance in optical flow estimation, demonstrating its potential for advancing the field.

Through extensive experimentation, we have discovered that deep learning-based optical flow methods, traditionally used for 3D data with continuous change fields like video frame interpolations, can be effectively applied to medical image augmentation. In our research, we treat MRI slices as analogous to video frames and leverage optical flow estimation to approximate missing slices. This approach has demonstrated promising results, enabling the generation of interpolated MRI slices. By utilizing optical flow techniques, we aim to enhance the completeness and quality of MRI datasets, ultimately improving analysis and diagnosis in medical imaging applications

3. Method

This section presents the SOFNet architecture, designed based on the joint refinement of optical flow and UNet [37], serving as its backbone. This combination creates an efficient encoder–decoder framework tailored for brain MRI slice interpolation. Furthermore, we introduce the objective functions, including the network data flow, optical flow, and feature combined architecture. These components are crucial in guiding our model to achieve excellent performance in MRI slice interpolation.

3.1. SOFNet

We propose SOFNet to address the challenges of using optical flow in low-resolution (LR) 3D MRI to high-resolution (HR) 3D MRI data augmentation. Fig. 2 illustrates the workflow of our SOFNet model. Like a traditional UNet [37], SOFNet follows a similar processing procedure. However, to generate the intermediate flow representation, adjacent slices as dual inputs to the generator. This configuration allows the generator to produce corresponding frames in the target domain and interpolated slices of these two input slices.

Our approach incorporates the cost volume module (CVM) to capture hidden inter-slice movement in the MRI data. By leveraging the CVM, our SOFNet model can generate more realistic slices, resulting in improved performance metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index). Including the CVM enhances the accuracy and fidelity of the generated slices, ultimately leading to superior image quality assessment metrics.

In addition to the increased quality of the generated intermediate optical flow, we employ the Laplacian loss L_{warp} for flow refinement. Besides, the proposed SOFNet is supervised by multiple losses, feature estimation loss L_{geo}, and slice-interpolation reconstruction loss L_{rec}. In the following sections, we explain how we utilize inter-slice attention to extract features for MRI slice interpolation. We then delve into the structure of the whole pipeline.

The proposed model utilizes a simultaneous translation approach. It takes two images, A_0 and A_1, and translates them to context feature using the pyramid encoder. This process is illustrated as follows. In our experiments, we select the adjacent frame of each MRI slice as the paired frame for 6x augmentation. The model incorporates an encoder phase, extracting a pyramid of features from each frame. The refinement process follows a coarse-to-fine approach, gradually improving bilateral intermediate flow fields and the residual intermediate features. This refinement is achieved through a flow estimate deconvolution layer and a feature estimate deconvolution layer. The process continues until reaching the highest level of the pyramid, resulting in the final output. The overall architecture of our SOFNet model is depicted in Fig. 2.

In SOFNet, we employ an encoder with shared parameters for the paired input to extract contextual representation from the input frames A_0 and A_1. The encoder is composed of blocks consisting of 3×3 convolutions, along with additional 3×3 convolutions with strides of 1 and 2, respectively. This design results in four levels of features with decreasing spatial size and increasing channel dimensions (32, 48, 72, and 96), following the encoder architecture used in [34]. This multi-scale feature extraction allows SOFNet to efficiently capture contextual information and facilitate accurate flow estimation and high-quality intermediate frame synthesis. The extracted features can be represented as A_0^k for A_0 and A_1^k for A_1, where ($k \in \{1, 2, 3, 4\}$).

3.2. CVM

These explorations assess visual similarity by correlating high-dimensional features encoded by a convolutional neural network. The resulting cost volume, which represents the visual similarity between pairs of pixels, plays a crucial role in supporting optical flow estimation. Among the works, RAFT [26], PWC-Net [22], and Flow-former [36] employees the cost volume for flow estimation. However, the utilization of cost information in these approaches lacks effectiveness. We modified it and utilized it for MRI imaging augmentation. Fig. 3 illustrates how the inter-slice cost volume acquires inter-frame features. Now suppose we have the extracted feature of two extracted features, denoted as A_0 and $A_1 \in \mathbb{R}^{H\times W\times C}$. Specifically, we focus on obtaining the enhancing cost volume block of A_0 for brevity. For any region, which is denoted as $A_0^{i,j} \in \mathbb{R}^{N\times N\times C}$ in A_0, we use it and its closest spatial neighbors $A_1^{n_{i,j}} \in \mathbb{R}^{N\times N\times C}$ in A_1, where n represents the neighborhood window size(we set n=8 in our experiment), and (i, j) refers to the (ith, jth) window in A_0, to generate the query and keys respectively:

$$\mathbf{Q}_0^{i,j} = \mathbf{A}_0^{i,j}\mathbf{W}_Q \tag{1}$$

$$\mathbf{K}_1^{n_{i,j}} = \mathbf{A}_1^{n_{i,j}}\mathbf{W}_K \tag{2}$$

In this process, we employ linear projection matrices W_Q and W_K of dimensions $\mathbb{R}^{C\times\hat{C}}$ to perform a dot product between $\mathbf{Q}_{i,j}^0$ and each position of $\mathbf{K}_1^{n_{i,j}}$. This operation is followed by applying the softmax function, resulting in the generation of the attention map $S_{0\to1}^{i,j} \in \mathbb{R}^{N\times N}$. Each value in the attention map represents the degree of similarity between $A_0^{i,j}$ and its neighboring elements.

$$\mathbf{S}_{0\to1}^{n_{i,j}} = \mathbf{SoftMax}(\mathbf{Q}_0^{i,j}(\mathbf{K}_1^{n_{i,j}})^T/\sqrt{\hat{C}}) \tag{3}$$

Next, we create coordinate maps $\mathbf{M}^0$, $\mathbf{M}^\mathbf{1} \in \mathbb{R}^{H\times W}$ that represents the relative position of each location in the entire image. The values in $\mathbf{M}^0$ indicate the relative positions, ranging from 1 for the top left to −1 for the bottom right. While $\mathbf{M}^1$ is the transpose of $\mathbf{M}^0$. And $\mathbf{M} \in \mathbb{R}^{H\times W\times 2}$ is the concatenation of $\mathbf{M}^0$, $\mathbf{M}^\mathbf{1}$. As is illustrated in Fig. 3. Using the coordinate map, we weigh the neighboring coordinates to estimate the corresponding position of $\mathbf{A}_0^{i,j}$ in $\mathbf{I}_1$. The correlation vector $\mathbf{Vol}_{0\to1}^{i,j} \in \mathbb{R}^2$ of $\mathbf{A}_0^{i,j}$ is then generated by subtracting the original position of $\mathbf{A}_0^{i,j}$ from the estimated position in $\mathbf{I}_1$.

$$\mathbf{Vol}_{0\to1}^{i,j} = \mathbf{S}_{0\to1}^{n_{i,j}}\mathbf{M}^{n_{i,j}} \tag{4}$$

The information contained in $\mathbf{Vol}_{0\to1}^{i,j}$ serves as an explicit prior for flow estimation, while the correlation of A_1 can be represented by $\mathbf{Vol}_{1\to0}^{i,j}$. To generate the combined correlation feature, we pass $\mathbf{Vol}_{0\to1}^{i,j}$ through a 1×1 convolutional layer, but the number of channels is halved. This convolutional layer simplifies the calculation process and aids in extracting meaningful correlation features. The formulation for generating the combined correlation feature is as follows:

$$\mathbf{Vol}^{i,j} = \mathbf{Conv}(\mathbf{Cat}[\mathbf{Vol}_{0\to1}^{i,j}, \mathbf{Vol}_{1\to0}^{i,j}]) \tag{5}$$

Overall, with extracted feature A_0^k, A_1^k, we can calculate cost volume $\mathbf{Vol}^k$ in the above way, where $k \in \{1, 2, 3, 4\}$.

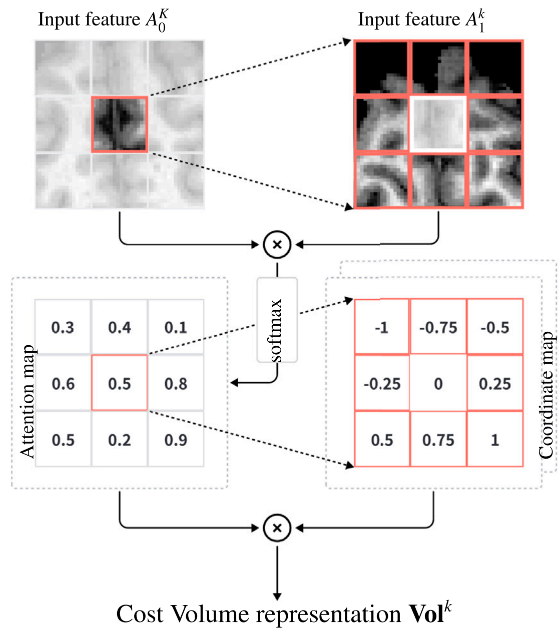

Fig. 3. Illustration of the computation of cost volume.

3.3. Coarse-to-fine flow/feature estimation

In the above section, we tried to calculate cost volume information through the module. We then gradually refine intermediate flow fields through 4 decoders by wrapping pyramid features following the idea of UNet [37] and IFRNet [34]. Compared to IFRNet, which estimated the intermediate feature and flow using the shared decoder, we calculate them separately, aiming for a more accurate flow expression. In addition, flow estimation can be done more manageable with the help of the calculated cost volume $\mathbf{Vol}^k$.

As depicted in Fig. 2, $F_{t\to0}^3$, $F_{t\to1}^3$ calculated by flow decoder by taken $\mathbf{Vol}^4$ as input. Moreover, $\hat{A}_t^3$ and M^3 can be calculated by feature decoder by taking A_0^4, A_0^4 and T as input, where T refers to time step hint, which is a one-channel matrix for slice interpolation. Except the above, the calculation of F^k and A_0^k where $k \in \{0, 1, 2\}$ can be formulated as:

$$F_{t\to0}^k, F_{t\to1}^k = \mathbf{Flow}(Cat[\mathbf{Vol}^{k-1}, F_{t\to0}^{k-1}, F_{t\to1}^{k-1}]) \tag{6}$$

$$\begin{aligned} &W_0, W_1 = w(A_0^{k-1}, F_{t\to0}^{k-1}), w(A_1^{k-1}, F_{t\to1}^{k-1}) \\ &R^k, M^k = \mathbf{Feature}(Cat[R_t^{k-1}, M^{k-1}, W_0, W_1]) \end{aligned} \tag{7}$$

The term **Flow** alludes to the flow prediction block, elucidated in Fig. 2. This entity encompasses a CNN block endowed with two 3×3 convolutional layers housing 16 hidden channels, in tandem with a subsequent PixelShuffle [42] operation featuring a downscale factor of $r = 2$. Conversely, the designation **Feature** signifies the residual feature estimation network. This constituent comprises a convolutional layer coupled with a deconvolutional layer, the latter characterized by a stride of 2 for effective upsampling. Additionally, the matrix M is obtained through a sigmoid function, ensuring values between 0 and 1. The output of these functions includes the estimated mask M^k, flows $F_{t\to1}^k$, $F_{t\to1}^k$, and the residual feature R^k, respectively. Finally, the calculation process for the intermediate compositing $\hat{A}_t^k$ is the same as IFRNet [34]:

$$\hat{A}_t^k = M^k \odot W_0 + \overline{M} \odot W_1 + R^k \tag{8}$$

In the above formulation, $\odot$ denotes element-wise multiplication. $\overline{M}$ represents the negation of M relative to 1, indicating the inverse importance or weight of the elements.

Table 1
Comparisons of Different Super-resolution Methods. The PSNR and SSIM on NFBS, IXI, and HCP by Current State-of-the-Art Methods based on SISR, Video-Frame interpolation, MRI/CT super-resolution, and Our SOFNet. We evaluated this method on IXI, HCP, and NFBS datasets. The best results are in bold.

Method	IXI		HCP		NFBS		Params (M)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
RCAN [38]	29.56	0.9391	27.50	0.8815	28.04	0.9391	15.3
EDSR [39]	26.92	0.8792	29.53	0.9375	27.11	0.9319	38.4
SwinIR [40]	27.43	0.8833	29.06	0.9288	24.40	0.8860	11.6
ELAN-light [41]	28.57	0.8834	31.21	0.9402	29.01	0.9336	**0.6**
IFRNet [34]	30.92	0.9340	33.24	**0.9410**	**29.80**	0.9423	5.0
SOFNet	**32.30**	**0.9477**	**34.36**	0.9366	29.50	**0.9449**	3.8

3.4. Loss function

We followed the idea of IFRNet [34], and the feature geometric alignment loss was conducted on SOFNet. For the input of $\hat{A}_0$, $\hat{A}_1$, A_t, we try to employ the same parameter shared encoder to extract a pyramid of features $\hat{A}_t^k$ from the ground truth A_t, and use it to regularize the reconstructed feature $\hat{A}_t^k$ in multi-scale feature representations.

$$L_{geo} = \sum_{k=1}^{3} L_{cen}(A_t^k, \hat{A}_t^k) \tag{9}$$

where L_{cen} refers to census loss, the soft Hamming distance is calculated between census-transformed corresponding feature maps with 3 × 3 patches in a channel-by-channel manner, which was first introduced by UnFlow [43].

To enhance the accuracy of optical flow estimation in the coarse-to-fine procedures, we introduce the warp loss L_{warp}, supervised by ground truth extracted features A^k. Instead of directly using L_1 loss across features, we conducted the $L1$ loss between the Laplacian pyramids of the warped frames and the ground truth, as introduced in previous work [44]. This loss function measures the discrepancy in high-frequency details and helps improve the fidelity and realism of the generated frames. This approach allows for implicit supervision of the flow estimation process. By directly comparing the warped features with the ground truth features, we encourage the model to generate more accurate optical flow fields that align with the desired features. This indirect supervision contributes to refining the optical flow estimation and improves the model's overall performance.

$$L_{warp} = \sum_{i=1}^{3} L_{lap}(\hat{A}_t^k, A_t^k) \tag{10}$$

In slice interpolation tasks, we also use a combination of L_1 and L_{cen} for inter-slice reconstruction alignment, denoted as L_{rec}. This approach further improves the accuracy of the intermediate slice reconstruction process. Finally, by combining the geometric loss, warp loss, and the feature coarse-to-fine loss, the entire objective function of the SOFNet generator is expressed as:

$$L_G = \lambda_g L_{geo} + \lambda_w L_{warp} + L_{rec} \tag{11}$$

where λ_g and λ_w control the importance of different terms, were setting to 0.01, 0.01 respectively.

4. Experiments

This section evaluates and discusses the performance of our SOFNet. Suffering from the problem of high-resolution MRI imaging is costly and time-consuming, and the conditions of MRI slices often have wide variations, which dramatically degrade the generalization of a computer-aid diagnosis system trained on a single domain. This section tackles the slice interpolation problem for MRI slices using our SOFNet. We test the super-resolution results evaluated by PSNR and SSIM. After that, we test the effect of the cost volume optical flow coarse-to-fine approach and the assistant warp loss through an ablation study.

4.1. Datasets preparation and training details

We evaluate our interpolation-based method on the following datasets covering diverse brain MRI domains for a comprehensive comparison. Common super-resolution metrics, such as PSNR and SSIM, are adopted to evaluate slice augmentation quantitatively. In our work, to maximize the results in domains that have large slice thickness over the axial as many clinical datasets, we processed HCP, NFBS, and IXI datasets over the axial to match the axial axis low-resolution MRI dataset's slice thickness. Detailed, we dropped the intermediate five slices every seven slices for downsampling. In this study, MRI data were obtained from multi-centers. All data of HCP and IXI datasets were skull stripped with the ITK software (https://itk.org/). T1-weighted MRI slices were adopted in our experiment. Then SOFNet was performed on IXI, NFBS, and HCP datasets, and PSNR/SSIM results were calculated.

(1) **NFBS** [2]: Composed of skull-stripped Brain MRI slices, features 125 samples with an in-plane resolution of 0.94 mm and a slice thickness of 1.19 mm. Our study employed 100 T1-weighted samples for model training, while the rest were reserved for testing and evaluation.
(2) **IXI** [45]: Contains 600 MRIs from normal, T1-weighted healthy subjects, boasting an in-plane resolution of 0.94 mm and a slice thickness of 1.2 mm. Our study employed 125 randomly chosen samples, 100 samples for training and the rest for testing and evaluation.
(3) **HCP** [46,47]: It includes T1-weighted MRIs with an in-plane resolution of 0.7 mm and slice thickness over axial, sagittal, and coronal axes. For our work, we randomly selected 125 samples from the HCP dataset for model training and evaluation.

The proposed algorithm is implemented in PyTorch using the AdamW optimizer. The training involves 200 epochs with a batch size of 2, utilizing four NVIDIA 3090ti GPUs. The initial learning rate is set to 2×10^{-4} and gradually decays to 1×10^{-5} following a cosine attenuation schedule. During training, various data augmentation techniques are applied, including random flipping, rotating, reversing sequence order, and random cropping of patches with a size of 224 × 224. These augmentations are consistently applied to frame triplets throughout the training process. The primary objectives of our training are twofold: to augment the dataset for slice interpolation and enhance the robustness and clarity of the interpolated middle slices. We evaluate the performance of our model in two scenarios: data augmentation and domain discrepancy of augmented MRIs. Overall, our approach yields promising results in data augmentation, demonstrating the effectiveness of our proposed algorithm.

4.2. MRI augmentation baseline

In this subsection, the performance and the number of parameters employed by the proposed network are compared to that of the state-of-the-art light-weight super-resolution and frame-interpolation networks, respectively, RCAN [38], EDSR [39], SwinIR [40], ELAN-light [41], IFRNet [34]. The performance of the proposed and several

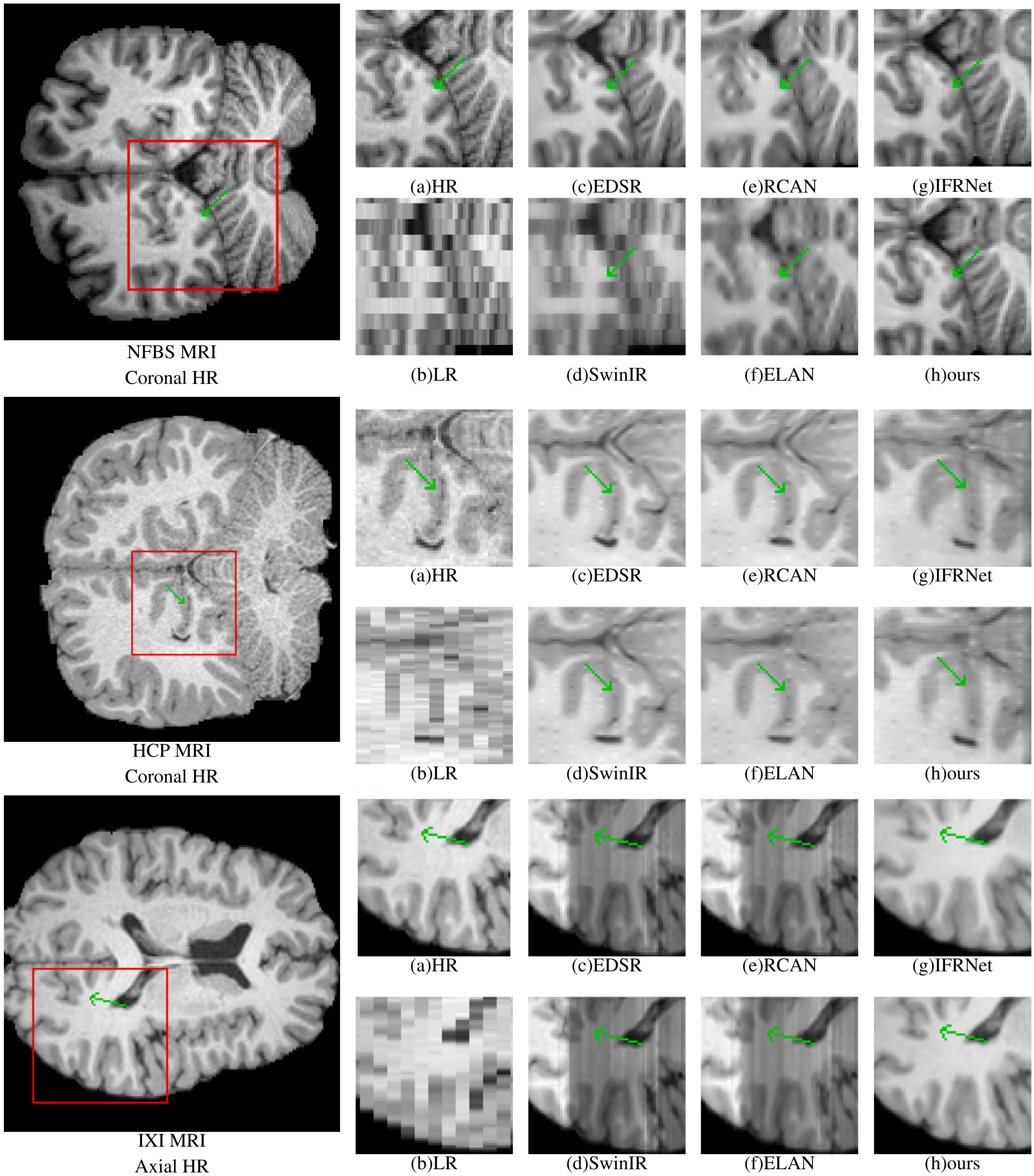

Fig. 4. Super-resolution results presentation. (a) The Original MRI slices with high-resolution. (b) The down-sampled MRI slices with low-resolution. (c)–(g) are the super-resolution slices produced by EDSR, SwinIR, RCAN, ELAN, and IFRNet trained by NFBS [2], IXI [45] and HCP [46,47] datasets. (h) Our super-resolution results. These results correspond to the model corresponding to the highest PSNR(dB)/SSIM in Table 1.

state-of-the-art light-weight super-resolution schemes on three datasets, namely, NFBS, HCP, IXI, with scaling factors of Peak Signal-to-Noise Ratio(PSNR) value and SSIM [48]. As discussed in the above sections, we tried to explore the effectiveness of networks based on large-scale MRI data augmentation. HCP, IXI, and NFBS datasets are processed into three scales over the axial axis. The distance between slices is 4.200 mm, 5.625 mm, and 5.999 mm, respectively. As shown in Table 2, we first evaluated the performance of the interpolation-generated image quality of IFRNet and our SOFNet with PSNR and SSIM metrics on axial axis. With the integration of the Cost Volume Module (CVM) for flow estimation, the outcomes achieved by SOFNet exhibit gains of 0.38 dB and 1.12 dB in comparison to the IXI dataset on the HCP and NFBS datasets, respectively. The enhanced performance can be attributed to the CVM's ability to encapsulate inherent motion features through its attention mechanism. When coupled with the preceding coarse flow estimations, this synergistic approach culminates in more refined flow representations, consequently enhancing the overall efficacy of the SOFNet framework.

Also, the proposed generator generates more realistic MRI slices with higher PSNR/SSIM than the classical SISR method on the Sagittal and Coronal axis. The number of parameters of each of the lightweight super-resolution schemes used for comparison is given in Table 1. It is seen from this table that the proposed network employs around 1.2M fewer parameters than IFRNet does, which has the second-best

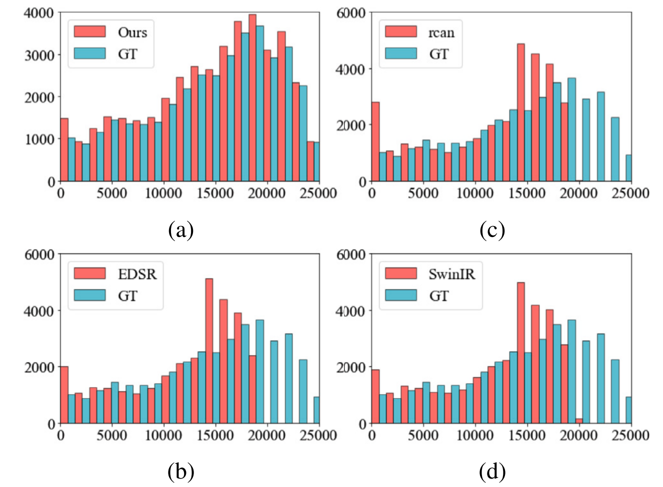

Fig. 5. Histograms of MRI slices augmentation comparison between our optical-flow and SISR methods on HCP dataset. (a) Visualization of histogram using SOFNet. (b–d) results produced by RCAN, EDSR and SwinIR respectively.

Table 2
Quantitative results on HCP, NFBS and IXI datasets. Shows the PSNR(dB)/SSIM value of the interpolated slices compared to original GT slices.

Domain	IFRNet [34]		SOFNet	
	PSNR	SSIM	PSNR	SSIM
HCP	34.81	**0.9444**	**35.10**	0.9400
NFBS	29.67	0.9290	**30.15**	**0.9356**
IXI	30.59	0.9330	**31.24**	**0.9365**

performance. Thus, based on the results shown in Table 2, the proposed SOFNet outperforms the state-of-the-art lightweight super-resolution schemes when the networks' performance and complexity are considered. Fig. 4 shows the visual quality of the validated datasets from the HCP, IXI and NFBS dataset super-resolved by the current best method and the proposed SOFNet. Illustrated in this figure, noticeable discrepancies arise when applying ELAN and other SISR techniques, leading to distorted edges that substantially deviate from the original image's characteristics, particularly under extensive super-resolution scaling. In contrast, the edges obtained through the utilization of SOFNet and IFRNet exhibit a superior semblance to the ground truth, even in evaluations conducted along the Coronal axis. The outcomes achieved by Optical flow methods demonstrate heightened realism and enhanced reliability for subsequent medical scrutiny and clinical assessment. This phenomenon can be attributed to the inherent capacity of optical-flow based approaches to effectively leverage inter-frame relationships, distinguishing them from traditional SISR methods. This advantageous ability is posited to yield more faithful reconstructions of intracranial structures. As a result, our comparative analysis delves into the sagittal and coronal axes, building upon the insights derived from axial axis slice interpolation results. By the way, it is seen in Table 1 that the difference between PSNR values of the images super-resolved by SOFNet compared with IFRNet on the IXI dataset is more than 0.65 (dB). Moreover, when trained by NFBS data, the PSNR values of the image produced by SOFNet and IFRNet is 0.48 (dB). And for HCP dataset, the improvement is 0.29 (dB).

4.3. Optical-flow and SISR comparison

Brain magnetic resonance imaging (MRI) is a three-dimensional dataset of sagittal, coronal, and axial views. In low-resolution MRI acquisitions, cost considerations often result in inadequate data acquisition along one dimension, leading to high resolution in one dimension and lower resolution in the other. Previous approaches, mainly relying on single-image super-resolution (SISR) methods, have yet to exploit the three-dimensional structure inherent in brain MRI data fully. Therefore, we propose an alternative strategy inspired by video processing techniques, aiming to generate three-dimensional data that exhibits higher similarity to the ground truth than SISR methods, particularly regarding internal brain structures.

To verify our hypothesis and evaluate our proposed method's reliability, we downsampled brain MRI slices from the same individual and employed various restoration methods. Since brain MRI slices are grayscale images where the grayscale values are crucial in classifying different physiological structures, we compared the discrepancy between these methods and the ground truth using histogram analysis. As shown in Fig. 5, our proposed method exhibits the slightest domain discrepancy compared to the ground truth, whereas the results obtained from SISR methods diverge significantly from the original data. Consequently, we infer that our method effectively reconstructs the missing internal brain structures absent in the input data.

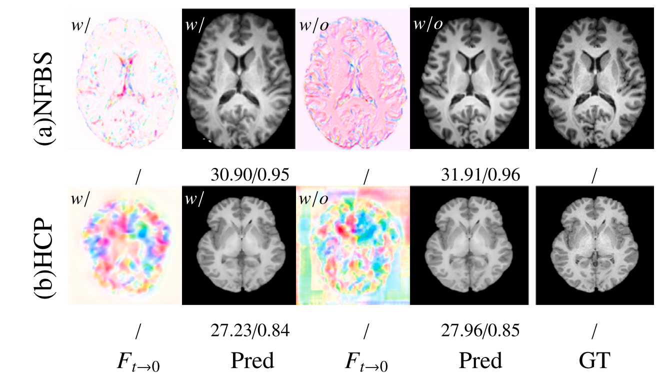

Fig. 6. Visual comparison of intermediate flow and predicted frame of SOFNet w/o and w/ CVM. One of the middle slices and corresponding flows of the test set between adjacent slices are selected. The first row is selected from NFBS, and the second is selected from HCP. The notation in the last row shows the PSNR/SSIM differences.

4.4. Ablation study

To assess the effectiveness of our proposed approaches, we conducted an ablation study on the NFBS and HCP datasets. The study evaluated the contributions of different components of our SOFNet model, including the cost volume flow extraction method and the warp loss L_{warp}. The results and the effectiveness of the ablation study for SOFNet are presented in Table 3. In the ablation study, we built a modified model by removing the cost volume $\mathbf{Vol}^k$ generation module and flow refinement loss L_{warp}, then directly estimated optical flow using the extracted features of A_0 and A_1. This modified model was trained using the reconstruction loss L_{rec}, the feature alignment loss L_{geo}, and L_{warp}. Then, we also drop the L_{war} to testify its effect of flow refinement. The remaining parts of the network were kept unchanged.

The experimental findings, as evidenced in Table 3, underscore the pivotal role played by the cost volume module (CVM) in precise optical flow estimation and the attainment of advanced performance within the SOFNet framework. The integration of both CVM and the warp loss L_{warp} yields elevated PSNR and SSIM values. This amalgamation not only enhances the accuracy of optical flow estimation but also elevates the overall efficacy of SOFNet. The outcome is a collection of reconstructed images characterized by heightened quality and enhanced fidelity. We are convinced that the windows-based attention metric for inter-frame motion information extraction offers substantial utility in refining model outcomes. Furthermore, the utilization of pre-extracted features contributes significantly to the establishment of robust feature representations in medical images. For the warp loss L_{warp}, we followed the basic function architecture as designed in [44], and in addition, we

Table 3
Ablation study of SOFNet. A means the CVM and B means the L_{warp}.

	Setup	IXI		HCP	
		PSNR	SSIM	PSNR	SSIM
1	w/o A & B	29.82	0.8521	35.17	0.8860
2	w/o A	30.46	0.8710	35.49	0.8921
3	w/o B	31.42	0.9434	35.97	0.9382
4	Ours	**31.54**	**0.9547**	**36.18**	**0.9417**

set the Gaussian kernel size to 5 × 5 and the levels of the Laplacian pyramid to 3, 2, and 1, respectively, for feature $\hat{A}_t^k$(for $k \in 0, 1, 2$).

In Fig. 6, it is evident that the CVM effectively minimizes differences between the features of the warped frames and the ground truth, leading to a purified optical flow estimation. This cleaner and more accurate estimation contributes to higher PSNR values for specific generated images, improving the network's overall performance.

In conclusion, the ablation study confirms the effectiveness of our proposed components in enhancing the performance of SOFNet for MRI slice interpolation.

5. Conclusion and future work

In this paper, we present a framework, SOFNet, for the data augmentation of Brain MRI-based data. To our best knowledge, this is the first work to address the problem of MRI super-resolution using the optical flow method. We evaluate the super-resolution performance of our SOFNet on low-resolution natural MRIs. Benefiting from our cost volume and optical flow estimation calculation and coarse-to-fine feature building, SOFNet can effectively and efficiently capture the feature structure. The experimental results demonstrate that our model can significantly complement the missing structures between brain MRIs. In the future, we plan to improve the performance SOFNet from two aspects. Though SOFNet achieved a significant speedup than traditional SISR methods like RCAN, and EDSR, the optical flow calculation between adjacent slices was still computation and memory-intensive compared to those light-weight CNN-based models. In the future, we will explore more efficient implementations or approximations of optical flow for the brain and other MRI slices.

Declaration of competing interest

No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

In this work, we propose a novel end-to-end optical-flow network that operates on Brain MRI data and introduces a more effective CVM model for flow refinement. In our research, we have achieved improved results by designing an optical-flow-based encoder-decoder coarse-to-fine network. To enhance the extraction of optical flow and ensure the quality of generated intermediate slices, we utilize a cost volume module (CVM) to generate a representation of the cost volume. This cost volume representation, combined with intermediate feature estimation, enhances the accuracy and fidelity of the optical flow estimation process in our approach. Through extensive experimentation, we have discovered that deep learning-based optical flow methods, traditionally used for 3D data with continuous change fields like video frame interpolations, can be effectively applied to medical image augmentation. In our research, we treat MRI slices as analogous to video frames and leverage optical flow estimation to approximate missing slices. This approach has demonstrated promising results, enabling the generation of interpolated MRI slices. By utilizing optical flow techniques, we aim to enhance the completeness and quality of MRI datasets, ultimately improving analysis and diagnosis in medical imaging applications.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the National Science Foundation of China under Grants 62272298, 62101346 and 62301330, the National Key Research and Development Program of China under Grant 2022YFC2407000, Clinical Special Program of Shanghai Municipal Health Commission (Grant number: 20204Y0199), the Interdisciplinary Program of Shanghai Jiao Tong University under Grants YG2023LC11,YG2022ZD007 and YG2022QN089, and the Guangdong Basic and Applied Basic Research Foundation under Grants 2021A1515011702 and 2022A1515110101. B. Wei, X. Liu, Y. Wen and X. Qi contribute equally to this paper.

References

[1] F. Shi, J. Cheng, L. Wang, P.-T. Yap, D. Shen, LRTV: MR image super-resolution with low-rank and total variation regularizations, IEEE Trans. Med. Imaging 34 (12) (2015) 2459–2466.
[2] B. Puccio, J.P. Pooley, J.S. Pellman, E.C. Taverna, R.C. Craddock, The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data, Gigascience 5 (1) (2016) s13742–016.
[3] C. Meurée, P. Maurel, J.-C. Ferré, C. Barillot, Patch-based super-resolution of arterial spin labeling magnetic resonance images, Neuroimage 189 (2019) 85–94.
[4] S. Jain, D.M. Sima, F. Sanaei Nezhad, G. Hangel, W. Bogner, S. Williams, S. Van Huffel, F. Maes, D. Smeets, Patch-based super-resolution of MR spectroscopic images: application to multiple sclerosis, Front. Neurosci. 11 (2017) 13.
[5] Z. Zhou, G. Li, G. Wang, A hybrid of transformer and CNN for efficient single image super-resolution via multi-level distillation, Displays 76 (2023) 102352.
[6] R. Xu, X. Kang, C. Li, H. Chen, A. Ming, DCT-FANet: DCT based frequency attention network for single image super-resolution, Displays 74 (2022) 102220.
[7] N. Han, L. Zhou, Z. Xie, J. Zheng, L. Zhang, Multi-level U-net network for image super-resolution reconstruction, Displays 73 (2022) 102192.
[8] K. Zeng, H. Zheng, C. Cai, Y. Yang, K. Zhang, Z. Chen, Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network, Comput. Biol. Med. 99 (2018) 133–141.
[9] J. Du, L. Wang, Y. Liu, Z. Zhou, Z. He, Y. Jia, Brain MRI super-resolution using 3D dilated convolutional encoder–decoder network, IEEE Access 8 (2020) 18938–18950.
[10] C.-H. Pham, A. Ducournau, R. Fablet, F. Rousseau, Brain MRI super-resolution using deep 3D convolutional networks, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 197–200.
[11] L. Snoek, M.M. van der Miesen, T. Beemsterboer, A. Van Der Leij, A. Eigenhuis, H. Steven Scholte, The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses, Sci. Data 8 (1) (2021) 85.
[12] J. Wang, Y. Chen, Y. Wu, J. Shi, J. Gee, Enhanced generative adversarial network for 3D brain MRI super-resolution, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 3627–3636.
[13] S. You, B. Lei, S. Wang, C.K. Chui, A.C. Cheung, Y. Liu, M. Gan, G. Wu, Y. Shen, Fine perceptive GANs for brain MR image super-resolution in wavelet domain, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[14] D. Sun, S. Roth, M.J. Black, A quantitative analysis of current practices in optical flow estimation and the principles behind them, Int. J. Comput. Vis. 106 (2014) 115–137.
[15] A. Bruhn, J. Weickert, C. Schnörr, Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods, Int. J. Comput. Vis. 61 (2005) 211–231.
[16] M.J. Black, P. Anandan, A framework for the robust estimation of optical flow, in: 1993 (4th) International Conference on Computer Vision, IEEE, 1993, pp. 231–236.
[17] B.K. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1–3) (1981) 185–203.
[18] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, T. Brox, Flownet: Learning optical flow with convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2758–2766.
[19] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, Flownet 2.0: Evolution of optical flow estimation with deep networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2462–2470.
[20] A. Ranjan, M.J. Black, Optical flow estimation using a spatial pyramid network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4161–4170.

[21] D. Sun, X. Yang, M.-Y. Liu, J. Kautz, Models matter, so does training: An empirical study of cnns for optical flow estimation, IEEE Trans. Pattern Anal. Mach. Intell. 42 (6) (2019) 1408–1423.
[22] D. Sun, X. Yang, M.-Y. Liu, J. Kautz, Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8934–8943.
[23] T. Hui, X. Tang, C.L. Change Loy, A lightweight convolutional neural network for optical flow estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8981–8989.
[24] T.-W. Hui, X. Tang, C.C. Loy, A lightweight optical flow CNN—Revisiting data fidelity and regularization, IEEE Trans. Pattern Anal. Mach. Intell. 43 (8) (2020) 2555–2569.
[25] G. Yang, D. Ramanan, Volumetric correspondence networks for optical flow, Adv. Neural Inf. Process. Syst. 32 (2019).
[26] Z. Teed, J. Deng, Raft: Recurrent all-pairs field transforms for optical flow, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 402–419.
[27] M. Hofinger, S.R. Bulò, L. Porzi, A. Knapitsch, T. Pock, P. Kontschieder, Improving optical flow on a pyramid level, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII, Springer, 2020, pp. 770–786.
[28] H. Xu, J. Yang, J. Cai, J. Zhang, X. Tong, High-resolution optical flow from 1d attention and correlation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10498–10507.
[29] S. Jiang, D. Campbell, Y. Lu, H. Li, R. Hartley, Learning to estimate hidden motions with global motion aggregation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9772–9781.
[30] F. Zhang, O.J. Woodford, V.A. Prisacariu, P.H. Torr, Separable flow: Learning motion cost volumes for optical flow estimation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10807–10817.
[31] Y. Zheng, M. Zhang, F. Lu, Optical flow in the dark, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6749–6757.
[32] W. Yan, A. Sharma, R.T. Tan, Optical flow in dense foggy scenes using semi-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13259–13268.
[33] Z. Huang, X. Pan, R. Xu, Y. Xu, G. Zhang, H. Li, et al., Life: Lighting invariant flow estimation, 2021, arXiv preprint arXiv:2104.03097.
[34] L. Kong, B. Jiang, D. Luo, W. Chu, X. Huang, Y. Tai, C. Wang, J. Yang, IFRNet: Intermediate feature refine network for efficient frame interpolation, 2022, arXiv:2205.14620.
[35] G. Zhang, Y. Zhu, H. Wang, Y. Chen, G. Wu, L. Wang, Extracting motion and appearance via inter-frame attention for efficient video frame interpolation, 2023, arXiv:2303.00440.
[36] Z. Huang, X. Shi, C. Zhang, Q. Wang, K.C. Cheung, H. Qin, J. Dai, H. Li, FlowFormer: A transformer architecture for optical flow, 2022, arXiv:2203.16194.
[37] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, 2015, arXiv:1505.04597.
[38] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, 2018, arXiv:1807.02758.
[39] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, 2017, arXiv:1707.02921.
[40] J. Liang, J. Cao, G. Sun, K. Zhang, L.V. Gool, R. Timofte, SwinIR: Image restoration using swin transformer, 2021, arXiv:2108.10257.
[41] X. Zhang, H. Zeng, S. Guo, L. Zhang, Efficient long-range attention network for image super-resolution, 2022, arXiv:2203.06697.
[42] W. Shi, J. Caballero, F. Huszár, J. Totz, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[43] S. Meister, J. Hur, S. Roth, UnFlow: Unsupervised learning of optical flow with a bidirectional census loss, 2017, arXiv:1711.07837.
[44] S. Niklaus, F. Liu, Softmax splatting for video frame interpolation, 2020, arXiv:2003.05534.
[45] A. Alansary, Ixi Dataset: Brainwwor Development Organization, Imperial College London, 2016.
[46] D.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S.W. Curtiss, et al., The Human Connectome Project: a data acquisition perspective, Neuroimage 62 (4) (2012) 2222–2231.
[47] D.C. Van Essen, S.M. Smith, D.M. Barch, T.E. Behrens, E. Yacoub, K. Ugurbil, W.-M.H. Consortium, et al., The WU-Minn human connectome project: an overview, Neuroimage 80 (2013) 62–79.
[48] Z. Wang, Image Quality Assessment : From Error Visibility to Structural Similarity (Ph.D. thesis), 2004.