Neurocomputing 573 (2024) 127197

Contents lists available at ScienceDirect

Neurocomputing

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

# Enhancement of guided thermal image super-resolution approaches

Patricia L. Suárez [a,*], Dario Carpio [a], Angel D. Sappa [a,b]

[a] ESPOL Polytechnic University, FIEC, CIDIS, Campus Gustavo Galindo, 09-01-5863, Guayaquil, Ecuador
[b] Computer Vision Center, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Communicated by J. Yu

*Keywords:*
Thermal-like image
Brightness channel
Contrastive GAN
Guided super-resolution

ABSTRACT

Guided image processing techniques are widely used to extract meaningful information from a guiding image and facilitate the enhancement of the guided one. This paper specifically addresses the challenge of guided thermal image super-resolution, where a low-resolution thermal image is enhanced using a high-resolution visible spectrum image. We propose a new strategy that enhances outcomes from current guided super-resolution methods. This is achieved by transforming the initial guiding data into a representation resembling a thermal-like image, which is more closely in sync with the intended output. Experimental results with upscale factors of ×8 and ×16, demonstrate the outstanding performance of our approach in guided thermal image super-resolution obtained by mapping the original guiding information to a thermal-like image representation.

## 1. Introduction

Over the past two decades, the usage of thermal imagery has significantly increased due to cost reductions and improved availability of thermal cameras [1]. However, despite the growing adoption of thermal cameras, there are still limitations on image resolution imposed by the technology. While higher-resolution thermal cameras exist, they often rely on more expensive actively cooled technology, leading to the prevalent use of uncooled thermal cameras in most applications due to their affordability. To overcome these resolution limitations, Super Resolution (SR) techniques, originally developed for the visual-optical spectrum, can be adapted for the thermal domain. The goal of SR approaches is to reconstruct a High Resolution (HR) image from one or more Low Resolution (LR) input images. Both traditional algorithm-based [2] and machine learning-based [3] SR methods can be employed.

Traditional approaches typically address this challenge within a multi-image framework, known as Multi-Frame Super Resolution (MFSR) [4]. In contrast, some methods learn correspondences between low- and high-resolution image patches, which are then applied to a new LR image to recover its most probable HR version [5]. Furthermore, machine learning-based approaches, also known as single-frame super-resolution (SFSR), are known to recover the HR image by leveraging a set of training examples, often known as example-based single-image SR [6]. Although MFSR and SFSR approach achieve acceptable results at ×2, ×3, and even in certain scenarios at ×4, super-resolution beyond ×4 is limited by the scarce information in the provided LR image. In light of these conditions, the path has been opened for guided super-resolution approaches, in which the process of super-resolving the thermal image is "guided" by the information provided by an image of the same scene acquired with a low-cost camera in the visible spectrum, which may be low-cost but high-resolution. The guidance strategy is not new in the image processing field it has already been explored to solve different problems during the last decades. For instance, we can find guided-based solutions for enhancement [7], image filtering [8], super-resolution [9], just to mention a few. Regarding the super-resolution problem, different guidance SR techniques have been proposed to tackle depth-map SR, infrared SR, thermal SR, hyperspectral SR, and some others; Section 2 presents details of state-of-the-art approaches proposed in the literature for guided SR in all these contexts.

Most of the guided super-resolution approaches mentioned above focus on developing novel architectures to efficiently extract and integrate features from the HR guiding image toward the LR-guided image during the super-resolution process, without taking into account whether the guiding and guided images are from the same or different domains. On the contrary to all these approaches, the current work is focused on this fact trying first to map the guiding image towards the guided domain, to perform a more efficient guidance task. We hypothesize that as more similar the domain of the input images is as better the guidance process will be. This hypothesis is validated by testing nine different state-of-the-art approaches using as a guidance an HR thermal-like image, instead of the given HR visible spectrum image. In all the tested approaches and datasets, better results are obtained with the proposed strategy.

* Corresponding author.
*E-mail addresses:* plsuarez@espol.edu.ec (P.L. Suárez), dncarpio@espol.edu.ec (D. Carpio), asappa@espol.edu.ec, asappa@cvc.uab.es (A.D. Sappa).

https://doi.org/10.1016/j.neucom.2023.127197
Received 28 July 2023; Received in revised form 7 December 2023; Accepted 26 December 2023
Available online 30 December 2023

The main contributions of the manuscript can be summarized as follows:

- A novel strategy to improve state-of-the-art guided super-resolution approaches is proposed. It is based on the use of HR thermal-like images as a guidance, this pseudo-thermal image is obtained from an HR visible spectrum image.
- An in-depth evaluation of the proposed strategy is performed with nine different architectures of the state-of-the-art, in challenging scenarios (i.e., ×8 and ×16) from three different datasets.

The manuscript is organized as follows. Section 2 presents works related to guided super-resolution approaches. Section 3 presents the proposed strategy. Experimental results and comparisons with different implementations and datasets are given in Section 4. Finally, conclusions are presented in Section 5.

## 2. Related works

In recent years, deep learning-based approaches have become the common way to tackle most computer vision problems. For instance, in [10] the authors propose a novel approach for the super-resolution of hyperspectral images related to medical imaging, terrestrial and remote sensing. Contrary to existing methods based on observations of multiple scenes, their framework leverages transfer learning from natural images, using a deep convolutional neural network to learn low-to-high resolution mapping. Additionally, they employ collaborative non-negative matrix factorization to improve performance. Also focused in the remote sensing field, in Qin et al. [11] the authors propose a method for image super-resolution using a multi-scale convolutional neural network called MSCNN. The proposed method takes advantage of multiscale image features to improve the extraction of high-frequency features and reconstruct high-resolution images with rich details. The influence of image super-resolution on aerial scene classification has been a topic of interest in the field of remote sensing and computer vision. The impact of SR to improve the classification of aerial scenes is studied in [12], with special emphasis on the analysis of different state-of-the-art SR algorithms, including traditional methods and deep learning-based approaches.

More recently, diffusion models have been also considered for tackling the super-resolution problem. In [13], the authors address the low inference speed of diffusion-based image super-resolution methods by significantly reducing the number of diffusion steps, thereby eliminating the need for post-acceleration during inference and its associated performance deterioration. The method constructs a Markov chain that transfers between the high-resolution image and the low-resolution image by shifting the residual between them, substantially improving the transition efficiency. Additionally, an elaborate noise schedule is developed to flexibly control the shifting speed and the noise strength during the diffusion process.

Following the state of the art, different CNN-based architectures have been proposed for the guided super-resolution problem. Therefore, a comprehensive review of the state-of-the-art architectures designed for the guided super-resolution problem is presented below, and recent advances in the field are described. The review highlights the integration of deep learning architectures, attention mechanisms, multiscale fusion strategies, and synthesis into guided super-resolution models.

These advancements enable the effective utilization of guidance information for improved super-resolution results. In addition to guided thermal image super-resolution, this section also reviews approaches that are focused on guided super-resolution of depth maps. The following approaches provide an overview of notable works in each of these domains.

One of the first guided thermal super-resolutions has been presented in [14], the authors propose a novel approach where the guided SR is tackled as a pixel-to-pixel mapping, from the guided HR image to the domain of the LR source image. This mapping process is parametrized as a multi-layer perceptron, whose weights are learned by minimizing the distance between the downsampled HR target image and the given LR source image. Also in [15] the authors propose an architecture capable of restoring high-quality images from sequences of noisy, misaligned, and low-resolution RAW bursts. This architecture introduces the Burst Super-Resolution Transformer (BSRT), a framework that enhances the extraction of inter-frame information and the overall reconstruction process. The BSRT incorporates a pyramid flow-guided deformable convolution network (Pyramid FG-DCN) in conjunction with Swin Transformer Blocks and Groups as the main backbone. Another technique to address super-resolution methods for Synthetic Aperture Radar (SAR) images with large-scale factors is proposed by [16], this method utilizes co-registered HR optical images to guide SAR image reconstruction named as Optical-Guided Super-Resolution Network (OGSRN). This architecture comprises two sub-networks: SAR image Super-Resolution U-Net (SRUN) and SAR-to-Optical Residual Translation Network (SORTN). The training process involves SAR image reconstruction using SRUN and a residual learning process based on an attention module with channel and spatial mechanisms and finally, a translation process to obtain the optical images using SORTN. Another guided super-resolution approach is presented in [17]; it uses a specialized optimization layer that adapts during the learning process. This layer operates on a graph representation, capturing the relationships between pixels in the image. By learning the potentials of this graph, the approach incorporates contextual information from a guide image while ensuring faithful reconstruction of the high-resolution target from the low-resolution source. The idea is to incorporate information from a guide image to ensure that the high-resolution output accurately spatially matches the low-resolution image. Unlike existing methods, the authors treat the source as a constraint, resulting in sharper and more natural-looking images.

On the contrary to previous works, in [18], the authors address the challenge of limited image resolution in thermal imaging systems used in UAVs. To overcome this limitation, the authors propose the multiconditioned guidance network (MGNet), which uses high-resolution visible images to enhance the super-resolution of thermal UAV images. Note that HR images are rich in information such as distinct appearance, semantic details, and edge features that are useful for the SR process. To leverage this information, the authors introduce a multicue guidance module MGM to effectively integrate information from visible images to guide the process of thermal UAV image super-resolution.

Recently, [19] introduced a novel approach called Dual-IRT-GAN, which is a Generative Adversarial Network (GAN), specifically developed to tackle simultaneous tasks of super-resolution and defect detection in infrared thermography. The visibility of flawed areas in the resulting high-resolution images is enhanced by using defect-aware attention maps derived from segmented defect images. To perform the training process the researchers use a large dataset containing generated thermal images of composite materials with defects of various types, sizes, and locations to train the Dual-IRT-GAN model. Continuing with the literature review, in [20], the authors present the techniques and results obtained from a guided thermal image super-resolution challenge, proposed at the PBVS-CVPR 2023 Workshop. The challenge consists of generating a ×8 super-resolved thermal image using as guidance the corresponding high-resolution visible spectrum image. According to the authors, 74 teams have been initially registered showing the interest of the research community in this topic.

Unlike existing methods, that primarily focus on enhancing edges, in [21] the authors propose the HTI-Net, which takes a system-level perspective and optimizes the neural network structure for image super-resolution reconstruction. The inspiration behind HTI-Net comes from the inherent similarity between thermal particles and image pixels, leading to the proposal of a heat-transfer-inspired network based on heat transfer theory. To achieve improved feature reuse through the

integration of multiple information, the researchers employ finite difference theory. They utilize a second-order mixed-difference equation to redesign the residual network (ResNet). Additionally, by deriving a pixel value flow equation from the thermal conduction differential equation, HTI-Net effectively uncovers deep potential feature information. Following guided approaches, [22] presents an architecture that combines guided anisotropic diffusion with a deep convolutional network. Combining the edge enhancement capabilities of diffusion with the contextual reasoning of deep neural networks improves matching with the source image. With this architecture, improved performance is obtained at larger scales, such as the ×32 scale.

Focusing on guided depth map super-resolution, [23] presents an edge-based technique, where edges from color images are used as guidance during the SR process. The authors address the challenge of inconsistency between color edges from the guiding images and depth discontinuities on the LR depth maps. This problem generates texture copy artifacts and blurring depth discontinuities in restored depth maps. The paper proposes a robust optimization framework for color-guided depth map restoration t0.0hat uses a robust penalty function to model the smoothness term of the model. The proposed method is shown to be robust against the inconsistency between color edges and depth discontinuities, even when using simple guidance weight. On the other hand, in [24] a novel deep network named DepthSR-Net is introduced for guided depth map super-resolution. This architecture is constructed based on a residual U-Net deep network architecture, incorporating hierarchical features to guide the process of residual learning. Similarly, in [25] a technique that is focused on depth super-resolution by combining an internal smoothness prior and an external gradient consistency constraint in the graph domain is introduced. It aims to reconstruct a high-resolution depth map based on a low-resolution observation, aided by a corresponding high-resolution color image. The method uses a pyramid structure to extract multi-scale features from the high-resolution color image and employs a deep dense-residual network to enhance the intensity-guided depth map. Continuing with the use of RGB images, in [26] an approach for achieving both rapid and high-quality hierarchical depth-map super-resolution (HDS) by using an HR RGB image to guide bilateral filtering of the depth map is presented. The authors extend the HDS model to a Classification-based Hierarchical Depth-map Super-resolution (C-HDS) model, implementing a context-aware trilateral filter to mitigate the influence of unreliable neighbors on the missing depth information.

## 3. Proposed strategy

Multi-modality integration has emerged as an exciting area of research, prompting the community to explore innovative techniques that take advantage of the coexistence of different sources of information. This integration can be achieved through various approaches, including information fusion and guided methods. Information fusion involves combining inputs to create a unified representation (for example, [35, 36]), while guided approaches use one modality to guide the processing of another (for example, filtering [33]). In the super-resolution guided context [37], the guiding modality plays a crucial role in improving the resolution and quality of the target modality.

In the current work, we introduce a new strategy designed to improve the performance of state-of-the-art guided super-resolution techniques, which are based on the usage of a high-resolution image as guidance. In general, this guidance is an HR visible spectrum image, in our case we propose to use a representation of this HR visible spectrum image similar to the LR image (i.e., thermal image) being super-resolved. Fig. 1(*bottom*) depicts an illustration of the proposed strategy. Our approach involves using synthetic data, particularly pseudo-thermal images, as a reference to guide the process. The main goal is to demonstrate that these synthetic images can give better results than those obtained when super-resolution is guided by high-resolution images in the visible spectrum. This section begins by explaining the changes proposed to our previous approach presented in [38], where a cycled adversarial network with multiple loss functions is used to obtain a pseudo-thermal representation. Next, the nine state-of-the-art guided super-resolution approaches evaluated in the current work are briefly described.

### 3.1. Thermal like synthetizing

Generative models have been extensively explored to obtain synthetic representations for different computer vision applications, like infrared image colorization (e.g., [39,40]), estimation of vegetation indexes [41] also include the generation of synthetic face representations [42], lightweight image translations [43], medical image synthesis [44], among others.

A recent proposal introduced a novel approach to generate thermal-like representations from low-cost visible images [38] referred to as synthesized thermal images in the current work. The main objective is to create synthetic images that closely resemble real thermal images, providing valuable information about the objects in the scene. These synthetic images can contribute to the goals of other computer vision algorithms. The generation of synthetic thermal-like representations opens up new possibilities in thermal image processing. These synthetic representations provide a cost-effective alternative to acquiring actual thermal images, allowing for broader access to thermal-related information. Moreover, the ability to generate thermal image-like representations from visible images enhances the synergy between different imaging modalities and facilitates the integration of thermal information into existing machine vision algorithms.

Additionally, alternative solutions can be provided to the low availability of data in this spectrum with these thermal data generators, which is sometimes a real challenge. The generation of synthetic thermal representations provides a practical solution to mitigate this limitation. Researchers can use these representations to expand their data sets and improve the robustness and generalization of computer vision models, especially when real thermal data is scarce or expensive to acquire.

In this work, the cyclic GAN architecture proposed in [38] is considered (see illustration in Fig. 2), which uses as an input a RGB image converted to the HSV color space from which only the brightness channel (H) is taken. This image (H channel) is transferred to a pseudothermal image domain. These generated synthetic images are used as a guide in all the guided SR approaches evaluated in the current work. The cyclic GAN architecture combines multiple loss functions to improve the stability and convergence quality of the GAN network. A brief explanation of each of these loss functions used in the model is given below. One of these loss functions is relativistic loss, which is effective in generating high-dimensional data. It achieves this by encouraging generated samples to closely resemble real samples, avoiding model saturation and speeding up the training process. In addition to the relativistic adversarial loss, we incorporate contrastive and identity loss functions. These losses play a pivotal role in preserving the structural and semantic information present in the input images. They ensure that the generated thermal images maintain essential features and details, safeguarding the overall image representation's integrity. The integration of these loss functions within the generative cycle GAN architecture not only ensures the creation of high-quality synthetic thermal images but also fosters a strong resemblance to real thermal images. By harnessing the capabilities of these loss functions, we effectively bridge the gap between different image domains, facilitating the generation of accurate and realistic thermal representations from the H channel of the given image. The usage of these loss functions provides an optimal framework for minimizing errors in network predictions within the generative cycle GAN architecture. By thoughtfully considering their mathematical formulations and incorporating them into the training process, we guide the network to prioritize essential visual features, preserve structural and semantic information,

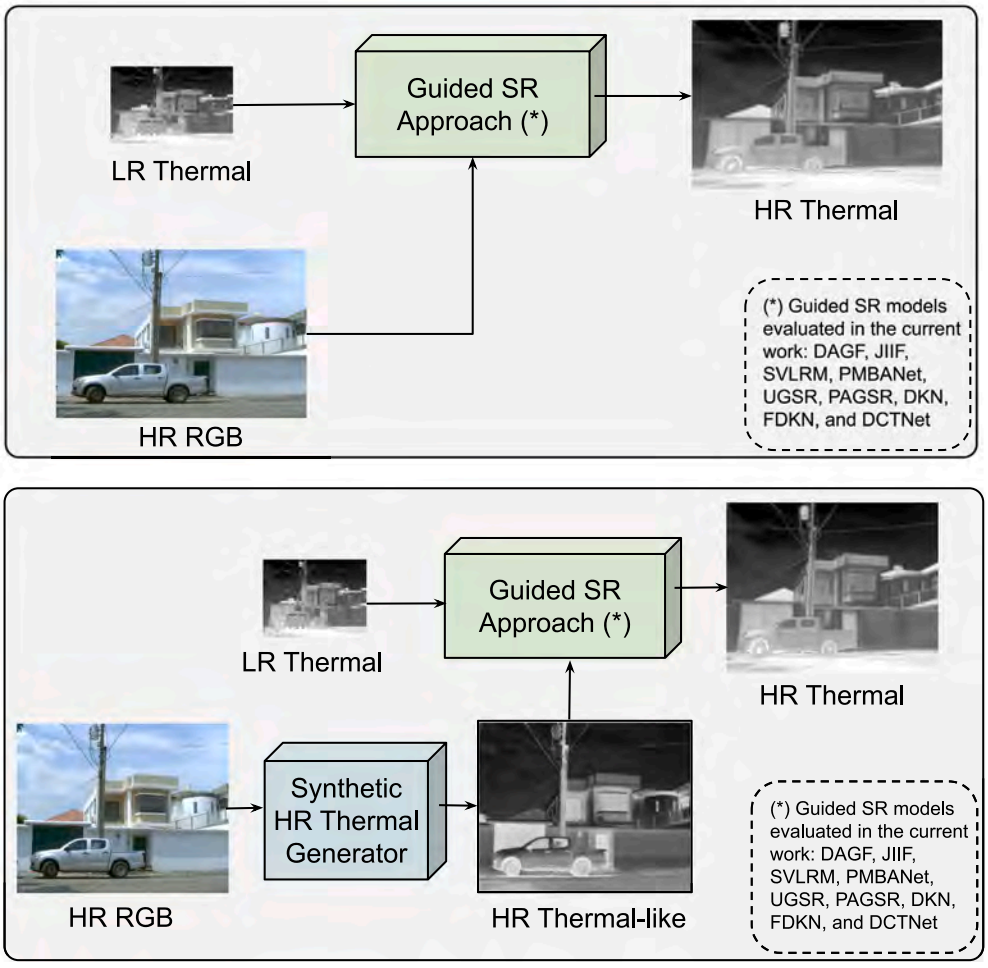


**Fig. 1.** Guided thermal image super-resolution approaches evaluated in the current work: DAGF [27], JIIF [28], SVLRM [29], PMBANet [30], UGSR [31], PAGSR [32], DKN [33], FDKN [33], and DCTNet [34] following a: (*top*) Classical scheme; (*bottom*) Proposed strategy.

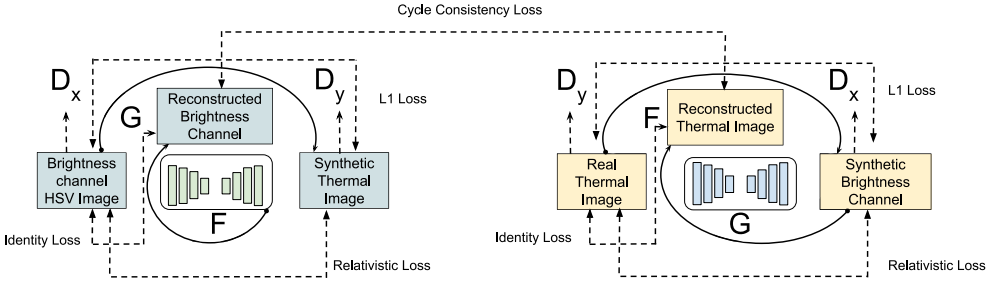


**Fig. 2.** Cycle GAN Architecture for Thermal Synthetized images proposed.

and generate thermal images that are visually compelling and virtually indistinguishable from real thermal images.

The relativistic loss is defined as follows:

$$L_D^{RGAN} = \mathbb{E}_{(x_r,x_f)\sim(\mathbb{P},\mathbb{Q})}\left[f\left(C\left(x_r\right)-C\left(x_f\right)\right)\right], \tag{1}$$

$$L_G^{RGAN} = \mathbb{E}_{(x_r,x_f)\sim(\mathbb{P},\mathbb{Q})}\left[g\left(C\left(x_f\right)-C\left(x_r\right)\right)\right], \tag{2}$$

where $f$ and $g$ are mappings that transform the confidence of the discriminator in classifying a sample as real or fake to a scalar value and $x_r$ and $x_f$, represent the real and fake images, respectively. Contrastive loss has also been incorporated to minimize the dissimilarity between similar pairs of data points and maximize the dissimilarity between dissimilar pairs within a given dataset. In accordance with [45], this loss can be defined as:

$$\mathcal{L}_{\text{contrastive}}(\hat{Y},Y) = \sum_{l=1}^{L}\sum_{s=1}^{S_l} \ell_{\text{contr}}\left(\hat{v}_l^s, v_l^s, \bar{v}_l^s\right), \tag{3}$$

the shape of the tensor, denoted as $V_l \in \mathbb{R}^{S_l \times D_l}$, is determined by the specific structure and characteristics of the model. Here, $S_l$ represents the number of spatial locations of the tensor. To refer to a ground truth and predicted feature vector within the tensor, the notation $v_l^s \in \mathbb{R}^{D_l}$ and $\hat{v}_l^s \in \mathbb{R}^{D_l}$ is used, indicating the $D_l$-dimensional feature vector at the $s$th spatial location. Conversely, $\bar{v}_l^s \in \mathbb{R}^{(S_l-1)\times D_l}$ refers to the collection of feature vectors at all other spatial locations apart from $s$.

Building upon the principles proposed in [38], an additional loss known as identity loss is incorporated into the model. This loss is very important to assess the disparity between the features extracted

from the real and generated images. By minimizing the intensity loss, the generator network is motivated to generate outputs that not only possess a visually realistic appearance but also exhibit similar features and structures as the real images. The use of identity loss serves as a valuable tool in training the generator network to comprehend the underlying characteristics of the real images and replicate them in the generated outputs. This loss function encourages the generator to learn and preserve the essential attributes, textures, and details that are present in real thermal images. Consequently, the generated synthetic thermal images demonstrate a higher fidelity and maintain the structural integrity of the original images. By incorporating identity loss into the training process, the model becomes proficient in capturing the distinctive features and nuances of real thermal images. This ensures that the generated outputs possess not only a realistic appearance but also exhibit similar patterns, structures, and semantic content to their corresponding real counterparts. The identity loss acts as a guiding force, enabling the generator network to generate synthetic thermal images that not only visually resemble the real images but also retain their corresponding pixel and semantic information. This loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{\text{identity}}\ (G,F) &= \mathbb{E}_{c\sim P_{\text{data}}\ (c)}\left[\|F(c)-c\|\right] \\ &+ \mathbb{E}_{n\sim P_{\text{data}}\ (n)}\left[\|G(n)-n\|\right],\end{aligned} \tag{4}$$

where $G$ and $F$ correspond to mapping functions that generate the synthetic images $F(G(x))$ and $G(Y(z))$ respectively and $c$ and $n$ correspond to a real image from the source and target domains respectively. Additionally, in the current work, a cycle consistency loss is introduced to further enhance the translation results. This loss plays a crucial role in ensuring that the mapping between the domains remains consistent and bijective. In other words, when an image is translated from domain A to domain B and then back to domain A, the resulting image should be identical to the original image from domain A. It is defined as:

$$\begin{aligned}\mathcal{L}_{cycle}(G,F) &= \mathbb{E}_{x}\sim_{p\ \text{data}_{(x)}}[\|F(G(x))-x\|_1] \\ &+ \mathbb{E}_{y}\sim_{p\ \text{data}_{(y)}}[\|G(F(y))-y\|_1],\end{aligned} \tag{5}$$

where $G$ and $F$ correspond to mapping functions that generate the reconstructed images $F(G(x))$ and $G(Y(z))$ respectively. Also, $x$ and $y$ correspond to real images. The cycle consistency loss encourages $F(G(x)) \approx x$ real and $G(Y(z)) \approx y$ real. This loss allows for generating high-quality images that are both realistic and semantically meaningful. Finally, the multiple loss functions implemented in our model can be defined as:

$$\begin{aligned}\mathcal{L}_{\text{final}}\ &= \lambda_1\mathcal{L}_{\text{RGAN}}(G,D,X,Y)+\lambda_2\mathcal{L}_{\text{cont}}\ (G,H,X) \\ &+\lambda_3\mathcal{L}_{\text{cont}}(G,H,Y)+\lambda_4\mathcal{L}_{\text{ident}(G,F)}+\lambda_5\mathcal{L}_{\text{cycle}(G,F)},\end{aligned} \tag{6}$$

where $\lambda_i$ are empirically defined. Finally, in the current work, another change is proposed to enhance the efficiency of the thermal-like image generator model presented in [38]. We propose changing the default value of the beta1 parameter in the Adam optimizer from 0.9 to 0.72. This modification is aimed at enhancing the importance of current gradient information and achieving a balance between stability and efficiency during image generation. The change in the beta1 parameter leads to a reduction in the weight of historical gradient information, which improves the performance of the Adam optimizer. By making this modification, the model achieves an optimal hyper-parameter combination for effective convergence and high-quality thermal-like image generation.

### 3.2. Guided super-resolution models

This section briefly details the nine guided SR approaches evaluated in the current work. Note that some of the approaches have been originally proposed for guiding SR of depth maps, while others are for thermal images. To evaluate all of them in a common framework just thermal images are considered as input, despite the fact some of them were proposed for thermal or depth map images. All the approaches from the state-of-the-art included in this section have the corresponding code provided by the authors. All these approaches are used in the experimental results section of this work.

Through this section, the notation given by the reviewed paper will be used to avoid misunderstandings with the original papers, but note that different terms have been used for the same content (i.e., source, guide, guidance, target, input image, guided).

**Deep Attentional Guided Filtering — DAGF:** One of the approaches in the current work is the one presented in [27]. In this work, the authors affirm that in the realm of image filtering, most existing methods focus on constructing filter kernels solely from the guidance itself, neglecting the interdependence between the guidance and the target. Those methods often lead to the presence of undesirable artifacts, as there are typically significant variations in edges between two images. Recognizing this challenge, the authors propose a framework called deep attentional-guided image filtering. This framework see Fig. 3, aims to fully leverage the complementary information present in both the guidance and target images.

The proposed method incorporates an attentional kernel learning module that generates two sets of filter kernels: one from the guidance image and the other from the target image. These dual sets of filter kernels are then intelligently combined by capturing the pixel-wise dependency between the two images. This adaptive integration process allows for a more precise and accurate transfer of structural information from the guidance to the target. With this multi-scale guided image filtering module is possible to progressively generate the filtering result with the constructed kernels in a coarse-to-fine manner. Additionally, a multi-scale fusion strategy is introduced to reuse the intermediate results in the coarse-to-fine process. This module progressively generates the filtering result using the constructed kernels in a coarse-to-fine manner. It enables a comprehensive analysis of the image at different scales, ensuring that the filtering process captures both global and local details. To maximize efficiency and enhance overall performance, a multi-scale fusion strategy is employed to reuse intermediate results obtained during the coarse-to-fine process.

**Joint Implicit Image Function for Guided Depth Super-Resolution — JIIF:** According to the authors, the existing methods for guided super-resolution often face limitations in terms of model capability or interpretability. To overcome these challenges, the authors propose an approach [28] to formulate a guided super-resolution as a neural implicit image interpolation problem, adopting a general image interpolation framework. Fig. 4 depicts the proposed network architecture. They introduce a Joint Implicit Image Function (JIIF) representation, which enables the simultaneous learning of both interpolation weights and values. The JIIF representation characterizes the target image domain by using spatially distributed local latent codes extracted from both the input (guided) image and the guide image. The concept of implicit neural representation revolves around the use of a deep implicit function (DIF) to map continuous coordinates to signals within a specific domain. To facilitate knowledge sharing across different input observations, an encoder is employed to extract latent codes from the input, thereby enabling the conditioning of the DIF to the current observation. Consequently, a scene or image can be represented by a collection of local latent codes distributed across the input domain's coordinates, providing valuable information for various downstream tasks such as semantic segmentation and super-resolution.

Additionally, the authors propose to learn the interpolation weights concurrently. In the neural implicit interpolation part, they consider the interpolation at each query pixel as a graph problem. To create a smooth output, DIF uses a weighted average of the predictions from nearby points. This can be thought of as a neural network performing implicit interpolation. The weights and values of the average are learned using a deep implicit function. This function is able to learn the relationships between the latent codes and the pixel values, and it

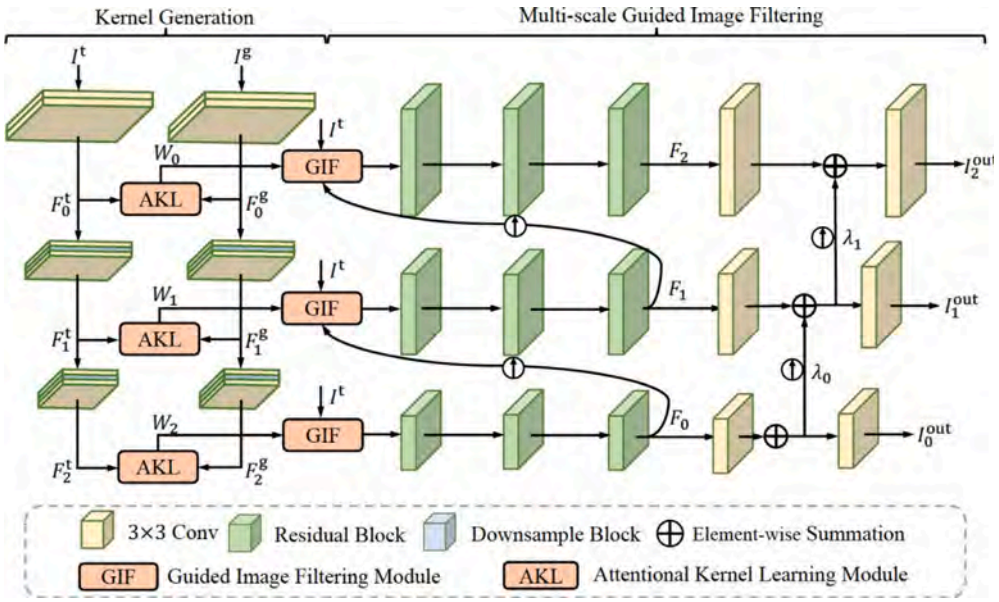


**Fig. 3.** DAGF architecture.
*Source:* Illustration from [27].

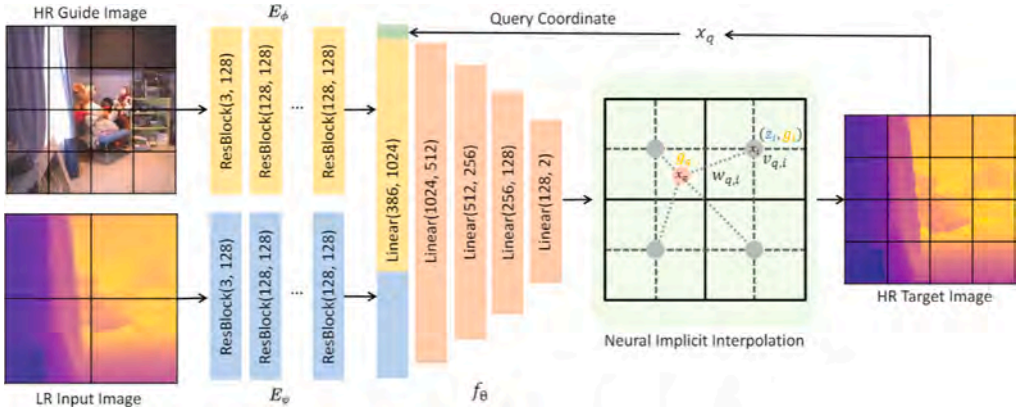


**Fig. 4.** JIIF architecture.
*Source:* Illustration from [28].

allows the decoder to produce high-quality images even when the LR input image is very low-resolution.

Finally, the proposed architecture offers a method that combines the advantages of implicit neural representation and image interpolation, paving the way for improved guided super-resolution with enhanced interpretability and model capability. These Graph Convolution Networks (GCNs) have emerged as a powerful approach for addressing problems associated with graph-structured data.

**Spatially Variant Linear Representation Models for Joint Filtering — SVLRM:** In contrast to existing methods that rely on locally linear models or hand-designed objective functions, in [29] the authors present a novel joint filtering algorithm based on a spatially variant linear representation model (SVLRM). Contrary to traditional techniques that directly predict the target image using a deep convolutional neural network, the proposed method employs a deep CNN to estimate the spatially variant linear representation coefficients. These coefficients effectively capture the structural information present in both the guidance and input images, and are subsequently utilized to generate the desired target image. Fig. 5 depicts the efficacy of the architecture in a diverse range of applications, including depth/RGB image upsampling and restoration, flash/no-flash image deblurring, natural image denoising, and scale-aware filtering. Also, the proposed method substantially enhances the clarity of a flash/no-flash image deblurring scenario. The introduction of the SVLRM for joint filtering provides a powerful approach for representing the target image. This method offers increased flexibility and adaptability in capturing the core structural characteristics of the images. Additionally, the researchers have developed an efficient optimization technique that leverages a deep convolutional neural network constrained by the SVLRM. This enables precise estimation of the spatially variant linear representation coefficients, which effectively capture the intricate structural details present in both the input image and the guidance image. As a result, the algorithm can accurately determine whether specific structures should be transferred to the target image. The proposed algorithm has undergone extensive evaluations, demonstrating its exceptional performance in various applications such as depth/RGB image upsampling and restoration, flash/no-flash image deblurring, natural image denoising, and scale-aware filtering.

**PMBANet: Progressive multi-branch aggregation network for scene depth super-resolution — PMBANet:** In the context of depth map super-resolution, there are significant challenges due to the ill-posed nature of the inverse problem. Reconstructing depth boundaries

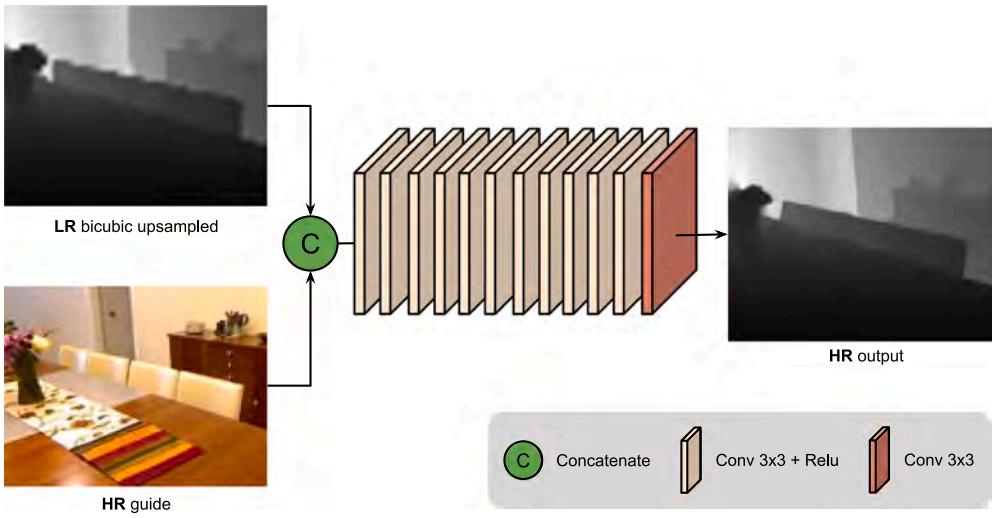


**Fig. 5.** SVLRM architecture.
*Source:* According to [29].

accurately, especially at high magnification factors, is particularly difficult. Additionally, downsampling degradation often leads to severe destruction of depth regions on fine structures and small objects in the scene. To address these challenges, a novel approach called the progressive multi-branch aggregation network (PMBANet) is proposed in [30]. The PMBANet is built upon stacked progressive multi-branch aggregation (MBA) blocks, as depicted in Fig. 6, it aims to comprehensively tackle the aforementioned difficulties and progressively restore the degraded depth map. Each MBA block consists of multiple parallel branches, each serving a specific purpose.

The first branch called the reconstruction branch (RB), is introduced as prior knowledge to assist in recovering depth details. This branch uses attention-based error feed-forward/back modules. It iteratively addresses the downsampling errors and refines the depth map by employing the attention mechanism in the feed-forward/-back process. This mechanism progressively highlights informative features at depth boundaries, leading to improved depth map quality. The second branch, known as the guidance branch (GB), is divided into a multi-scale branch and a color branch. The multi-scale branch learns a representation that focuses on objects of different scales, enabling better feature extraction. On the other hand, the color branch uses auxiliary color information to regularize the depth map. By leveraging the internal structural correlation between depth and color, the color branch contributes to enhancing the accuracy of depth reconstruction. To fuse and select the most discriminative features from all the branches, a fusion block is introduced. This block adaptively combines the information obtained from each branch, allowing the network to effectively leverage the diverse features learned in parallel.

**Toward Unaligned Guided Thermal Super-Resolution — UGSR:** The UGSR-Base CNN model proposed by Gupta et al. [31], has been designed for unaligned guided super-resolution, and incorporates two approaches to effectively mitigate the misalignment present in the input images. These methods are known as unaligned guided super-resolution with feature-space alignment (UGSR-FA) and unaligned guided super-resolution with misalignment estimation (UGSR-ME). The first unaligned guided super-resolution algorithm, UGSR-FA, introduces a feature-space alignment loss that maximizes the spatial correlation between feature maps from the low-resolution thermal and high-resolution visible images. This loss encourages spatial alignment and acts as a regularizer to ensure that the network compensates for the misalignment between the feature maps. The second algorithm, UGSR-ME, approximates the misalignment between the input and guide images by jointly performing alignment correction and super-resolution in an end-to-end manner. It incorporates a misalignment estimation block that takes the thermal image and a stack of warped guides as inputs and computes an optimal translation map by formulating the task as a classification problem. The proposed network aims to enable guided super-resolution from unaligned low-resolution thermal images, eliminating the need for pixel-to-pixel alignment between the thermal and guide images. The base network incorporates dense blocks and self-attention modules to enhance the merging of cross-domain features at a global level (see Fig. 7).

The models are trained with $L1$ loss and an edge loss term that encourages sharpness in the reconstructed images. Furthermore, UGSR-FA incorporates a feature-alignment loss, labeled as LFA, while UGSR-ME employs the $L0$ norm on the depth map. The UGSR-FA method addresses the blurriness and edge ambiguity issues that arise when fusing features from misaligned thermal images and guiding images to reconstruct the high-resolution image. Simply using a pixel-wise loss like $L1$ does not effectively reduce blurriness since the misalignment occurs in the feature space, and the back-propagated gradients may not perceive the spatial shift as the cause of blurriness. To overcome this, additional constraints are imposed on the network to accommodate the misalignment. The alignment of features in the feature space is achieved by introducing a loss function that enhances the cross-correlation between features. UGSR-ME addresses the problem through the estimation and correction of misalignment before performing guided super-resolution. However, explicitly estimating a dense misalignment map can be highly challenging and is not necessary for guided super-resolution. To avoid the additional complexity, a model is proposed that incorporates alignment correction as part of the super-resolution process by estimating the misalignment in an end-to-end manner and applying the estimated misalignment to the guide image.

**Pyramidal Edge-Maps and Attention-based Guided Thermal Super-Resolution — PAGSR:** The approach presented in [32] focuses on guided super-resolution of thermal images using visible images as guidance. The method utilizes pyramidal edge maps to mitigate artifacts in the enhanced images. These edge maps are obtained from multiple hierarchical levels, which are derived by merging previously extracted edge features. These extracted edge features, called gedges, contain high-frequency orientation information. The spatial resolution edges obtained are then compared to that of the low-resolution thermal image using an average pooling layer. Attention mechanisms are employed to integrate these edge maps into the super-resolution network. All the spatial information is integrated into attention blocks, which are responsible for fusing the orientation information with the thermal super-resolution network. For each edge map, a set of merging

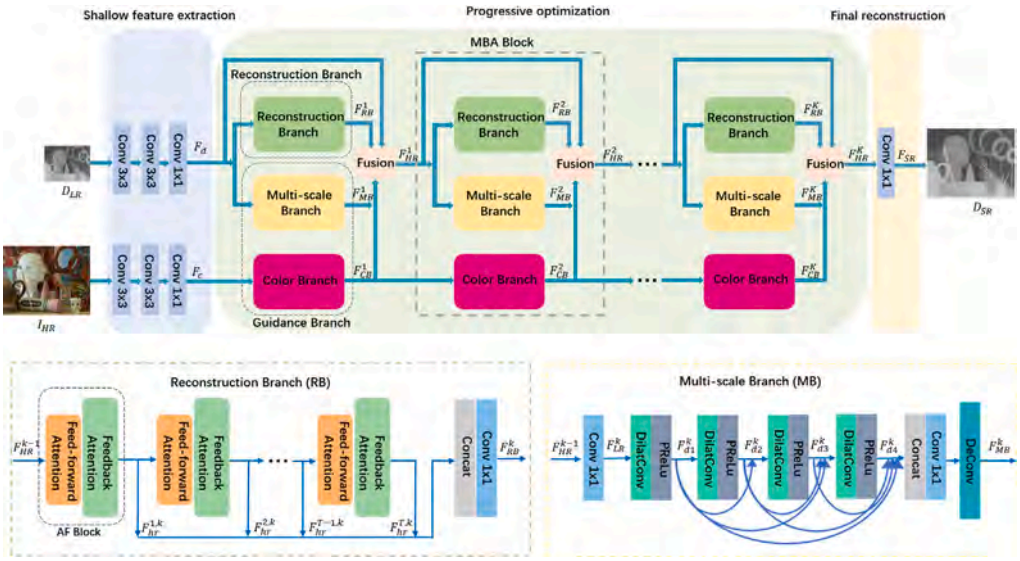


**Fig. 6.** PMBANet architecture.
*Source:* Illustration from [30].

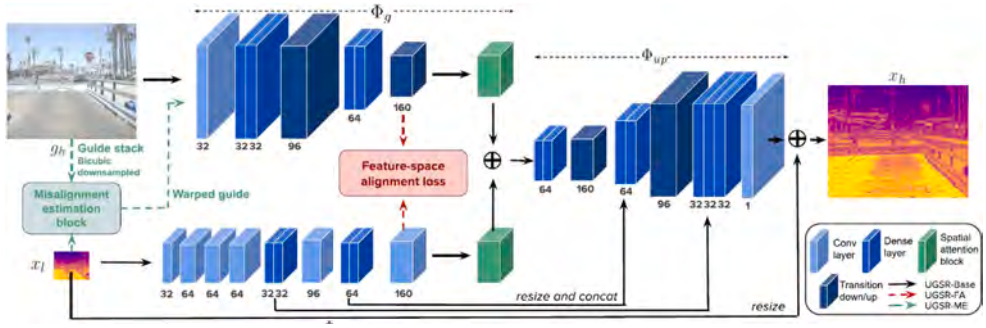


**Fig. 7.** UGSR architecture.
*Source:* Illustration from [31].

subblocks is created, which establishes the connections in the thermal super-resolution network.

The main challenge addressed by this approach is to effectively extract high-frequency details from the guide image and adaptively integrate them with the thermal image to produce a visually pleasing and artifact-free reconstructed image. By applying this method, the texture of objects in the guided super-resolution images is enhanced. The architecture, as shown in Fig. 8, illustrates the low-resolution thermal image and the edges used as inputs to the process. The proposed method uses a guide image that has a higher resolution compared to the input thermal image. The guide image contains valuable high-frequency details that can improve the super-resolution process. To effectively integrate these details, an adaptive blending module is employed, which takes advantage of edge maps extracted from the visible image at different pyramidal levels.

**Deformable Kernel Networks for Joint Image Filtering — DKN:** In [33], the authors introduce a new approach to joint image filtering by revisiting the guided weighted average framework. They argue that current CNN-based methods, which employ spatially-invariant kernels, have limitations in encoding structural details that vary with image location in both the guidance and target images. To overcome this limitation, the authors propose a data-driven approach that explicitly utilizes spatially-variant kernels, similar to classical approaches using the weighted average. This network consists of two parts: the first one is in charge of learning spatially varying kernel weights and the second one is the spatial sampling that is compensated with the normal grid. Features are extracted from individual targeting and target images. A two-stream CNN network is used where each subnet is in charge of one of two images, with different feature maps used to estimate the corresponding kernel weights and offsets. After, a weighted average is calculated using the learned kernel weights and the sample locations calculated from the offsets to obtain a residual image. The dual monitoring information for the weights and offsets in this model learns these parameters by directly minimizing the discrepancy between the network output and a reference image. In particular, the constraints on the regression of weight and offset, mean and sigmoid subtraction layers, specify how kernel weights and offsets behave and guide the learning process avoiding loss resolution. For the weight regression, it is applied a sigmoid layer that makes all elements greater than 0 and less than 1. The filtering result is obtained by combining the residuals with the target image. The key innovation is the use of a deformable kernel network (DKN), a CNN architecture designed to learn the sampling locations of neighboring pixels and their corresponding kernel weights for each pixel. This approach to image restoration uses spatially-variant kernels, which are learned in a data-driven way, which allows for an adaptive and sparse neighborhood system for each pixel. This can be more effective than a hand-designed kernel, as it can better adapt to the specific needs of the pixel being restored. The top of Fig. 9 shows the DKN architecture of the guided filtering. The model is fully convolutional and learned end-to-end, using element-wise multiplication and dot product operations. Reshaping and residual connections are also used. Additionally, the authors propose a more efficient alternative to

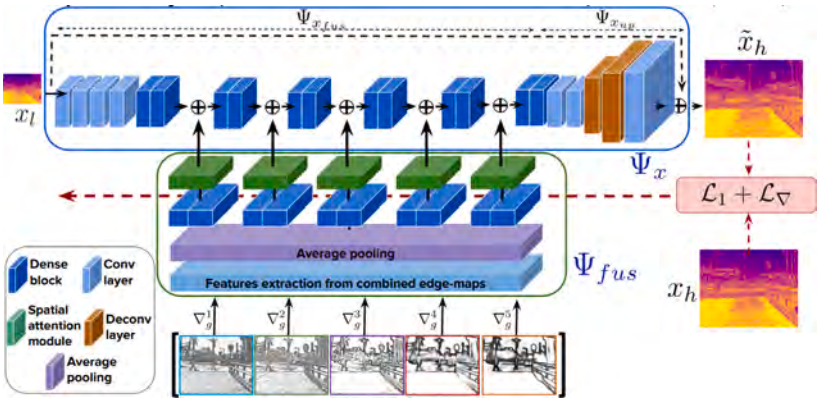


**Fig. 8.** Pyramidal edge-maps and attention-based architecture for guiding thermal image super-resolution. *Source:* Illustration from [32].

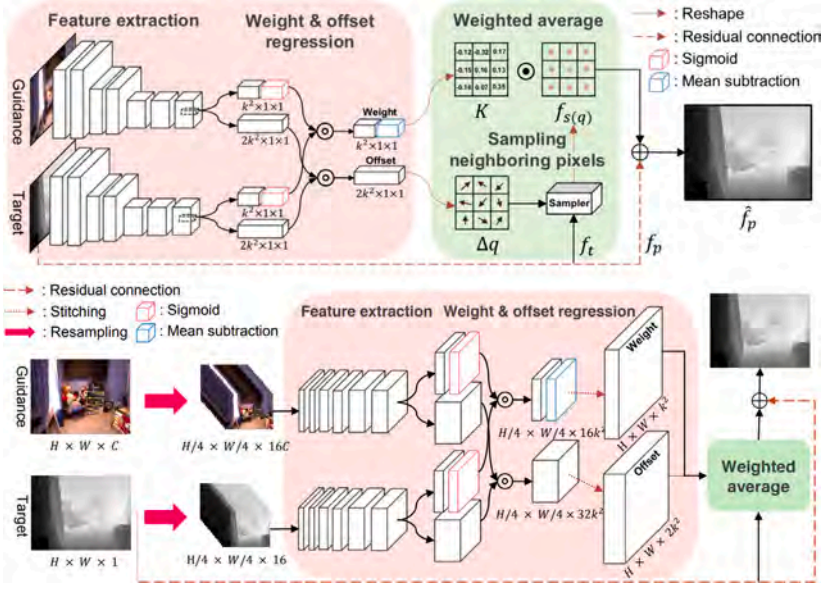


**Fig. 9.** (*top*) DKN architecture, (*bottom*) FDKN architecture.
*Source:* Illustrations from [33].

the deformable kernel network, called FDKN as shown at the bottom of Fig. 9. This lightweight model is obtained by removing DownConv layers while maintaining the same receptive field size as DKN. The filtering output can be obtained in a single forward pass by splitting the input images into $n$ subsampled and shifted parts, stacking them into new target and guidance images, and then using the target image to filter the guidance image. The target image has n channels, while the guidance image has $nC$ channels, where $C$ represents the number of channels in the guidance image. FDKN achieves a comparable effective receptive field to DKN, but with significantly fewer parameters due to the reduced input image resolution and shared weights across channels. The individual channels are then combined to create the final upsampled image. Even though FDKN has more network parameters than DKN, it still demonstrates a 17 times speedup compared to DKN while maintaining competitive performance.

**Discrete Cosine Transform Network for Guided Depth Map Super-Resolution — DCTNet:** In [34], the authors propose a discrete cosine transform network (DCTNet), to perform guided depth super-resolution, which is inspired by coupled dictionary learning and physically-based modeling. DCTNet comprises four components: semi-coupled feature extraction (SCFE), guided edge spatial attention (GESA), a discrete cosine transform (DCT) module, and a depth reconstruction (DR) module, (see Fig. 10). The SCFE leverages the correlation between intensity edges in RGB images and depth discontinuities while preserving unique properties in both modalities. The depth reconstruction module in this schema has the task of generating a high-resolution depth map by using a feature map, which is obtained from the DCT module. The purpose is to predict the detailed depth information based on the extracted features. The DCT module enhances explainability by utilizing DCT to solve an optimization model and acquire depth map features guided by RGB features. This module utilizes DCT to solve a well-designed optimization model for GDSR and integrates it into the deep learning model as a module. It acquires high-resolution depth map features driven by RGB features in the multichannel feature domain. Notably, this is the first known usage of DCT to restore degraded depth maps, and the fitting parameters in the DCT module are made learnable to enhance the flexibility of the model.

The proposed model introduces semi-coupled residual blocks to exploit the correlation between the intensity edge in RGB images and the depth discontinuities in the depth map images. These blocks combine the information from both images to produce a more accurate restoration. These blocks aim to preserve unique properties such as detailed texture and segment smoothness in both modalities. Each convolutional layer within these blocks is divided into two parts. One part focuses on extracting depth-shared information from RGB images, while the other part extracts unique information from the depth and

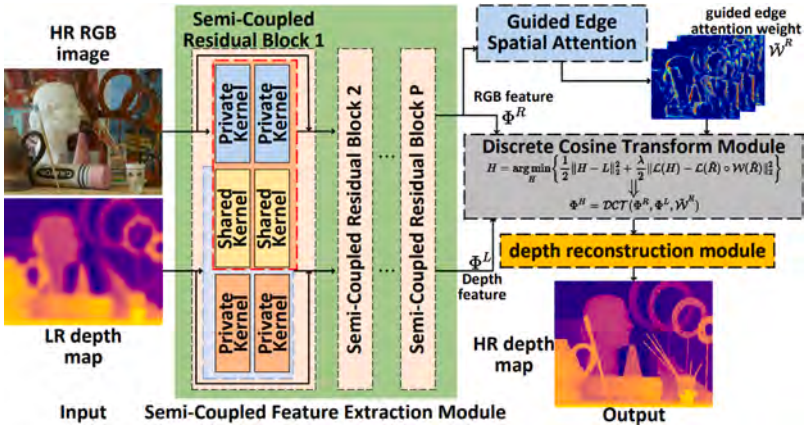


**Fig. 10.** Guided depth super-resolution architecture.
*Source:* Illustration from [34].

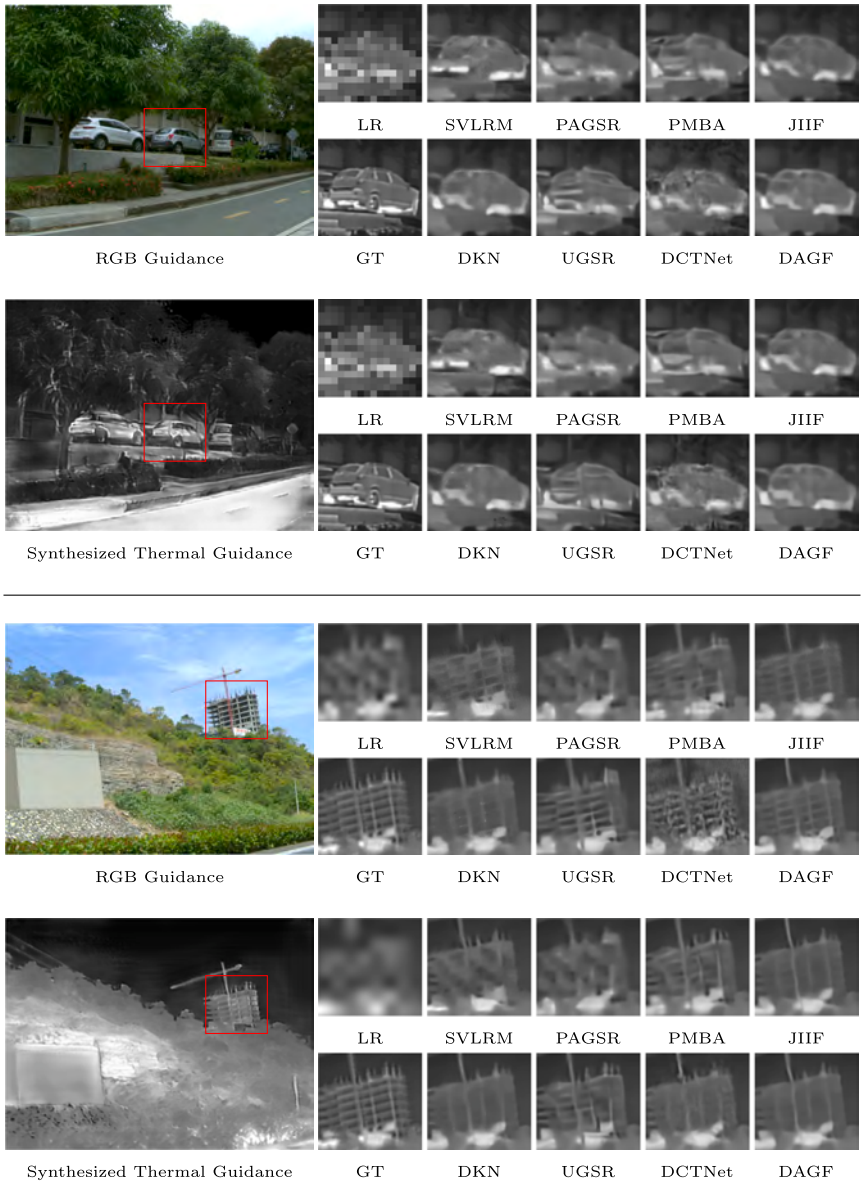


**Fig. 11.** Results on Thermal Stereo testing set when RGB guidance and synthesized thermal guidance are considered with ×8 super-resolution factor.

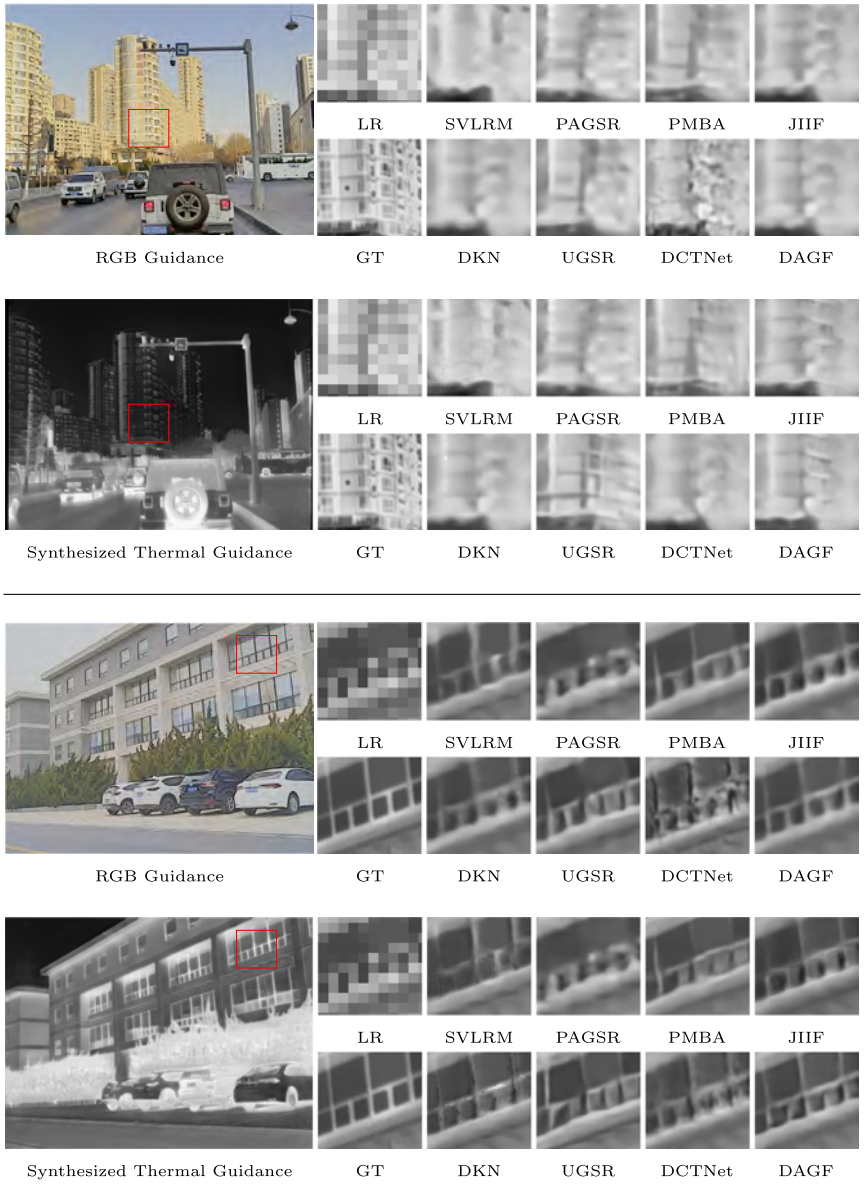


**Fig. 12.** Results on M3FD testing set when RGB guidance and synthesized thermal guidance are considered with ×8 super-resolution factor.

RGB images separately. The parameters in the private kernel are not shared, allowing the feature extractor with semi-coupled blocks to effectively extract feature information for guided depth super-resolution from input image pairs.

To address the issue of over-transferred texture details in RGB images, the GESA module employs RFANet's Enhanced Spatial Attention block. This block highlights the edges in RGB images that are relevant for guided depth super-resolution. Activating intensity edges associated with depth discontinuities facilitates the adaptive transfer of texture structure in guided imagery.

### 3.3. Dataset

The proposed strategy has been assessed using three cross-spectral datasets. The first dataset, referred to as M3FD, was recently introduced for image fusion in Liu et al. [46]. This dataset was used to train the thermal image-like image generator described in Section 3.1. The M3FD dataset consists of a substantial collection of 4500 pairs of aligned visible and infrared images captured using a binocular optical and infrared sensor. These images possess a resolution of 1024 × 768 pixels and exhibit diverse scenes encompassing various environments such as roads, campuses, streets, forests, and more. Furthermore, the dataset covers different lighting conditions including daytime, nighttime, and overcast scenarios, providing a comprehensive representation of real-world scenarios. In the experiments, a subset of 3000 image pairs from the M3FD dataset was used for training the proposed thermal image generator, while 890 pairs were set aside for validation purposes. The remaining images in the dataset were exclusively reserved for model testing, ensuring an unbiased evaluation of performance. By employing this dataset, the proposed approach was trained to generate thermal image-like representations from the visible images, facilitating subsequent super-resolution processes.

In addition to the M3FD dataset, the Flir ADAS V2 [47] and the Thermal Stereo datasets [20] have been also used to evaluate the generalization capabilities of the proposed strategy. The Flir ADAS V2 dataset is a set of annotated images that are used for training autonomous driving systems. The dataset was acquired via a thermal and visible camera pair mounted on a vehicle. Thermal images were acquired with a Teledyne FLIR Tau 2 13 mm f/1.0. Visible images were captured with a Teledyne FLIR BlackFly S BFS-U3-51S5C (IMX250) camera. Time-synced capture was executed by Teledyne FLIR's Guardian software;

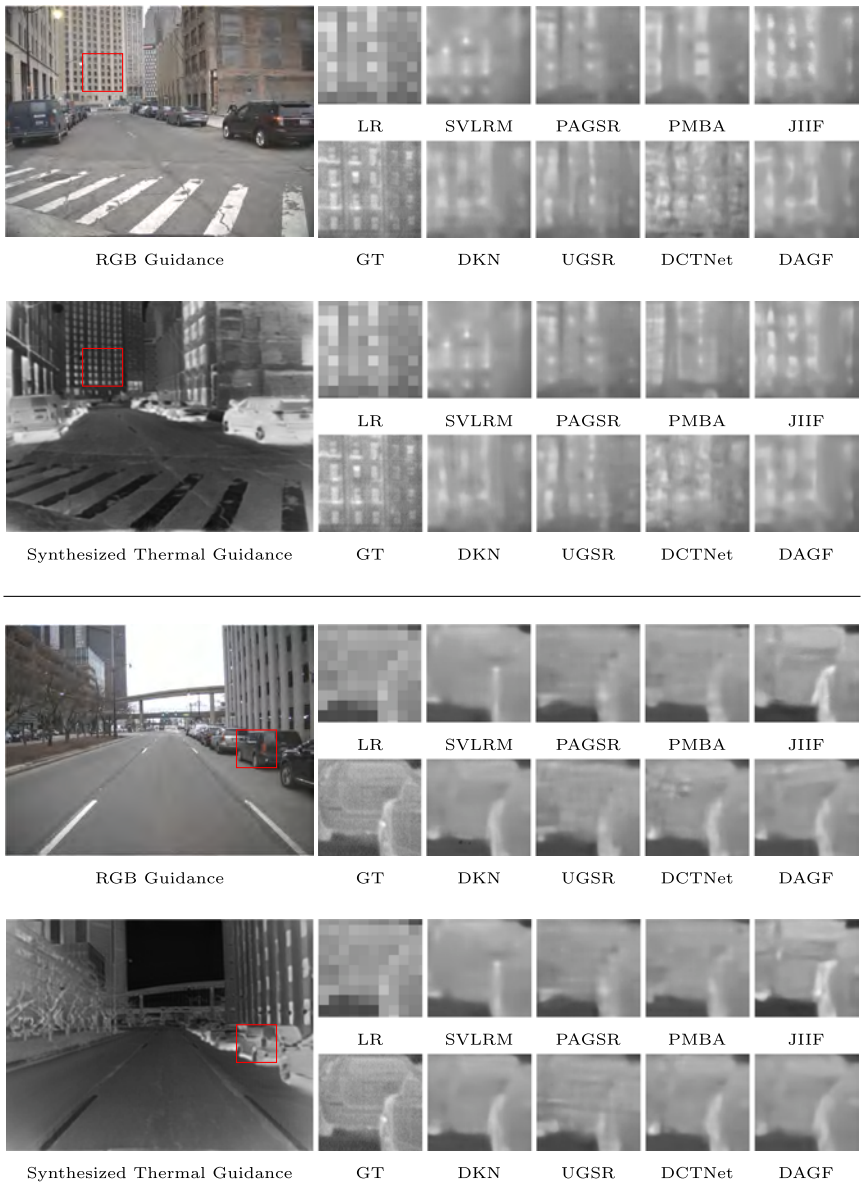


**Fig. 13.** Results on Flir ADAS V2 testing set when RGB guidance and synthesized thermal guidance are considered with ×8 super-resolution factor.

the dataset consists of 9233 pairs of RGB-Thermal images. On the other hand, the Thermal Stereo dataset was acquired using Basler and TAU2 cameras and consists of 200 pairs of images, comprising both visible and their corresponding thermal images. To ensure accurate alignment between the two modalities, the images were registered using the Elastix method [48], resulting in well-aligned pairs with a resolution of 640 × 480 pixels. This unique dataset provided a valuable opportunity to assess the strategy's performance on a different set of cross-spectral images, enabling a comprehensive evaluation of its effectiveness in various scenarios.

## 4. Experimental results

This section presents a comprehensive evaluation of the state-of-the-art guided super-resolution approaches introduced in Section 3.2. Our evaluation encompasses two distinct strategies: (*i*) using given high-resolution RGB images as guidance during the training; and (*ii*) considering HR thermal image-like counterparts. To conduct quantitative and qualitative evaluations, we employed the Thermal Stereo, M3FD and Flir ADAS V2 datasets. Before evaluation, all images were pre-processed by resizing them to a resolution of 640 × 480 pixels. For the generation of low-resolution thermal images, we applied a downsampling using bicubic interpolation on the HR thermal images. To train the guided super-resolution methods, we divided the Thermal Stereo dataset into three subsets: 160 image pairs for training, 30 image pairs for validation, and 10 image pairs for testing.

Our evaluation focused on the nine state-of-the-art guided super-resolution methods presented in Section 3.2. These methods were evaluated using the proposed strategy, which involves training on both visible and synthesized thermal images, with a scale factor of ×8 and ×16. To assess the performance of these methods, we employed SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio).

Tables 1, 2 and 3 present the results obtained through the evaluation of each guided super-resolution approach using a scale factor of ×8 for Thermal Stereo, M3FD and Flir ADAS V2 datasets respectively. Similarly, Tables 4, 5 and 6 show the results obtained with a scale factor of ×16 for Thermal Stereo, M3FD and Flir ADAS V2 datasets—in the ×16 scale factor case just seven approaches are evaluated since the code provided for PAGSR [32] and UGSR [31] does not include the ×16 upsampling. All tables highlight the improvements achieved in terms

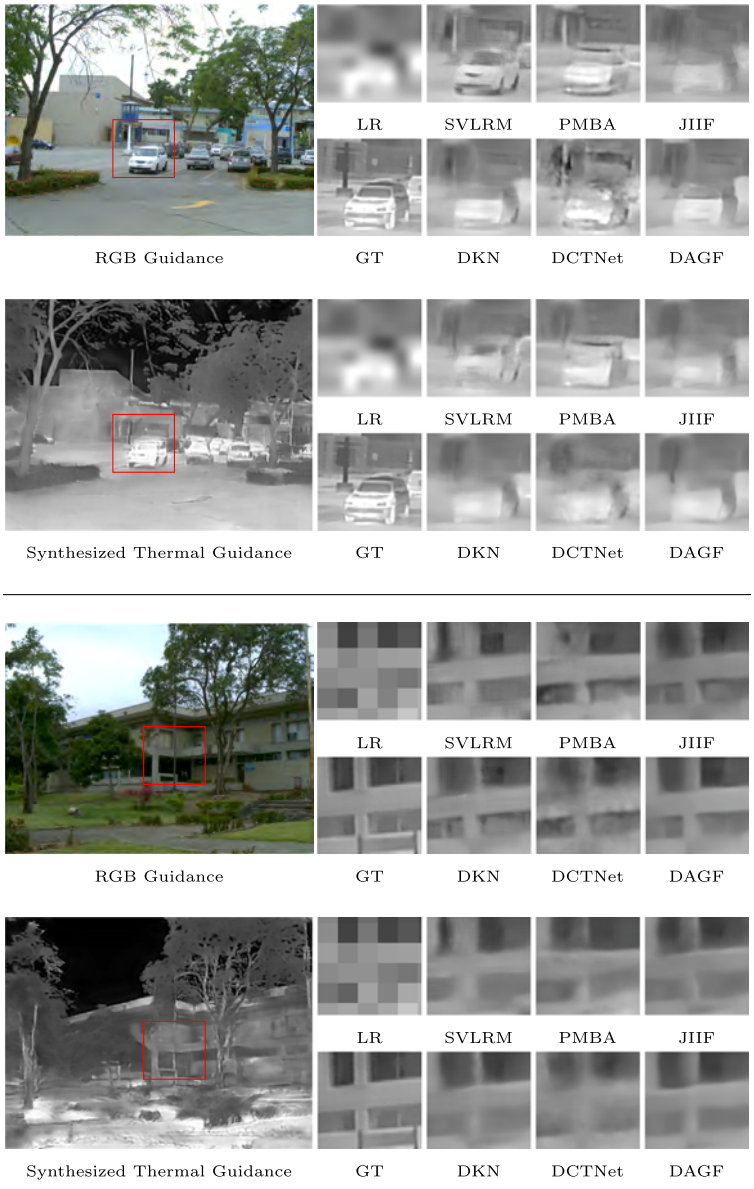


**Fig. 14.** Results on Thermal Stereo testing set when RGB guidance and synthesized thermal guidance are considered with × 16 super-resolution factor.

**Table 1**
Results of the guided super-resolution approaches evaluated in the current work with Thermal Stereo Dataset, a ×8 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 27.05 | 0.7840 | 27.46 | 0.8202 | 1.5% | 4.6% |
| PAGSR [32] | 26.47 | 0.8008 | 26.47 | 0.8008 | 0.0% | 0.0% |
| PMBA [30] | 27.38 | 0.8242 | 27.86 | 0.8374 | 1.7% | 1.6% |
| JIIF [28] | 26.56 | 0.7879 | 26.77 | 0.8109 | 0.8% | 2.9% |
| DKN [33] | 29.11 | 0.8254 | 29.32 | 0.8313 | 0.7% | 0.7% |
| FDKN [33] | 29.05 | 0.8215 | 29.20 | 0.8276 | 0.5% | 0.7% |
| UGSR [31] | 26.83 | 0.8099 | 27.20 | 0.8241 | 1.4% | 1.8% |
| DCTNet [34] | 23.57 | 0.6729 | 28.34 | 0.8037 | 20.3% | 19.4% |
| DAGF [27] | 29.22 | 0.8268 | 29.51 | 0.8354 | 1.0% | 1.0% |

of quantitative metrics when employing synthesized thermal images as guidance. To facilitate analysis and provide a clearer understanding of the results, the last column of the tables presents the percentage improvement achieved when using synthetic thermal imaging as a guide compared to using visible spectrum imaging. It is important to mention that the training of the guided super-resolution model with synthetic images has been carried out with the Thermal Stereo data set, however, to validate the proposed strategy, the analysis has been expanded by testing two additional data sets, the M3FD and Flir ADAS V2. The obtained results demonstrate that the model trained using the Thermal Stereo data set is robust and generalizable. Since better results can be seen even when validated with other data sets such as M3FD and Flir ADAS V2. Qualitatively, the proposed strategy exhibits superior performance in terms of contour details compared with the approaches guided by visible spectrum images. To provide a visual representation of the evaluation results, Figs. 11–13 present results obtained from a sample of images in the testing set. Each super-resolution method's performance is shown based on both guiding strategies, using a scale factor of ×8.

Figs. 14–16 present comparisons for a sample of images in the testing set, using a scale factor of ×16. In this figures we can appreciate the enhanced reconstruction of structural details when employing synthesized thermal images as guidance.

The results obtained in the experiments for a scale factor of ×8 and ×16, allow us to demonstrate that the model trained with the synthetic

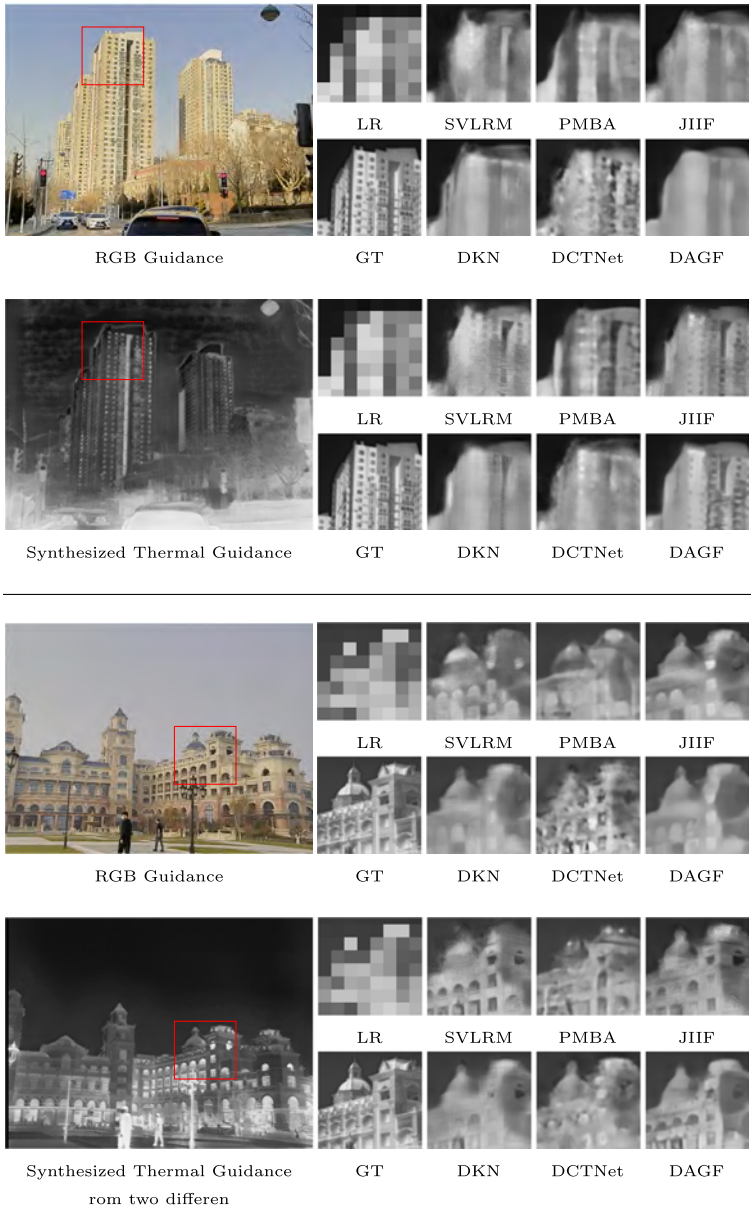


**Fig. 15.** Results on M3FD testing set when RGB guidance and synthesized thermal guidance are considered with ×16 super-resolution factor.

**Table 2**
Results of the guided super-resolution approaches evaluated in the current work with M3FD Dataset, a ×8 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 25.65 | 0.7579 | 25.67 | 0.7898 | 0.1% | 4.2% |
| PAGSR [32] | 24.89 | 0.7736 | 24.89 | 0.7736 | 0.0% | 0.0% |
| PMBA [30] | 26.08 | 0.7961 | 26.31 | 0.8011 | 0.9% | 0.6% |
| JIIF [28] | 26.03 | 0.7646 | 26.38 | 0.7672 | 1.3% | 0.3% |
| DKN [33] | 28.13 | 0.8038 | 28.48 | 0.8119 | 1.2% | 1.0% |
| FDKN [33] | 28.02 | 0.7992 | 28.22 | 0.8049 | 0.7% | 0.7% |
| UGSR [31] | 24.70 | 0.7596 | 25.07 | 0.7706 | 1.5% | 1.4% |
| DCTNet [34] | 20.92 | 0.5636 | 25.19 | 0.7569 | 20.4% | 34.3% |
| DAGF [27] | 28.04 | 0.8006 | 28.56 | 0.8128 | 1.9% | 1.5% |

**Table 3**
Results of the guided super-resolution approaches evaluated in the current work with Flir ADAS V2 Dataset, a ×8 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 27.25 | 0.6550 | 27.32 | 0.6565 | 0.3% | 0.2% |
| PAGSR [32] | 26.56 | 0.6610 | 26.56 | 0.6610 | 0.0% | 0.0% |
| PMBA [30] | 27.20 | 0.6608 | 27.82 | 0.6758 | 2.3% | 2.3% |
| JIIF [28] | 27.03 | 0.6723 | 27.39 | 0.6786 | 1.3% | 0.9% |
| DKN [33] | 28.98 | 0.6702 | 29.01 | 0.6713 | 0.1% | 0.2% |
| FDKN [33] | 29.00 | 0.6703 | 29.01 | 0.6706 | 0.0% | 0.0% |
| UGSR [31] | 26.85 | 0.6582 | 27.34 | 0.6672 | 1.8% | 1.4% |
| DCTNet [34] | 28.45 | 0.6591 | 28.92 | 0.6702 | 1.7% | 1.7% |
| DAGF [27] | 28.88 | 0.6679 | 29.09 | 0.6733 | 0.8% | 0.8% |

images of the thermal stereo dataset has been able to generalize the process of restoring the quality of the images so effectively that the same model has been used to evaluate the super-resolution of the images from M3FD and Flir ADAS V2 datasets. These findings highlight the efficacy of the proposed strategy in enhancing guided super-resolution outcomes.

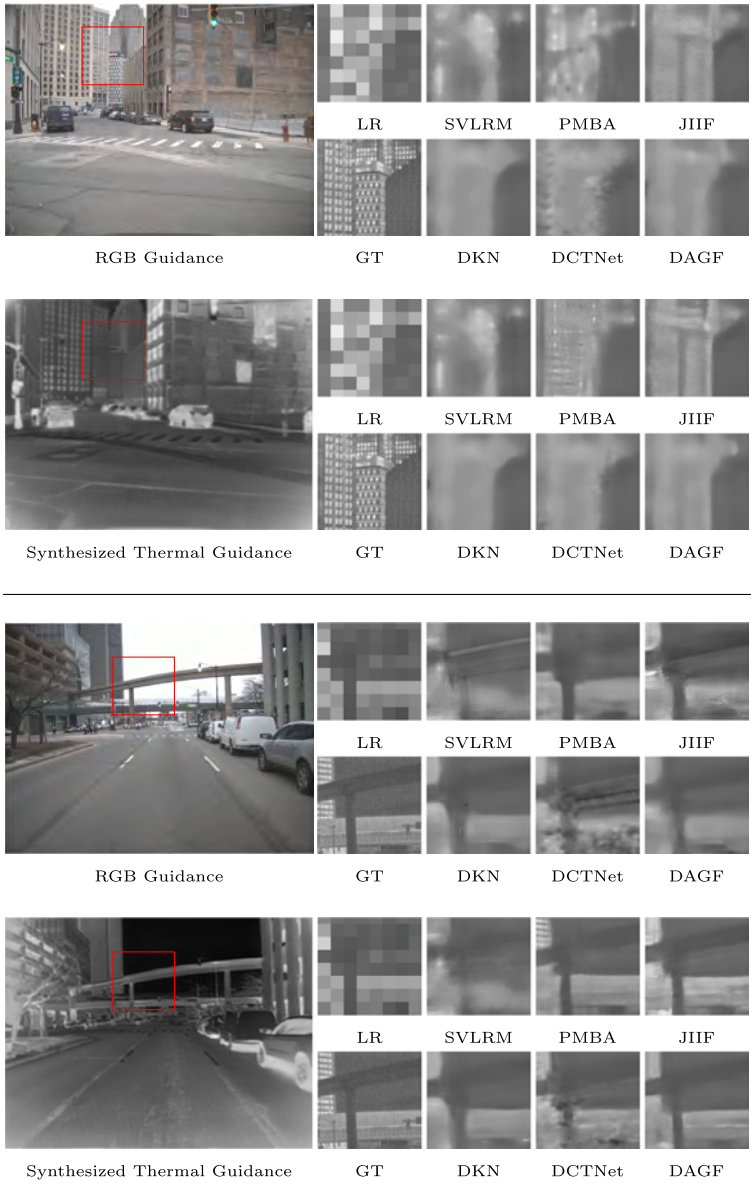


**Fig. 16.** Results on Flir ADAS V2 testing set when RGB guidance and synthesized thermal guidance are considered with ×16 super-resolution factor.

**Table 4**
Results of the guided super-resolution approaches evaluated in the current work with Thermal Stereo Dataset, a ×16 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 22.71 | 0.6429 | 25.37 | 0.7489 | 7.3% | 16.5% |
| PMBA [30] | 23.14 | 0.6966 | 24.87 | 0.7685 | 7.5% | 10.3% |
| JIIF [28] | 23.75 | 0.7248 | 24.69 | 0.7592 | 4.0% | 4.7% |
| DKN [33] | 25.33 | 0.7273 | 26.11 | 0.7674 | 3.1% | 5.5% |
| FDKN [33] | 25.35 | 0.7296 | 26.07 | 0.7641 | 2.8% | 4.7% |
| DCTNet [34] | 23.21 | 0.6498 | 25.18 | 0.7330 | 8.5% | 12.8% |
| DAGF [27] | 26.11 | 0.7581 | 26.58 | 0.7778 | 1.8% | 2.6% |

Table 7 presents the details of each of the super-resolution models evaluated in the current work (number of parameters, inference time, training time, and memory usage).

**Table 5**
Results of the guided super-resolution approaches evaluated in the current work with M3FD Dataset, a ×16 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 21.46 | 0.6282 | 22.27 | 0.7102 | 3.8% | 13.1% |
| PMBA [30] | 21.76 | 0.6866 | 22.60 | 0.7264 | 3.8% | 5.8% |
| JIIF [28] | 22.66 | 0.6743 | 23.50 | 0.7037 | 3.7% | 4.4% |
| DKN [33] | 24.28 | 0.7116 | 24.73 | 0.7336 | 1.8% | 3.1% |
| FDKN [33] | 24.15 | 0.7067 | 24.58 | 0.7311 | 1.8% | 3.5% |
| DCTNet [34] | 21.16 | 0.5842 | 23.19 | 0.6947 | 9.6% | 18.9% |
| DAGF [27] | 24.48 | 0.7228 | 25.15 | 0.7445 | 2.7% | 3.0% |

## 5. Conclusions

The present study provides empirical evidence that guided image processing techniques, specifically in the context of guided thermal image super-resolution, can be enhanced by incorporating guiding information that overlaps with the guided domain. This observation emphasizes the importance of aligning the guiding information with the

**Table 6**
Results of the guided super-resolution approaches evaluated in the current work with Flir ADAS V2 Dataset, a ×16 scale factor is considered.

| Methods | RGB guidance | | Synt. guidance | | Improvement on | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SVLRM [29] | 24.72 | 0.6102 | 25.15 | 0.6181 | 1.7% | 1.3% |
| PMBA [30] | 24.88 | 0.6207 | 25.64 | 0.6344 | 3.1% | 2.2% |
| JIIF [28] | 23.74 | 0.6126 | 24.48 | 0.6307 | 3.1% | 2.9% |
| DKN [33] | 26.80 | 0.6280 | 26.83 | 0.6282 | 0.1% | 0.0% |
| FDKN [33] | 26.70 | 0.6256 | 26.78 | 0.6269 | 0.3% | 0.2% |
| DCTNet [34] | 25.72 | 0.6005 | 26.43 | 0.6164 | 2.8% | 2.7% |
| DAGF [27] | 26.60 | 0.6238 | 26.91 | 0.6301 | 1.1% | 1.0% |

**Table 7**
Information about the training process for each guided super-resolution approach evaluated in the current work.

| Methods | # Param. (M) | Training Time (h) | Inf. Time (s) | Inf. Memory (MiB) |
|---|---|---|---|---|
| SVLRM [29] | 0.37 | 5.038 | 0.0732 | 10 239 |
| PAGSR [32] | 0.18 | 1.983 | 0.4673 | 6605 |
| PMBA [30] | 46.04 | 3.366 | 0.2019 | 4819 |
| JIIF [28] | 10.83 | 4.283 | 0.8929 | 7629 |
| DKN [33] | 1.16 | 2.950 | 0.4587 | 15 595 |
| FDKN [33] | 0.69 | 0.514 | 0.3571 | 3444 |
| UGSR [31] | 2.17 | 1.455 | 0.7246 | 7533 |
| DCTNet [34] | 0.48 | 1.656 | 0.3322 | 3325 |
| DAGF [27] | 2.44 | 1.383 | 0.2564 | 7605 |

target domain to achieve superior results. The experiments conducted in this work focus on generating synthesized thermal images that closely resemble real thermal images. These synthesized images serve as effective guidance in the image processing pipeline, facilitating the super-resolution process. By using synthesized images as guidance, the evaluated models are able to acquire the necessary information for super-resolution tasks with ease. As a future research endeavor, the scope of this study will be extended to explore guided depth super-resolution and guided denoising image processing. By expanding the investigation to these areas, additional insights can be gained into the potential benefits and limitations of guided image processing approaches. This expanded exploration will contribute to advancing the field of image enhancement and provide a deeper understanding of the capabilities of guided techniques in various image-processing tasks.

## CRediT authorship contribution statement

**Patricia L. Suárez:** Investigation, Writing – original draft, Writing – review & editing, Methodology, Validation. **Dario Carpio:** Resources, Writing – original draft, Writing – review & editing. **Angel D. Sappa:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Angel Sappa reports financial support was provided by the Air Force Office of Scientific Research under award number FA9550-22-1-026.

## Data availability

Data will be made available on request.

## Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research, USA under award number FA9550-22-1-0261; and partially supported by the Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033, Spain and by “ERDF A way of making Europe”; the “CERCA75046 Programme / Generalitat de Catalunya, Spain”; and the ESPOL, Ecuador project CIDIS-12-2022.

## References

[1] R. Gade, T.B. Moeslund, Thermal cameras and applications: A survey, Mach. Vis. Appl. 25 (2014) 245–262.
[2] A.W. Van Eekeren, K. Schutte, L.J. Van Vliet, Multiframe super-resolution reconstruction of small moving objects, IEEE Trans. Image Process. 19 (11) (2010) 2901–2912.
[3] Z. Wang, J. Chen, S.C. Hoi, Deep learning for image super-resolution: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2020) 3365–3387.
[4] T. Kato, H. Hino, N. Murata, Double sparsity for multi-frame super resolution, Neurocomputing 240 (2017) 115–126.
[5] W. Sun, Y. Zhang, Attention-guided dual spatial-temporal non-local network for video super-resolution, Neurocomputing 406 (2020) 24–33.
[6] W. Yu, Z. Li, Q. Liu, F. Jiang, C. Guo, S. Zhang, Scale-aware frequency attention network for super-resolution, Neurocomputing (2023) 126584.
[7] S. Zhuo, X. Zhang, X. Miao, T. Sim, Enhancing low light images using near infrared flash images, in: Proceedings of the IEEE International Conference on Image Processing, 2010.
[8] T. Shibata, M. Tanaka, M. Okutomi, Misalignment-robust joint filter for cross-modal image pairs, in: Proceedings of the IEEE International Conference on Computer Vision, 2017.
[9] Z. Fan, T. Dan, B. Liu, X. Sheng, H. Yu, H. Cai, SGUNet: Style-guided UNet for adversely conditioned fundus image super-resolution, Neurocomputing 465 (2021) 238–247.
[10] H. Wu, L. Zhang, J. Ma, Remote sensing image super-resolution via saliency-guided feedback GANs, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–16.
[11] X. Qin, X. Gao, K. Yue, Remote sensing image super-resolution using multi-scale convolutional neural network, in: Proceedings of the 11th UK-Europe-China Workshop on Millimeter Waves and Terahertz Technologies. Vol. 1, UCMMT, 2018.
[12] M. Ramzy Ibrahim, R. Benavente, D. Ponsa, F. Lumbreras, Unveiling the influence of image super-resolution on aerial scene classification, in: Iberoamerican Congress on Pattern Recognition, Springer, 2023, pp. 214–228.
[13] Z. Yue, J. Wang, C.C. Loy, Resshift: Efficient diffusion model for image super-resolution by residual shifting, 2023, arXiv.
[14] R.d. Lutio, S. D'aronco, J.D. Wegner, K. Schindler, Guided super-resolution as pixel-to-pixel transformation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
[15] Z. Luo, Y. Li, S. Cheng, L. Yu, Q. Wu, Z. Wen, H. Fan, J. Sun, S. Liu, BSRT: Improving burst super-resolution with swin transformer and flow-guided deformable alignment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.
[16] L. Yanshan, Z. Li, X. Fan, C. Shifu, OGSRN: Optical-guided super-resolution network for SAR image, Chin. J. Aeronaut. 35 (5) (2022) 204–219.
[17] Z. Zhong, X. Liu, J. Jiang, D. Zhao, X. Ji, Deep attentional guided image filtering, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[18] Z. Zhao, Y. Zhang, C. Li, Y. Xiao, J. Tang, Thermal UAV image super-resolution guided by multiple visible cues, IEEE Trans. Geosci. Remote Sens. (2023).
[19] L. Cheng, M. Kersemans, Dual-IRT-GAN: A defect-aware deep adversarial network to perform super-resolution tasks in infrared thermographic inspection, Composites B 247 (2022).
[20] R.E. Rivadeneira, A.D. Sappa, B.X. Vintimilla, D. Bin, L. Ruodi, L. Shengye, Z. Zhong, X. Liu, J. Jiang, C. Wang, Thermal image super-resolution challenge results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2023.
[21] M. Zhang, Q. Wu, J. Guo, Y. Li, X. Gao, Heat transfer-inspired network for image super-resolution reconstruction, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[22] N. Metzger, R.C. Daudt, K. Schindler, Guided depth super-resolution by deep anisotropic diffusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.
[23] W. Liu, X. Chen, J. Yang, Q. Wu, Robust color guided depth map restoration, IEEE Trans. Image Process. 26 (2017) 315–327.
[24] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, P. Han, Hierarchical features driven residual learning for depth map super-resolution, IEEE Trans. Image Process. 28 (5) (2018) 2545–2557.
[25] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, W. Gao, Depth super-resolution via joint color-guided internal and external regularizations, IEEE Trans. Image Process. 28 (4) (2018) 1636–1645.

[26] Y. Qiao, L. Jiao, W. Li, C. Richardt, D. Cosker, Fast, high-quality hierarchical depth-map super-resolution, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 4444–4453.
[27] Z. Zhong, X. Liu, J. Jiang, D. Zhao, X. Ji, Deep attentional guided image filtering, IEEE Trans. Neural Netw. Learn. Syst. (2023) 1–15.
[28] J. Tang, X. Chen, G. Zeng, Joint implicit image function for guided depth super-resolution, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021.
[29] J. Pan, J. Dong, J.S. Ren, L. Lin, J. Tang, M.-H. Yang, Spatially variant linear representation models for joint filtering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[30] X. Ye, B. Sun, Z. Wang, J. Yang, R. Xu, H. Li, B. Li, PMBANet: Progressive multi-branch aggregation network for scene depth super-resolution, IEEE Trans. Image Process. 29 (2020) 7427–7442.
[31] H. Gupta, K. Mitra, Toward unaligned guided thermal super-resolution, IEEE Trans. Image Process. 31 (2021) 433–445.
[32] H. Gupta, K. Mitra, Pyramidal edge-maps and attention based guided thermal super-resolution, in: Proceedings of European Conference on Computer Vision Workshops, Springer, 2020.
[33] B. Kim, J. Ponce, B. Ham, Deformable kernel networks for joint image filtering, Int. J. Comput. Vis. 129 (2) (2021) 579–600.
[34] Z. Zhao, J. Zhang, S. Xu, Z. Lin, H. Pfister, Discrete cosine transform network for guided depth map super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.
[35] H. Xu, M. Gong, X. Tian, J. Huang, J. Ma, CUFD: An encoder–decoder network for visible and infrared image fusion based on common and unique feature decomposition, Comput. Vis. Image Underst. 218 (2022).
[36] B. Meher, S. Agrawal, R. Panda, L. Dora, A. Abraham, Visible and infrared image fusion using an efficient adaptive transition region extraction technique, Eng. Sci. Technol. Int. J. 29 (2022).
[37] B. Dogan, S. Gu, R. Timofte, Exemplar guided face image super-resolution without facial landmarks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[38] P.L. Suárez, Á.D. Sappa, Toward a thermal image-like representation, in: Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2023.
[39] P.L. Suárez, A.D. Sappa, B.X. Vintimilla, Infrared image colorization based on a triplet DCGAN architecture, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.
[40] A. Mehri, A.D. Sappa, Colorizing near infrared images through a cyclic adversarial approach of unpaired samples, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[41] P.L. Suárez, A.D. Sappa, B.X. Vintimilla, R.I. Hammoud, Image vegetation index through a cycle generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[42] J.C. Peterson, S. Uddenberg, T.L. Griffiths, A. Todorov, J.W. Suchow, Deep models of superficial face judgments, Proc. Natl. Acad. Sci. 119 (17) (2022).
[43] J. Zhao, F. Lee, C. Hu, H. Yu, Q. Chen, LD-GAN: Lightweight domain-attention GAN for unpaired image-to-image translation, Neurocomputing 506 (2022) 355–368.
[44] J. Wang, G. Xie, Y. Huang, J. Lyu, F. Zheng, Y. Zheng, Y. Jin, FedMed-GAN: Federated domain translation on unsupervised cross-modality brain image synthesis, Neurocomputing 546 (2023) 126282.
[45] A. Andonian, T. Park, B. Russell, P. Isola, J.-Y. Zhu, R. Zhang, Contrastive feature loss for image prediction, in: Proceedings of the IEEE International Conference on Computer Vision, 2021.
[46] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, Z. Luo, Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.
[47] FLIR, Thermal, Dataset, FREE teledyne FLIR thermal dataset for algorithm training, 2023, URL https://www.flir.com/news-center/. (Accessed on 07 November 2023).
[48] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P. Pluim, Elastix: A toolbox for intensity-based medical image registration, IEEE Trans. Med. Imaging 29 (1) (2009) 196–205.

**Patricia L. Suárez** graduated as a Computer Systems Engineer in 2002, obtained a Master's degree in Management Information Systems in 2008 from ESPOL, and a Ph.D. in applied computer science in 2020 at ESPOL. She received the Best Work Award at the CVPR-PBVS workshop in 2017. She has 12 years of experience in the professional field of computing. Since 2006, Patricia has actively participated as a professor at different universities in Ecuador. Additionally, since 2015 she is a researcher in the field of multispectral image processing and representation at the CIDIS research laboratory of ESPOL, Guayaquil, Ecuador. She has been a member of research teams in national and international projects (Meta, US Air Force Office of Scientific Research (AFOSR)).

**Darío Carpio** received his Computer Science Engineering degree from the ESPOL Polytechnic University, Guayaquil, Ecuador, in 2020. Currently, he is a member of the CIDIS research center. He works on super-resolution thermal and visible spectrum imaging, specializing in the research of new guided SR architectures.

**Angel Domingo Sappa** (IEEE S'94-M'00-SM'12) received the Electromechanical Engineering degree from the National University of La Pampa, General Pico, Argentina, in 1995, and the Ph.D. degree in Industrial Engineering from the Polytechnic University of Catalonia, Barcelona, Spain, in 1999. In 2003, after holding research positions in France, the U.K., and Greece, he joined the Computer Vision Center, in Barcelona, where he currently holds a Senior Scientist position. Since 2016 he has been a full professor at the ESPOL Polytechnic University, Guayaquil, Ecuador, where he leads the computer vision team at the CIDIS research center; he is the director of the Electrical Engineering Ph.D. program.