

ThermISRnet: an efficient thermal image super-resolution network

Heena Patel^{DOI},^a Vishal Chudasama^{DOI},^a Kalpesh Prajapati^{DOI},^a
Kishor P. Upla^{DOI},^{a,b,*} Kiran Raja^{DOI},^b Raghavendra Ramachandra^{DOI},^b
and Christoph Busch^{DOI},^b

^aSardar Vallabhbhai National Institute of Technology, Electronics Engineering Department,
Surat, India

^bNorwegian University of Science and Technology, Gjøvik, Norway

Abstract. The prime limitation of optical sensors is the need for external sources of illumination while capturing the scene. This prevents them from recognizing objects in extreme conditions, such as insufficient illumination or severe weather (e.g., under fog or smoke). The thermal imaging sensors have been introduced to circumvent this deficiency, which acquires the image based on thermal radiation emitted by the objects. The technological advancement in thermal imaging enables the visualization of objects beyond the visible range that promotes its use in many principal applications, such as military, medical, agriculture, etc. However, hardware point of view, the cost of a thermal camera is prohibitively higher than that of an equivalent optical sensor. This led to employ software-driven approaches called super-resolution (SR) to enhance the resolution of given thermal images. We propose a deep neural network architecture referred to as “ThermISRnet” as the extension of our earlier winner architecture in the Perception Beyond the Visible Spectrum (PBVS) thermal SR challenge. We use a progressive upscaling strategy with asymmetrical residual learning in the network, which is computationally efficient for different upscaling factors such as $\times 2$, $\times 3$, and $\times 4$. The proposed architecture consists of different modules for low- and high-frequency feature extraction along with upsampling blocks. The effectiveness of the proposed architecture in ThermISRnet is verified by evaluating it with different datasets. The obtained results indicate superior performance as compared to other state-of-the-art SR methods. © 2021 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.OE.60.7.073101](https://doi.org/10.1117/1.OE.60.7.073101)]

Keywords: thermal image super-resolution; convolutional neural network; image reconstruction; Perception Beyond the Visible Spectrum super-resolution challenge.

Paper 20210168 received Feb. 13, 2021; accepted for publication Jun. 9, 2021; published online Jul. 1, 2021.

1 Introduction

The optical camera senses the reflected energy from the object being observed, and hence they typically rely on the external source of illumination. The technological advancement of optical sensors has led to obtain very high-resolution (HR) images, prominently for high-end applications such as object recognition, tracking, activity monitoring, etc. However, one of the leading barrier concerned to optical imagery is the requirement of an external source of illumination, which may not be available in many situations such as night-time and/or severe weather conditions (e.g., foggy and rainy). Alternatively, the thermal camera’s choice provides the radiation emitted by the objects in the range of long-wavelength infrared (LWIR) spectrum and hence they are independent of illumination or background clutters. They are passive sensors and measure the infrared radiation in terms of temperature above absolute zero.¹ The recent progress in thermal imaging has made many real-world applications possible¹ in different domains, such as military,^{1,2} medical,^{3,4} pedestrian and person detection,^{5–7} fire detection,⁸ visual odometry,^{9,10} gas detection,¹¹ city planning,¹² building defect detection,¹³ precision farming,^{14,15} and maritime.¹⁶

*Address all correspondence to Kishor P. Upla, kishorupla@gmail.com

Despite the ability to capture images in challenging conditions, thermal cameras often come with limited spatial resolution compared to RGB cameras, which typically provide megapixels of resolution. This is due to the lower spectral width of thermal sensors than that of visible cameras. In the case of thermal imaging, the instantaneous-field-of-view must be broad enough to sense sufficient energy at the sensor's output, making their spatial resolution coarser. Alternatively, increasing the sensor size directly impacts the cost, making the technology non-affordable. Additionally, to achieve accurate thermal measurement, infrared detectors usually are encapsulated in individual vacuum packages, which is a time-consuming and expensive process.¹⁷ Due to these constraints, a thermal camera's cost is prohibitively higher than the one with a similar resolution using visible imaging. To deal with the limitations posed by thermal imaging hardware, it is necessary to supplement HR thermal imaging with an economical and effective alternative. This motivates us to use a software-driven approach called super-resolution (SR), which enhances the spatial resolution of low-resolution (LR) thermal images. We propose an efficient SR approach to super-resolve the LR images obtained from the thermal camera in this direction.

In the recent years, the development of deep models [especially convolutional neural network (CNN)] motivated by the availability of large-scale datasets and better computation capability has made remarkable SR performance and achieved the different benchmarks. A vast amount of work has been reported for achieving super-resolved images from classical RGB cameras.¹⁸ Motivated by such works, a set of recent works have focused on super-resolving the thermal images.^{16,19–22} In a continued effort in this direction, we present an approach based on deep learning (using CNN) to super-resolve the LR images obtained from the thermal camera. The proposed thermal SR architecture is computationally efficient and obtains promising SR results, which we refer hereafter as “ThermISRnet” in the remainder of this paper. This is an extension of our winner architecture in the Perception Beyond the Visible Spectrum (PBVS) thermal SR challenge.²³ As a novel contribution compared to earlier works, we employ a direct upscaling strategy with residual learning to obtain SR from an LR thermal image.

Figure 1 shows the SR images obtained using the EDSR²⁵ and our previous works, i.e., TherISuRNet²⁶ and the winner of PBVS Challenge²³ in addition to the proposed method for upscaling factor $\times 4$ on PBVS validation dataset. It can be observed that the proposed method gains considerable improvement in terms of preserving high-frequency details over the previous works, including EDSR method. We also note the quantitative assessment in terms of peak

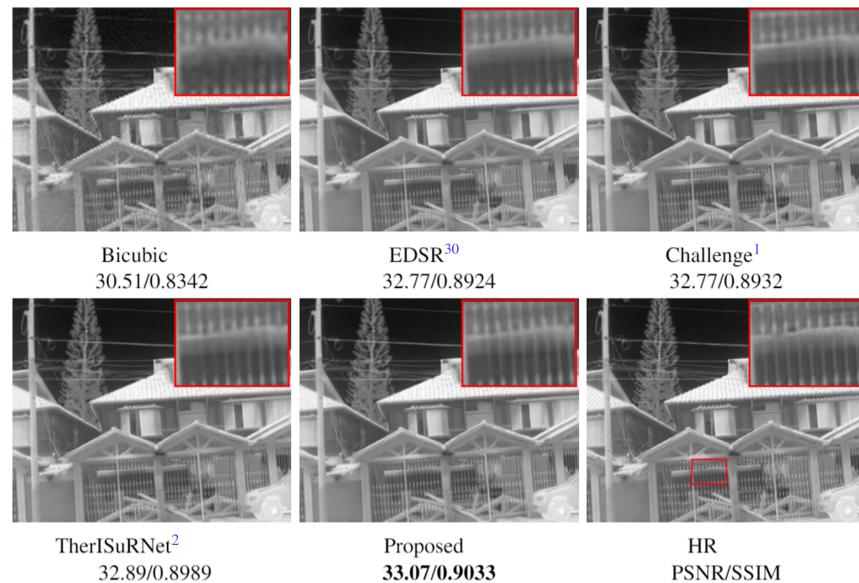


Fig. 1 The thermal SR performance of the proposed method (ThermISRnet) along with state-of-the-art existing methods for upscaling factor $\times 4$ on PBVS validation dataset.²⁴ Here, boldface values of PSNR and SSIM indicate best values among all.

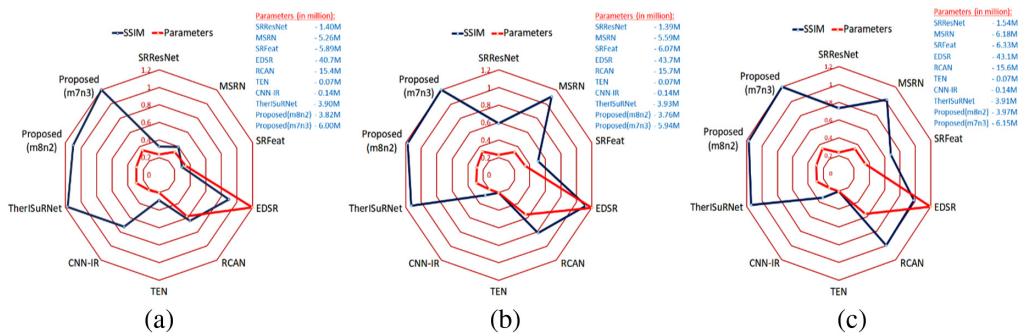


Fig. 2 The radar chart of SSIM and number of training parameters required to train different methods for (a) Domo, (b) Axis, and (c) GT testing datasets for upscaling factor of $\times 2$, $\times 3$, and $\times 4$, respectively.

signal-to-ratio (PSNR), and structural similarity measure index (SSIM) alongside SR results to gauge the proposed method's performance over the other methods. It is worth mentioning here that the proposed approach (i.e., ThermISRnet) exhibits superior values of PSNR and SSIM than the other SR techniques, including our previous works. Additionally, in Fig. 2, we depict the computational complexity of different SR methods in addition to the proposed method. One can see that the proposed ThermISRnet requires approximately 6M number of parameters in order to obtain superior performance, which is considerably lower than required in many other state-of-the-art methods. Further, in an effort toward generalization, we validate the proposed architecture by training and testing on disjoint datasets in order to evaluate the efficiency of the proposed method. We present both the qualitative and quantitative results by comparing the obtained SR results with other state-of-the-art visible image SR algorithms^{25,27–30} as well as thermal SR methods^{19,20} in the experimental section later in this paper. The key contributions of this work can be summarized as below.

- This work proposes an enhanced CNN architecture for super-resolving thermal images, which we refer to as ThermISRnet.
- The proposed approach is computationally efficient needing approximately 6M parameters to obtain a thermal SR image and is robust for different upscaling factors such as $\times 2$, $\times 3$, and $\times 4$.
- The work validates the generalizability and robustness of the proposed method by training it on PBVS thermal training dataset²⁴ and testing it on two disjoint datasets, such as FLIR³¹ and KAIST³² which are unseen during the training.

Section 2 reviews different SR methods for both thermal and visible images in the paper. Identifying the limitations, we present the proposed approach in Sec. 3 along with the experimental validation in Sec. 4. With the extensive analysis of results in the same section, we conclude the work in Sec. 5.

2 Related Work

The image SR is a classical problem in the computer vision community, yet it is an open research problem and challenging. The different SR techniques can be broadly categorized as single-image SR (SISR) and multi-image SR (MISR). The task of SISR is more challenging than MISR as it has one single LR observation to perform the SR task. The earlier works for SISR can be roughly classified as interpolation-based, traditional, and deep learning-based SR techniques. It has to be noted that interpolation-based SR methods do not add any extra information in the upscaled LR image, making it of limited use in real-life. Following the early work on the SR approach by Tsai and Huang,³³ several traditional SR methods have been proposed using the principle of reconstruction.^{34,35} Another set of traditional SR works have employed the concept of patch-based self-similarity from LR and HR images pairs.³⁶ Sparse representation was further explored to create dictionaries for both LR and HR images to achieve even better SR images.³⁷

Exploiting the recent advancements in deep learning, numerous recent works have been employed deep learning approaches to obtain better SR results simply by learning the mapping between LR and HR pairs from available large datasets.^{38,39} This new direction has motivated works explicitly in various domains such as spectral imaging⁴⁰ and medical imaging,⁴¹ including the thermal imaging applications.^{19–22}

Dong et al.³⁸ proposed the first CNN-based SR approach termed as super-resolution convolutional neural network (SRCNN). In the following work, VDSR network⁴² showed significant improvement over the SRCNN by increasing the network depth from 3 to 20 convolutional layers. VDSR method adopts the global residual learning (GRL) paradigm to predict the difference between the bicubic upsampled LR image and original ground truth HR image instead of the actual pixel value to achieve fast convergence speed. Inspired by these works,^{38,42} many works on image SR have been published in Refs. 43 and 44, which use bicubic interpolation to pre-upsample input LR image and then apply a deep network and thus it increases the computational costs.

While other works are based on a post-UP strategy for upscaling of input LR observation.^{45,46} Deep Laplacian pyramid-based SR network (LapSRN) is described by Lai et al.³⁹ in which the sub-band residuals of HR images are progressively reconstructed at multiple pyramid levels. Recently, many SR approaches using CNN, such as SRF-M,²⁸ MSRN,²⁹ EDSR,²⁵ and RCAN,³⁰ obtained state-of-the-art performance for visible LR images. The generative adversarial networks (GANs)⁴⁷ are further used as unsupervised learning models for achieving SR image in recent years. Ledig et al.²⁷ proposed SISR using GAN called SRGAN, which serves as a new state-of-the-art with impressive performance using a deep residual network (ResNet) with skip connection.⁴⁸ Following the initial works, many works on SR based on GAN model have been reported recently in Refs. 49 and 50.

The success of deep learning for SR of visible images was further extended for thermal and/or infrared images. The first CNN approach for thermal SR referred to as thermal enhancement network (TEN) was reported in Ref. 19, which was based on the SRCNN model.³⁸ It has to be noted that the TEN method¹⁹ employed RGB images in the training process due to the unavailability of large-scale thermal image dataset. On a similar idea, Marivani et al.²² obtained SR of near-infrared (NIR) images using RGB images as auxiliary information. Furthermore, Rivadeneira et al.⁵¹ use the thermal images dataset in the training process and conclude that performance of SR is better if the CNN network is trained on the thermal images instead of visible images as done in Refs. 19 and 22. Bhattacharya et al.²⁰ propose two CNN models for denoising and SR for maritime infrared images (CNN and CNN-IR). Recently, He et al.¹⁶ use the cascaded CNN architectures in order to obtain SR for upscaling factor $\times 8$. They use two-level CNN architectures in their approach, in which the first level was used to restore the structure-related information and second CNN network level was utilized to obtain finer details in the thermal images. Recently, Kansal and Nathan⁵² obtain SR of thermal images for different upscaling factors. The authors use Sobel loss in addition to L_1 and SSIM in order to preserve edge details in the SR image. Lastly, Mandanici et al.⁵³ obtained SR of thermal imagery using the concept of MISR approach.

Inspired from SRGAN,²⁷ Liu et al.²¹ use GAN model to obtain SR of the given thermal image. The SR thermal image in their approach was obtained by utilizing the different information such as resolution, scene, and field of view of corresponding RGB image in the training process. Similarly, Guei and Akhloufi⁵⁴ use the DCGAN model⁵⁰ to obtain SR of NIR and LWIR images for upscaling factor $\times 4$. In Ref. 55, authors utilize conditional GAN to enhance the contrast of given infrared image, which is capable of removing background noise present in infrared images. Furthermore, Rivadeneira et al.²⁴ released a dataset of thermal image SR and perform SR of a thermal image using CycleGAN⁴⁹ for upscaling factor $\times 2$.

In addition to thermal SR, many works also focus on the enhancement of the thermal images. For instance, authors in Ref. 56 use the CNN network to improve the contrast between target and background in the testing image. Additionally, Lee et al.⁵⁷ propose infrared image enhancement based on the brightness of the RGB images. The authors trained their network on RGB images and obtained the residual thermal image at the output of the CNN network. The final enhanced thermal image is obtained after adding a residual image with the input thermal image based on VDSR.⁴²

2.1 Constraints Noted from Related Works

With the detailed review of different thermal SR methods, we note the following constraints with existing works.

- All the present thermal SR methods (i.e., in Refs. 16, 19, and 54) are fixed to a particular upscaling factor limiting the applicability in real-life use cases where multiple upscaling factors may be necessary.
- The approaches proposed for SR of thermal images are computationally expensive as they require a large number of parameters (i.e., in Refs. 16 and 49).
- To the best of our knowledge, the robustness of thermal SR methods has not been tested in cross-database setting (i.e., in Refs. 19 and 51). Most of these works employ the same dataset and split them in training and testing set, limiting the insights on the generalizability of proposed approaches.

In an attempt to overcome the above-mentioned limitations, we devise a CNN-based computationally efficacious architecture, which we refer to as ThermISRnet for the SR of thermal images for upscaling factors $\times 2$, $\times 3$, and $\times 4$. The novel design of ResBlock assists the network to obtain a fewer number of parameters compared to the other state-of-the-art methods. The robust nature of the proposed method is demonstrated by validating it on the different datasets, which are not used in the training process.

3 Proposed Approach: ThermISRnet

Figure 3 depicts the architecture of the proposed method for the task of thermal image SR for different upscaling factors. The thermal LR image (i.e., I^{LR}) is applied as an input to the architecture and its corresponding SR image for upscaling factors $\times 2$, $\times 3$, and $\times 4$ (i.e., $I_{\times 2}^{SR}$, $I_{\times 3}^{SR}$, and $I_{\times 4}^{SR}$) is obtained. To extract rich features from the given input thermal LR image, the same is passed through low- and high-frequency feature extraction modules. The upsampling block is then utilized to accomplish the desired scaled feature maps, and finally, SR image is reconstructed through the reconstruction module in the architecture. The GRL is further employed in the reconstructed image to generate the final SR image. Our architecture consists of four following modules, which are designed to fulfil specific tasks:

- low-frequency feature extraction (LFE) module,
- high-frequency feature extraction (HFE) module,
- upsampling (UP) block, and
- reconstruction (REC) module.

The importance and detailed description related to each of the above module is discussed in the following sections.

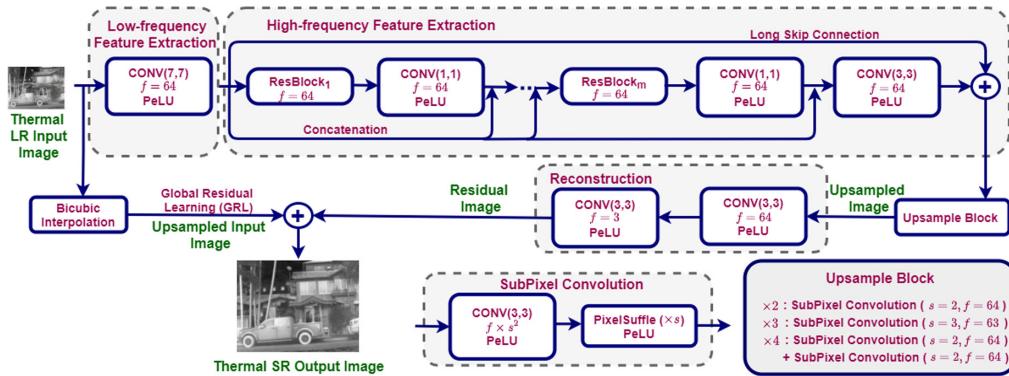


Fig. 3 The network architecture design of the proposed model (ThermISRnet). Here, s and f indicate the scale factor and feature maps, respectively.

3.1 Low-Frequency Feature Extraction

The thermal LR image is first applied to the LFE module, which includes a single convolutional layer to extract features from LR observation. Here, the kernel size, number of features, and stride values in this convolution layer are set to 7, 64, and 1, respectively. We use parametric exponential linear unit (PeLU) activation function in the architecture, which efficiently improves the learning performance at each layer. The parameters associated with PeLU activation at different layers are learned to make appropriate activation shape to improve the SR performance.⁵⁸ The low-frequency features are extracted from the input LR thermal image (I^{LR}) that can be represented as

$$I_{LFE} = f_{LFE}(I^{LR}), \quad (1)$$

where $f_{LFE}(\cdot)$ represents the function of the LFE module.

3.2 High-Frequency Feature Extraction

Due to the thermal sensors' coarser spatial resolution, the acquired data lack high-frequency details. To extract those high-frequency details such as edges and structures, we pass the feature maps from the LFE module to HFE module composed of m number of residual blocks (ResBlocks). Further, the output generated from the LFE module is fed forwarded to the input of each ResBlock in the HFE module to improve the learning fidelity of the network. Additionally, the output of each ResBlock is followed by the convolutional layer to control the number of feature maps as depicted in Fig. 3.

The novel element in the proposed architecture is the design of ResBlock, which is responsible for preserving the high-frequency details in the SR image and reducing the number of parameters associated with the network. The architectural design of the same is depicted in Fig. 4. It comprises one convolution layer with a kernel size of 1 followed by three parallel branches of concatenated blocks (CBs). In each parallel branch, we deploy n number of CBs (i.e., $n = 3$). The different CBs in each ResBlock are densely connected to retain meaningful features as displayed in Fig. 4 and each CB employs several convolution layers followed by a channel attention (CA) module. The use of the CA module in CBs is to perform adaptive rescaling of the channel-wise features. The CA module's effectiveness is also analyzed by conducting experiments and discussed later in the section of ablation study. Such a design of CB improves the performance of the network by extracting richer features. The features obtained from each branch of CBs are concatenated further and then passed through one convolution layer of kernel

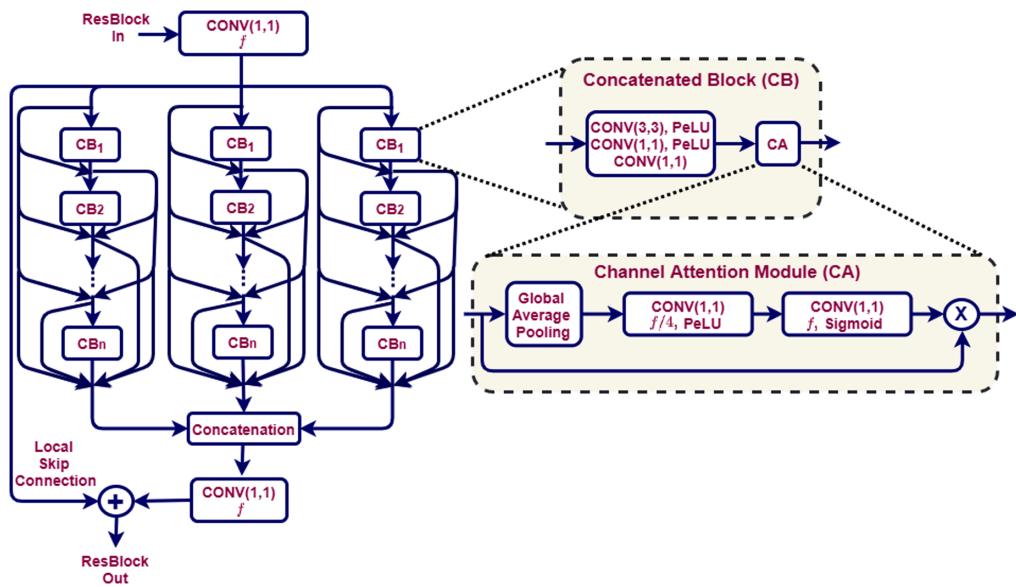


Fig. 4 The design of the ResBlock used in the HFE module of the proposed model.

size 1. This convolution layer works as a transition layer and thus, yields the desired number of feature maps. Additionally, to eliminate the problem of vanishing gradient, we deploy a local skip connection in each ResBlock (see Fig. 4). Alongside local skip connection, a long skip connection is also deployed in the HFE module as depicted Fig. 3, which is followed by one convolution layer with a filter size of 3. Thus, the output feature maps obtained from the HFE module can be represented as

$$I_{\text{HFE}} = f_{\text{HFE}}(I_{\text{LFE}}), \quad (2)$$

where $f_{\text{HFE}}(\cdot)$ denotes the operation of the HFE module.

3.3 Upsampling

The upsampling block is utilized to upscale the feature maps to the desired scaling factors (i.e., $\times 2$, $\times 3$, and $\times 4$). We adopt different upsampling strategies in order to perform SR operation, which corresponds to different upsampling factors. To obtain SR image with scaling factor of 2 (i.e., $I_{\times 2}^{\text{SR}}$), single upsample block is utilized, which comprises sub-pixel convolution with 64 feature maps. While obtaining output SR image with scaling factor 3 (i.e., $I_{\times 3}^{\text{SR}}$), the sub-pixel convolution with scaling factor ($s = 3$) and 63 feature maps is employed. Further, the upsample block of $\times 2$ is recycled to obtain the SR image with scaling factor $\times 4$ (i.e., $I_{\times 4}^{\text{SR}}$). Mathematically, this can be defined as

$$I_{\text{UP}(\times 2, \times 3, \times 4)} = f_{\text{UP}(\times 2, \times 3, \times 4)}(I_{\text{HFE}}), \quad (3)$$

where $f_{\text{UP}(\times 2, \times 3, \times 4)}(\cdot)$ denotes the function of the upsample block for the scaling factor $\times 2$, $\times 3$, and $\times 4$.

3.4 Reconstruction

The output obtained from the UP module with the desired scaling factor is passed through the REC module to generate the output thermal SR image. This module comprises of two convolution layers with a kernel size of 3. The final output SR image generated from the REC module can be represented as

$$I_{\text{residual}(\times 2, \times 3, \times 4)}^{\text{SR}} = f_{\text{Rec}(\times 2, \times 3, \times 4)}(I_{\text{UP}(\times 2, \times 3, \times 4)}), \quad (4)$$

where $f_{\text{Rec}(\times 2, \times 3, \times 4)}$ indicates the REC function of the REC module.

Further, we use GRL in the network architecture to defeat the barrier associated with network training stability. The input thermal LR image (i.e., I^{LR}) is applied to the bicubic interpolation layer. In this layer, the thermal LR image is interpolated with scaling factor of $\times 2$, $\times 3$, and $\times 4$ according to the SR operation. The GRL also helps to learn the identity function for LR image I^{LR} in addition to stabilization of the training process.⁵⁹ Finally, the SR image (I^{SR}) is obtained with upscaling factor of $\times 2$, $\times 3$, and $\times 4$ as represented in Eq. (5),

$$I_{\times 2, \times 3, \times 4}^{\text{SR}} = I_{\text{residual}(\times 2, \times 3, \times 4)}^{\text{SR}} + I_{\text{GRL}(\times 2, \times 3, \times 4)}^{\text{SR}}. \quad (5)$$

3.5 Comparison with Winner Network of PBVS Challenge

In comparison with our winner architecture of the PBVS challenge²³ (Team Name: MLCV –Lab_SVNIT_NTNU), following extensions have been employed in this work in order to improve the SR performance.

- Inspired from literature, the design of ResBlock is modified in terms of dense connections to extract rich features from the thermal LR observation.
- Further, we employ GRL in the proposed model, which helps to learn the identity function for LR image as well as to stabilize the training process.⁵⁹ The effectiveness of GRL is also

justified experimentally, and the same is mentioned in the section of ablation study later in the paper.

- In this extended work, we train the network with the weighted combination of two-loss functions: L_1 and SSIM; whereas in PBVS challenge, we used only L_1 -based loss function to train the network. To validate the importance of such weighted loss function utilized in the proposed method, we have carried out different experiments on individual loss functions and prove its effectiveness.
- Further, the ablation study has been appended in this work to study the sensitivity of the different parameters associated with the network design. This ablation study proves the importance of the structure design, loss functions, activation function, etc., of the proposed method.
- To verify the generalizability of the proposed method, we test it on cross-validation datasets such as FLIR and KAIST, whereas the PBVS challenge architecture was explored on a single dataset within the context of the challenge alone.
- Additionally, for a fair comparison of the proposed method with the other existing state-of-the-art methods, we have retrained all those methods as per our training strategy and dataset and show the thermal SR results.

4 Experimental Analysis

The proposed method's efficacious design is validated with numerous experiments on different datasets for upsampling factors $\times 2$, $\times 3$, and $\times 4$. All the experiments have been carried out on a computer with Intel Xeon(R) CPU E5-2620 v4 processor at 2.10 GHz $\times 32$ running on a 128 GB RAM and two NVIDIA Quadro P5000 with 16 GB GPUs. The detailed description related to different experiments and its associated analysis are discussed further in the following sections.

4.1 Training Details and Hyperparameter Settings

To train the proposed network (i.e., ThermISRnet), we use a dataset of thermal images released in PBVS challenge.²⁴ This dataset consists of images acquired using three different cameras. It has a number of indoor and outdoor images acquired under various lighting conditions (i.e., morning, afternoon, and night) and objects (i.e., buildings, cars, humans, and vegetation). The cameras were mounted on a rig that minimizes the baseline distance between the optical axis such that the acquired images are almost registered. The specifications of the cameras utilized in the generation of the PBVS dataset is mentioned in Table 1. This dataset consists of 1021 thermal images obtained from three distinct thermal cameras at different resolutions (LR, mid-resolution, and HR), resulting in 3063 thermal images. Hence, it has three sub-parts : Domo, Axis, and GT corresponding to upsampling factors $\times 2$, $\times 3$, and $\times 4$, respectively. Each sub-dataset includes 951 number of training images. An augmentation with 180-deg rotation, horizontal flipping, and warp affine operations are performed in the dataset. Further, the bicubic interpolation is employed to downsample the LR thermal images with Gaussian noise of 0 mean and standard

Table 1 The specifications of FLIR, KAIST, and PBVS datasets.

Dataset	Resolution	Brand camera	FOV	Focal length	Native resolution	Number of thermal images
PBVS	Low	Axis Domo P1290	35.4	4 mm	160 \times 120	1021
	Mid	Axis Q2901-E	35	9 mm	320 \times 240	1021
	High	FC-6320 FLIR	32	19 mm	640 \times 512	1021
FLIR	High	IR Tau2	37	13 mm	640 \times 512 ^a	>14,000
KAIST	Mid	FLIR-A35	39	2.4 m	320 \times 256	>41,500

^aIndicates the HR images have been cropped to 640 \times 480.

deviation of 10. We train the proposed architecture up to 5×10^4 number of iterations with Adam optimizer. The learning rate and batch size are set to 0.001 and 16, respectively, based on the empirical validation. We train the model with a weighted combination of L_1 and SSIM loss functions and the same is represented in Eq. (6) as

$$L_{\text{SR}} = \lambda_0 L_1 + \lambda_1 \text{SSIM}, \quad (6)$$

where λ_0 and λ_1 are the weighting factors assigned to L_1 loss and SSIM loss, respectively.

4.2 Testing Details

We use three different datasets to check the generalizability of the proposed network: PBVS challenge (i.e., Domo, Axis, and GT),²⁴ FLIR,³¹ and KAIST³² validation datasets for upscaling factor of $\times 2$, $\times 3$ and $\times 4$. The description of these validation datasets is provided in Table 1. The details of the PBVS challenge dataset are already discussed in the earlier section. From this dataset, the validation images (i.e., 50 number of images) are separated for each resolution (i.e., Domo, Axis, and GT). The dimensions of input thermal LR image in terms of width \times height for Domo (i.e., for $\times 2$ scaling factor), Axis (i.e., for $\times 3$ scaling factor), and GT (i.e., for $\times 4$ scaling factor) datasets are 80×60 , 110×80 , and 160×120 , respectively. Furthermore, the FLIR dataset is acquired using IR Tau2 camera in driving condition (i.e., video is recorded at 30 Hz), and dataset sequences are sampled at 2 frames/s or 1 frame/s. 60% images have been captured for daytime, and remaining images have been acquired in night-time conditions. Among these images, 1366 images have been separated for validation purpose. We have used these validation images to validate our proposed model. While in the case of the KAIST dataset, the FLIR-A35 camera is used to acquire more than 41,500 thermal images with a resolution of 320×256 . The frame rate of the camera is equal to 20 frames/s. The hardware was mounted on the roof of a car, and then it is used for capturing ego-centric images of the traffic scenes. In particular, they captured various scenes at day and night-time to consider changes in light conditions. The KAIST dataset is reserved for training as well as the validation and testing processes. We have chosen 500 random images from the KAIST validation dataset. Additionally, the input LR image dimensions in FLIR and KAIST datasets are 320×240 , 210×160 , and 160×120 for upscaling factor $\times 2$, $\times 3$, and $\times 4$, respectively.

Further, an adaptive histogram equalization technique⁶⁰ is used to enhance the KAIST validation dataset's images. The corresponding LR pair images of testing datasets are generated by adding Gaussian noise with mean value 0 and standard deviation of 10 followed by corresponding downsampling operation via bicubic interpolation for upscaling factors $\times 2$, $\times 3$, and $\times 4$. The architecture presented in PBVS challenge by our team MLCV-Lab_SVNIT_NTNU has been extended in this work both in terms of architecture and training. Therefore, we compare the proposed model in this work with our earlier architecture in the PBVS challenge and the other recently proposed state-of-the-art methods. The quantitative and qualitative assessments consist of the comparison of SR thermal results obtained using the proposed method with many state-of-the-art visible image SR methods, such as SRResNet,²⁷ MSRN,²⁹ EDSR,²⁵ SRFeat-M,²⁸ and RCAN,³⁰ and with the recently proposed methods on thermal SR approaches, such as TEN,¹⁹ CNN-IR,²⁰ and TherISurNet.²⁶ These methods are retrained on the PBVS challenge dataset with the same training strategy adopted in the proposed method for a fair comparison. Further, the PSNR and SSIM metrics are utilized for quantitative analysis. These measures are calculated in YCbCr color space where we eliminate four boundary pixels of Y -channel image as suggested in Refs. 25, 27, and 59.

4.3 Ablation Study

To analyze the sensitivity of the different components used in the proposed method (i.e., ThermISRnet), we have conducted many experiments in the ablation study. The quantitative results of these experiments are depicted in Table 2. We demonstrate the effectiveness of proposed loss functions, chosen values of λ_i in Eq. (6) of weighted loss function and PeLU activation function in our method. Additionally, we also consider the different architectural design

Table 2 The comparison of proposed methods on GT ($\times 4$) validating dataset with three different scenarios. Here, the best values are mentioned in bold font.

	PSNR↑	SSIM↑
Loss function		
L_2	34.4344	0.9056
L_1	34.4800	0.9087
SSIM	34.4301	0.9097
Proposed [Eq. (6)]	34.5075	0.9102
Values of λ_i in Eq. (6)		
$\lambda_0 = \lambda_1$ ($\lambda_0 = \lambda_1 = 10$)	34.5265	0.9094
$\lambda_0 > \lambda_1$ ($\lambda_0 = 5, \lambda_1 = 0.5$)	34.5501	0.9093
Proposed [$\lambda_0 < \lambda_1$ ($\lambda_0 = 1, \lambda_1 = 10$)]	34.5075	0.9102
Activation function		
ReLU	29.2313	0.8838
LReLU	34.4308	0.9097
PReLU	34.4308	0.9093
eLU	34.4264	0.9095
Proposed (PeLU)	34.5075	0.9102
Network design		
Without CA	34.4186	0.9099
Without GRL	34.2680	0.9094
Proposed (with CA and GRL)	34.5075	0.9102

aspects such as CA and GRL in the proposed network design to learn its effectiveness. The performance of each component evaluated with different number of ResBlocks, i.e., $m = 7$, and number of CBs, i.e., $n = 3$, in the proposed method is depicted in Table 2. We indicate the SR performance on GT testing dataset for upscaling factor $\times 4$ in terms of PSNR and SSIM metrics. The importance of each component is discussed at length in the following paragraphs.

- **Effectiveness of proposed weighted combination loss function:** In the proposed method, we employ a weighted combination of L_1 and SSIM loss functions to train the model. To understand the effect of such losses on the proposed model's SR performance, we train the proposed model with different loss functions such as L_2 , L_1 , and SSIM individually. One can note from Table 2 that the performance achieved using individual loss function is ineffective compared to the weighted sum of L_1 and SSIM loss functions, which gains noticeable improvement in PSNR and SSIM measurements over the individual ones.
- **Importance of λ_i values in Eq. (6):** We set the values of λ_i , $i = 0, 1$ in the equation of weighted loss function as 1 and 10, respectively. These values are set empirically; however, to learn its effectiveness in the proposed method, we have trained it with a different combination of λ_i and depict the results in Table 2. We consider three different cases: $\lambda_0 = \lambda_1$, $\lambda_0 > \lambda_1$, and $\lambda_0 < \lambda_1$. In each of above case, again we change the different combinations of λ_i ; for instance, in $\lambda_0 = \lambda_1$, we have analyzed the performance with λ_i equal to 0.1, 0.5, 1, 5, and 10 and note the result, which is best among all in Table 2. Similarly, for the second case, we have selected $\lambda_0 = 5, \lambda_1 = 0.5$; $\lambda_0 = 10, \lambda_1 = 1$; and $\lambda_0 = 100$ and these

combinations are reversed for third case. In the proposed method, the chosen values of weighting constants, i.e., $\lambda_0 = 1$, $\lambda_1 = 10$, perform best among all the above combinations. This validates the selected values of weighting factors i.e., λ_i , in Eq. (6) of the proposed method.

- **Importance of PeLU activation function:** To see the effectiveness of the PeLU activation function in the proposed method, we have trained the proposed method with different activation functions, such as ReLU,⁶¹ LReLU,⁶² PReLU,⁶³ and eLU,⁶⁴ and measurements are depicted in Table 2. From that, it can be noticed that the PeLU activation function outperforms over others. Thus, it is useful to improve the SR performance in the proposed network.
- **Effectiveness of GRL and CA:** The GRL and CA modules are also employed in the proposed method, which play prominent roles in enhancing the performance of the network architecture. We demonstrate their importance by conducting two additional experiments where the proposed network is trained without using those modules, and quantitative measures such as PSNR and SSIM metrics are noted. Here, we observe significant improvement with GRL and CA modules as one can notice from Table 2, which validates their importance in the architecture of the proposed method.

4.4 Fidelity of Thermal SR Images

The qualitative and quantitative comparison of the proposed method and the other existing state-of-the-art methods are presented in this section. We consider the recently proposed other existing methods, such as MSRN,²⁹ SRFat,²⁸ EDSR,²⁵ RCAN,³⁰ TEN,¹⁹ CNN-IR,²⁰ and TherISurNet,²⁶ for the comparison of the proposed method to verify the SR performance on single thermal LR image. Additionally, we also compare SR performance with our previous work published in PBVS challenge as a winner architecture.²³

4.4.1 Quantitative evaluation

We first quantitatively compare the proposed ThermISRnet with the winner architecture in the PBVS thermal SR challenge²³ as depicted in Table 3. The number of ResBlocks (i.e., m) and CBs (i.e., n) in the challenge was set to 6 and 3 in the architecture, respectively. As stated earlier, to enhance the SR performance further, in this work, we have increased the values of m and n and note the quantitative improvement achieved in Table 3. Along with the change in the values of m and n , we have also modified the proposed method's loss functions to improve the performance. The PBVS challenge architecture utilized L_1 loss only while the proposed module in this work introduces the SSIM loss along with the L_1 loss. Further, the PBVS CVPR workshop-2020 challenge architecture was tested on PBVS challenge validation dataset²³ only. We demonstrate its performance on two additional datasets, such as FLIR and KAIST in this work, to prove generalizability of the proposed method. It is worth noting from Table 3 that proposed model with $m = 7$ and $n = 3$ outperforms over the winner architecture of PBVS challenge in all the cases of PBVS validation dataset.

It is also noticed from Table 3 that our model of PBVS challenge obtains slightly better results on FLIR dataset. However, the proposed model with $m = 7$ and $n = 3$ performs better on KAIST dataset for all upscaling factors except scaling factor of $\times 2$. By inspecting the performance of the different modules presented in Table 3, we have chosen the number of ResBlocks, i.e., $m = 7$ and CBs i.e., $n = 3$ in the proposed architecture. The efficiency of the proposed method (with $m = 7$ and $n = 3$) is analyzed by comparing it further with the current state-of-the-art methods, which is described in the following subsequent sections.

Further quantitative comparisons in terms of PSNR and SSIM measures are obtained using different state-of-the-art and proposed methods depicted in Table 4 for different upscaling factors. The highest values of PSNR and SSIM metrics are highlighted in bold while the second-highest values are highlighted in italics. It can be observed from the table that the proposed model attains better SSIM measure in all cases (i.e., on different testing datasets with upscaling factor of $\times 2$, $\times 3$, and $\times 4$) with large margin than other competing models.

Table 3 The quantitative comparison of the proposed method along with our winner architecture of PBVS thermal SR challenge²³ with different values of m and n on different validation datasets in terms of PSNR and SSIM metrics. Here, the bold and italic fonts indicate highest and second-highest values among the different methods, respectively.

	Dataset	Metrics	PBVS challenge ($m = 6, n = 3$)	Proposed ($m = 8, n = 2$)	Proposed ($m = 7, n = 3$)
x2	Domo	PSNR	33.5853	33.5283	33.6015
		SSIM	0.8982	0.9006	0.9017
	FLIR	PSNR	35.2890	35.1930	<i>35.2284</i>
		SSIM	0.8695	0.8716	0.8722
	KAIST	PSNR	37.8025	37.6092	<i>37.6559</i>
		SSIM	0.9473	0.9469	0.9476
	Axis	PSNR	32.9898	33.0320	33.0330
		SSIM	0.9013	0.9046	0.9052
	FLIR	PSNR	32.4710	32.4538	32.4342
		SSIM	0.8326	0.8346	0.8350
	KAIST	PSNR	34.0250	34.1425	34.2931
		SSIM	0.8963	0.8982	0.8998
x4	GT	PSNR	34.4925	34.4746	34.5402
		SSIM	0.9073	0.9107	0.9112
	FLIR	PSNR	30.8650	30.7692	30.8123
		SSIM	0.7728	0.7775	0.7781
	KAIST	PSNR	32.6507	32.9487	32.9589
		SSIM	0.8778	0.8822	0.8827

In contrast to other thermal SR methods (i.e., TEN¹⁹ and CNN-IR²⁰), the proposed model outperforms those methods with large margin except for the case of KAIST validation dataset with scaling factor $\times 2$ where SR method CNN-IR²⁰ performs better. However, it is worth noting that the proposed method utilizes a comparable number of trainable parameters against MSRN²⁹ and SRFat-M²⁸ methods (approximately 60% to 85% lesser number of training parameters than that of RCAN³⁰ and EDSR²⁵). To verify this, we compare the SSIM measure obtained using the other existing methods and the proposed model on the number of training parameters for different validation datasets. The graphs of the same are depicted in Figs. 2, 5, and 6, for PBVS, FLIR, and KAIST datasets, respectively. The graphs present the number of trainable parameters and values of SSIM for all the methods, including the proposed approach. For the sake of illustration, the values are normalized and scaled in the range of 0.2 to 1.2. The actual parameters for all the methods are also presented alongside each figures for convenience of the reader.

It is observed here that the proposed method with $m = 8$ and $n = 2$ [i.e., mentioned with proposed (m8n2) in Figs. 2, 5, and 6] results in lesser number of training parameters on $\times 2$ and $\times 3$ upscaling factors and obtains comparable SSIM measure on those testing datasets as compared to the recently proposed TherISurNet method.²⁶ However, the proposed method (i.e., $m = 8, n = 2$) outperforms the recent MSRN, SRFat, RCAN, and EDSR models by obtaining better SSIM measure with significantly less number of training parameters. Further, by increasing the number of training parameters than that of TherISurNet model, the proposed method with $m = 7$ and $n = 3$ [i.e., mentioned with proposed (m7n3) in Figs. 2, 5, and 6] achieves better SSIM performance for upscaling factor of $\times 2$, $\times 3$, and $\times 4$.

Table 4 The quantitative comparison of the proposed method along with other state-of-the-art methods on different validation datasets in terms of PSNR and SSIM metrics. The bold and italic fonts indicate the highest and second-highest values among the different methods, respectively.

	Dataset	Metrics	Bicubic	SRResNet ²⁷	MSRN ²⁸	SRFeat ²⁸	EDSR ²⁵	RCAN ³⁰	TEN ¹⁹	CNN-IR ²⁰	TherISRNet ²⁶	Proposed
x2	Domo	PSNR	32.1229	33.0817	33.1215	33.1253	33.5248	33.3144	33.1919	33.5272	33.6559	33.6075
		SSIM	0.8751	0.8905	0.8927	0.8916	0.8983	0.8955	0.8915	0.8964	0.9014	0.9017
	FLIR	PSNR	34.3019	34.8267	34.9860	34.9806	35.2352	35.0518	35.0352	35.2632	35.2955	35.2284
		SSIM	0.8488	0.8651	0.8665	0.8660	0.8698	0.8684	0.8657	0.8687	0.8720	0.8722
	KAIST	PSNR	37.1974	37.3715	37.5627	37.5051	37.7663	37.5993	37.5356	37.8287	37.7233	37.6559
		SSIM	0.9319	0.9444	0.9449	0.9449	0.9458	0.9467	0.9462	0.9455	0.9473	0.9474
x3	AXIS	PSNR	30.3577	32.5174	33.1015	32.4329	33.1278	32.8011	33.1311	32.2217	32.9803	33.0330
		SSIM	0.8032	0.8913	0.9031	0.8894	0.9035	0.8965	0.8824	0.8843	0.9036	0.9052
		PSNR	30.3373	32.2763	32.4962	32.2582	32.5487	32.3345	32.0931	32.1912	32.3202	32.4342
	FLIR	PSNR	30.3373	32.2763	32.4962	32.2582	32.5487	32.3345	32.0931	32.1912	32.3202	32.4342
		SSIM	0.7475	0.8273	0.8331	0.8263	0.8342	0.8307	0.8215	0.8232	0.8332	0.8350
		PSNR	32.3202	34.0937	34.1822	33.7220	34.3233	34.1102	33.9656	34.1729	34.1499	34.2931
x4	KAIST	SSIM	0.8332	0.8971	0.8978	0.8905	0.8991	0.8972	0.8958	0.8950	0.9000	0.8998
		PSNR	32.6657	33.1240	34.4718	34.1245	34.4852	34.4200	33.6230	33.7723	34.4956	34.5402
		SSIM	0.8625	0.9018	0.9076	0.9007	0.9068	0.9072	0.8910	0.8938	0.9101	0.9112
	GT	PSNR	30.1153	30.3533	30.7161	30.7513	30.8986	30.8275	30.5943	30.6758	30.8108	30.8123
		SSIM	0.7467	0.7551	0.7702	0.7683	0.7730	0.7728	0.7625	0.7656	0.7769	0.7781
		PSNR	32.4649	32.0788	32.9730	32.8661	33.0546	32.9962	32.5402	32.7842	32.6999	32.9589
	SSIM	0.8707	0.8652	0.8799	0.8773	0.8795	0.8804	0.8758	0.8758	0.8790	0.8827	

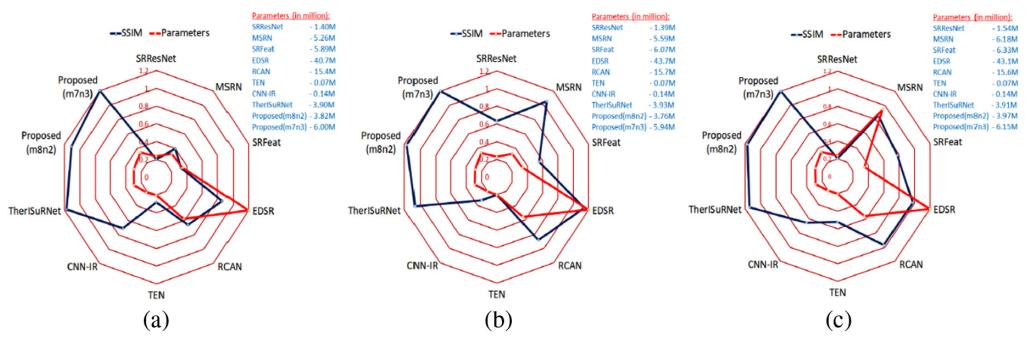


Fig. 5 The radar chart of SSIM and number of training parameters required to train different SR methods for FLIR testing dataset for upscaling factors of (a) $\times 2$, (b) $\times 3$, and (c) $\times 4$, respectively.

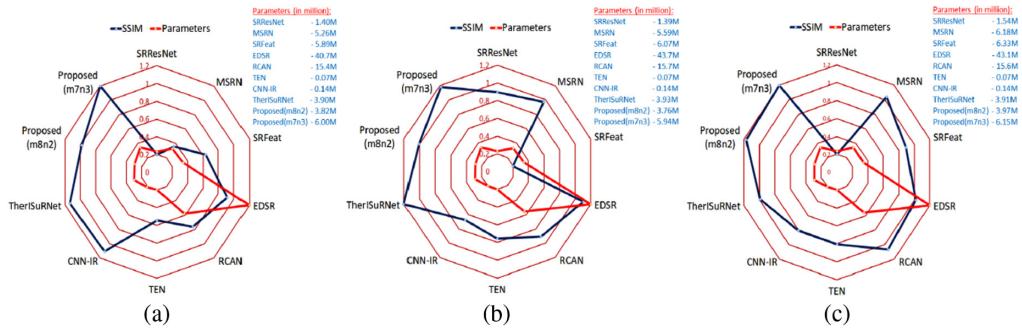


Fig. 6 The radar chart of SSIM and number of training parameters required to train different methods for KAIST testing dataset for upscaling factors of (a) $\times 2$, (b) $\times 3$, and (c) $\times 4$, respectively.

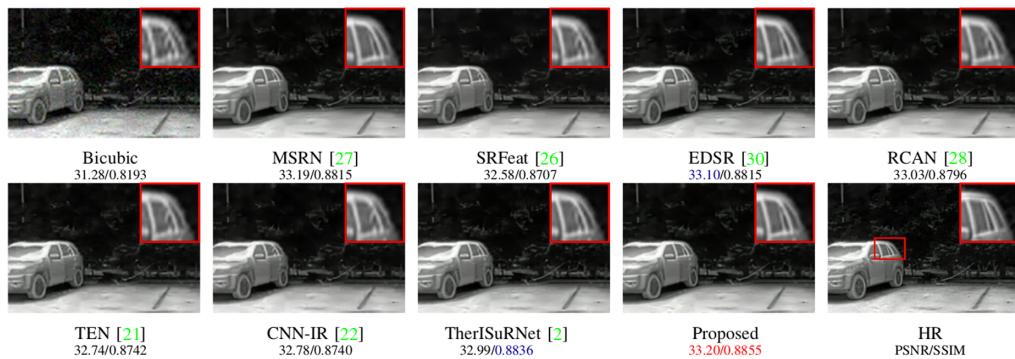
To validate the proposed model in terms of computational complexity, an additional experiment is conducted in which the inference time required by the proposed model (i.e., ThermISRnet) along with the existing state-of-the-art SR models for upscaling factor $\times 4$ is recorded. The inference test time indicates the time taken by the algorithm to execute the applied input in second. It is one of the most crucial aspects for deploying a deep network into a real-time processing hardware. Lowering the inference time superior is the execution speed. For a fair comparison, all the models have been tested on a common platform (i.e., computer with Intel i7-6850K processor at 3 GHz $\times 12$, 64 GB RAM and NVIDIA TITAN X Pascal 12 GB GPU), and the corresponding inference time is measured. This comparison is depicted in Table 5 where we also denote the value of SSIM measure along with inference time obtained using the different methods on PBVS, FLIR, and validation datasets. From this table, one can observe that the proposed model attains less inference time than that of EDSR,²⁵ RCAN,³⁰ and TherISuRNet²⁶ models and gains better SSIM measure than those models. Further, the proposed approach sets high inference time than that of SRResNet,²⁷ MSRN,²⁹ SRFeat,²⁸ TEN,¹⁹ and CNN-IR;²⁰ however, the proposed model obtains better SSIM measure than the above models. Although the complexity of the proposed method (i.e., 6M parameters) results in poor performance in terms of inference time compared to few methods, the same is helpful to gain better quantitative and qualitative performance.

4.4.2 Qualitative evaluation

In addition to quantitative comparison, the qualitative comparison is also carried out on PBVS challenge (i.e., Figs. 7–9), FLIR (i.e., Figs. 10–12), and KAIST (i.e., Figs. 13–15) validation datasets for the upscaling factor of $\times 2$, $\times 3$, and $\times 4$. For a comprehensive comparison, PSNR and SSIM values are noted at the bottom of each result in all the above figures. The red colored font highlights the result obtained with highest values while the blue colored font indicates the second-highest measure. In Fig. 7, we show the visual comparison on PBVS challenge validation dataset on scaling factor of $\times 2$. It can be observed from Fig. 7 (see zoomed-in patches) that the

Table 5 The inference time of proposed and other existing methods on PBVS, FLIR, and KAIST validation dataset with scaling factor $\times 4$. Here, the highest measures are indicated with bold font.

Dataset	PBVS		FLIR		KAIST	
	Methods	Inference time (s)	SSIM	Inference time (s)	SSIM	Inference time (s)
SRResNet ²⁷	0.0694	0.9018	0.0471	0.7551	0.0485	0.8652
MSRN ²⁹	0.0708	0.9076	0.0491	0.7702	0.0493	0.8799
SRFeat ²⁸	0.0858	0.9007	0.0610	0.7683	0.0620	0.8773
EDSR ²⁵	0.1991	0.9068	0.1680	0.7730	0.1698	0.8795
RCAN ³⁰	0.1760	0.9072	0.1170	0.7728	0.1196	0.8804
TEN ¹⁹	0.0435	0.8910	0.0227	0.7625	0.0239	0.8758
CNN-IR ²⁰	0.0609	0.8938	0.0426	0.7656	0.0439	0.8758
TherISuRNet ²⁶	0.1727	0.9101	0.1148	0.7769	0.1171	0.8790
Proposed	0.1627	0.9112	0.1020	0.7781	0.1045	0.8827

**Fig. 7** The qualitative comparison of different thermal SR approaches obtained using PBVS challenge validation dataset²⁴ for scaling factor $\times 2$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at top right corner of each SR result.**Fig. 8** The qualitative comparison of different thermal SR approaches obtained using PBVS challenge validation dataset²⁴ for scaling factor $\times 3$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at top right corner of each SR result.

proposed method performs better than the other state-of-the-art methods and obtains the highest PSNR SSIM values among the others. The qualitative comparison on scaling factor of $\times 3$ is depicted in Fig. 8. Here, one can inspect by looking at the zoomed-in patches that the proposed model generates better SR solutions than other models and preserves better texture details than

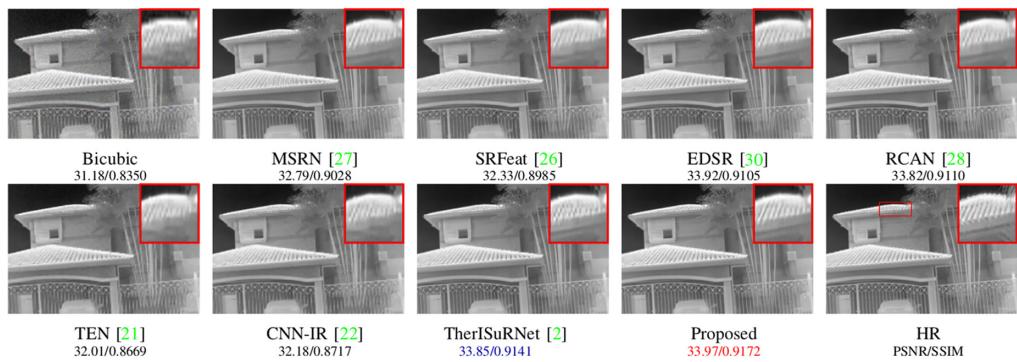


Fig. 9 The qualitative comparison of different thermal SR approaches obtained using PBVS challenge validation dataset^[24] for scaling factor $\times 4$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at top right corner of each SR result.



Fig. 10 The qualitative comparison of different thermal SR approaches obtained using FLIR validation dataset^[31] for scaling factor $\times 2$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom left corner of each SR result.

other models. Also, the proposed method exhibits the highest PSNR and SSIM values than that of existing models. Similarly, the visual comparison on PBVS validation dataset with a scaling factor of $\times 4$ is presented in Fig. 9. It can be noticed from the zoomed-in patches of Fig. 9 that the proposed model achieves better SR result than other methods and obtains better PSNR and SSIM measurements.

Further, the qualitative comparison on FLIR dataset for scaling factor $\times 2$ is displayed in Fig. 10. In this figure, the proposed method gains improvement in preserving details in the SR observation than other models. However, the proposed model attains comparable PSNR value compared to TherISURNet and EDSR models while it produces the highest SSIM measure

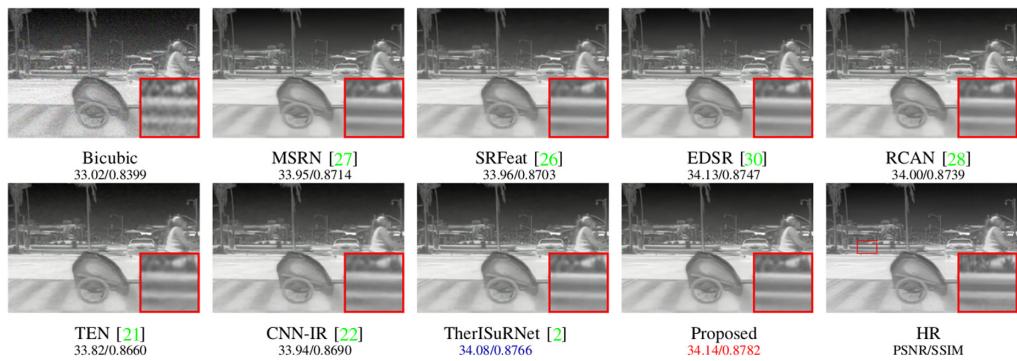


Fig. 11 The qualitative comparison of different thermal SR approaches obtained using FLIR validation dataset^[31] for scaling factor $\times 3$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom right corner of each SR result.

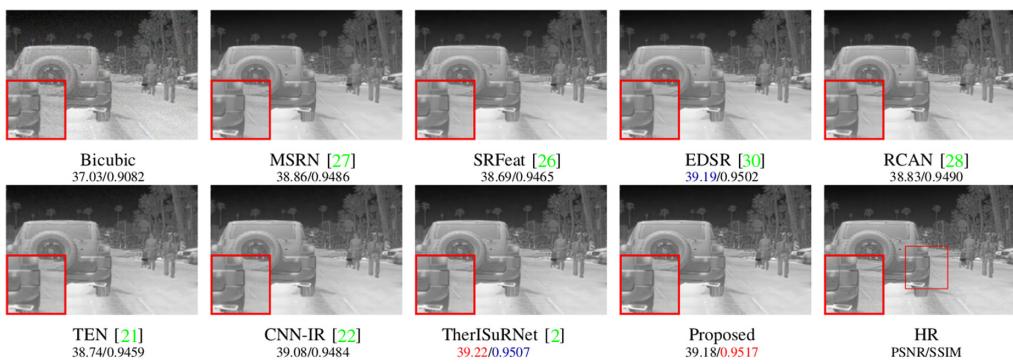


Fig. 12 The qualitative comparison of different thermal SR approaches obtained using FLIR validation dataset³¹ for scaling factor $\times 4$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom left corner of each SR result.

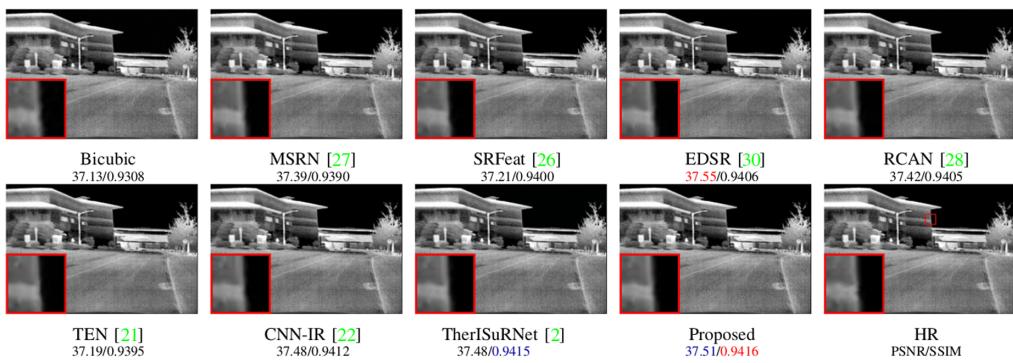


Fig. 13 The qualitative comparison of different thermal SR approaches obtained using KAIST validation dataset³² for scaling factor $\times 2$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom left corner of each SR result.

among other state-of-the-art methods. Figure 11 depicts the comparison of the proposed method with other existing methods on scaling factor of $\times 3$. As indicated in the figure, the proposed model generates better SR result than that of other models (see enlarged patches in Fig. 11). Also, the proposed model excels in terms of PSNR and SSIM measures, among others. Similarly, the comparison of the same validation dataset with a scaling factor of $\times 4$ is illustrated in Fig. 12. Here, similar to earlier results, by looking at the zoomed-in patches, one can assert that the proposed model preserves more texture details than the existing models and attains comparable PSNR value with EDSR and TherISURNet models. However, the proposed model earns the highest value of SSIM measure as compared to the other models.

Figure 13 displays the visual comparison on KAIST dataset with scaling factor $\times 2$. The visual inspection of results from the proposed model exhibits superior SR results over other state-of-art methods. Further, it can also be noticed here that the proposed model earns the second-highest PSNR value and sets the highest SSIM measure when compared to others (i.e., see Fig. 13). Additionally, the comparison of the scaling factor of $\times 3$ is depicted in Fig. 14. Here, it is important to note that the proposed model outperforms to the existing state-of-the-art methods (see zoomed-in patches) with improvement in PSNR and SSIM values. Finally, Fig. 15 shows the SR results obtained on scaling factor $\times 4$. By looking at the enlarged patches of different SR results, it can be asserted that the proposed model produces SR result with better perceptual details than that of other models. The proposed model achieves highest PSNR measures with an exception for the EDSR model. While the proposed model achieves comparable PSNR measure, it notes the highest performance on SSIM measures than other models.

It can be seen from Figs. 7–15 that the proposed model exhibits SR results with better conserved high-frequency details and also achieves noticeable quantitative measures than competing methods on all validation datasets.

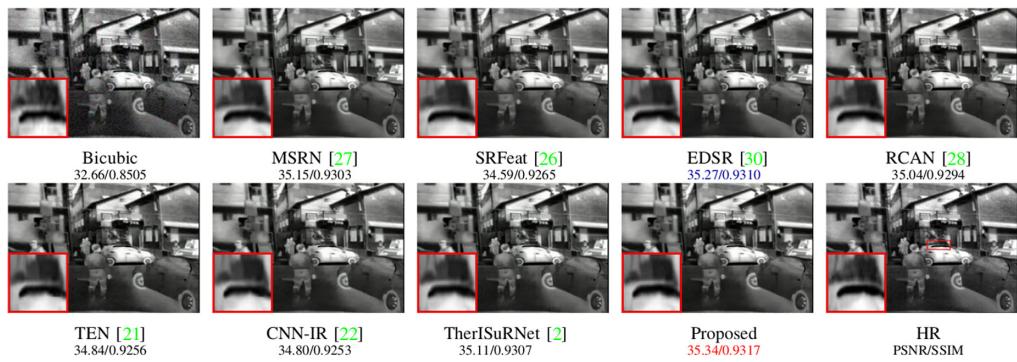


Fig. 14 The qualitative comparison of different thermal SR approaches obtained using KAIST validation dataset³² for scaling factor $\times 3$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom left corner of each SR result.

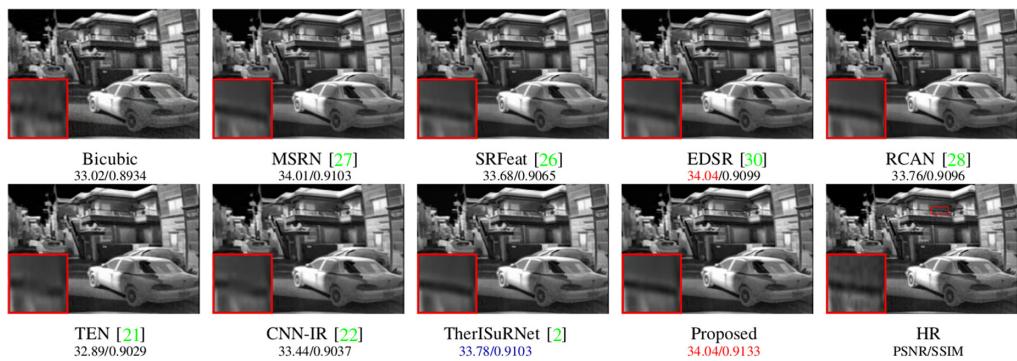


Fig. 15 The qualitative comparison of different thermal SR approaches obtained using KAIST validation dataset³² for scaling factor $\times 4$. The zoomed-in patches of small region highlighted with red border in HR image are displayed at bottom left corner of each SR result.

5 Conclusion

A computationally efficient CNN-based architecture (referred to as ThermISRnet) is presented for thermal image SR for different upscaling factors such as $\times 2$, $\times 3$, and $\times 4$. The progressive upscaling with asymmetrical learning strategy and GRL are utilized in the proposed model. The proposed network is optimized with weighted combination of L_1 and SSIM losses to improve the SR performance. To verify the efficacy of the proposed method for generalizability, numerous experiments have been conducted on different datasets with cross-dataset settings. The proposed approach has clearly shown an improvement over PBVS challenge result as well as other competitive state-of-the-art SR methods in terms of both qualitative and quantitative assessments. The obtained SR results from the proposed method could be deployed in many night-vision applications, such as object detection and recognition where the resolution of the images is challenging.

Acknowledgments

Authors are thankful to Science and Engineering Research Board (SERB), a statutory body of Department of Science and Technology (DST), Government of India for providing hardware support for conducting experiments (ECR/2017/003268).

References

- R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Mach. Vision Appl.* **25**, 245–262 (2014).

2. A. C. Goldberg, T. Fischer, and Z. I. Derzko, "Application of dual-band infrared focal plane arrays to tactical and strategic military problems," *Proc. SPIE* **4820**, 500–514 (2003).
3. H. Qi and N. A. Diakides, "Thermal infrared imaging in early breast cancer detection—a survey of recent research," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc. (IEEE Cat. No. 03CH37439)*, Vol. 2, pp. 1109–1112 (2003).
4. L. Zhang, H. Guo, and Z. Li, "Application of medical infrared thermal imaging in the diagnosis of human internal focus," *Infrared Phys. Technol.* **101**, 127–132 (2019).
5. C. Herrmann, M. Ruf, and J. Beyerer, "CNN-based thermal infrared person detection by domain adaptation," *Proc. SPIE* **10643**, 1064308 (2018).
6. X. Zhao et al., "Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification," *Pattern Recognit.* **48**(6), 1947–1960 (2015).
7. J. Baek et al., "Efficient pedestrian detection at nighttime using a thermal camera," *Sensors* **17**(8), 1850 (2017).
8. B. C. Arrue, A. Ollero, and J. R. M. de Dios, "An intelligent system for false alarm reduction in infrared forest-fire detection," *IEEE Intell. Syst. Appl.* **15**, 64–73 (2000).
9. P. V. K. Borges and S. Vidas, "Practical infrared visual odometry," *IEEE Trans. Intell. Transp. Syst.* **17**, 2205–2213 (2016).
10. T. Mouats et al., "Multispectral stereo odometry," *IEEE Trans. Intell. Transp. Syst.* **16**, 1210–1224 (2015).
11. A. Prata and C. Bernardo, "Retrieval of volcanic ash particle size, mass and optical depth from a ground-based thermal infrared camera," *J. Volcanol. Geotherm. Res.* **186**(1), 91–107 (2009).
12. G. Bitelli et al., "Aerial thermography for energetic modelling of cities," *Remote Sens.* **7**(2), 2152–2170 (2015).
13. M. Fox, S. Goodhew, and P. D. Wilde, "Building defect detection: external versus internal thermography," *Build. Environ.* **105**, 317–331 (2016).
14. K. Ribeiro-Gomes et al., "Uncooled thermal camera calibration and optimization of the photogrammetry process for UAV applications in agriculture," *Sensors* **17**(10), 2173 (2017).
15. P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Vegetation index estimation from monospectral images," *Lect. Notes Comput. Sci.* **10882**, 353–362 (2018).
16. Z. He et al., "Cascaded deep networks with multiple receptive fields for infrared image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.* **29**, 2310–2322 (2019).
17. A. Rogalski, P. Martyniuk, and M. Kopytko, "Challenges of small-pixel infrared detectors: a review," *Rep. Prog. Phys.* **79**, 046501 (2016).
18. Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.* (2020).
19. Y. Choi et al., "Thermal image enhancement using convolutional neural network," in *IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, pp. 223–230 (2016).
20. P. Bhattacharya, J. Riechen, and U. Zölzer, "Infrared image enhancement in maritime environment with convolutional neural networks," in *Proc. 13th Int. Joint Conf. Comput. Vision, Imaging and Comput. Graphics Theory and Appl.*, Vol. 4 (2018).
21. S. Liu et al., "Infrared image super resolution using GAN with infrared image prior," in *IEEE 4th Int. Conf. Signal and Image Process.*, pp. 1004–1009 (2019).
22. I. Marivani et al., "Multimodal image super-resolution via deep unfolding with side information," in *27th Eur. Signal Process. Conf.*, pp. 1–5 (2019).
23. R. E. Rivadeneira et al., "Thermal image super-resolution challenge-pbvs 2020," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 96–97 (2020).
24. R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, "Thermal image super-resolution: a novel architecture and dataset," in *Int. Conf. Comput. Vision Theory and Appl.*, pp. 1–2 (2020).
25. B. Lim et al., "Enhanced deep residual networks for single image super-resolution," in *IEEE Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 1132–1140 (2017).
26. V. Chudasama et al., "TherISuRNet—a computationally efficient thermal image super-resolution network," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 86–87 (2020).

27. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4681–4690 (2017).
28. S.-J. Park et al., "SRFeat: single image super-resolution with feature discrimination," in *Proc. Eur. Conf. Comput. Vision*, pp. 439–455 (2018).
29. J. Li et al., "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vision*, pp. 517–532 (2018).
30. Y. Zhang et al., "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vision*, pp. 286–301 (2018).
31. "Free flir thermal dataset for algorithm training," <https://www.flir.in/oem/adas/adas-dataset-form/>.
32. S. Hwang et al., "Multispectral pedestrian detection: benchmark dataset and baseline," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.* (2015).
33. R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Adv. Comput. Vision and Image Process.*, pp. 317–339 (1984).
34. M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.* **6**, 1646–1658 (1997).
35. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.* **5**, 996–1011 (1996).
36. D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *IEEE 12th Int. Conf. Comput. Vision*, pp. 349–356 (2009).
37. J. Yang et al., "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**, 2861–2873 (2010).
38. C. Dong et al., "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2016).
39. W. Lai et al., "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 2599–2613 (2019).
40. H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning—a review," in *Sixth Int. Conf. Image Process. Theory, Tools and Appl.*, IEEE, pp. 1–6 (2016).
41. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).
42. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1646–1654 (2016).
43. J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1637–1645 (2016).
44. Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *IEEE Conf. Comput. Vision and Pattern Recognit.* (2017).
45. W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1874–1883 (2016).
46. C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," *Lect. Notes Comput. Sci.* **9906**, 391–407 (2016).
47. I. Goodfellow et al., "Generative adversarial nets," in *Adv. Neural Inf. Process. Syst.*, Z. Ghahramani et al., Eds., Curran Associates, Inc., pp. 2672–2680 (2014).
48. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
49. J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE Int. Conf. Comput. Vision* (2017).
50. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv:1511.06434 (2015).
51. R. E. Rivadeneira et al., "Thermal image superresolution through deep convolutional neural network," *Lect. Notes Comput. Sci.* **11663**, 417–426 (2019).

52. P. Kansal and S. Nathan, "A multi-level supervision model: a novel approach for thermal image super resolution," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 426–431 (2020).
53. E. Mandanici et al., "A multi-image super-resolution algorithm applied to thermal imagery," *Appl. Geom.* **11**(3), 215–228 (2019).
54. A.-C. Guei and M. Akhloufi, "Deep learning enhancement of infrared face images using generative adversarial networks," *Appl. Opt.* **57**, D98–D107 (2018).
55. X. Kuang et al., "Single infrared image enhancement using a deep convolutional neural network," *Neurocomputing* **332**, 119–128 (2019).
56. Z. Fan et al., "Dim infrared image enhancement based on convolutional neural network," *Neurocomputing* **272**, 396–404 (2018).
57. K. Lee et al., "Brightness-based convolutional neural network for thermal image enhancement," *IEEE Access* **5**, 26867–26879 (2017).
58. L. Trottier et al., "Parametric exponential linear unit for deep convolutional neural networks," in *16th IEEE Int. Conf. Mach. Learn. and Appl.*, IEEE, pp. 207–214 (2017).
59. M. S. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: single image super-resolution through automated texture synthesis," in *IEEE Int. Conf. Comput. Vision*, IEEE, pp. 4501–4510 (2017).
60. S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vision Graph. Image Process.* **39**(3), 355–368 (1987).
61. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Int. Conf. Mach. Learn.* (2010).
62. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, Vol. **30**, No. 1, p. 3 (2013).
63. K. He et al., "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 1026–1034 (2015).
64. A. Shah et al., "Deep residual networks with exponential linear unit," in *Proc. Third Int. Symp. Comput. Vision and Internet*, pp. 59–65 (2016).

Heena Patel is pursuing PhD from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. She received her BE degree in electronics and communications engineering from Gujarat Technological University, India, and her ME degree in communication systems from Sarvajanik College of Engineering, India. Her research interests include image enhancement, domain translation, image super-resolution, and computer vision applications using deep learning algorithms.

Vishal Chudasama received his bachelor's degree from Maharaja Sayajirao University, Vadodara, India and his master's degree in communication system from Dharmsinh Desai University, Nadiad, India. He is pursuing PhD from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. His research interests include image processing, and deep learning with application to super-resolution, object detection and recognition, low-resolution face detection and recognition, medical imaging, and biometrics.

Kalpesh Prajapati is pursuing PhD at the Sardar Vallabhbhai National Institute of Technology, Surat, India. He received his master's degree in automatic control and robotics from the Maharaja Sayajirao University of Baroda, India and bachelor's degree in electronics and communication from Dharmsinh Desai University, Nadiad, India. His research interests include image enhancement, single-image super-resolution, image quality assessment, unsupervised learning, and medical imaging.

Kishor P. Upla is an assistant professor in Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. He worked as an ERCIM postdoctoral fellow with NTNU, Gjøvik, Norway. He received his PhD from Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India. His areas of interest include signal and image processing, low-resolution face recognition, biometric, and multispectral and hyperspectral image analysis. He is also a member of European Association for Biometrics.

Kiran Raja received his PhD in computer science from the Norwegian University of Science and Technology, Norway, where he is faculty member with the Department of Computer Science. His main research interests include statistical pattern recognition, image processing, and machine learning with applications to biometrics, security and privacy protection. He was/is participating in EU projects SOTAMD, iMARS, and other national projects. He is a member of European Association of Biometrics.

Raghavendra Ramachandra is currently appointed as a full professor with the Institute of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology, Gjøvik, Norway. His main research interests include deep learning, machine learning, data fusion schemes, and image/video processing, with applications to biometrics, multimodal biometric fusion, human behavior analysis, and crowd behavior analysis. He was/is also involved in various conference organizing and program committees and serving as an associate editor for various journals.

Christoph Busch is member of the Norwegian University of Science and Technology. He holds a joint appointment with HDA, Germany. For the German BSI, he has been the Coordinator for numerous biometric projects. He was/is a partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD and iMARS. He is also a principal investigator in ATHENE and is a co-founder of the European Association for Biometrics. He has co-authored more than 500 technical papers. He is a convenor of WG3 in ISO/IEC SC37.