TherISuRNet - A Computationally Efficient Thermal Image Super-Resolution Network

Vishal Chudasama[1], Heena Patel[1], Kalpesh Prajapati[1], Kishor Upla[1, 2],
Raghavendra Ramachandra[2], Kiran Raja[2], Christoph Busch[2]
[1]Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India.
[2]Norwegian University of Science and Technology (NTNU), Gjøvik, Norway.
{vishalchudasama2188, hpatel1323, kalpesh.jp89, kishorupla}@gmail.com,
{raghavendra.ramachandra, kiran.raja, christoph.busch}@ntnu.no

Abstract

Human perception is limited to perceive the objects beyond the range of visible wavelengths in the Electromagnetic (EM) spectrum. This prevents them to recognize objects in different conditions such as poor illumination or severe weather (e.g., under fog or smoke). The technological advancement in thermographic imaging enables the visualization of objects beyond visible range which enables it's use in many applications such as military, medical, agriculture, etc. However, due to the hardware constraints, the thermal cameras are limited with poor spatial resolution when compared to similar visible range RGB cameras. In this paper, we propose a Super-Resolution (SR) of thermal images using a deep neural network architecture which we refer to as TherISuRNet*. We use 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* TherISuRNet *is verified by evaluating it with different datasets. The obtained results indicate superior results as compared to other state-of-the-art SR methods.*

1. Introduction

The human visual system can perceive ae scene in the visible spectrum which spans approximately from 380nm to 720nm. Built on the principles of the human visual system, RGB cameras in principle sense the reflected energy from the objects in the scene to capture an image. However, during night time or in severe weather conditions, the limited visible light leads to a captured image with almost no details when a regular RGB camera is used. In such a situation, an external illumination can be employed such that reflection can be captured to a certain degree. While this is a reasonable approach, the captured image may be sub-optimal due to inherent limitations of quantum efficiency of the regular RGB cameras.

Alternatively, to represent the objects beyond the human's perception capability and in extreme visibility conditions, thermal imaging can be employed. Thermal imaging enables the visualization during nighttime or even in the presence of fog or smoke. Thermal cameras are passive sensors which sense the infrared radiation emitted by all objects with a temperature above absolute zero [9]. Due to this, they are invariant against the complex conditions such as lack of illumination or severe weather conditions and also they do not require any external source of illumination. The recent technological advancement in thermal imaging has made many real-world applications possible [9] such as in military [11], medical [34], pedestrian & person detection [16], visual odometry [3], maritime [15], etc.

Despite the ability to capture the image in challenging conditions, thermal cameras often come with limited spatial resolution as compared to that of RGB cameras which typically provide mega-pixels of resolution. The spatial resolution of the thermal sensors cannot be extended beyond a certain limit due to limitations of Signal to Noise Ratio (SNR) of the sensor area in the cameras. On the other hand, increasing the size directly increases the cost making the technology non-affordable and thereby makes it difficult to increase the spatial resolution. Furthermore, in order to achieve accurate thermal measurement, infrared detectors are normally encapsulated in individual vacuum packages which is a time consuming and expensive process [38]. Due to these constraints, the cost of a thermal camera is comparatively higher than the one with similar resolution for visible imaging.

In order to deal with the limitations posed by the hardware, it is necessary to supplement High-Resolution (HR) thermal imaging with an algorithmic solution. One direct implication is the efforts to super-resolve the thermal images which are typically of Low-Resolution (LR) in nature. A vast amount of work has been reported for achieving super-resolved images from classical RGB cameras [44]. Motivated by such works, a set of recent works have focused on super-resolving the thermal images [4, 2, 27, 29, 15]. In a continued effort in this direction, we present a new approach based on deep learning (using Convolutional Neural Network (CNN)) to super-resolve the LR images obtained from the thermal camera. Our approach is motivated by the promising results by CNNs coupled with availability of large scale datasets and good computation capability. In this work, we propose the Super-Resolution (SR) architecture for thermal images using CNNs which is computationally efficient with promising results. We refer to proposed network as *TherISuRNet* in the remainder of this paper. Unlike previous works, we use a progressive upscaling strategy with residual learning in order to obtain SR from an LR thermal image. Further, in an effort towards 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 [22, 31, 24, 25, 47] as well as thermal SR methods [4, 2]. The key contributions of this work can be summarized as below:

- This work proposes a new network architecture for super-resolving thermal images.
- The proposed approach is computationally efficient needing only 3.91M parameters to a obtain thermal SR image, and is also 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 dataset [36] and testing it on two disjoint datasets such as FLIR [1] and KAIST [17] which are unseen during the training.

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

2. Related Work

The visible image SR is a classical problem, yet it is a challenging and open research problem in the computer vision community. The different SR techniques can be broadly categorised as single-image SR (SISR) and multi-image SR (MISR). The task of SISR is more challenging as compared to MISR as it has one single LR observation in order to perform the SR task. The earlier works for SISR can be roughly classified as interpolation-based SR, traditional SR and deep learning based SR.

Following the early work on SR approach by Tsai and Huang [43], a number of traditional SR methods have been proposed using the principle of reconstruction [7, 39]. 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. A set of traditional SR works have employed the concept of patch based self-similarity from LR and HR images pairs [10]. Sparse representation was further explored to create dictionaries for both LR and HR images to achieve even better SR images [45]. Exploiting the recent advancements in deep learning, a number of 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 [5, 21]. This new direction has motivated works explicitly in various domains such as spectral imaging [32], medical imaging [26] including the thermal imaging applications [4, 2, 27, 29].

Dong et al. [5] proposed the first CNN based SR approach termed as Super-Resolution Convolutional Neural Network (SRCNN). In a following work, VDSR network [18] showed significant improvement over the SRCNN by increasing the network depth from 3 to 20 convolutional layers. To achieve fast convergence speed, VDSR method adopts the global residual learning paradigm to predict the difference between the bicubic upsampled LR image and original ground truth HR image instead of the actual pixel value. Inspired by these works [18, 5], many works on image SR have been published in [19, 41] which use bicubic interpolation to pre-upsample input LR image and then apply a deep network which increases the computational costs.

While other works are based on post-upscaling strategy for upscaling of input LR observation [40, 6]. Deep Laplacian Pyramid based SR Network (LapSRN) is described by Lai et al. [21] in which the sub-band residuals of HR images are progressively reconstructed at multiple pyramid levels. Recently, many SR approaches using CNN such as SRFeat-M [31], MSRN [24], EDSR [25] and RCAN [47] obtained state-of-the-art performance for visible LR images. The Generative

Adversarial Networks (GANs) [12] are further used as unsupervised learning models for achieving SR image in the recent years. Ledig et al. [22] proposed single image SR using GAN called as *SRGAN* which serves as a new state-of-the-art with impressive performance using a deep residual network (ResNet) with skip connection [14]. Following the initial works, many works on SR based on GAN model have been reported recently in [48, 35].

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 as *Thermal Enhancement Network (TEN)* was reported in [4] which was based on the SRCNN model [5]. It has to be noted that the TEN method [4] employed RGB images in the training process due to unavailability of large scale thermal image dataset. On the similar idea, Marivani et al. [29] obtained SR of Near-Infrared (NIR) images by using RGB images as a auxiliary information. Furthermore, Rivadeneira et al. [37] 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 [4, 29]. Bhattacharya et al. [2] propose two CNN models for denoising as well as SR for maritime infrared images. Recently, He et al. [15] 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 first level was used to restore the structure related information and second CNN network level was utilized to obtain fine details in the thermal images. Lastly, Mandanici et al.[28] obtained SR of thermal imagery using the concept of multi-image SR (MISR) approach. In addition to thermal SR, many works also focus on enhancement of the thermal images. For example, authors in [8] use the CNN network to improve the contrast between target and background in the testing image. Additionally, Lee et al. [23] propose infrared image enhancement based on the brightness of the RGB images. They trained their network on RGB images and obtain the residual thermal image at the output of CNN network. The final enhanced thermal image is obtained after adding residual image with the input thermal image as based on VDSR [18].

Inspired from SRGAN [22], Liu et al. [27] use GAN model to obtain SR of 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 et al. [13] use the DCGAN model [35] to obtain SR of NIR and Long-Wavelength Infrared (LWIR) images for upscaling factor $\times 4$. In [20], authors utilize conditional GAN to enhance contrast of given infrared image which is capable to remove background noise present in infrared images. Furthermore, Rivadeneira et al. [36] released a dataset of thermal image SR and perform SR of thermal image using CycleGAN [48] for upscaling factor $\times 2$.

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., [4, 13, 15]) are fixed to a particular upscaling factor limiting the applicability in real-life use cases.
- The approaches proposed for SR of thermal images are computationally inefficient due to large amount of parameters (i.e., [15, 48]).
- The robustness of thermal SR methods has not been tested in cross-database setting (i.e., [4, 37]). Most of these works employ the same dataset and split them in training and testing set limiting the insights on generalizability of proposed approaches.

3. Proposed Methodology

Noting the limitations, we present a new architecture for the task of SR of thermal images specifically targeting to generalizability and various upscaling factors. Fig. 1 depicts the architecture in which the thermal LR image (i.e., I^{LR}) is applied as an input to the network to obtain it's corresponding SR image for upscaling factors of $\times 2$, $\times 3$ and $\times 4$ (i.e., $I^{SR}_{\times 2}$, $I^{SR}_{\times 3}$ and $I^{SR}_{\times 4}$). We specifically employ the progressive upscaling with residual asymmetrical learning in the proposed architecture. The LR thermal image is first passed through feature extraction module to extract effective features from the thermal image. This module is followed by upsampling module for factor $\times 2$. This process is repeated in order to produce final SR image with desired upscaling factor. The proposed network consists following four modules which are designed for specific tasks:

- Low-frequency Feature Extraction module (referred as LFE),
- High-frequency Feature Extraction module-1 (referred as HFE_1),
- High-frequency Feature Extraction module-2 (referred as HFE_2)and
- Reconstruction module (referred as REC).

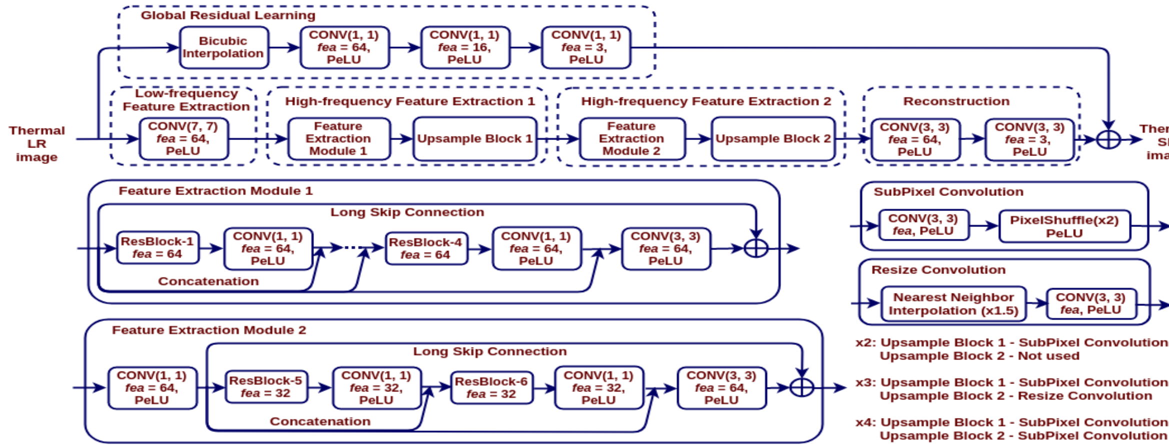

Figure 1: The network architecture design of the proposed model-TherISuRNet. Here, *fea* indicates the number of feature maps.

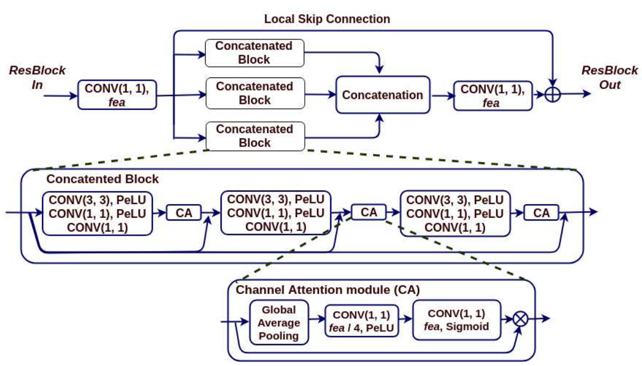

Figure 2: The design of the residual block used in the feature extraction module of the proposed model.

3.1. Low-frequency Details Extraction

The LR image is first passed through the LFE module which has one convolution layer with kernel size of 7 and a feature maps of size 64 with the use of a stride of 1. Here, Parametric Exponential Linear unit (PeLU) is used as activation function in the proposed model. As the parameters of PeLU are learned to make proper activation shape at each convolution layer, learning of activation at different layers using PeLU improves the performance in our architecture [42]. The LFE module extracts the low-frequency details from the LR thermal image which can be represented as,

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

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

3.2. High-frequency Details Extraction

While low-frequency details are handled by LFE, high frequency details pertaining to edges and structures from the feature maps are obtained by passing the output from LFE module which has two high-frequency feature extraction modules (i.e., HFE_1 and HFE_2). Both modules consist of a feature extraction module and upsampling block. The feature extraction module has several residual blocks connected via concatenation operation and one long skip connection (see Fig. 1). The design of the residual block used in feature extraction module is depicted in Fig. 2. It employs one convolution layer with kernel size of 1 which is followed by three parallel branches of concatenated blocks. These concatenated block consist of several convolution layers and Channel Attention (CA) modules. Inspired by [47], the CA module is adopted to re-scale the channel-wise features adaptively. Such structure of concatenated block helps to learn features of the thermal image in an effective manner. The obtained feature maps from three parallel concatenated blocks are further concatenated and then passed through one convolution layer with kernel size of 1 which acts as transition layer and produces the desired number of feature maps. After each residual block, a short skip connection is used to reduce the vanishing gradient problems.

We use different number of residual blocks in each feature extraction module in order to perform an asymmetrical learning. First feature extraction module comprises four residual blocks while second feature extraction module uses only two residual blocks. The feature maps from the feature extraction module is passed to upsample block in order to upscale the feature maps to the desired scale factor. We use different upscaling strategies in upsample blocks. In case of $I^{SR}_{\times 2}$

(i.e., SR with factor 2), only single upsample block 1 is used which is made up of sub-pixel convolution operation with factor 2 (i.e., as depicted in Fig. 1) while for $I^{SR}_{\times 3}$ (i.e., SR with factor 3), the sub-pixel convolution with factor 2 and resize convolution with factor 1.5 are used in upsample block 1 and upsample block 2, respectively. Here, use of resize convolution in second upsample block is to perform overall upscaling of factor 3 for the given thermal LR observation. In the case of upscaling factor ×4 (i.e., $I^{SR}_{\times 4}$), we use sub-pixel convolution in both upsample blocks. The output feature maps of the HFE_2 module is represented as,

$$I_{HFE_2} = f_{HFE_2}(f_{HFE_1}(I_{LFE})). \quad (2)$$

Here, f_{HFE_1} and f_{HFE_2} denote the function of the HFE_1 and HFE_2 modules.

3.3. SR Image Reconstruction

Given the feature maps of the LR image obtained through LFE, HFE_1 and HFE_2 modules, the final thermal SR image is reconstructed through the reconstruction module (REC). Specifically, this module has two convolution layers to obtain the residual SR image and it can be indicated in Equation (3) as,

$$I^{SR}_{residual} = f_{REC}(I_{HFE_2}), \quad (3)$$

where, f_{REC} indicates the reconstruction function of the REC module.

Additionally, we also implement the Global Residual Learning (GRL) in which input LR observation ($i.e., I^{LR}$) is passed through a bicubic interpolation layer followed by three convolution layers with kernel size of 1 which produce the super-resolved image I^{SR}_{GRL}. Here, the LR observation is interpolated with factor of 2, 3 and 4 as per the corresponding SR operation. Such learning (i.e., GRL) helps the network to learn the identity function for I^{LR} and it also stabilizes the training process. Finally, the network generates the SR image (I^{SR}) at an upscaling factor ×2, ×3, ×4 as given in Equation (4) as,

$$I^{SR}_{\times 2, \times 3, \times 4} = I^{SR}_{residual} + I^{SR}_{GRL}. \quad (4)$$

4. Experimental Analysis

In order to validate the proposed method, different experiments[1] have been conducted and it's detailed analysis is presented in this section.

[1] All the experiments are performed on a computer with Intel Xeon(R) CPU E5-2620 v4 processor @2.10GHz ×32 running on a 128GB RAM and two NVIDIA Quadro P5000 with 16GB GPUs.

4.1. Hyper-parameter Settings

The proposed method is trained on PBVS challenge training dataset [36] which has three training sub-dataset modules called as *Domo*, *Axis* and *GT* for ×2, ×3 and ×4 upscaling factors, respectively. Each sub-dataset module in PBVS challenge dataset consists of 951 training images. These images are augmented using horizontal flipping, 180^o rotation and warp affine operation. The LR thermal images in PBVS challenge dataset are generated by adding Gaussian noise with mean 0 and standard deviation 10 followed by down-sampling operation via bicubic interpolation. The proposed model is trained up to 5×10^4 iterations with a batch size of 16 and it is optimized using Adam optimizer with an initial learning rate of 1×10^{-4}. Furthermore, the proposed model is trained using the weighted combination of three reconstruction loss functions: L_1, Structural Similarity Index Measure (SSIM) and Contextual (CX) [30] instead of single reconstruction loss function as indicated by Equation (5).

$$L_{SR} = 10L_1 + 10SSIM + 0.1CX. \quad (5)$$

The proposed model along with the other state-of-the-art SR methods are tested on three different datasets: PBVS challenge (i.e., Domo, Axis and GT) [36], FLIR [1] and KAIST [17] validation datasets for upscaling factor of ×2, ×3 and ×4. The PBVS challenge validation dataset consists 50 number of validating images. However, the FLIR validation dataset [1] comprises 1366 number of thermal images. The KAIST validation dataset [17] is generated by randomly selecting a number of 500 thermal images from their complete testing dataset and then these images are enhanced using adaptive histogram equalization technique [33]. The images of FLIR and KAIST datasets correspond to a size of 640×512 which are resized into 640×480 as HR images. The corresponding LR pair images of testing datasets for upscaling factors ×2, ×3 and ×4 are generated by adding Gaussian noise with mean 0 and standard deviation of 10 followed by corresponding bicubic downsampling operation.

The qualitative and quantitative evaluations of the proposed method are performed by comparing the thermal SR results with the state-of-the-art visible image SR techniques such as SRResNet [22], SRFeat-M [31], MSRN [24], EDSR [25] and RCAN [47] as well as recently proposed thermal SR approaches (TEN [4] and [2]). For the fair comparison, the SR results of those methods are generated by re-training them on the same training dataset of the proposed method with same training strategy. Furthermore, for quantitative analysis, we use different measures such as Peak Signal to Noise Ratio (PSNR) and SSIM. These measurements are calculated after removing the four boundary pixels of Y-channel images in YCbCr color space. Additionally, we also use Learned Perceptual Image Patch Similarity (LPIPS) metric [46] which measures the perceptual similarity between SR and HR images. A lower value of LPIPS indicates a better perceptual quality of SR image.

4.2. Result Analysis

In this sub-section, we present the detailed analysis of the SR performance of the proposed model along with other state-of-the-art SR methods on upscaling factor of ×2, ×3 and ×4. First, we present the ablation study on the proposed model and then SR performance of the proposed model is described.

4.2.1 Ablation Study

In order to see the effectiveness of the proposed CNN architecture, different experiments related to the selection of various components have been carried out and are reported in Table 1. Here, we consider three cases: loss functions, activation functions and network design. The SR performance is compared in terms of PSNR, SSIM and LPIPS measures on GT testing dataset for upscaling factor ×4. First, in order to understand the importance of the weighted loss function used to train the model, the proposed model is trained with different loss functions and it's corresponding measurements are mentioned in the Table 1. It can be noticed here that the proposed model trained using the proposed loss function (Equation (5)) obtains comparable PSNR and SSIM measures with best LPIPS measures than other similar loss functions. In order to understand the importance of PeLU activation function [42], the proposed model is also trained with PReLU and ReLU activation functions. From Table 1, it can be observed that the PeLU activation function helps the model to obtain better PSNR, SSIM and LPIPS measures. Additionally, we show the effectiveness of Channel Attention (CA) and Global Residual Learning (GRL) modules by conducting experiments on the proposed model without CA module as well as without GRL and simple GRL (i.e., only bicubic interpolation layer) modules. From Table 1, one can observe that the proposed model with CA module and proposed GRL justify in terms of better PSNR, SSIM and LPIPS measures.

4.2.2 Parameters and Computational Efficiency

In order to check the computational efficiency of the proposed method in terms of number of parameters with respect to SSIM, we plot the number of parameters vs SSIM in Fig. 3 for Domo, Axis and GT testing datasets for upscaling factor of ×2, ×3 and ×4, respectively. Here, one can observe that the proposed method obtains better SSIM measures with large margin for Domo and GT testing dataset than that of other existing state-of-the-art SR methods with considerable reduction in the number of parameters. In case of Axis testing dataset, the proposed model obtains comparable performance with that of EDSR model. However, the proposed model sets such performance with approximately 90% less number of training parameters than that of EDSR model.

Table 1: The comparison of proposed methods on GT (×4) validating dataset with three different scenarios. Here, the best values are mentioned in bold font text.

	PSNR↑	**SSIM↑**	**LPIPS↓**
Loss Function			
CX	22.7004	0.2613	0.743
L_2	34.4211	0.9055	0.199
L_1	**34.5325**	0.9077	0.200
SSIM	34.4120	**0.9113**	0.194
L_1 + SSIM	34.5299	0.9099	0.193
TherISuRNet (Equation (5))	34.4956	0.9101	**0.190**
Activation Function			
PReLU	34.4101	0.9091	0.187
eLU	34.4348	0.9094	0.191
TherISuRNet PeLU	**34.4956**	**0.9101**	**0.190**
Network Design			
Without CA	34.4445	0.9094	0.191
Without GRL	34.4572	0.9101	0.193
Simple GRL	34.4782	0.9101	0.190
TherISuRNet GRL	**34.4956**	**0.9101**	**0.190**

4.2.3 Fidelity of Thermal SR Images

The quantitative comparison in terms of PSNR, SSIM and LPIPS measures obtained for the state-of-the-art along with the proposed methods are presented in Table 2 for PBVS challenge, FLIR and KAIST datasets for upscaling factor of ×2, ×3 and ×4. Here, the highest value of PSNR and SSIM metrics is highlighted with red color font while the second highest values are with blue color font. Since lower value of LPIPS indicates better perceptual quality, the same is indicated with red colored font while the second lowest value is represented with blue color font in the Table 2. From this table, one can notice that the proposed model obtains better PSNR, SSIM and LPIPS measures in most cases of upscaling factor ×2, ×3 and ×4 for three different testing datasets with large margin than that of other models except that of the EDSR model [25] where it obtains comparable performance. However, it is also worth to mention that the proposed method obtains this SR performance with approximately 35% to 90% reduction in the trainable parameters than that

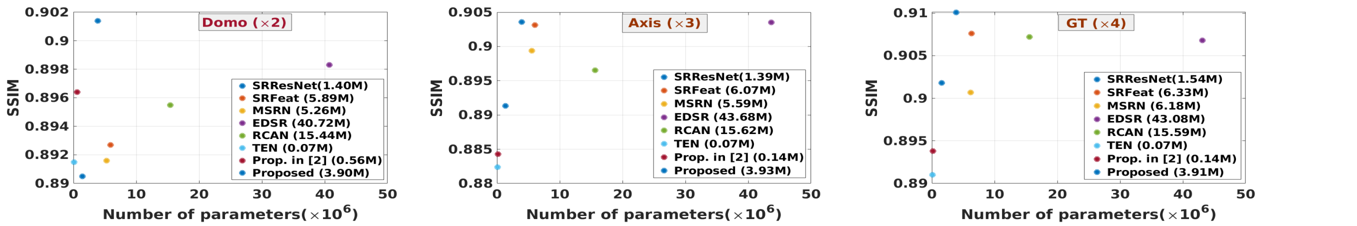

Figure 3: The effect of SSIM value vs. number of training parameters required to train different methods for Domo, Axis and GT testing datasets for upscaling factor of 2, 3 and 4, respectively.

Table 2: The quantitative comparison of the proposed method along with other state-of-the-art methods on different validation datasets in terms of PSNR, SSIM and LPIPS metrics.

	Dataset	Metrics	Bicubic	SRResNet	MSRN	SRFeat	EDSR	RCAN	TEN	Prop. in [2]	TherISuRNet
×2	**Domo**	PSNR	32.1229	33.0817	33.1215	33.1253	33.5248	33.3144	33.1919	33.5272	33.6559
		SSIM	0.8751	0.8905	0.8927	0.8916	0.8983	0.8955	0.8915	0.8964	0.9014
		LPIPS	0.201	0.158	0.147	0.163	0.142	0.152	0.155	0.162	0.145
	FLIR	PSNR	34.3019	34.8267	34.9860	34.9806	35.2352	35.0518	35.0352	35.2632	35.2955
		SSIM	0.8488	0.8651	0.8665	0.8660	0.8698	0.8684	0.8657	0.8687	0.8720
		LPIPS	0.276	0.232	0.222	0.244	0.219	0.235	0.236	0.260	0.221
	KAIST	PSNR	37.1974	37.3715	37.5627	37.5051	37.7663	37.5993	37.5356	37.8287	37.7233
		SSIM	0.9319	0.9444	0.9449	0.9458	0.9467	0.9462	0.9455	0.9473	0.9474
		LPIPS	0.205	0.105	0.101	0.100	0.098	0.098	0.107	0.113	0.098
×3	**AXIS**	PSNR	30.3577	32.5174	33.1015	32.4329	33.1278	32.8011	33.1311	32.2217	32.9803
		SSIM	0.8032	0.8913	0.9031	0.8894	0.9035	0.8965	0.8824	0.8843	0.9036
		LPIPS	0.415	0.157	0.156	0.175	0.154	0.161	0.166	0.165	0.147
	FLIR	PSNR	30.3373	32.2763	32.4962	32.2582	32.5487	32.3345	32.0931	32.1912	32.3202
		SSIM	0.7475	0.8273	0.8331	0.8263	0.8342	0.8307	0.8215	0.8232	0.8332
		LPIPS	0.663	0.331	0.334	0.334	0.331	0.328	0.414	0.337	0.338
	KAIST	PSNR	32.3202	34.0937	34.1822	33.7220	34.3233	34.1102	33.9656	34.1729	34.1499
		SSIM	0.8332	0.8971	0.8978	0.8905	0.8991	0.8972	0.8958	0.8950	0.9000
		LPIPS	0.477	0.237	0.274	0.264	0.266	0.256	0.250	0.243	0.243
×4	**GT**	PSNR	32.6657	33.1240	34.4718	34.1245	34.4852	34.4200	33.6230	33.7723	34.4956
		SSIM	0.8625	0.9018	0.9076	0.9007	0.9068	0.9072	0.8910	0.8938	0.9101
		LPIPS	0.383	0.229	0.194	0.210	0.202	0.204	0.221	0.221	0.190
	FLIR	PSNR	30.1153	30.3533	30.7161	30.7513	30.8986	30.8275	30.5943	30.6758	30.8108
		SSIM	0.7467	0.7551	0.7702	0.7683	0.7730	0.7728	0.7625	0.7656	0.7769
		LPIPS	0.533	0.399	0.395	0.404	0.397	0.402	0.410	0.418	0.401
	KAIST	PSNR	32.4649	32.0788	32.9730	32.8661	33.0546	32.9962	32.5402	32.7842	32.6999
		SSIM	0.8707	0.8652	0.8799	0.8773	0.8795	0.8804	0.8758	0.8758	0.8790
		LPIPS	0.355	0.259	0.274	0.278	0.278	0.282	0.280	0.287	0.273

of MSRN [24] and SRFeat-M [31], RCAN [47] and EDSR [25] SR methods. In the case of comparison with other thermal SR methods (i.e., [4, 2]), the proposed TherISuRNet model outperforms to these methods with large margin except for the case of KAIST validation dataset where SR method in [2] performs slightly better.

Finally, to see the qualitative improvement achieved in the proposed method, we display the SR results obtained using the proposed and other existing state-of-the-art visible SR methods (i.e., SRResNet [22], SRFeat [31], MSRN [24], RCAN [47] and EDSR [25]) and thermal SR methods (i.e., [4, 2]) in Fig. 4 for the upscaling factor of ×2, ×3 and ×4 of a single image of PBVS challenge validation dataset and for upsclaing factor ×4 of single image of FLIR and KAIST validation thermal datasets due to space constraint. The quantitative measurements (i.e., SSIM and LPIPS values) corresponding to that sample image are also depicted at the bottom of each SR results. From Fig. 4, it can be observed that the proposed model obtains better high frequency details along with better quantitative measures than that of the other competing methods for all testing datasets.

5. Conclusion

In this paper, we proposed a computationally efficient SR approach for thermal images using CNN architecture. We use progressive upscaling with asymmetrical strategy and residual learning in the proposed architecture for different upscaling factors such as ×2, ×3 and ×4. The potential of the proposed method is

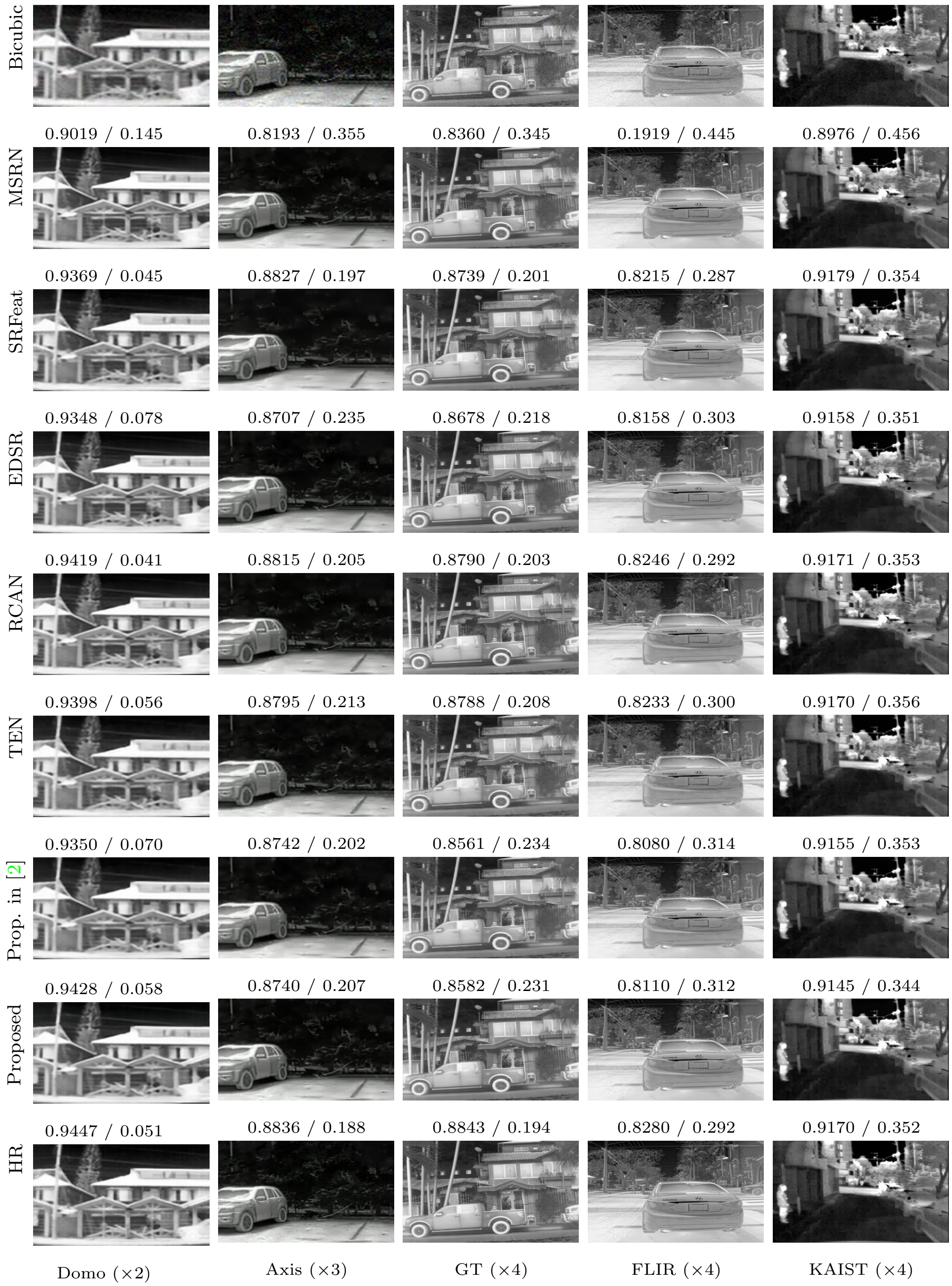

Figure 4: The qualitative comparison on PBVS challenge validation dataset on scaling factor ×2, ×3,×4, FLIR and KAIST validation datasets on scaling factor ×4.

verified by conducting different experiments on various datasets, specifically in cross-dataset settings as a step towards generalizability. The proposed approach has clearly shown an improvement over other competitive state-of-the-art SR methods in terms of both qualitative and quantitative assessments.

Acknowledgment

This work was supported by ERCIM, who kindly enabled the internship of Kishor Upla at NTNU, Gjøvik.

References

[1] Free flir thermal dataset for algorithm training. https://www.flir.in/oem/adas/adas-dataset-form/. 2, 5

[2] Purbaditya Bhattacharya, Jörg Riechen, and Udo Zölzer. Infrared image enhancement in maritime environment with convolutional neural networks. In *VISIGRAPP*, 2018. 2, 3, 5, 7, 8

[3] P. V. K. Borges and S. Vidas. Practical infrared visual odometry. *IEEE Transactions on Intelligent Transportation Systems*, 17(8):2205–2213, Aug 2016. 1

[4] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon. Thermal image enhancement using convolutional neural network. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 223–230, Oct 2016. 2, 3, 5, 7

[5] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, Feb 2016. 2, 3

[6] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, pages 391–407, Oct 2016. 2

[7] M. Elad and A. Feuer. Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658, Dec 1997. 2

[8] Zunlin Fan, Duyan Bi, Lei Xiong, Shiping Ma, Linyuan He, and Wenshan Ding. Dim infrared image enhancement based on convolutional neural network. *Neurocomputing*, 272:396 – 404, 2018. 3

[9] Rikke Gade and Thomas B. Moeslund. Thermal cameras and applications: A survey. *Mach. Vision Appl.*, 25(1):245–262, Jan. 2014. 1

[10] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE 12th International Conference on Computer Vision*, pages 349–356, Sep. 2009. 2

[11] Arnold C. Goldberg, Theodore Fischer, and Zenon I. Derzko. Application of dual-band infrared focal plane arrays to tactical and strategic military problems. In Bjorn F. Andresen, Gabor F. Fulop, and Marija Strojnik, editors, *Infrared Technology and Applications XXVIII*, volume 4820, pages 500 – 514. International Society for Optics and Photonics, SPIE, 2003. 1

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. 3

[13] Axel-Christian Guei and Moulay Akhloufi. Deep learning enhancement of infrared face images using generative adversarial networks. *Appl. Opt.*, 57(18):D98–D107, Jun 2018. 3

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[15] Z. He, S. Tang, J. Yang, Y. Cao, M. Ying Yang, and Y. Cao. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8):2310–2322, Aug 2019. 1, 2, 3

[16] Christian Herrmann, Miriam Ruf, and Jürgen Beyerer. CNN-based thermal infrared person detection by domain adaptation. In Michael C. Dudzik and Jennifer C. Ricklin, editors, *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, volume 10643, pages 38 – 43. International Society for Optics and Photonics, SPIE, 2018. 1

[17] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baselines. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 5

[18] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE CVPR*, pages 1646–1654, June 2016. 2, 3

[19] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1637–1645, June 2016. 2

[20] Xiaodong Kuang, Xiubao Sui, Yuan Liu, Qian Chen, and Guohua Gu. Single infrared image enhancement using a deep convolutional neural network. *Neurocomputing*, 332:119 – 128, 2019. 3

[21] W. Lai, J. Huang, N. Ahuja, and M. Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11):2599–2613, Nov 2019. 2

[22] Christian Ledig, Lucas Theis, Ferenc Huszár, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2, 3, 5, 7

[23] Kyungjae Lee, Junhyeop Lee, Joosung Lee, Sangwon Hwang, and Sangyoun Lee. Brightness-based convolutional neural network for thermal image enhancement. *IEEE Access*, 5:26867–26879, 2017. 3

[24] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 517–532, 2018. 2, 5, 7

[25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolutaion. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. 2, 5, 6, 7

[26] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017. 2

[27] S. Liu, Y. Yang, Q. Li, H. Feng, Z. Xu, Y. Chen, and L. Liu. Infrared image super resolution using gan with infrared image prior. In *IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 1004–1009, July 2019. 2, 3

[28] Emanuele Mandanici, Luca Tavasci, Francesco Corsini, and Stefano Gandolfi. A multi-image super-resolution algorithm applied to thermal imagery. *Applied Geomatics*, 11(3):215–228, 2019. 3

[29] I. Marivani, E. Tsiligianni, B. Cornelis, and N. Deligiannis. Multimodal image super-resolution via deep unfolding with side information. In *27th European Signal Processing Conference (EUSIPCO)*, pages 1–5, Sep. 2019. 2, 3

[30] Roey Mechrez, Itamar Talmi, Firas Shama, and Lihi Zelnik-Manor. Maintaining natural image statistics with the contextual loss. In *Asian Conference on Computer Vision*, pages 427–443. Springer, 2018. 5

[31] Seong-Jin Park, Hyeongseok Son, Sunghyun Cho, Ki-Sang Hong, and Seungyong Lee. Srfeat: Single image super-resolution with feature discrimination. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 439–455, 2018. 2, 5, 7

[32] Henrik Petersson, David Gustavsson, and David Bergstrom. Hyperspectral image analysis using deep learning - a review. 12 2016. 2

[33] Stephen M Pizer, E Philip Amburn, John D Austin, et al. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 5

[34] H. Qi and N. A. Diakides. Thermal infrared imaging in early breast cancer detection-a survey of recent research. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, volume 2, pages 1109–1112 Vol.2, Sep. 2003. 1

[35] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 3

[36] Rafael E. Rivadeneira, Angel D. Sappa, and Boris X. Vintimilla. Thermal image super-resolution: a novel architecture and dataset. In *International Conference on Computer Vision Theory and Applications*, pages 1–2, 2020. 2, 3, 5

[37] Rafael E Rivadeneira, Patricia L. Suarez, Angel D Sappa, and Boris X Vintimilla. Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*, pages 417–426. 3

[38] A Rogalski, P Martyniuk, and M Kopytko. Challenges of small-pixel infrared detectors: a review. *Reports on Progress in Physics*, 79(4):046501, mar 2016. 1

[39] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5(6):996–1011, June 1996. 2

[40] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, June 2016. 2

[41] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, 2017. 2

[42] Ludovic Trottier, Philippe Gigu, Brahim Chaib-draa, et al. Parametric exponential linear unit for deep convolutional neural networks. In *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 207–214. IEEE, 2017. 4, 6

[43] Roger Y. Tsai and Thomas S. Huang. Multiframe image restoration and registration. In *Advances in computer vision and image processing*, pages 317–339, 1984. 2

[44] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey, 2019. 2

[45] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, Nov 2010. 2

[46] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6

[47] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 2, 4, 5, 7

[48] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 3