HAL
open science

Edge focused super-resolution of thermal images

Yannick Zoetgnande, Jean-Louis Dillenseger, Javad Alirezaie

To cite this version:

Yannick Zoetgnande, Jean-Louis Dillenseger, Javad Alirezaie. Edge focused super-resolution of thermal images. International Joint Conference on Neural Networks, Jul 2019, Budapest, Hungary. pp.1-8, 10.1109/IJCNN.2019.8852320 . hal-02270646

HAL Id: hal-02270646

https://hal.science/hal-02270646v1

Submitted on 26 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Edge focused super-resolution of thermal images

Yannick Wend Kuni Zoetgnande[a], Jean-Louis Dillenseger[a], and Javad Alirezaie[b]

[a]Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France
[b]Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, M5B 2K3, Canada
{*yannick.zoetgnande,jean-louis.dillenseger*}*@univ-rennes1.fr*
{*javad*}*@ryerson.ca*

***Abstract*—In this work, a super-resolution method is proposed for indoor scenes captured by low-resolution thermal cameras. The proposed method is called Edge Focused Thermal Super-resolution (EFTS) which contains an edge extraction module enforcing the neural networks to focus on the edge of images. Utilizing edge information, our model, based on residual dense blocks, can perform super-resolution for thermal images, while enhancing the visual information of the edges. Experiments on benchmark datasets showed that our EFTS method achieves better performance in comparison to the state-of-the-art techniques.**

***Keywords*—Thermal image, Neural network, Super-resolution, Edge extraction**

I. Introduction

Thermal infrared images are used in the field of indoor surveillance even if they are subject to many drawbacks such as infrared reflection, infrared halo, and noise. Unlike their visible counterparts, they can be used in daily and nightly situations even if they are more expensive. Recently, a few manufacturers proposed some very low price thermal cameras, e.g., the FLIR Lepton 2, but with a low 80×60 pixels resolution. However, this little resolution induces a loss of accuracy that can alter the efficiency of the final application.

One solution to address this induced low accuracy is to increase the size of images, so adding more information. However, such process, so-called super-resolution, is an ill-posed inverse problem. Indeed, an infinite number of high-resolution images can correspond to the same low-resolution image. When only one image is used during this process, it is called a single image super-resolution (SISR).

Generally, there are three ways to perform SISR: interpolation based methods [1], model-based optimization methods [2, 3] and learning-based methods [4, 5]. As in many other fields, deep learning based methods (a sub-set of learning based methods) have achieved better results than other methods regarding quality. In [6], the authors are the first to propose a convolutional neural network to perform super-resolution. Their network, SRCNN, is composed of 3 modules: features extraction and representation, non-linear mapping and reconstruction. Authors in [7] proposed VDSR which is a very deep network inspired by VGGNet [8]. The difficulty of training is by-passed by global residual learning and gradients clipping. Handling multiple scales helps their model to better generalize and gives better results than a single scale model. To reduce the number of parameters, a recursive neural network called DRCN is proposed in [9]. In [10], the authors proposed to handle denoising, deblocking and multi-scale super-resolution at the same time. But as most of the previous super-resolution methods, they preprocess the low-resolution image by applying bicubic resizing which can introduce some artifacts and increase the computation time.

To encounter the bicubic resizing disadvantages, many works propose to upscale the image at the final stage of the super-resolution process. In [11], the authors presented a scheme to add a sub-pixel convolution layer at the late stage of the network. In [12] a similar idea was proposed by the researchers using deconvolution layer. In [13] authors proposed LapSRN, a multi-scale super-resolution neural network that gradually upscales the image by a factor of 2. To improve the perceptual quality of the super-resolved image, authors in [14] proposed SRGAN using two sub-pixel convolutional layers to upscale the input by 4.

It is also possible to improve the quality of reconstructed images by increasing the depth of networks. But a more complex model is much more difficult to be trained. In [15], the authors proposed DenseNet where the features in each dense block are propagated. Short path connections counteract the vanishing gradients. Deconvolution layers are integrated into the late stage to upscale the low-resolution image. In [16], the authors, inspired by DenseNet, proposed a dense residual network, called RDN. Their network is composed of four parts: shadow features extraction, residual dense blocks, dense fusion module, and an up-scaling module.

While there are many published works for visible (RGB) images, the thermal images have received less attention. Authors in [17] proposed TEN, a thermal enhancement network inspired by SRCNN with three convolutional layers. In [18] authors used two inputs for their networks: a low resolution near infrared and a high-resolution visible image. In [19], the authors proposed a model CNN with skipped connections. Inspired by VDSR, in [20] authors proposed a two convolutional neural network using 20 + 10 layers. The low-resolution image is gradually upscale from 1 to 2 and then to 8.

In most of these reports, the authors used bicubic degradation. But in real-world applications, the degradation is more complicated. Moreover, thermal cameras point spread function can be different of visible cameras point spread function. In [21], authors proposed a network handling multiple models of degradation for visible images. They state that blind model

cannot work well in real applications. The input of their network is the low-resolution image and a degradation map. For known blur kernel and noise, the degradation map is estimated through a dimensionality stretching. Given that real images do not have ground truth, authors performed a grid search to estimate the degradation settings with good visual quality. Such a scheme will be quite challenging for real-time applications. Moreover, for the thermal images, it is difficult to assess the visual quality of the reconstructed images. However, this assumption is not always valid for actual imaging device sensors. Therefore, once generating synthetic low-resolution images for training, we must consider a wide range of noise and blurring artifacts as possible. Unlike previously published super-resolution methods, here we propose a blind model the Edge Focused Thermal Super-resolution (EFTS) to perform single image super-resolution for thermal images. Our model is based on residual dense block preceded by an edge extraction module which focuses the reconstruction on the edge enhancement. Our contributions are threefold:

- First, we investigated to find the best combination of edge operators (Sobel, Kirsch, Laplace, Prewitt) to obtain better results
- Second, we proved that the edge extraction module helps our model to output a reconstructed image with more enhanced edges.
- Third, we showed that in the context of thermal images, very deep networks tend to over-fit given that thermal images contain less pixel variance than their visible counterparts.

The paper is structured as follows: the details of our degradation model and the proposed network are given in Section II. In Section III, results including PSNR, SSIM and Edge Preservation Index are presented. Finally, Section IV concludes the paper and presents some perspectives.

II. Approach

A. Degradation model

There are multiple degradation models reported in [22], however, the basic model is given below:

$$I_{lr}(m,n) = d\left(h\left(w\left(I_{hr}\left(x,y\right)\right)\right)\right) + \sigma\left(m,n\right) \tag{1}$$

where I_{lr} and I_{hr} are respectively low-resolution and high-resolution images, w is a warping function, h is a blurring function, d is a down-sampling operator, and σ is additive noise. This equation has been modified or simplified in many situations.

1) Blur kernel: The blur kernel η has a significant impact on the reconstruction of the image. The most common blur kernel is isotropic Gaussian with kernel size and a standard deviation [23]. In certain conditions, it is also possible to consider anisotropic blur kernel where x and y standard deviations are different [24]. When there is long exposure time, in [25], the authors use more complex blur models such as motion blurring.

The estimation of the blur kernel is essential for the reconstruction. If the estimated blur kernel is smaller than the ground truth blur kernel, reconstructed images are smooth, and if it is bigger than the ground truth blur kernel, there are some artifacts in the reconstructed images.

2) Noise: Low-resolution thermal cameras are more sensitive to noise than higher resolution ones. Infrared images drawbacks such as halo and noise are highlighted. Traditionally, noise is assumed to be Gaussian. To deal with noise, there are many solutions. One way is first to perform denoising and then implement super-resolution. But the denoising process causes the loss of some image information. It is also possible to perform super-resolution before denoising, but this process becomes computationally costly given that it is completed with a high-resolution image. To deal with these drawbacks, denoising, and super-resolution can be performed jointly.

3) Why learn a blind model for thermal images?: In [21], to determine the blur kernel in real-world applications, authors used a visual quality assessment. For real-time applications, this is quite complicated. Moreover, if human eyes can compare the quality of two visible reconstructed images, such a task will be more complicated for thermal images. Even high-resolution thermal images contain noise, blurring, and artifacts.

One could anticipate that the degradation model can be learned only one time for a given thermal camera. But depending on the luminosity, the heat, the distance between the cameras and objects, the quality of the image can be up- or downgraded. For all these reasons, it is difficult to determine the degradation model given that such degradation model is not fixed. This is why we are using a blind model for our network.

Rather than performing SISR for a specific type of degradation model values (blur kernel η and noise σ), we want our network to be as general as possible.

This is why for each image we have generated synthetic low resolution images by randomly selecting a degradation model values to have both enough training images and enough generalization. These values are selected from $\{\eta_{min}, \eta_{max}\} \times \{\sigma^2_{min}, \sigma^2_{max}\}$.

B. Proposed network

Thermal images are highly texture-less. Even high-resolution images are more affected by noise than their visible counterparts. There are also other drawbacks such as infrared halo effects and history effects. Two thermal images taken by the same sensor can be very different. These images suffer from low signal-to-noise ratio (SNR).

However, what is often less sensitive to the variation of temperature or measure error are edges. One of the challenges of super-resolution is to reconstruct salient edges. So, we have integrated a mean to enhance edges in the thermal image through an edge extraction module.

The proposed EFTS model (Fig. 1) is composed of four modules: 1) edge extraction module (EEM) (Fig. 2), 2) shallow feature extractor (SFE), 3) non-linear mapping module (NMM) and finally 4) an upscaling module (UM).

1) Edge extraction module: Image edge detection is one of the basic fields in image processing. To detect edges in

an image I, a kernel is generally convoluted with this image. There are three types of edge operators: classical, Zero crossing (Laplacian of Gaussian), Gaussian and colored edge detectors.

In [26], the authors highlight the advantages and disadvantages of each type of edge detector. The Laplacian of Gaussian (LoG) is a well-known operator that can find correct places of edges, but it does not work well at corners and edges. The main advantage of LoG is that it is noise tolerant, given that a Gaussian kernel first blurs the image. There are classical operators such as Sobel [27], Prewitt [28] and Kirsch [29]. If these methods are simple and can detect edges in many orientations, they would be sensitive to noise. Zero crossing operator such as Laplacian [30] can detect edges and their orientations using fixed characteristics in all direction. On the other hand, Gaussian and colored edge detectors are complex and time-consuming.

The primary objective of the edge extraction module is not to denoise the input image but rather to represent the noise and blurring effect. So, the network would receive additional information as input. Edges have a crucial factor in a thermal image, and these edges have an essential impact on segmentation [31].

All the operators as mentioned earlier respond differently to noise and blurring and can represent many ways to see the same scene. This difference can bring more information to our network. But how to combine this information to improve the network? We have tried the following operators: Prewitt, Sobel, Laplacian, Kirsch and their combinations (Sobel-Kirsch-Laplacian, Sobel-Kirsch-Prewitt, and Kirsch-Prewitt-Laplacian).

The EEM module takes the original low-resolution degraded image as input and output Υ.

$$\begin{aligned}\Upsilon &= EEM(I_{dlr}) \\ &= \Gamma_1 \otimes \Gamma_2 \otimes \cdots \otimes \Gamma_n\end{aligned} \tag{2}$$

where Γ_i is the ith edge extractor, n the number of edge extractors and $\otimes$ is the concatenation operator.

Then F_{EM} is concatenated with the original low resolution degraded image I_{dlr}. So we have:

$$\Lambda = I_{dlr} \otimes \Upsilon \tag{3}$$

2) Shallow feature extractor module: For shallow features extraction, we use one convolutional layer as proposed by [32]. This is also a difference in our model compared to RDN. Given that F' is composed of very different information we first fuse them in a 1×1 convolutional neural layer. So, we have:

$$\begin{aligned}\Psi_0 &= Fu\left[\Lambda\right] \\ \Psi_1 &= SF\left[\Psi_0\right]\end{aligned} \tag{4}$$

where Fu is a 1×1 convolutional layer and SF a 3×3 convolution layer.

3) Non-linear mapping: The non-linear mapping module allows learning the non-linear mapping between I_{dlr} and I_{hr}.

This part is inspired by RDN [16]. Their dense residual network can extract hierarchical features through contiguous memory, local feature fusion, local residual learning, and global residual learning. Their model is also based on DenseNet [15] and MemNet [33].

The global residual learning induces:

$$\zeta = \Psi_0 + \Lambda \tag{5}$$

where

$$U = DFu\left(\Phi_1 \otimes \Phi_2 \otimes \cdots \otimes \Phi_D\right) \tag{6}$$

Where DFu expresses dense feature fusion composed of a 1×1 convolutional layer followed by a 3×3 convolutional layer, Φ_i the output of the ith residual block and D the number of residual blocks. The output of the ith residual block (Ω_i) is defined as follows:

$$\Phi_i = \Omega_i\left(\Omega_{i-1}\left(\ldots \Omega_1\left(\Psi_1\right)\ldots\right)\right) \tag{7}$$

where $\Omega_i = \Omega_{i-1} + B_i$. B_i is the ith block containing C (Convolutional + ReLU) layers followed by a 1×1 convolutional layer. The output of B_i is defined as follows:

$$B_i = Fu_i(FC_{i,0} \otimes FC_{i,1} \cdots \otimes FC_{i,N}) \tag{8}$$

where Fu_i is a 1×1 convolutional layer and

$$FC_{i,j} = ReLU\left(FC_{i,j-1} \otimes FC_{i,j-2} \otimes \cdots \otimes FC_{i,0}\right) \tag{9}$$

where $ReLU$ is the non-linear activation function and $FC_{i,j}$ the jth 3×3 convolutional of the ith block. The input of the first residual block is Ψ_1.

Each convolutional layer FC with input ι has bias b and weights W in such a way that:

$$FC = W \times \iota + b \tag{10}$$

4) Upsaling module: The upscaling module is inspired by ESPCN [11]. It is followed by a 1×1 convolutional layer and a 3×3 convolutional layer.

III. Experiments

A. Settings

1) Dataset: The generalization of the deep network model very much depends on the data. Our main goal is indoor surveillance, and this is why we focused only on indoor thermal dataset such as one reported in [34]. The resolution of images is 512×512 (cropped from 1024×640). The dataset [34] is composed of various scenes, cameras views and sequences. Two indoor situations are considered: atrium-test (Atrium), lab1-test-seq1 (Lab1).

We used various views and sequence from Lab1 to construct our training dataset. Thus, we end up having 894 images. For testing, we use 30 images from multiple perspectives and sequences of Atrium.

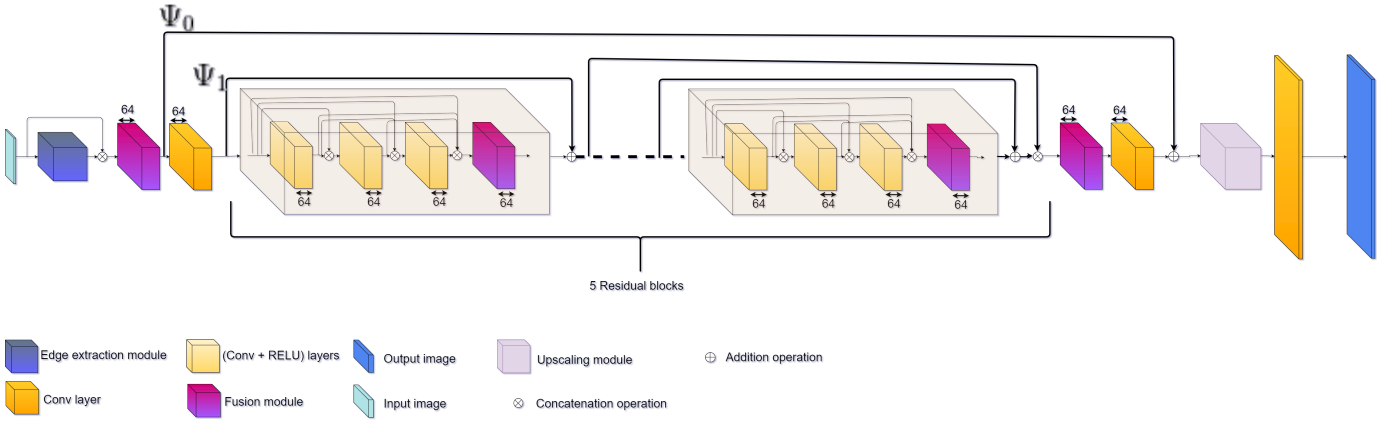

Figure 1: Edge Focused Thermal Super-resolution (EFTS)

2) Degradation model: For each high resolution image, we randomly select a blur kernel standard deviation $\eta \in \{\eta_{min}, \cdots, \eta_{max}\}$ and Gaussian noise variance $\sigma^2 \in \{\sigma^2_{min}, \cdots, \sigma^2_{max}\}$. To implement the degradation model, we down-sampled a high resolution blurred image with an additive noise.

3) Training settings: We follow the settings of [32], so we have extracted 32×32 images from each low-resolution degraded image. We used a stride of 16. We also proceed with data augmentation by randomly flipping and rotating the images. The batch size is set to 16, and we train the network for 100 epochs. For each epoch, we have 680 iterations. Our model EFTS is implemented on top of Tensorflow, and the initial learning rate is set to $1e-4$. To update the weights we used Adam optimizer.

4) Comparison methodology: Given that, we knew the ground truth (the high-resolution image) we evaluated the impact of our model by three metrics: the Peak Signal to Noise Ratio (PSNR), the Structural Similarity Index (SSIM) and the Edge Preservation Index (EPI) [35].

We took 30 images from Atrium, and for each image we generated 456 low resolutions images with blur $\eta \in [0.2, 4]$ and Gaussian noise $\sigma^2 \in [5, 50]$. We had so 13680 degraded low-resolution images. For each methods we computed the average of the metrics over these 13680 images.

Even if the SSIM/EPI metrics are supposed to be related to the perceived quality, we also made some qualitative visual comparisons between the highly resolved image and the reconstructed one. To better highlight the impact of the reconstruction methods to the edge preservation, we create an Edges Map of each image using the Sobel operator. This allowed us to make some qualitative visual inspection of the edge preservation or degradation.

B. Depth of the network

With a deeper non-linear mapping model we should normally be able to obtain better results. In [16], the authors show that using big values of D (number of residual blocks) and C (number of convolutional layers per residual block) the performance of the network is better. In their implementation, they use $D = 16$ and $C = 8$. However, for real-time execution purpose, we tried smaller values of D and C. We have tried the following combinations $D3C1$, $D5C3$ and $D7C5$, but grid search also could be performed.

Table I illustrates the performance of each one of the network depths. It is noticeable that $D5C3$ outperforms both $D3C1$ and $D7C5$. When the number of layers increases, the number of network parameters also increases. Given very similar training data, a highly complex function fits the training data better than a less complex one. Such complex function will memorize the training data, over-fitting, and the model performs poorly on the unseen data resulting in high generalization error. So, these results are overall due to the type of our dataset. Our focus is the indoor scene super-resolution of people. Such scenes are limited to human shapes and contain less information than their visible counterparts.

	D3C1	D5C3	D7C5
PSNR/SSIM	39.07/0.9549	**39.37/0.9588**	38.99/0.9573

Table I: Average PSNR and SSIM of 3 combinations of D (number of residual blocks) and C (number of convolutional layers). The best two results are highlighted in bold and underlined, respectively.

C. Edge extraction module

We have investigated different types of combinations of edge operators to see which one is more suitable for super-resolution. We compare SKL (Sobel, Kirsch, Laplace), SKP (Sobel, Kirsch, Prewitt) and KPL (Kirsch, Prewitt, Laplace). We use the same type of experiment as in section III-B.

Table II illustrates that the model SKL outperform SKP and KPL. KPL gives second best SSIM while regarding PSNR, SKP gives second best results.

The fact that SKL gives better results than SKP can be explained by the fact that Prewitt operator is derived from Sobel. So Prewitt operator does not bring more information to the network than Sobel operator. In SKP model, Sobel and Prewitt's operators are bringing almost the same kind of information about the edges.

KPL is very close to SKL in therms of SSIM, but the difference is more noticeable regarding PSNR. The main difference between these models is that Prewitt replaces Sobel. In [36], the authors reported that although Prewitt is similar to Sobel, there are differences in their spectral responses. As shown in table II, our results demonstrate that noise suppression characteristics are better with Sobel than with Prewitt.

For all these reasons, we used SKL (Fig. 2). As illustrated by this figure, we first extracted edges using the three operators. For Sobel and Kirsh operators, we have computed the edge magnitudes. The output of the edge extraction module is the concatenation of the results of the three operator.

For our network, we use the model designed in Fig. 1. We use five residual blocks with 3 (convolutional+ RELU) layers in each. For all convolutional layers, the kernel size is 3×3 except fusion layers which kernel size is set to 1×1.

	SKL	SKP	KPL
PSNR/SSIM	**39.39/0.9588**	39.37/0.9576	39.21/0.9586

Table II: Average PSNR and SSIM of 3 combinations of edge operators (S Sobel, K Kirsch, L Laplacian and P Prewitt). The best two results are highlighted in bold and underlined, respectively.

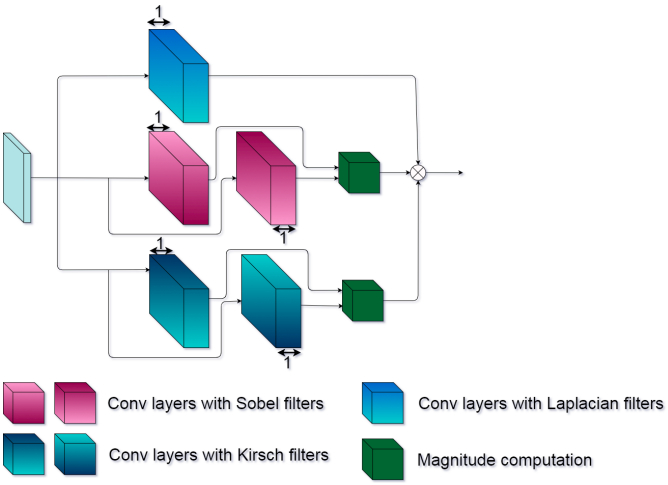

Figure 2: Edge extraction module

D. Comparison with state-of-the-art methods

We evaluated our proposed EFTS method against other existing state-of-the-art techniques in the literature. First, we compared our method with the methods developed for thermal images (TEN [17], CNN [19]) and as well as those which are developed for visible images (VDSR [7], LapSRN [13], RDN [16]). We use the same training dataset for all these models and the parameters they used in their respective papers for a fair comparison.

Most of these models source codes were available online except TEN [17] and CNN [19]. For these models, we have implemented their codes based on what is reported in [17, 19]. These re-implementations achieved expected Super-resolution results similar as those reported in the original articles. For RDN, we used the same number of blocks as for ours, that is to say, $D = 5$ and $C = 3$. All comparisons are made for a scale of $\times 4$.

Table III shows quantitative comparisons with methods TEN and CNN (for thermal images) and VDRS, LapSRN and RDN (for visible images). As evaluation metrics, we use PSNR and SSIM for images degraded by different values of noise and blur kernel. Among thermal image-based methods, CNN gets the closer results to EFTS while TEN is diverging. The performance of RDN is very close to EFTS. Such results can be explained by the fact that EFTS uses residual blocks like RDN, but its edge extraction module performs better in comparison. TEN is the shallower network with only 4 layers tends to provide less reconstruction quality than Bicubic interpolation for the values $(1, \sqrt{5})$ and $(2, \sqrt{5})$. Overall, EFTS is performing better than CNN with a noticeable amount of dB PSNR in most of the cases. TEN is diverging, giving sometimes worst results compared to bi-cubic interpolation.

While PSNR and SSIM are essential for comparison between the original and reconstructed images, the PSNR/SSIM between the Edges Map of these two images also bring more information in the quality of the reconstruction. Many thermal image applications are based on edge extraction. Therefore, we evaluated the performance of our method by computing the PSNR/SSIM of the Edges Maps. As shown in table III and table V the performance of EFTS is also better than RDN. EFTS and RDN performance are more stable to degradation variations than TEN and CNN. In these methods PSNR decreases almost for 2 dB from $[1, \sqrt{5}]$ to $[3, \sqrt{35}]$. The Edge Preservation Index (EPI) calculates the number of edges preserved in an image after applying each method to the original low-resolution image. Table IV confirms the results reported in table III and table V. It is noticeable that EFTS and RDN results are very close, but EFTS still outperforms TEN and CNN.

Figure 3. shows qualitative comparisons between EFTS, RDN, CNN and TEN. It is noticeable, in the reconstructed images and their Edges Maps, that EFTS is more able to enhance edges than the other methods. The edges extracted by RDN are comparable to edges extracted by EFTS, but we can see that our model responds equally to edges and does not enhance certain parts of edges while weakening other parts. some artifacts can be seen, while our reconstructed images contain no artifacts. TEN and CNN results proved that these methods are not suitable for edge-based thermal applications with this degradation model.

E. Application on a real low-resolution camera

The primary goal of our model is to apply super-resolution in real-world applications with very low resolution thermal images. This is why we have acquired indoor images using Lepton 2. The resolution of such images is 80×60. Fig 4. shows the qualitative results super-resolution of such images. The originals images are very noisy and practically unusable. By

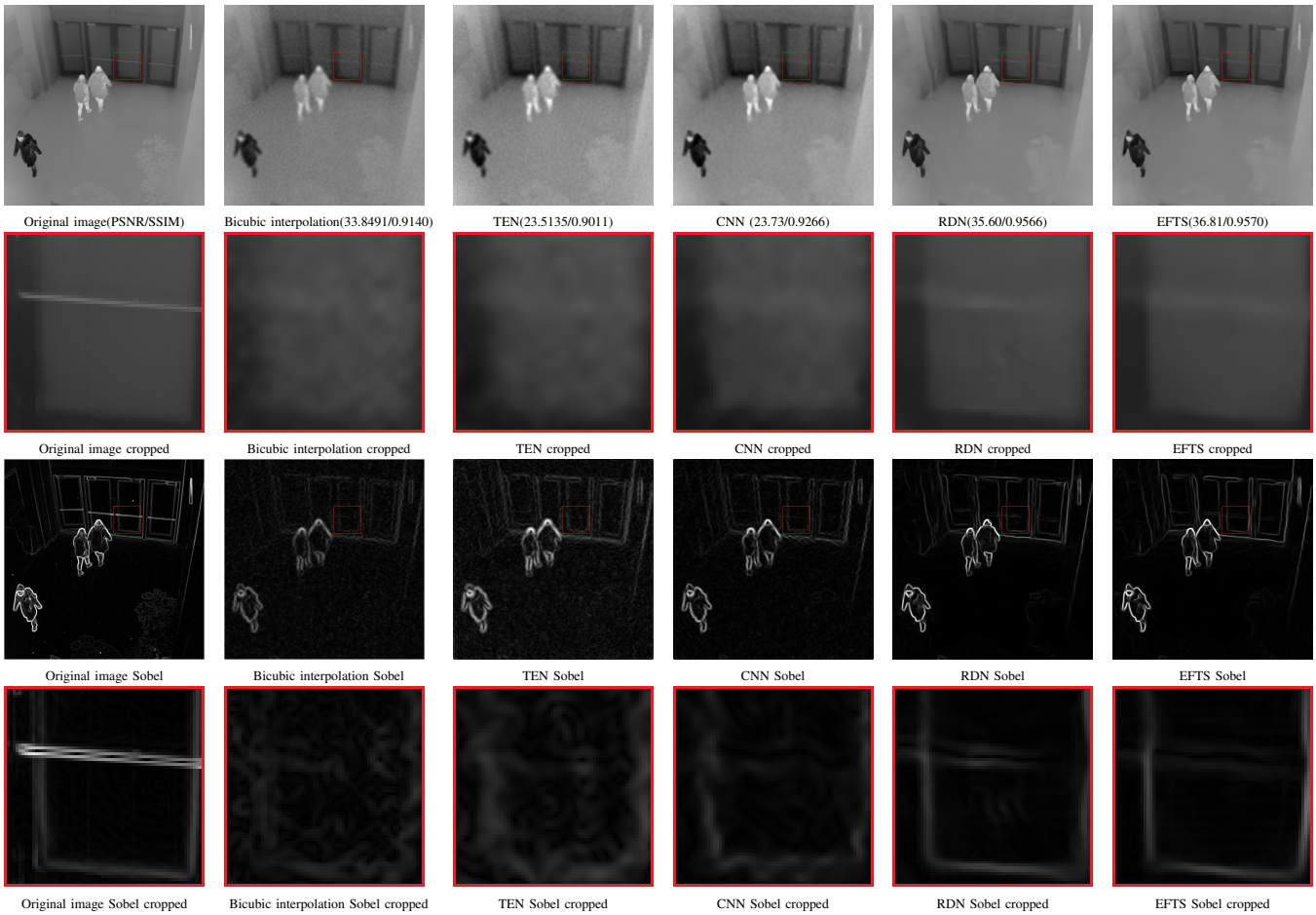

Figure 3: SISR using a blur kernel of 3 and a noise of 50

Degradation		Methods						
η	σ^2	Bicubic	VDSR [7]	LapSRN [13]	RDN [16]	TEN [17]	CNN [19]	EFTS
1	5	39.27/0.9499	39.99/0.9670	39.51/0.9601	41.22/0.9703	37.42/0.9602	40.30/0.9659	**41.65/0.9718**
	35	35.50/0.8447	38.58/0.9503	38.45/0.9398	40.34/0.9643	36.56/0.9293	39.19/0.9498	**40.71/0.9655**
2	5	38.41/0.9339	39.49/ 0.9667	38.77/0.9526	39.71/ 0.9661	37.90/0.9486	40.00/ 0.9551	**41.19/0.9677**
	35	34.90/0.8149	38.08/0.9484	37.86/0.9301	39.59/0.9573	36.94/0.9143	38.98/0.9278	**40.28/0.9592**
3	5	36.28/0.9145	38.06 /0.9540	36.90/0.9398	38.80/ 0.9597	37.27/0.9337	37.87/0.9397	**39.20/0.9614**
	35	33.80/0.7847	36.76/0.9436	36.29/0.9144	38.17/0.9506	36.40/0.8968	37.23/0.9233	**38.68/0.9518**

Table III: Comparison of EFTS vs state-of-the-art methods in terms of PSNR/SSIM. The best two results are highlighted in bold and underlined, respectively.

performing super-resolution, we want to get more information about the scene.

Fig 4. points out super-resolution applied without ground truth. We used the same trained network as in earlier sections. It is easily noticeable that EFTS allows reconstructing more details than CNN and TEN. The output images of these later methods contain some artifacts. CNN provides better results than TEN and the reconstructed image is less blurred.

Training our model with several degradation model settings allowed our network to generalize better. So such a network can be used for very low-resolution images (80×60).

IV. Conclusion

We proposed a network to perform thermal image super-resolution to handle several kinds of degradation via a single model. Unlike previous thermal image super-resolution methods, we use residual blocks and above all an edge extraction that allows us to obtain stronger reconstructed edges. Moreover, we not only evaluated the performance of our proposed model based on PSNR/SSIM but also assessed the PSNR/SSIM of Edges Maps and Edge Preservation Index. All these evaluations metrics confirm that the edge extraction module improves the

Degradation		Methods						
η	σ^2	Bicubic	VDSR [7]	LapSRN [13]	RDN [16]	TEN [17]	CNN [19]	EFTS
1	5	0.9369	0.9598	0.9435	0.9619	0.9546	0.9562	**0.9620**
	35	0.9350	0.9591	0.9347	0.9610	0.9536	0.9555	**0.9615**
2	5	0.9505	0.9593	0.9464	0.9606	0.9532	0.9548	**0.9609**
	35	0.9453	0.9582	0.9377	0.9592	0.9522	0.9537	**0.9594**
3	5	0.9509	0.9579	0.9479	0.9578	0.9519	0.9528	**0.9585**
	35	0.9471	0.9561	0.9393	0.9566	0.9509	0.9521	**0.9567**

Table IV: Comparison of EFTS vs state-of-the-art methods in terms of EPI. The best two results are highlighted in bold and underlined, respectively.

Degradation		Methods						
η	σ^2	Bicubic	VDSR [7]	LapSRN [13]	RDN [16]	TEN [17]	CNN[19]	EFTS
1	5	24.99/0.8234	26.48/0.8196	25.54/0.7554	27.42/0.8315	25.61/0.7432	26.54/0.8040	**27.64/0.8375**
	35	23.92/0.6329	26.06/0.7449	25.40/0.6866	27.04/0.8123	25.22/0.5969	26.31/0.7523	**27.27/0.8171**
2	5	24.95/0.6335	26.17/0.8040	24.87/0.6956	26.98/0.8189	24.70/0.6695	25.52/0.7360	**27.13/0.8219**
	35	23.88/0.4386	25.73/0.7275	24.79/0.6192	26.35/0.7945	24.42/0.5171	25.38/0.6806	**26.49/0.7959**
3	5	24.10/0.4733	25.68/0.7876	24.16/0.6318	26.09/0.7992	23.91/0.6069	24.40/0.6676	**26.39/0.8010**
	35	23.35/0.2580	25.35/0.7352	24.10/0.5482	25.51/0.7648	23.70/0.4545	24.32/0.6133	**26.10/0.7767**

Table V: Comparison of EFTS vs state-of-the-art methods in terms of PSNR/SSIM of the edge maps. The best two results are highlighted in bold and underlined, respectively.

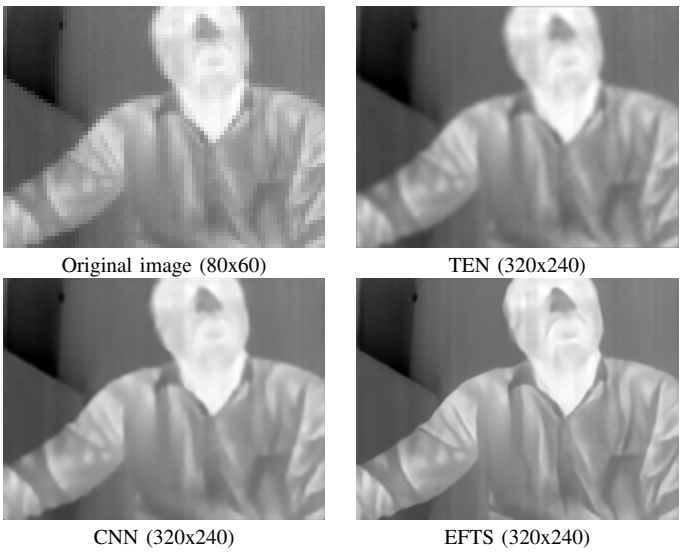

Figure 4: Super-resolution of low-resolution thermal image of a person sit in front of the camera

results.

The edge extraction module is composed of three edge extractors (Sobel, Prewitt, Laplacian) that are concatenated with the original low-resolution image and is fused to extract shallow features. The results on real very low-resolution images acquired from Lepton2 show that we can significantly enhance the resolution of such type of images.

Here, our degradation models included isotropic blurring and Gaussian noise; however, we should consider that thermal images are also affected by other degradation models such as motion blur. To increase our network generalization and real-world performance, we must take into account such degradation. Moreover, in the context of indoor surveillance, it is possible to associate two thermal sensors together or a thermal sensor with another type of sensor. It could be possible to use disparity to enhance thermal image resolution even further.

Acknowledgments

This work was supported by Mitacs and Campus France through the Globalink Research Internship program (grant no. IT10912).

References

[1] D. Gottlieb and C.-W. Shu, "On the Gibbs phenomenon and its resolution," *SIAM review*, vol. 39, no. 4, pp. 644–668, 1997.

[2] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[3] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *JOSA A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[4] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[5] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[6] 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*, vol. 38, pp. 295–307, 2 2016.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 295–307, 11 2015.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[9] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1637–1645, 2016.

[10] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[11] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883, 2016.

[12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*, pp. 391–407, Springer, 2016.

[13] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5835–5843, 10 2017.

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network.," in *CVPR*, vol. 2, p. 4, 2017.

[15] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4809–4817, 2017.

[16] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[17] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon, "Thermal image enhancement using convolutional neural network," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 223–230, 10 2016.

[18] T. Y. Han, Y. J. Kim, and B. C. Song, "Convolutional neural network-based infrared image super resolution under low light environment," in *Signal Processing Conference (EUSIPCO), 2017 25th European*, pp. 803–807, IEEE, 2017.

[19] P. Bhattacharya, J. Riechen, and U. Zölzer, "Infrared image enhancement in maritime environment with convolutional neural networks.," in *VISIGRAPP (4: VISAPP)*, pp. 37–46, 2018.

[20] Z. He, S. Tang, J. Yang, Y. Cao, M. Y. 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*, pp. 1–1, 2018.

[21] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 6, 2018.

[22] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Machine Vision and Applications*, vol. 25, pp. 1423–1468, 8 2014.

[23] W. Dong, L. Zhang, G. Shi, and X. Li, "Non locally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, pp. 1620–1630, 4 2013.

[24] G. Riegler, S. Schulter, M. Ruther, and H. Bischof, "Conditioned regression models for non-blind single image super-resolution," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 522–530, IEEE, 12 2015.

[25] G. Boracchi and A. Foi, "Modeling the performance of image restoration from motion blur," *IEEE Transactions on Image Processing*, vol. 21, pp. 3502–3517, 8 2012.

[26] M. Sharifi, M. Fathy, and M. T. Mahmoudi, "A classified and comparative study of edge detection algorithms," in *Proceedings - International Conference on Information Technology: Coding and Computing, ITCC 2002*, pp. 117–120, 2002.

[27] W. Gao, X. Zhang, L. Yang, and H. Liu, "An improved sobel edge detection," in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, vol. 5, pp. 67–71, IEEE, 2010.

[28] B. S. Lipkin, *Picture Processing and Psychopictorics.* Elsevier Science, 1970.

[29] G. S. Robinson, "Color edge detection," *Optical Engineering*, vol. 16, p. 165479, 10 1977.

[30] L. J. van Vliet, I. T. Young, and G. L. Beckers, "A nonlinear laplace operator as edge detector in noisy images," *Computer Vision, Graphics, and Image Processing*, vol. 45, pp. 167–195, 2 1989.

[31] K. Hajebi and J. S. Zelek, "Structure from infrared stereo images," *2008 Canadian Conference on Computer and Robot Vision*, pp. 105–112, 5 2008.

[32] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*, vol. 1, p. 4, 2017.

[33] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4539–4547, 2017.

[34] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 201–208, IEEE, 6 2014.

[35] F. Sattar, L. Floreby, G. Salomonsson, and B. Lovstrom, "Image enhancement based on a nonlinear multiscale method," *IEEE Transactions on Image Processing*, vol. 6, pp. 888–895, 6 1997.

[36] D. Adlakha, D. Adlakha, and R. Tanwar, "Analytical comparison between Sobel and Prewitt edge detection techniques," *International Journal of Scientific & Engineering Research*, 2016.