Channel Split Convolutional Neural Network (ChaSNet) for Thermal Image Super-Resolution

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

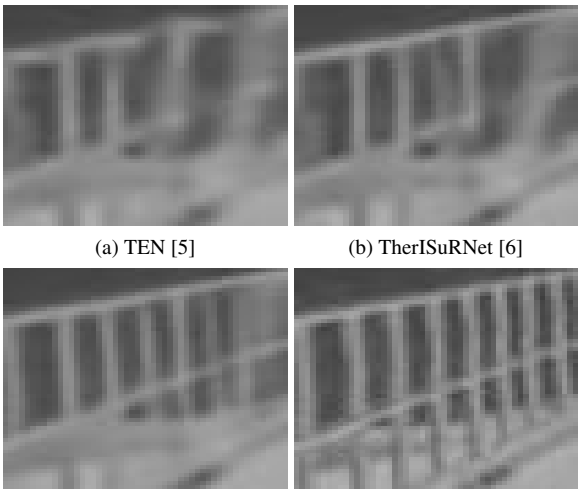
Abstract

*The ability of thermal sensors to penetrate smoke, mist, dust and aerosol makes them attractive for deployment in essential applications in military, medical, agricultural and animal ecology over the regular optical cameras. However, unlike optical imaging devices, the most sophisticated commercial thermal imaging sensor does not match the megapixel imaging ability. The Low-Resolution (LR) images from thermal sensors can be enhanced through a software-driven solution called Super-Resolution (SR). A number of works have been proposed to employ deep networks for SR task; however, they are overloaded with redundant features due to the deep architecture. This paper introduces a Channel Splitting-based Convolutional Neural Network (*ChasNet*) for thermal image SR eliminating the redundant features in the network. The use of channel splitting extracts the versatile features from Low-Resolution (LR) thermal image, helping to preserve high-frequency details in the SR images. We demonstrate the applicability proposed network for SR task in two different scenarios organized in the PBVS-2021 Thermal SR Challenge, consisting of noise elimination (Track-1) and domain shifting (Track-2). The efficacy is justified by comparing the SR results with other state-of-the-art thermal SR techniques in qualitative and quantitative metrics. A set of extensive experiments separately analyzes the importance of each block in the proposed architecture. The code of this work is also published online*[1].

1. Introduction

The optical camera senses the reflected light from the scene being observed within the visible range of the electromagnetic (EM) spectrum. Despite the flexibility of obtaining high-resolution images from optical sensors, their deployment in night-time and conditions under the presence of smoke, mist, dust, etc., results in demotivating performance.

[1]https://github.com/kalpeshjp89/ChasNet



(a) TEN [5] (b) TherISuRNet [6]
(c) Proposed (d) HR

Figure 1: The visual assessment of the thermal SR images obtained using the different CNN based methods for upscaling factor $\times 4$.

In such conditions, thermal cameras can be employed which sense the thermal radiation emitted by the objects in the range of Long Wavelength Infrared (LWIR) spectrum. Such passive sensors measure the infrared wavelength in terms of temperature above absolute zero [11] and has found applications in enormous use in military [12], agriculture [37], maritime surveillance [16], medical [29], industry [1, 28], urban development [4, 9] amongst others. However, thermal sensors are limited to a coarser spatial resolution as compared to visible sensors due to the lower span of the LWIR range. Thus, the most sophisticated commercial thermal imaging device does not match the megapixels provided by optical cameras [11]. An alternative solution in such a case is to use software-driven algorithms to enhance the spatial resolution of the scene being captured by the thermal sensors, which is usually referred to as *Super-Resolution* in the computer

vision community [38, 19, 10, 17].

Although a number of approaches are proposed for SR of images, the recent technological advancements in computing hardware, memory devices, availability of large data for various problems and latest innovations related to deep learning networks has led to the use of Convolutional Neural Network (CNN) for SR. Dong et al. [7] first introduced CNN based SR solution (referred to as SRCNN) using a neural network of three convolutional layers. Such a shallow network beats the performance of traditional methods and shows its effectiveness for SR over classical methods. To further improve the quality of SR images, abundant CNN based SR techniques have been published in recent time [8, 41, 22, 27, 39, 40]. However, these works are limited to the RGB data, where three different channels are provided with reasonable spatial resolution. Motivated from the success of SR for visible images, SR approaches are also explored for thermal images in the recent works [5, 3, 23, 25, 16].

On the other hand, most deep networks extract highly redundant features due to the deeper nature of networks due to multiple layers extracting similar features. We propose a novel Channel Split Convolutional Neural Network (ChaSNet) for the thermal image SR task to address this limitation and simultaneously provide better SR results. The proposed approach employs the concept of channel splitting at different stages of the network, helping to extract versatile features in the given thermal image. The prime element in the proposed framework is the structure of Common Block (CB) connected in a residual manner using Residual Blocks (ResBlocks) to extract the relevant details from LR observation. The potential of the proposed method has been verified in two different SR scenarios released in PBVS-2021 Thermal SR Challenge (i.e., Track-1 noise elimination and Track-2 domain shifting) and it shows committed improvement as discussed in the experimental section later in the manuscript. A sample out from the proposed network is depicted in Fig. 1 where the SR results obtained using the recent techniques such as TEN [5] and TherISuRNet [6] (i.e., best paper and extension of winner architecture in PBVS-2020 Thermal SR Challenge [33]) are displayed alongside the proposed method. One can note the superior quality of the SR image obtained using the proposed method over the other networks by closely looking into Fig. 1. The key contributions of this work are:

- Proposes a novel CNN based architecture referred hereafter as *ChaSNet* for thermal image super-resolution. The work introduces a novel concept of channel splitting in the proposed framework. Such channel splitting is helpful to gain considerable improvement in the quality of SR results while reducing the redundant features.
- Additionally, this work proposes to use Channel Attention (CA) scheme with average and standard deviation pooling to find statistical information from the features.
- To tackle the problem of domain shifting where the exact paired data is unavailable for training, we propose a fusion approach to improve the SR performance significantly.
- Further, the SR performance of the proposed network is validated in the two different scenarios, for instance, noise elimination in SR results (Track-1) and domain shifting case (Track-2). The performance gain is illustrated through empirical validations obtained using the proposed method over the other existing state-of-the-art methods.
- Presents a detailed statistical study on the impact of various design considerations in the proposed network to exemplify the rationale of the choices made.

In the rest of the paper, the overview of relevant SR works on thermal images is presented in Section 2. The architecture and methodology of the proposed framework are described in Section 3. The experiment analysis, including training details, description of the dataset, ablation study, quantitative and qualitative analyses, are discussed in Section 4. The concluding remarks of the proposed work and a set of future works are presented in Section 5.

2. Related Works

The successful attempts of deep models to SR for visible images motivated many researchers to use it for thermal images. Inspired by SRCNN [7] (i.e., the first CNN based SR approach), Choi et al. [5] proposed to employ it for enhancement of thermal images and referred to it as Thermal Enhancement Network (TEN). Further, Marivani et al. [25] have proposed a CNN model trained on Near-Infrared (NIR) images along with RGB images to provide auxiliary information. However, the training of the model on RGB images limits its performance on thermal images, as demonstrated by Rivadeneira et al. [34] in which they trained their model explicitly on thermal images. Additionally, the idea of the sequential connection of two networks for thermal SR was presented by Bhatt et al.[3] where the first network was used for denoising of given infrared image and another network was employed for SR task on maritime infrared images. Similarly, based on the above cascade connections of two networks, the same idea was further explored by He et al.[16] to improve the quality of SR image for upscaling factor $\times 8$. Recently, TherISuRNet [6] was also proposed for SR of thermal images using asymmetric progressive learning strategy - a modified architecture which won first place in PBVS-2020 Thermal SR Challenge [33].

Further, Generative Adversarial Network (GAN) [13] was employed for SR of thermal images after the success

of SRGAN [20] for visible images. Liu et al. [23] proposed GAN based learning to train the deep network. Similarly, DCGAN model [30] is also exploited by Guei et al. [14] for NIR and LWIR images. Additionally, to further accelerate the research in SR of thermal images, Rivadeneira et al. [31] released a novel dataset consisting of thermal images acquired at different resolution factors. The first Thermal SR Challenge [33] was organized PBVS CVPR-2020 workshop based on this dataset. Multiple SR approaches such as TherISurNet [6] (winner of this challenge), multi-level supervision network [18] has been proposed under this challenge. Additionally, Rivadeneira et al.[32] utilized the features from each layers of sequential network using concatenation in order to improve the performance.

In a similar direction for thermal image super-resolution, we propose a novel approach named as *ChaSNet* to improve the quality of SR images. We introduce the concept of channel splitting in the proposed framework to enhance the performance of SR images. The details of the proposed approach are presented in the section below.

3. Proposed Method

A deeper neural network with a number of CNN layers with many feature maps in each CNN layer for SR occupies a larger space due to redundant information load. Such behaviour of deep models can be optimized to improve the performance. In this direction, we propose to employ channel splitting in the proposed network that distinguishes the features in the channel dimension. Different operations can be performed on the reduced feature maps leading to low redundancy and improving the SR performance. We propose two CNN models designed specifically for two scenarios of SR tasks as organised in PBVS 2021 Thermal SR challenge (i.e., Track-1 & Track-2)[2] based on the concept of channel splitting. In Track-1 of the challenge, it is required to perform super-resolution on LR thermal images with noise elimination for upscaling factor $\times 4$, while the Track-2 is designed for domain transfer including super-resolution with scale factor $\times 2$.

The architecture of the proposed approach for Track-1 is depicted in Fig. 2(a), which can be categorized into the following three modules based on their functionality:

- Shallow Feature Extraction (SFE) module,
- Deep Feature Extraction (DFE) module and
- Image Reconstruction (IRec) module

Initially, the input Low-Resolution (LR) image is passed through the SFE module, which is composed of two convolutional layers with a kernel size of 3×3 & C feature-maps and 5×5 with $2C$ feature-maps, respectively. The first convolutional layer with a smaller receptive field (i.e., with a kernel size of 3×3) extracts minutes details that correspond to smaller objects from the input LR image while the convolutional layer with larger kernel size (i.e. 5×5) learns the local information related to larger objects present in the thermal LR observation. Mathematically, the output of SFE module, i.e., X_{SFE} can be represented as,

$$X_{SFE} = \mathcal{F}_{SFE}(I_{LR}), \tag{1}$$

where, $\mathcal{F}_{SFE}$ represents the functionality of SFE module and I_{LR} denotes input thermal LR image.

The feature maps available from the SFE module (i.e., X_{SFE}) are separated into two streams (i.e., X^{in} and X^{skip}) each of size C channels first in the DFE module. The X^{in} feature maps are then passed through the sequence of Common Blocks (CBs) to extract rich features while the rest X^{skip} feature maps are bypassed and directly concatenated with the output of K number of CBs. Each CB unit takes C channel features and outputs $2C$ number of channels. The design of CB is depicted in Fig. 2(c), which is made up of three densely connected Channel Splitting Blocks (CSBs) followed by one convolutional layer of kernel 1×1. The architecture design of the CSB is displayed in Fig. 2(c), which consists of 1×1 convolutional layer followed by a channel splitting operation. Such channel splitting in CSB is required to perform two different operations correspond to densely connected convolutions and pass through a sequence of two Residual Blocks (ResBlocks). Finally, the output of both different paths is combined using concatenation at the end of the CSB unit. In Fig. 2(c)), we show the design of ResBlock composed of two convolutional layers followed by Pixel Attention (PA) and Channel Attention (CA) modules. Additionally, we also deploy local residual learning in the design of ResBlock to avoid the problem of vanishing gradient problem [15].

The mathematical formulation of each CB can be given as,

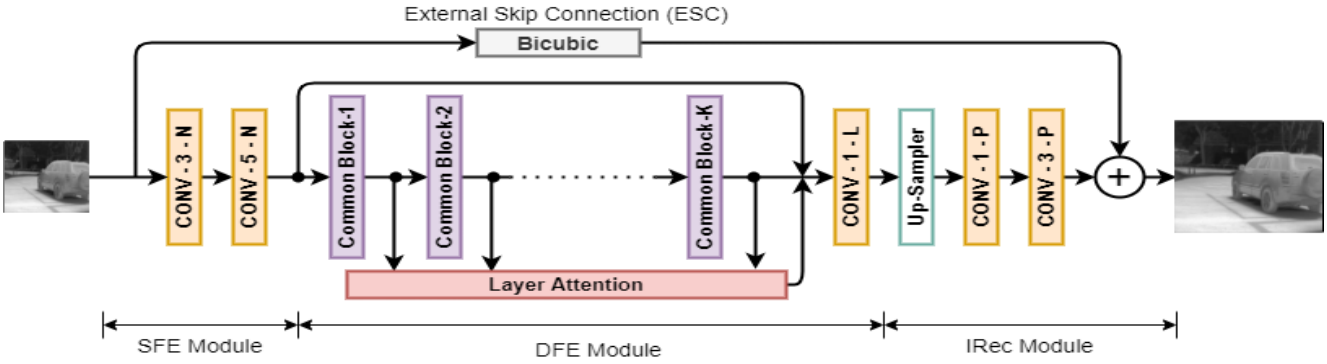
$$X^{skip}_{CB_i}, X^{in}_{CB_i} = \mathcal{S}(\mathcal{F}_{CB_i}(X^{in}_{CB_{i-1}})), i = 1, 2, \dots K. \tag{2}$$

Here, $\mathcal{F}_{CB_i}$ denotes the functionality of i^{th} CB and $\mathcal{S}$ represents channel splitting. From all such CB$_i$ (i.e., $i = 1, 2, ...K$), the feature $X^{in}_{CB_i}$ passed to next CB while other feature $X^{skip}_{CB_i}$ is bypassed through Layer Attention (LA) module. The operations performed by LA module can be expressed mathematically as,

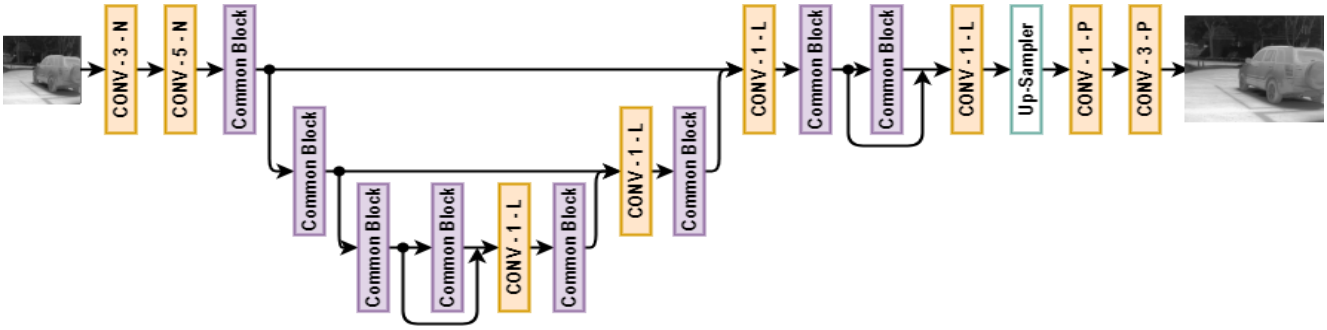
$$X_{LA} = \mathcal{F}_c(\mathcal{M}(X^{skip}_{CB_i}, \mathcal{F}_c(\dots, \mathcal{F}_c(\mathcal{M}(X^{skip}_{CB_2}, X^{skip}_{CB_1}))\dots))), \tag{3}$$

Where $\mathcal{M}(.,.)$ denotes the merging operation (i.e., concatenation) of the features along with channel dimension, after merging operation, one convolution layer with a kernel of 1×1 (i.e., noted by $\mathcal{F}_c(.)$) is applied which acts as a transition layer. Finally, the output feature maps of last CB (i.e.,

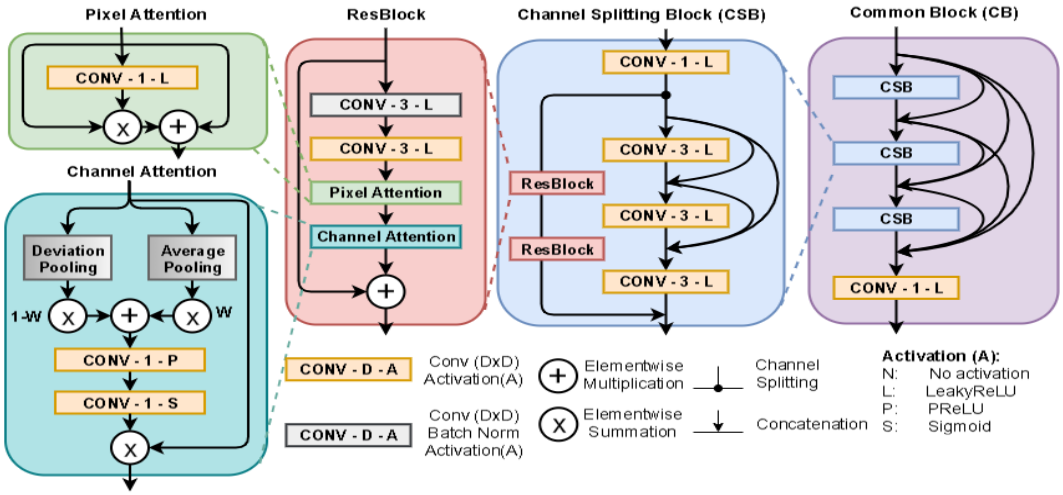
[2] https://pbvs-workshop.github.io/



(a) The proposed CNN framework for up-scaling $\times 4$ in Track-1 and for up-scaling $\times 2$ in Track-2 of PBVS-2021 Thermal SR Challenge.



(b) The proposed Unet based CNN model for up-scaling $\times 2$ in Track-2.



(c) The architecture design of Common Block (CB) of the proposed network.

Figure 2: The architecture design of the proposed ChaSNet for thermal image super-resolution.

$X^{in}_{CB_k}$), LA module (X_{LA}) and X^{skip} are concatenated together and passed through 1×1 convolutional layer in order to generate output feature maps of DFE module (i.e., X_{DFE}). This can be formulated mathematically as,

$$X_{DFE} = \mathcal{F}_c(\mathcal{M}(X^{skip}, X^{in}_{CB_k}, X_{LA})). \quad (4)$$

Further, The output feature maps from the DFE module are forwarded to the IRec module, which upscale the feature maps to the desired factor and then reconstruct the gray-scale thermal SR image. Here, we employ the pixel-shuffle based up-sampler to increase the spatial resolution of the feature maps [36]. Mathematically, the function of the up-sampler block can be expressed as,

$$X_{up} = \mathcal{F}_{PS}(\mathcal{F}_c(X_{DFE})), \quad (5)$$

Where $\mathcal{F}_{PS}$ denotes the pixel-shuffler operation for the factor of $\times 4$ and $\times 2$ for Track-1 and Track-2, respectively. Finally, the output feature maps of the up-sampler block are passed through two additional convolution layers in order to reconstruct the residual SR image. Further, we adopt the

Global Residual Learning (GRL) approach where the input LR image (i.e., I_{LR}) is upsampled via bicubic interpolation and then it is added with the residual output image obtained from the IRec module to generate the final thermal SR image. Such a GRL approach helps to stabilize the training of the network, and it also preserves the identity of the input image in the SR output. Mathematically, the final SR image can be formulated as,

$$I_{SR} = \mathcal{F}_{c_3}\big(\mathcal{F}_{c_1}(X_{up})\big) + B(I_{LR}), \tag{6}$$

where $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_3}$ denotes as convolution with kernel having the size of 1×1 and 3×3, respectively while B indicates the up-sampling operation using bicubic interpolation.

To tackle the track-2 scenario of PBVS-2021 Thermal SR Challenge where domain transfer is needed along with SR for upscaling factor 2, another framework is designed (i.e., depicted in Fig. 2(b)) which is inspired from UNet model [35]. The proposed network design consists of several CBs and many convolution layers connected in a UNet manner. By inspecting the LR and HR data of Track-2, we observe the uncertainty between the same, motivating us to employ the UNet type structure in the proposed method where most of the domain transfer tasks are performed at the bottom layers.

3.1. Loss Functions

For Track-1 and Track-2 cases of PBVS-2021 challenge[3], the network shown in Fig. 2(a) is trained using pixel-wise $\mathcal{L}_1$ loss between SR and HR images. However, due to inaccurate results of image registration between LR-HR pair images in Track-2 challenge dataset, we have employed semi-supervised paired data for domain shift. To deal with such dataset in Track-2, an adversarial learning is used to train the network shown in Fig. 2(b). In this setting, the generator network is trained using combination of pixel loss ($\mathcal{L}_1$), GAN loss ($\mathcal{L}_{GAN}$) and SSIM loss ($\mathcal{L}_{SSIM}$) which can be expressed as,

$$\mathcal{L}_G = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}^G_{GAN} + \lambda_3\mathcal{L}_{SSIM}. \tag{7}$$

Here, λ_1, λ_2 and λ_3 are the weighting constants set empirically in the training process. For adversarial learning, we use the Least Squared GAN (LSGAN) framework [24]. The GAN loss for generator network (i.e., $\mathcal{L}^G_{GAN}$) can be formulated as,

$$\mathcal{L}^G_{GAN} = \frac{1}{N}\sum\big(1 - D(I_{SR})\big)^2, \tag{8}$$

Here, N represents number of images in a batch and $D(\cdot)$ denotes the function of discriminator network which is trained using LSGAN framework with following loss function,

$$\mathcal{L}^D_{GAN} = \frac{1}{N}\sum\big((1 - D(I_{HR}))^2 + (D(I_{SR}))^2\big). \tag{9}$$

[3]https://pbvs-workshop.github.io/

4. Experimental Analysis

The details of the training setup of the proposed network followed by experimental analysis are elaborated in this section. The proposed networks are trained with Adam optimizer on the default β values and with the learning rate of 2×10^{-4} which is decayed by half at every 25% of a total number of iterations. We have trained the proposed modules up to 2×10^5 number of iterations with a batch size of 8. In training, the 192×192 patch from HR images and its corresponding patch from LR images are extracted randomly and then augmented using random horizontal flipping and rotation with 0^o and 90^o operations. The weighting constants λ_1, λ_2 and λ_3 associated to Equation (7) are set to 0.01, 0.1 and 0.5, respectively. Additionally, we also published the code of this work, and the same is available from https://github.com/kalpeshjp89/ChasNet.

4.1. Training and Testing datasets

The proposed network is trained on a thermal image dataset made available by Rivadeneira et al. [34]. However, in this work, we only used Flir (HR) and Axis (MR) datasets for the different tracks as organized in the PBVS-2021 challenge. These datasets contain 951, 50 and 20 (10 images for Track-1 and another 10 images for Track-2 individually) images for training, validation and testing purpose, respectively. In Track-1, the original Flir dataset is utilized as an HR dataset and corresponding LR data is created by bicubic down-sampling with a factor $\times4$ and degrading the same by Additive White Gaussian Noise (AWGN) with mean and standard deviation values of 0 and 10, respectively based on empirical trials.

For Track-2, the target (i.e., Axis (MR) dataset) is required with an upscaling factor of $\times2$ and match the same with the resolution of the Flir (HR) dataset. It is worth noting that the characteristics of both datasets are different, and there is a need to register them to perform the SR task. However, the available images in the dataset are not exactly registered with pixel-wise accuracy, which restricts their use for the training of the SR network. Hence, for Track-2, we perform image registration using SURF features [2] which results in semi-matched image pairs and later, those images are used for training the proposed methods. Further, the SR task in Track-2 is performed using two networks depicted in Fig. 2(a,b). The network presented in Fig. 2(b) is trained using the semi-paired dataset, which helps to convert the domain shift from Flir to Axis. However, due to the semi-paired nature of the registered dataset, the SR results from the network lack structure information. To improve the same, we also train the network displayed in Fig. 2(a), which uses Axis dataset as HR images and corresponding LR images are generated using bicubic downsampling.

Further, the proposed networks are tested compared with other state-of-the-art SR methods on the PBVS-2021 ther-

Table 1: The quantitative measurements for the experiments of ablation study.

Case	PSNR	SSIM
Common Block (CB) design		
w/o deviation pooling in CA	33.35	0.8973
w/o average pooling in CA	33.38	0.8974
w/o CA in ResBlock	33.36	0.8974
w/o PA in ResBlock	33.36	0.8972
w/o ResBlock in CSB	33.28	0.8960
w/o dense blocks in CSB	33.34	0.8971
w/o dense connection CB	33.36	0.8976
Network configuration		
w/o layer attention	33.37	0.8977
w/o ESC	33.08	0.8918
K:8, C:64	33.38	0.8978
K:12, C:64	33.30	0.8962
K:8, C:128 (Proposed)	**33.41**	**0.8983**

mal image super-resolution challenge dataset. The validation dataset for Track-1 and Track-2 has 50 images, while the testing dataset for both tracks consists of 10 images. For quantitative comparison, we measure the performance on the validation dataset because of the availability of the ground-truth images. The Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics are used to validate the SR performance, which is being calculated on the Y channel after removing s (i.e., scale factor) number of border pixels as suggested in [7]. Additionally, we show the qualitative performance on both validation and testing datasets to measure perceptual improvement using different thermal SR methods.

4.2. Ablation Study

The justification of different parameters settings used in the proposed methods are analyzed, and the same are discussed at length in this sub-section[4]. It includes different experiments based on the network design of various modules of Common Block (CB) and network configurations.

Design of Common Block (CB): We have conducted multiple experiments to check the sensitivity of the proposed network to the design of Common Block (CB). The SR outcomes of these experiments are compared in terms of PSNR and SSIM measures which are tabulated in Table 1. For the baseline network, we consider the proposed network having 64 channels and 8 common blocks (i.e., C=64 and K=8). Additionally, in the proposed network, we have modified the structure of the Channel Attention (CA) module by adopting deviation pooling parallel to the average pooling operation. An additional experiment is carried out in the ablation study to understand its effectiveness in the modified CA block. We train the base network with global average pooling and standard deviation pooling individually. Further, an experiment based on the training of baseline network without adopting CA module is also conducted. By looking at Table 1, one can note the importance of standard deviation pooling in the CA block. A gain of $+0.02dB$ in PSNR can be observed by adopting the CA block in the design of the CB unit. Additionally, one experiment is conducted in which the Pixel Attention (PA) block from ResBlock is removed, and its quantitative evaluation is depicted in Table 1. One can observe a gain of $+0.02dB$ in PSNR obtained by incorporating PA in the design of the CB module. To validate the importance of the Residual Block (ResBlock) in CB, the baseline network without ResBlock is trained, and the corresponding SR performance is noted in Table 1. Here, similar to the earlier experiments, we can notice the gain of $+0.1dB$ in PSNR value by using ResBlock in CB, which justifies our choice of ResBlock in the CB module. The effectiveness of densely connected convolutional layers in CSB block as displayed in the Fig. 2(c) is also examined, and a gain of $+0.04dB$ in PSNR value is obtained (see Table 1). Lastly, in the design of the common block, the effectiveness of dense connection is also analysed and presented in the same table where an additional gain of $+0.02dB$ in PSNR is achieved by adopting it in the design of the proposed method.

Network configurations: We also conduct a set of additional experiments to check the effectiveness of network configurations in the proposed network. The quantitative results from these experiments are presented in Table 1 in terms of PSNR and SSIM values. In the first experiment, we remove the Layer Attention (LA) module from the design of the proposed network (see in Fig. 2(a)) and observe that the proposed network with LA module obtains a gain of $+0.01dB$ in PSNR. Similarly, the use of External Skip Connection (ESC) in Fig. 2(a) has also been analysed and note that it helps to gain $+0.3dB$ in the PSNR value. At last, we also examine the effects of a different number of channels (C) and common blocks (K) in the proposed network; 1) K:8, C:64 2) K:12, C:64 and 3) K:8, C:128. From these three experiments, it can be noticed that the wider network (i.e., K:8, C:128) performs effectively than the deeper network (i.e., K:12, C:64) in the proposed network for thermal image super-resolution. This validates the use of a wider network in the design of the proposed network.

Additionally, we examine the effect of fusion of SR images from CNN based network (i.e. Fig. 2(a)) and GAN based network (i.e. Fig. 2(b)) in Track-2 task of PBVS-2021 challenge. In Fig. 3, we show the visual patches of SR images obtained by CNN based network, GAN based network, a fusion of SR images from both networks (i.e., Proposed), fusion with self-assemble (i.e., Proposed+) and semi-HR images. Due to the different characteristics in HR-LR pair, network fails to preserve the structure of the content in the SR images, and hence there is structure loss in SR images from the network trained using this imperfect data

[4]We train the network up to $50,000$ iterations for ablation analysis.

Within Domain Cross Domain Proposed Proposed+ Semi-HR

Figure 3: The visual comparison of SR results obtained using different method to validate the effects of fusion for Track-2 task of PBVS-2021 Thermal Image SR Challenge.

(see Cross-Domain results from the Fig. 3). However, the obtained gain by such network is partly domain shifted and hence, to improve the structural information, we employ another network that is trained using perfect HR data obtained by bicubic down-sampling. From Fig. 3, one can observe that the fusion of both networks generates better SR solutions (Fig. 3 mentioned with proposed); however, the fusion incorporating self-assembly (indicated as *Proposed+*) generates better visual SR images than that of others for the task of Track-2.

4.3. Quantitative Evaluation

The quantitative evaluation of the proposed method-*ChaSNet* has been verified by comparing it with other existing state-of-the-art methods in terms of PSNR and SSIM. For a fair comparison, we include different methods such as Bicubic, SRResNet [20], TEN [5], CNN-IR [3], SRFeat [26], RCAN [41], MSRN [21], EDSR [22], PBVS-2020 challenge winner [33] and TherISuRNet [6] methods. In Table 2, we depict the PSNR and SSIM values obtained using those methods on PBVS-2021 validation dataset for Track-1 - thermal image SR using upscaling factor $\times 4$ with noise elimination. *Proposed+* denotes the proposed method with the self-assembly approach. From this table, one can see that the proposed method obtains SR images with the highest quantitative fidelity in terms of PSNR and SSIM values over the other methods.

The above evaluation has been extended for the domain shifting case presented as Track-2 in the PBVS-2021 challenge. The proposed method incorporates the fusion of the SR results obtained by CNN and GAN-based networks. Due to the lack of true registered LR-HR pair for the Track-2 task, we demonstrate the quantitative evaluation of CNN-based networks only on LR images generated by bicubic down-sampling. Thus, the SR results of the proposed network (i.e., Fig. 2(a)) are compared with other methods in terms of PSNR and SSIM in Table 2. From this table, one can again notice the superiority of the proposed network on the synthetically generated LR thermal image for a factor of $\times 2$ over the other methods.

4.4. Qualitative Evaluation

We illustrate the qualitative improvement achieved using the proposed network over the other state-of-the-art methods in this section. Fig. 4 shows the SR results obtained on the Track-1 task where the noise elimination is required along with super-resolution for upscaling factor of $\times 4$. From these results, one can notice the preservation of high-frequency details in the SR results obtained using the proposed method compared to other state-of-the-art methods. The proposed method preserves the structural information in a better manner as compared to other methods. We can also observe the advantage of self-assembling in the thermal image super-resolution to improve the quality of SR, which is shown as *proposed+*. Further, we also illustrate the results from the proposed model (i.e., *ChaSNet*) for the Track-2 case where the task of domain shifting is intended with upscaling factor $\times 2$. The visual comparison of the SR results obtained using different methods on bicubically down-sampled Axis dataset is depicted in Fig. 5. From this, one can deduce that the proposed method performs better for domain shifting case over the other methods alongside quantitative improvement, as mentioned with each SR result in the same figure.

Table 2: The quantitative comparison of the proposed network with other existing thermal image super-resolution methods on PBVS-2021 challenge validation dataset. Highest two values are highlighted using red and blue colors respectively.

Method	**Flir** ($\times 4$)		**Axis** ($\times 2$)	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Bicubic	32.66	0.8625	34.74	0.9200
SRResNet [20]	33.12	0.9018	33.66	0.9229
MSRN [21]	34.47	0.9076	36.96	0.9471
SRFeat [26]	34.12	0.9007	-	-
EDSR [22]	34.48	0.9068	36.91	0.9466
RCAN [41]	34.42	0.9072	36.67	0.9438
TEN [5]	33.62	0.8910	36.10	0.9392
CNN-IR [3]	33.77	0.8938	36.66	0.9438
PBVS-2020 winner [33]	34.49	0.9073	-	-
TherISuRNet [6]	34.49	0.9101	36.76	0.9450
Proposed	34.86	0.9133	37.38	0.9509
Proposed+	34.90	0.9134	37.49	0.9518

5. Conclusion

Thermal imaging sensors, despite having various critical applications, are often faced with limited resolution. Deep networks have been proposed to super-resolve low-resolution images using software-driven approaches. How-

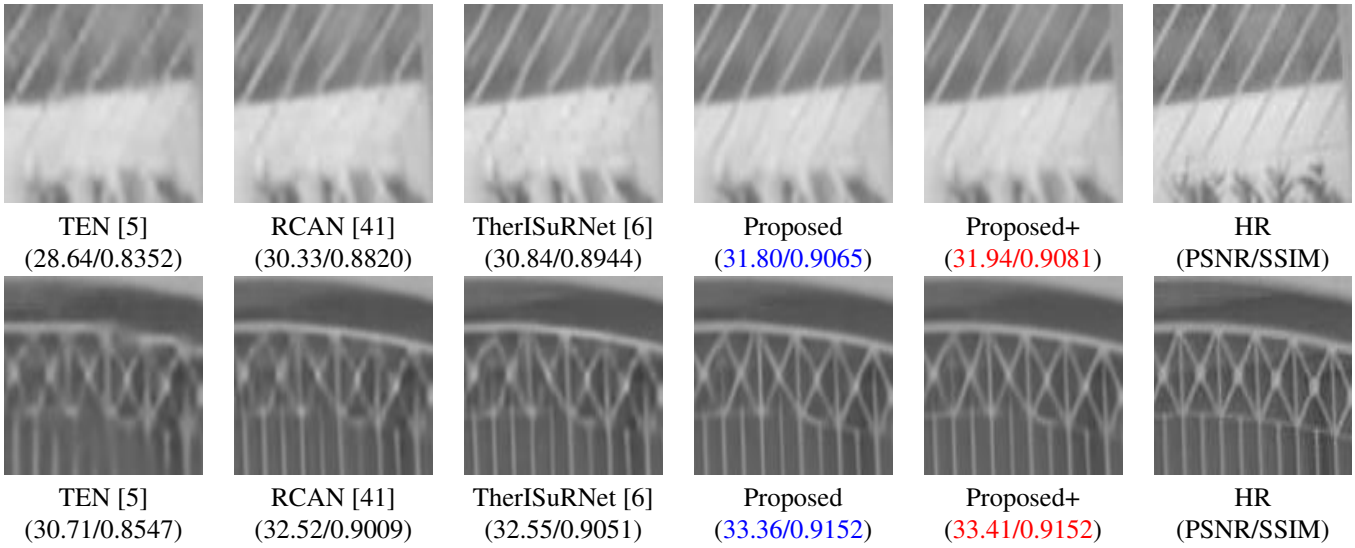


Figure 4: The visual comparison of the proposed network with other existing CNN based methods for thermal image super-resolution with factor $\times 4$ along with noise elimination on Flir dataset [34] (i.e., Track-1). The values of PSNR and SSIM values are mentioned along the result of each method where the top values are highlighted using red and blue colored fonts, respectively.

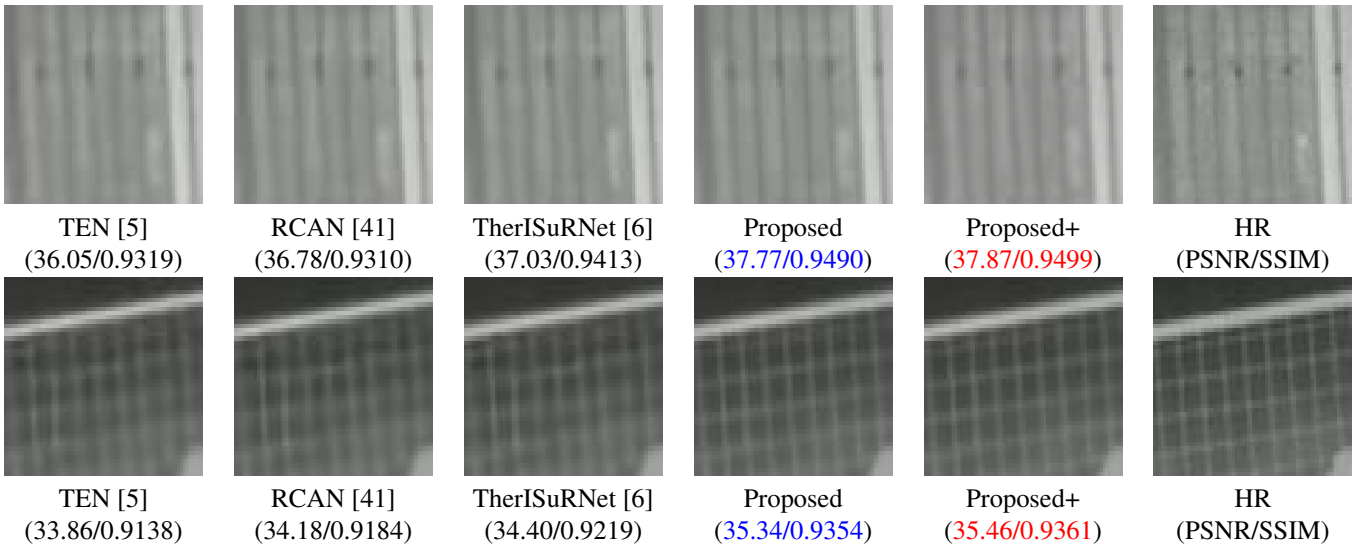


Figure 5: The visual comparison of the proposed network with other existing CNN based methods for thermal image super-resolution with factor $\times 2$ for domain shifting case on Axis dataset of PBVS-2021 thermal SR Challenge [34] (i.e., Track-2).

ever, the redundant emerging from deep networks can be bypassed directly using channel splitting in the architecture for super-resolution. This work has presented a new network for super resolving thermal images using a novel CNN framework leveraging channel splitting referred to as *ChaSNet*. We have proposed to employ Common Blocks structured in a residual manner to improve the SR performance. The performance of the proposed network is validated by comparing it with recent methods quantitatively and qualitatively, where the improvements are noted consistently. As a future work in this direction, one can adopt different losses alongside various training strategies in the GAN framework to improve the quality of SR images.

Acknowledgment

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 support for this research work (ECR/2017/003268).

References

[1] B. C. Arrue, A. Ollero, and J. R. Matinez de Dios. An intelligent system for false alarm reduction in infrared forest-fire

detection. *IEEE Intelligent Systems and their Applications*, 15(3):64–73, May 2000.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

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

[4] Gabriele Bitelli, Paolo Conte, Tamas Csoknyai, Francesca Franci, Valentina A. Girelli, and Emanuele Mandanici. Aerial thermography for energetic modelling of cities. *Remote Sensing*, 7(2):2152–2170, 2015.

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

[6] Vishal Chudasama, Heena Patel, Kalpesh Prajapati, Kishor P Upla, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Therisurnet-a computationally efficient thermal image super-resolution network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 86–87.

[7] 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.

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

[9] Matthew Fox, Steve Goodhew, and Pieter De Wilde. Building defect detection: External versus internal thermography. *Building and Environment*, 105:317 – 331, 2016.

[10] Gilad Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 30:12:1–12:11, 2011.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Jianchao Yang, J. Wright, T. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[18] P. Kansal and S. Nathan. A multi-level supervision model: A novel approach for thermal image super resolution. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 426–431, 2020.

[19] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981.

[20] 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.

[21] 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.

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

[23] 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 *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 1004–1009, July 2019.

[24] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

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

[26] 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.

[27] Kalpesh Prajapati, Vishal Chudasama, and Kishor Upla. A light weight convolutional neural network for single image super-resolution. *Procedia Computer Science*, 171:139–147, 2020.

[28] A.J. Prata and C. Bernardo. Retrieval of volcanic ash particle size, mass and optical depth from a ground-based thermal infrared camera. *Journal of Volcanology and Geothermal Research*, 186(1):91 – 107, 2009. Improved Prediction and Tracking of Volcanic Ash Clouds.

[29] 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.

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

[31] E Rivadeneira, Sappa Rafael Angel D, 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.

[32] Rafael Rivadeneira, Patricia Suarez, Angel Sappa, and Boris Vintimilla. *Thermal Image SuperResolution Through Deep Convolutional Neural Network*, pages 417–426. 08 2019.

[33] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Lin Guo, Jiankun Hou, Armin Mehri, Parichehr Behjati Ardakani, Heena Patel, Vishal Chudasama, Kalpesh Prajapati, et al. Thermal image super-resolution challenge-pbvs 2020. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 96–97, 2020.

[34] Rafael E Rivadeneira, Patricia L Su 'a rez, Angel D Sappa, and Boris X Vintimilla. Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*.

[35] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015.

[36] 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 *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, June 2016.

[37] Patricia L. Suárez, Angel D. Sappa, and Boris X. Vintimilla. Vegetation index estimation from monospectral images. In Aurélio Campilho, Fakhri Karray, and Bart ter Haar Romeny, editors, *Image Analysis and Recognition*, pages 353–362, Cham, 2018. Springer International Publishing.

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

[39] Chaofeng Wang, Zheng Li, and Jun Shi. Lightweight image super-resolution with adaptive weighted learning network. *arXiv preprint arXiv:1904.02358*, 2019.

[40] Xuehui Wang, Qing Wang, Yuzhi Zhao, Junchi Yan, Lei Fan, and Long Chen. Lightweight single-image super-resolution network with attentive auxiliary feature learning. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[41] 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.