Thermal Image Super-Resolution via Lightweight Efficient Channel Attention Network

Andrii Didenko[1] and Andrii Oliinyk[1]

Abstract

Single Image Super-resolution (SISR) methods are actively developed with the help of advancements in Convolution Neural Networks (CNNs) and attention mechanisms. Following the progress in RGB SISR methods, thermal image super-resolution methods (TISR) are beginning to adopt and implement these advancements. Despite showing prominent results, modern state-of-the-art SISR methods often have a large number of parameters, leading to a significant computational overhead and memory consumption and making it difficult to run these methods in real-time or on edge devices. To address these problems, we propose a parameter-efficient TISR model named LECAN, which consists of a stack of efficient channel-spatial attention blocks (ECSAB). Specifically, the ECSAB combines Pixel Attention (PA) with the proposed Efficient Contrast-aware Channel Attention (ECCA) to extract both spatial and channel-wise features while maintaining a low parameter count. Meanwhile, the Attentive Feature Fusion (AFF) mechanism effectively combines information from all blocks, capturing both low-level and high-level features. The qualitative and quantitative results show that the proposed method achieves superior results among same-size models while preserving the texture and patterns of the thermal image with a small number of parameters.

Keywords

Image super-resolution, thermal image, deep learning, attention mechanism, lightweight network

1. Introduction

The ability of thermography to capture light information beyond the visible scope has made it in demand in many spheres. Unlike thermal cameras, RGB cameras often struggle to capture images in low-light and bad weather conditions, making them highly sensitive and dependent tools. Moreover, thermal images are also used as an additional source of information, enhancing the overall process of data analysis. These advancements have made thermography a preferable instrument in many fields, including medicine [1], UAVs [2], agriculture [3], etc.

On the other hand, due to the high cost and complexity of producing high-resolution thermal cameras, the output infrared image is often low-resolution. This constraint makes analyzing thermal data difficult, making it harder to distinguish small details in the image. This, in turn, leads to a decrease in the quality of data analysis. To overcome this issue, image super-resolution (ISR) techniques can be used, that are invariant of the camera hardware.

With the development of CNNs, ISR task can be accurately solved with the help of deep learning techniques. These methods rely on the convolution operation, which can effectively extract patterns and textures of different complexity. On the other hand, to achieve high performance, these methods stack a sufficient amount of layers, making the overall size of the model large. This disadvantage makes it hard to integrate these models on edge devices or use them in real-time.

Efficient ISR methods help to decrease the size of the model while keeping the overall performance high. On the other hand, the development of such methods remains difficult due to the complexity of accuracy-size trade-off.

In this paper, we propose a novel architecture called Lightweight Efficient Channel Attention Network (LECAN) that is based on a combination of channel and spatial attention mechanisms. To keep the number of parameters low, we propose a combination of Contrast-aware and Efficient Channel Attention mechanisms. This structure helps to accurately extract fine details at different frequency levels.

The main contribution of the paper is:

[1]CMIS-2025: Eighth International Workshop on Computer Modeling and Intelligent Systems, May 5, 2025, Zaporizhzhia, Ukraine

an232did@gmail.com (A. Didenko); olejnikaa@gmail.com (A. Oliinyk)

0009-0009-9236-3936 (A. Didenko); 0000-0002-6740-6078 (A. Oliinyk)

- We propose a novel architecture called LECAN that consists its key components called Efficient Channel-Spatial Attention Blocks (ECSAB). The architecture of ECSAB allows to extraction spatial and channel features in parallel by dynamically recalibrating the importance of channels and pixels.
- We propose to use the Efficient Contrast-aware Channel Attention block (ECCA) for solving the TISR task. This idea combines the strength of contrast-aware feature extraction and the efficiency of ECA.
- The qualitative and quantitative results show that the proposed method demonstrates competitive results in different benchmarks with less number of parameters.

2. Related work

2.1. Image Super-Resolution methods

2.1.1. CNN-based

Image SR methods began rapid development since the CNNs became popular. SRCNN [4] was the first method to apply a CNN network to solve SR, consisting of 3 convolutional layers and becoming a state-of-the-art method at one time. However, 3 convolutional layers are not enough to learn a sufficient amount of patterns for reconstruction. The authors of VDSR [5] used 20 convolutional layers, showing significant improvements in results. Combining the advancements of residual learning [6] to increase the performance of very deep CNNs with Generative Adversarial Networks (GANs), the authors of SRGAN [7] proposed a GAN-based SR model to further boost the performance of SR methods. On the other hand, the training process of GAN-based is difficult due to hard convergence and mode collapse. To achieve better performance, the authors of EDSR [8] proposed to remove batch normalization in SrResNet [7] blocks. Despite the fact that the process of stacking layers can gradually improve the accuracy of SR models, it will eventually become inefficient to apply in real-world scenarios because of the large number of parameters. Addressing this problem, the authors of RCAN [9] proposed the use of the Channel Attention (CA) mechanism to adaptively enhance channel-wise features based on the interdependencies of the channels.

2.1.2. Efficient SR

To overcome the problem of considerable complexity of most state-of-the-art methods, efficient models can be used. These models try to minimize the overall complexity by removing unnecessary layers, reducing the number of parameters of basic building blocks, distilling large models, etc.

The authors of PAN [10] proposed a network with Pixel Attention (PA) mechanism, that aims to enhance spatial dependencies with fewer parameters. A^2N model [11] uses building blocks that consist of attention and non-attention branches weighted by a dynamic attention module, that helps to dynamically adjust the impact of each attention branch. IMDN [12] employs the channel-splitting strategy, in which one part of the channels is kept while the other part is passed on for further processing. This strategy allows to improve the performance while reducing the number of parameters. The authors of RFDN [13] successfully improved the idea behind IMDN by enhancing channel-splitting operation and reducing the number of parameters in convolutional layers. BSRN [14] further improves channel splitting idea

Although efficient models are able to produce an upscaled image quickly and with low computational costs, the quality of the output image remains low compared to the original image. This trade-off emphasizes the need to develop efficient models that can process images with high accuracy while remaining small in size.

2.2. Attention mechanism

Inspired by human attention, the attention mechanism in deep learning helps the model focus more on important parts of the input data. The Squeeze-and-Excitation (SE) [15] block was proposed to enhance features in channel dimension by assigning weights to each channel. CBAM [16] further expands this idea by assigning weights not only in the channels but also in the spatial dimensions. The

Efficient Channel Attention (ECA) is an improvement of the SE block that uses a one-dimensional convolutional block to reduce the model complexity. The self-attention mechanism that is used in Vision Transformers [18] also inspired the development of transformer-based models for low-level vision tasks. SwinIR [18] adopted the Swin Transformer [20] as a baseline by creating a Residual Swin Transformer Block with a convolutional layer and long residual connection. IPT [21] also applied ViT-based architecture and introduced multi-task pretraining for low-level vision. HAT [22] combines channel attention and window-based attention to enhance performance by activating more input pixels.

In this work, we endeavor to enhance the ISR model with the strength of the attention mechanism, while keeping the model efficient at the same time.

2.3. Thermal Image Super-Resolution

Due to the domain specificity of RGB and infrared images, which include visual and contextual differences, directly applying SR methods developed and trained for RGB images on infrared images may not be beneficial. Therefore, there is a need to design and train models directly for the TISR task.

The authors of the PSRGAN [23] proposed to use GAN-based model along with multistage transfer learning for solving TISR task. TherISuRNet [24] method consists of several residual blocks for extracting features and different frequency levels. The authors of MPRANet [25] proposed residual- and attention-based network with convolution of different kernel sizes. ChaSNet [26] uses channel-splitting technique to improve feature extraction. LISN [27] also uses a channel-splitting idea to reduce the number of parameters. LDANet [28] uses blocks with attention and non-attention branches weighted by dynamic attention modules.

Despite noticeable results in TISR, this area is still under-researched compared to other low-level vision tasks. This supports the need to develop robust methods directly for solving TISR task.

3. Method

This section describes the architecture of the building blocks and the overall proposed network for thermal image SISR, as well as the motivation for their implementation and usage.

3.1. Network Architecture

The architecture of the proposed networks (Figure 1) follows the extended standard structure for the SR models: shallow feature extractor (SFE), deep feature extractor (DFE), feature fusion (FF), and image reconstruction (IR, upsampling).

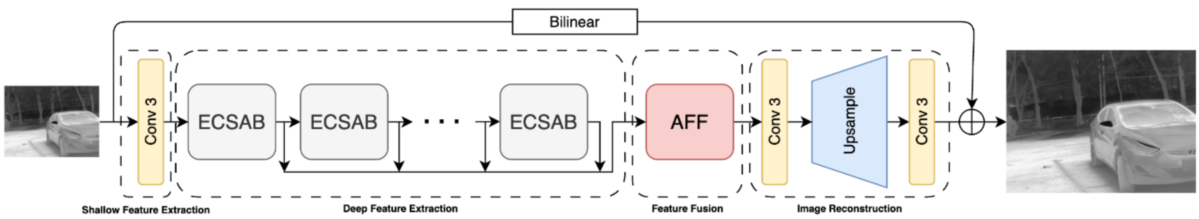

Figure 1: The architecture of the proposed method

The SFE is represented by a single convolutional layer with a kernel size of 3×3. Mathematically, the SFE is represented as follows (1):

$$x_{SFE} = f_{SFE}(I_{LR}), \tag{1}$$

where x_{SFE}– output features from SFE module;

I_{LR} – low-resolution image;

f_{SFE} – the function of the SFE module.

After shallow features are extracted, the output from the SFE module is then processed by the DFE module. The DFE module is the main part of the network which is responsible for extracting high-level features and complex patterns that might be useful for ISR. The DFE module consists of a stack of Efficient Channel-Spatial Attention Blocks (ECSAB), which will be explained in Section 3.2. The output of the DFE module is the stack of outputs of each ECSAB module that is then processed by the AFF module. Mathematically, the DFE is represented as follows (2):

$$x_{DFE}=f_{DFE}(x_{SFE})=\{f_{B_1}(x_{SFE});f_{B_2}(f_{B_1}(x_{SFE}));\ldots;f_{B_N}(f_{B_{N-1}}(\ldots f_{B_1}(x_{SFE})\ldots))\}, \qquad (2)$$

where x_{DFE}– outputs from DFE module;

f_{B_N} – the function of *N-th* ECSAB block.

The stacked outputs from the DFE module are then processed by the FF module to accurately fuse and extract the interdependencies between the output channels of each ECSAB block and the spatial features of the fused results. The FF module is represented by the AFF block, which is described in Section 3.3. Mathematically, the output of the FF module is expressed as follows (3):

$$x_{FF}=f_{FF}(f_{DFE}(f_{SFE}(I_{LR}))), \qquad (3)$$

where x_{FF} – output features from FF module;

f_{FF}– the function of the FF module.

Finally, the IR module reconstructs the high-resolution image by upscaling it with the desired factor. Figure 2 shows the IR block. The IR module consists of one IR block if upscaling factor is 2 and two IR block if upscaling factor is 4.

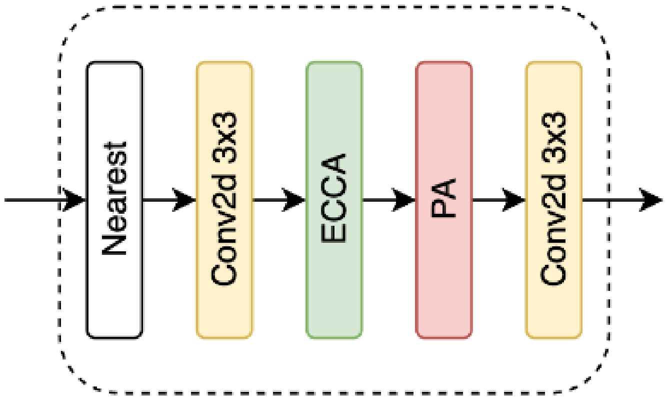

Figure 2: Image Reconstruction Block

IR block can be represented mathematically as follows (4):

$$x_{IR}=Conv(PA(ECCA(Conv(NN(x_{FF}))))), \qquad (4)$$

where x_{IR} – output from IR module;

NN – nearest neighbor interpolation.

The overall mathematical formula for ISR is the following (5):

$$I_{SR}=f_{IR}(f_{FF}(f_{DFE}(f_{SFE}(I_{LR}))))+B(I_{LR}), \qquad (5)$$

where I_{SR} – output upscaled image;

f_{IR} – the function of the IR module;

B – bilinear interpolation.

3.2. Efficient Contrast-aware Channel Attention Block

Simply combining the strength of the ECA [17] and CCA [12], we propose the ECCA block as part of the TISR method. First, the ECCA inherits parameter efficiency from the ECA block. Secondly, the contrast-aware part from CCA introduces a better refinement of textures and edges, allowing the model to capture information from low-, medium-, and high-level features. Following the original paper, the contrast-aware operation is the summation of each channel's standard deviation and mean. Figure 3 shows the structures of spatial-based attention (PA) and channel-based attention (including the proposed ECCA).

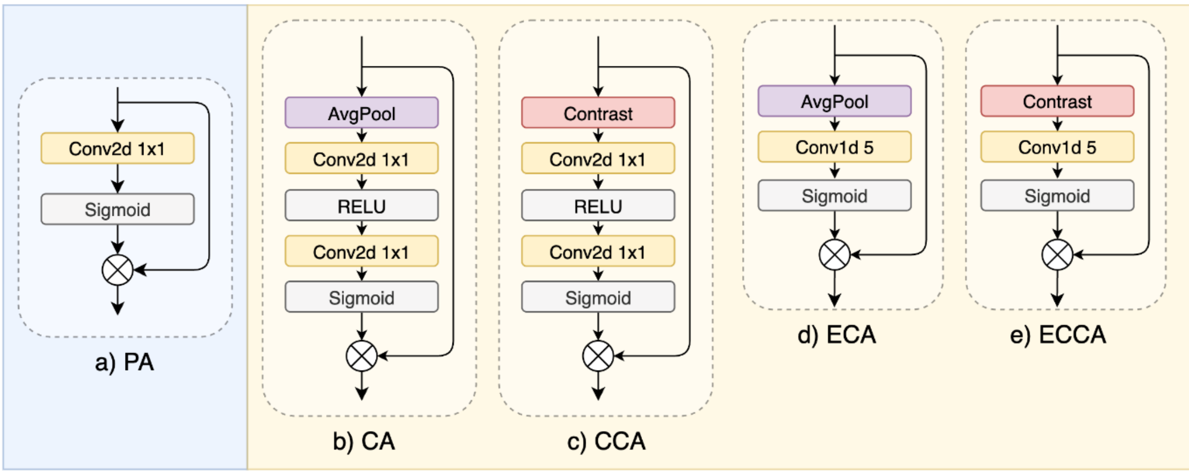

Figure 3:The visualization of attention mechanisms: a – Pixel Attention (PA), b – Channel Attention (CA), c – Contrast-aware Channel Attention (CCA), d – Efficient Channel Attention (ECA), e – Efficient Contrast-aware Channel Attention (ECCA).

Mathematically, ECCA can be expressed as follows (6):

$$x_{ECCA} = Sigmoid\left(Conv1d\left(mean_c\left(x_{in}\right) + std_c\left(x_{in}\right)\right)\right), \tag{6}$$

where x_{ECCA} – output from ECCA block;
x_{in} – input to the ECCA block;
$mean_c$ – per channel mean;
std_c – per channel standard deviation

3.3. Efficient Channel-Spatial Attention Block

The key component of the proposed method is the ECSA block and its combination of channel-wise and pixel-wise attention mechanisms. Specifically, this combination of two attention mechanisms allows the model to simultaneously focus on reweighting the feature maps along the channel dimension with the ECCA and emphasize the importance of individual pixels within the feature maps with the PA. Consequently, the network is able to leverage complementary information is across both spatial and channel dimensions. The parallel structure of applying attention mechanisms to the input features ensures independent behavior of extracting spatial and channel-wise dependencies with a further combination of the feature maps. Residual connections allow the preservation of input features' information for further layers. Figure 4 shows the architecture of the ECSA block.

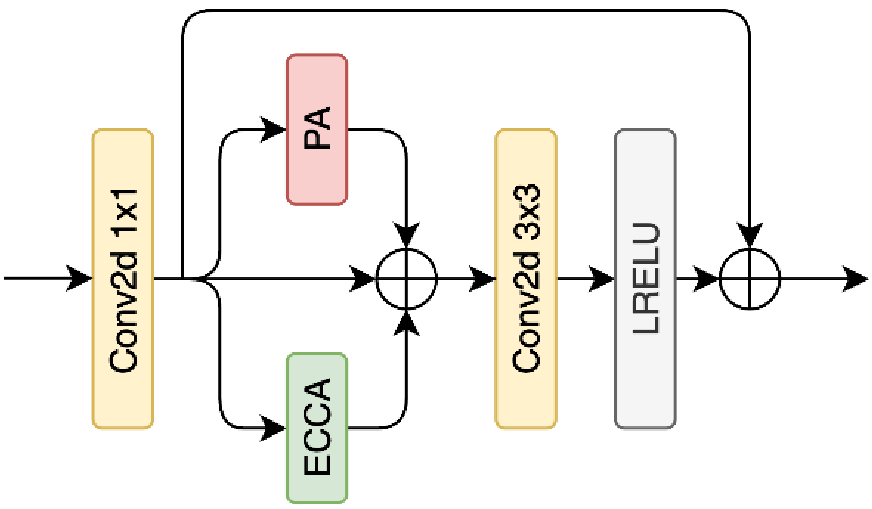

Figure 4: Architecture of the ECSA block

3.4. Attentive Feature Fusion

The AFF block serves as a additional part of the proposed model and its main goal is to efficiently fuse output features from different detailization levels. To do this, outputs from each ECSA block are concatenated along the channel dimension. Then, concatenated channels are processed with an ECCA block to extract dependencies across all blocks channel-wise. This operation allows the model to attend more to important information that might be spread along channels of different blocks, efficiently combining low-level and high-level features. Then, to reduce the number of parameters, the point-wise convolution is applied. In the end, the PA block is used to further process spatial information. In general, this module is based on the assumption that different levels of deep feature extraction might carry some portion of useful features and the attentive combination of features might improve the selection of this information. Figure 5 shows the architecture of the FFA block.

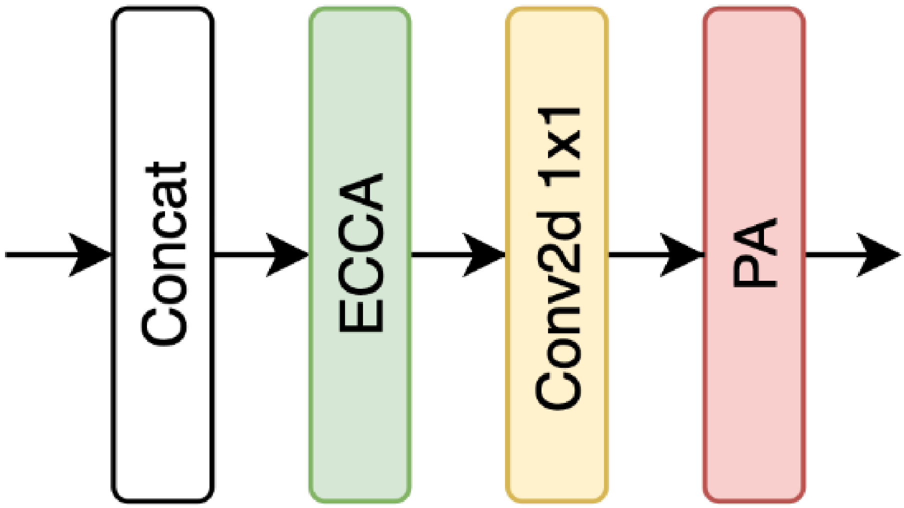

Figure 5: The architecture of the FFA block

4. Experimental analysis

The training setup for all models was the same to exclude the dependency of training parameters. AdamW [29] was used as an optimizer. The learning rate was set to 2e-4 with a MiltuStepLR scheduler that multiplies the learning rate by 0.5 at the following milestones: 50k, 65k, 80k, and 90k. The total number of iterations was set to 100k. During training, a patch of size 256×256 was randomly cropped from the HR image along with the corresponding patch from the LR image. The batch size during training was set to 8. Horizontal and vertical flips were used as data augmentation techniques, as well as random JPG compression with quality varying from 0.9 to 1. RSNR and SSIM were used as evaluation metrics. The experiments were conducted with PyTorch framework.

4.1. Training and testing datasets

For training, we used the Challenge dataset Помилка: джерело посилання не знайдено. This dataset consists of thermal images of three different resolutions: LR Domo, MR Axis, and HR FLIR. To create a training set for this task, we downsampled HR FLIR thermal images by the scales of 2 and 4. The resulting dataset contains 951 images for training and 50 images for validation.

For testing, we used a recent Challenge dataset (Challenge 2) [30]. This dataset contains 1000 images, where 900 images are provided for training and validation, while the other 100 images are used to evaluate entries for the challenge (ground-truth is hidden). To create a testing set, 900 GT images were downsampled by scales of 2 and 4.

We also used CVC-09: FIR Sequence Pedestrian Dataset [31] by randomly selecting 1000 GT images and downsampling them by scales of 2 and 4.

4.2. Ablation study

We conducted the ablation study to examine the effect of different attention mechanisms in the network. Specifically, we trained 4 models with CA, ECA, CCA, and ECCA blocks in the architecture. The results of the ablation study on Channel Attention type are shown in Table 1.

Table 1
The results of the ablation study on Channel Attention type. The top values are highlighted in red and blue respectively

Attention type	N Params (k)	PSNR (dB)	SSIM
CA	399	32,5775	0,9294
ECA	344	32,563	0,9292
CCA	399	32,5836	0,9296
ECCA	344	32,5894	0,9296

The results show that the contrast-aware part of the Channel Attention can improve the performance, while the usage of parameter-efficient Channel Attention reduces overall model complexity, keeping model's accuracy high.

4.3. Quantitative evaluation

We compared our proposed model with several state-of-the-art methods: SRCNN [4], BSRN [14], PAN [10], RFDN [13], A2N [11], and IMDN [12]. The quantitative evaluation shows that the proposed method achieves competitive results while remaining relatively small compared to other models. Tables 2-5 present quantitative results for each method, as well as the size of each model.

Table 2
The quantitative results on the Challenge 2 dataset with a scaling factor of 4. The top values are highlighted in red and blue respectively

Model	N Params (k)	PSNR (dB)	SSIM
SRCNN	8	31,0928	0,8959
BSRN	333	32,4592	0,9261
PAN	271	32,4990	0,9279
RFDN	530	32,5271	0,9276
A2N	1046	32,5683	0,9289
IMDN	696	32,649	0,9292
LECAN (proposed)	344	32,5894	0,9296

Table 3
The quantitative results on the Challenge 2 dataset with a scaling factor of 2. The top values are highlighted in red and blue respectively

Model	N Params (k)	PSNR (dB)	SSIM
SRCNN	8	41,6601	0,9746
BSRN	327	43,4645	0,9872
PAN	260	43,4757	0,9873
RFDN	524	43,3422	0,9867
A2N	1035	43,2968	0,9862
IMDN	687	43,5564	0,9867
LECAN (proposed)	336	43,5997	0,9874

Table 4
The quantitative results on the CVC-09 dataset with a scaling factor of 4. The top values are highlighted in red and blue respectively

Model	N Params (k)	PSNR (dB)	SSIM
SRCNN	8	37,4622	0,9075
BSRN	333	37,9992	0,9129
PAN	271	38,0288	0,9130
RFDN	530	38,0285	0,9129
A2N	1046	38,0571	0,9133
IMDN	696	38,0743	0,9135
LECAN (proposed)	344	38,0670	0,9136

Table 5
The quantitative results on the CVC-09 dataset with a scaling factor of 2. The top values are highlighted in red and blue respectively

Model	N Params (k)	PSNR (dB)	SSIM
SRCNN	8	41,2614	0,9418
BSRN	327	41,9301	0,9501
PAN	260	41,8375	0,9484
RFDN	524	41,7690	0,9474
A2N	1035	41,7778	0,9477
IMDN	687	41,7091	0,9466
LECAN (proposed)	336	41,8790	0,9491

4.4. Qualitative evaluation

The quantitative results show that the proposed method is able to reconstruct different patterns and textures of thermal images. The proposed model can Figures 6-9 present a qualitative analysis of LECAN comparing to state-of-the-art methods.

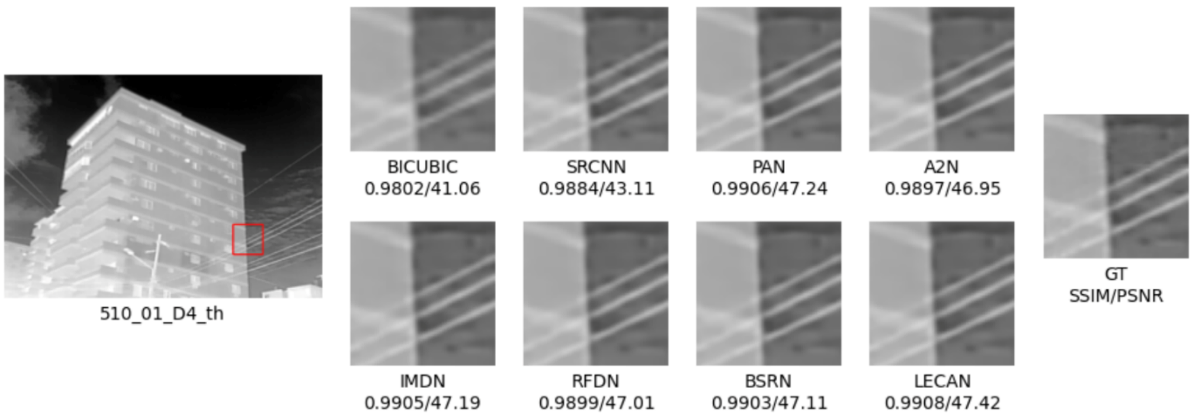

Figure 6: Qualitative results on the Challenge 2 dataset with a scaling factor of 2

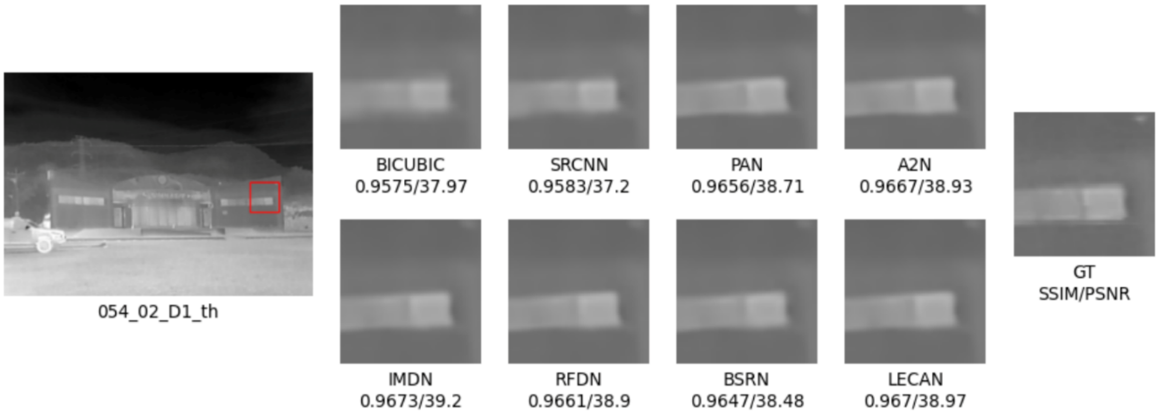

Figure 7: Qualitative results on the Challenge 2 dataset with a scaling factor of 4

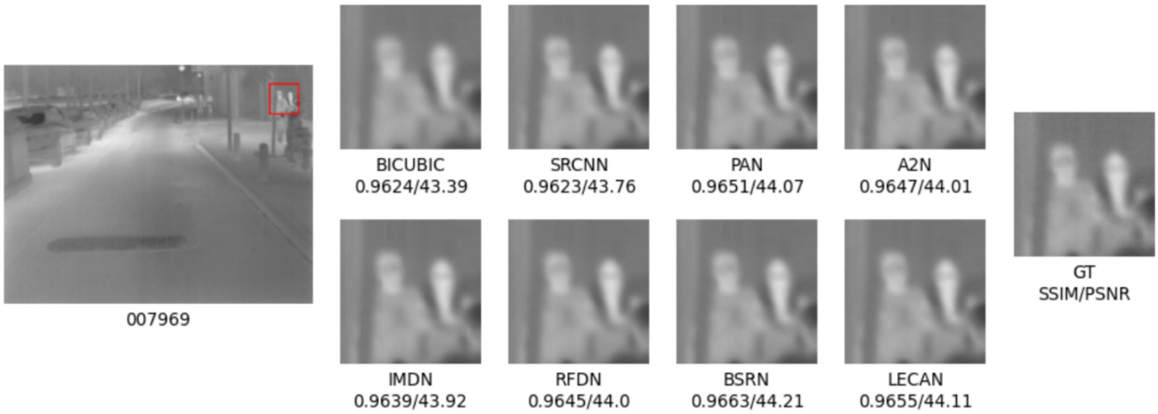

Figure 8: Qualitative results on the CVC-09 dataset with a scaling factor of 2

Figure 9: Qualitative results on the CVC-09 dataset with a scaling factor of 4

4.5. Inference speed evaluation

Inference evaluations were conducted on a server CPU AMD EPYC 7R32. The input image size is 256x256, and an average time of 10 runs was chosen. Table 6 lists the inference speed results in ms and the number of Floating Point Operations (FLOPs). The results show good trade-off between accuracy and inference speed.

Table 6
The quantitative inference results

Model	N Params (k)	Time (ms)	FLOPs (G)
SRCNN	8	118	4,26
BSRN	327	748	40,8
PAN	260	371	40
RFDN	524	418	68,47
A2N	1035	921	140,7
IMDN	687	433	89.8
LECAN (proposed)	336	559	48.6

5. Conclusion

In this paper, we propose LECAN for solving the TISR task. The proposed model consists of four main parts: shallow feature extraction, deep feature extraction, attentive feature fusion, and image reconstruction. The deep feature extraction consists of several ECSAB blocks. The ECSAB block efficiently combines channel and spatial attention mechanisms, where channel attention is represented by the Efficient Contrast-aware Channel Attention (ECCA) block, and spatial attention is represented by the Pixel Attention (PA) block. The combination of Contrast-aware and Efficient Channel Attention mechanisms allows to reduce the number of parameters and enhances the overall performance of the model. The qualitative and quantitative comparisons show that the proposed method demonstrates competitive results while maintaining a low parameter count. Further work can be aimed to improve the extraction of more complex features by enhancing attention mechanisms.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

[1] R. Ishimwe, K. Abutaleb, F. Ahmed, Applications of Thermal Imaging in Agriculture - A Review, Advances in Remote Sensing (2014), volume 3, 2014, pp. 128-140. doi:10.4236/ars.2014.33011.

[2] Q. Zhang., L. Zhou., J. An, Real-Time Recognition Algorithm of Small Target for UAV Infrared Detection, Sensors (2024), volume 10, 2024, doi:10.3390/s24103075.

[3] M. B. Rakhunde, S. Gotarkar, S. G. Choudhari, Thermography as a Breast Cancer Screening Technique: A Review Article, Cureus (2024), volume 14, 2024, doi:10.7759/cureus.31251.

[4] C. Dong, C. C. Loy, K. He, X. Tang, Image Super-Resolution Using Deep Convolutional Networks in: Proceedings of ECCV 2014, Zurich, 2014, pp. 184-199, doi:10.1007/978-3-319-10593-2_13.

[5] J. Kim, J. K. Lee, K. M. Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Networks, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016, pp. 1646-1654, doi:10.1109/CVPR.2016.182.

[6] K. He, X. Zhang, S. Ren. J. Sun, Deep Residual Learning for Image Recognition, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016, pp. 770-778, doi:10.1109/CVPR.2016.90.

[7] C. Ledig et al., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, in: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp. 105-114, doi:10.1109/CVPR.2017.19.

[8] B. Lim, S. Son, H. Kim, S. Nah. K. M. Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, in: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, 2017, pp. 1132-1140, doi:10.1109/CVPRW.2017.151.

[9] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image Super-Resolution Using Very Deep Residual Channel Attention Networks, in: Proceedings of ECCV 2018, Munich, 2018, pp 294–310, doi:10.1007/978-3-030-01234-2_18.

[10] H. Zhao, X. Kong, J. He, Y. Qiao, C. Dong, Efficient Image Super-Resolution Using Pixel Attention in: Proceedings of 2020 European Conference on Computer Vision, Glasgow, 2020, pp. 56-72, doi:10.1007/978-3-030-67070-2_3.
[11] H. Chen, J. Gu, Z. Zhang, Attention in Attention Network for Image Super-Resolution, 2021, doi:arXiv:2104.09497v3.
[12] Z. Hui, X. Gao, Y. Yang, X. Wang, Lightweight Image Super-Resolution with Information Multi-distillation Network, in: Proceedings of the 27th ACM International Conference on Multimedia, New York, 2019, pp. 2024–2032, doi:10.1145/3343031.3351084.
[13] J. Liu, J. Tang, G. Wu, Residual Feature Distillation Network for Lightweight Image Super-Resolution, in Proceedings of Computer Vision – ECCV 2020 Workshops, Glasgow, 2020, pp.41-55, doi:10.1007/978-3-030-67070-2_2.
[14] Z. Li et al., Blueprint Separable Residual Network for Efficient Image Super-Resolution, in: Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, 2022, pp. 832-842, doi:10.1109/CVPRW56347.2022.00099.
[15] J. Hu, L. Shen. G. Sun, Squeeze-and-Excitation Networks, in: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018, pp. 7132-7141, doi:10.1109/CVPR.2018.00745.
[16] S. Woo, J. Park, J. Lee, and I. S. Kweon, CBAM: Convolutional Block Attention Module, in: Proceedings of ECCV 2018 15th European Conference, Munich, 2018, pp.3-19, doi:10.1007/978-3-030-01234-2_1.
[17] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo. Q. Hu, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks, in: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 2020, pp. 11531-11539, doi:10.1109/CVPR42600.2020.01155.
[18] A. Dosovitskiy et al., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021 International Conference on Learning Representations, 2021, doi:10.48550/arXiv.2010.11929.
[19] J. Liang, J et. al., SwinIR: Image Restoration Using Swin Transformer, in: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, 2021, pp. 1833-1844, doi:10.1109/ICCVW54120.2021.00210.
[20] Z. Liu et al., Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, in: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, 2021, pp. 9992-10002, doi:10.1109/ICCV48922.2021.00986.
[21] H. Chen et al., Pre-Trained Image Processing Transformer, in: Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, 2021, pp. 12294-12305, doi:10.1109/CVPR46437.2021.01212.
[22] X. Chen, X. Wang, J. Zhou, Y. Qiao, C. Dong. Activating More Pixels in Image Super-Resolution Transformer, in: Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, 2023, pp. 22367-22377, doi:10.1109/CVPR52729.2023.02142.
[23] Y. Huang, Z. Jiang, R. Lan, S. Zhang, K. Pi, Infrared Image Super-Resolution via Transfer Learning and PSRGAN, IEEE Signal Processing Letters (2021), volume 28, 2021, pp. 982-986. doi:10.1109/LSP.2021.3077801.
[24] V. Chudasama et al., TherISuRNet - A Computationally Efficient Thermal Image Super-Resolution Network, in: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, 2020, pp. 388-397, doi:10.1109/CVPRW50498.2020.00051.
[25] H. Zhang, Y Hu, M. Yan, B. Ma, Thermal image super-resolution via multi-path residual attention network, Signal, Image and Video Processing (2023), volume 17, 2023, pp. 2073–2081. doi:10.1007/s11760-022-02421-x.
[26] K. Prajapati et al., Channel Split Convolutional Neural Network (ChaSNet) for Thermal Image Super-Resolution, in: Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, 2021, pp. 4363-4372, doi:10.1109/CVPRW53098.2021.00493.
[27] S. Lue et al., Infrared Image Super-Resolution via Lightweight Information Split Network, in: Advanced Intelligent Computing Technology and Applications ICIC 2024, volume 14869, Singapure, 2024. doi:10.1007/978-981-97-5603-2_24.

[28] H. Zhang, Y. Hu, Lightweight dynamic attention network for single thermal image super-resolution, Signal, Image and Video Processing (2024), volume 18, 2024, pp. 2195-2206. doi:10.1007/s11760-023-02886-4.
[29] I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, in: Proceedings of the 7th International Conference on Learning Representations, New Orleans, 2019. doi:10.48550/arXiv.1711.05101.
[30] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, Thermal image super-resolution: a novel architecture and dataset, in: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, volume 4, Valletta, 2020, pp. 111-119. doi:10.5220/0009173601110119.
[31] R. E. Rivadeneira et al., Thermal Image Super-Resolution Challenge Results-PBVS 2024, in: Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, 2024, pp. 3113-3122, doi:10.1109/CVPRW63382.2024.00317.
[32] Socarr´as, Y., Ramos, S., V´azquez, D., L´opez, A.M., Gevers, T.: Adapting pedestrian detection from synthetic to far infrared images. In: ICCV Workshops, vol. 3 (2013).