Thermal Image Super-Resolution Methods Using Neural Networks

Andrii Didenko[a], Andrii Oliinyk[a] and Sergey Subbotin[a]

[a] National University "Zaporizhzhia Polytechnic", Zhukovskoho street 64, Zaporizhzhia, 69063, Ukraine

Abstract

Thermography has gained popularity in many fields. Since the human eye cannot see the thermal spectrum and in low light, thermal image analysis has become an integral part of medicine, manufacturing, construction, and other industries. Most thermal cameras produce low-resolution images when analyzing the temperature of objects, which complicates the process of analyzing original thermogram. Therefore, the problem of improving the quality of thermal images is relevant today. With the development of artificial intelligence and deep learning technologies, new super-resolution methods are emerging. Such methods can also be adapted for thermal image processing. This work examines the performance of modern super-resolution methods in the thermal vision domain.

Keywords

Machine learning, thermography, deep learning, thermal image, neural network, super-resolution, image processing, computer vision

1. Introduction

At all temperatures above absolute zero, every object emits energy from its surface in the form of a spectrum of different wavelengths and intensities. The radiation we call visible light, forms a very small part of the electromagnetic spectrum. The infrared or thermal wavelength region extends over the range 0.7 to 1000 μm [1]. Infrared radiation can carry a lot of useful information about an object that is studied. However, it is invisible to the human eye, thus it is important to have special equipment to analyze infrared data.

Infrared thermography (IRT) has a lot of applications in different fields. It is fast and non-invasive method of diagnostics that is widely used in medicine to detect abnormal body temperature, to diagnose breast cancer, diabetes neuropathy, etc. [2]. It is also used in the architectural and civil engineering fields [3]. In manufacturing, for example, a combination of IRT and deep learning helps to detect cracks in steel plates [4]. IRT also finds its application in the condition monitoring of electrical equipment using machine learning algorithms [5]. Moreover, it can be also used for field phenomics of different tree species using unmanned aerial vehicles (UAVs) [6].

Despite having a lot of applications in different fields, IRT has its own drawbacks. The main issue with thermal imaging is the resolution of the output thermogram. While modern RGB cameras are able to produce high-resolution images, the majority of thermal cameras produce images of low spatial resolution. Besides, thermal cameras that can produce high-resolution thermograms cost much more than regular ones and are not affordable to ordinary users. This creates a need for thermal image super-resolution methods.

The goal of Super-Resolution (SR) methods is to recover a high-resolution image from one or more low-resolution input images [7]. A high-resolution image provides a higher pixel density and thus more detail about the original scene. The need for high-quality images often arises in the field of computer

vision to improve the efficiency of pattern recognition and image analysis. SR methods have been developing for decades and have come a long way from classical approaches to modern deep learning models.

The purpose of this work is the implementation and analysis of different modern SR methods in the thermal image domain.

2. Related Work

SR methods can be divided into several groups: interpolation methods, Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Transformers. Each of these methods has its own advantages and disadvantages but in practice, neural network based methods show better performance which usually overlaps their disadvantages.

2.1. Interpolation Methods

Interpolation methods are the simplest among SR methods. Among them are nearest neighbor interpolation, bilinear interpolation, and bicubic interpolation [8]. In the nearest neighbor method, each interpolated output pixel is assigned the value of the nearest sample point in the input image while bilinear interpolation is used to know values at random positions from the weighted average of the four closest pixels to the specified input coordinates. Bicubic is an advancement over the above interpolation and uses polynomials, cubic, or cubic convolution algorithms [8]. These methods are available in every image processing software. Despite their ease of implementation and clarity, the main disadvantage of the interpolation methods is the quality of the output image, thus they are usually used as baselines for more advanced SR methods.

2.2. CNNs

CNN is the main type of neural network that is used for image processing. It also has found its application in SR tasks. For the first time, the idea of using CNNs instead of classical approaches was introduced in [9]. In this method, the image is first upscaled using bicubic interpolation, then processed through a network consisting of three parts: image patch extraction, nonlinear mapping and reconstruction. Patch extraction represents patches of image as a multidimensional vector of features. Nonlinear mapping transforms each multidimensional vector into another multidimensional vector. At the reconstruction stage, the image patch representations are aggregated to generate the original enlarged image [9].

Method [10] is the improvement of the previous method that speedups image super-resolution. Unlike the method [9], the authors do not enlarge the image with bicubic interpolation before passing it to the neural network but work with the original image. In addition, the size of the convolutional layer filter was reduced to 5. Then, the number of filters is reduced to reduce the training load. At the nonlinear mapping stage, several convolutional layers are used instead of one. The next stage increases the number of filters, which improves the quality of image processing. The last step is the deconvolution, which produces an enlarged image.

Method [11] uses neural network with autoencoder architecture for image restoration which includes SR task. The authors propose a deep fully convolutional auto-encoder network, which is an encoding-decoding framework with symmetric convolutional-deconvolutional layers. The network is composed of multiple layers of convolution and de-convolution operators, learning end-to-end mappings from corrupted images to the original ones [11].

The ideas of the method [9] found their usage in thermal image SR in [12]. The authors examined the use of datasets from different domains to train the model. The results show that a model trained to upscale regular RGB images is better at enhancing thermal images than a model trained on thermal images.

In [13], the authors also studied the influence of the dataset domain on the quality of training and model results. The authors conducted experiments in such color models as grayscale, HSL, HSI, and

HSV. In addition, the architecture of the model was also inspired by the method [9]. According to the results, the network trained by the gray outperformed the one that used the lightness and intensity domains, but the networks based on the brightness domain provided better performance compared to the gray-based network [13].

In [14], on contrary to the above methods, the authors showed that training a model on a dataset of thermal images gives better results than training on a dataset of RGB images.

In [15] authors use a progressive upscaling strategy with asymmetrical residual learning [15] and also compare their work to SR methods that are commonly used in RGB image SR task.

2.3. GANs

GANs [16], as its name suggests, originally were used to generate images. GANs consist of two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake [16].

The architecture of GANs is widely used in SR tasks. For example, in [17] authors proposed a novel model for SR, called SRGAN. This method uses residual network called SrResNet as a generator. Besides, the authors propose a perceptual loss function which consists of an adversarial loss and a content loss.

An improvement of SRGAN is ESRGAN [18] which increases model performance and improves architecture and loss function of the previous method. In particular, they introduced a special block in generator model, called Residual-in-Residual Dense Block (RRDB). They also modified perceptual loss by using features before activation layers.

GANs found their application in thermal image SR task. For example, method [19] uses CycleGAN [20]. The generator is ResNet6 model and discriminator is PatchGAN, whole model is trained in unsupervised way. Besides, the authors also released dataset for training thermal image SR models, which consists of low-resolution, middle-resolution, and high-resolution thermal images.

2.4. Transformers

Recently, the Transformer neural network architecture has been increasingly used in deep learning solving different tasks. Transformer was first introduced for solving natural language processing (NLP) tasks [21], but then showed impressive performance in other domains. The core of Transfomer is the attention mechanism that helps neural network to memorize long sequences and focus on specific parts of the input.

In [22], the authors propose to use a Transformer, namely an encoder, to classify images, while preserving the original architecture as much as possible. This architecture is called a Vision Transformer (ViT). To do this, the image is divided into several patches (in this article, a patch is 16x16 pixels), so a sequence of patches is fed to the transformer's input. To preserve information about the location of the patches, information about the position of these patches relative to each other (positional coding) is added to the input sequence. In order for the model to learn to classify images, a representation of the image class to be learned during model training is also added to the input sequence. ViT shows better results in comparison to CNN models.

Despite having good performance, ViT has several important drawbacks. One of these drawbacks is the processing of high-resolution images, as the computational complexity increases quadratically with image size. In addition, the architecture of a ViT is poorly suited for solving other computer vision tasks, such as segmentation, since in this task it is important to distinguish image features at different scales.

To solve these problems Shifted-window Transformer (Swin Transformer) [23] was introduced. Swin Transformer has linear computation complexity as it computes self-attention not within every patch of the image but within patches in the local window. Then, to compute connections between each window, shifted window attention is used. It can thus serve as a general-purpose backbone for both image classification and dense recognition tasks [23].

Transformers can be also used for solving SR tasks. For example, SwinIR [24] is a transformer-based image restoration model mostly inspired by Swin Transformer. SwinIR consists of three parts: shallow feature extraction, deep feature extraction, and HQ image restoration. In turn, deep feature extraction block is a stack of special residual Swin Transformer blocks (RSTB) [24]. This architecture makes it possible to achieve excellent results in different image restoration tasks, such as SR, denoising, and artifact reduction.

3. Method

As each method has its own advantages and disadvantages, it was decided to conduct experiments with classical algorithms as baseline, CNNs, GANs and Transformer-based neural networks and then compare the results of these methods.

In particular, in this paper following neural networks are examined: SRGAN, SrResNet (generator of SRGAN), ESRGAN, RRDBNet (generator of ESRGAN), SwinIR on thermal images and RGB images domains, and SwinIRGAN (combination of SwinIR as generator and SRGAN discriminator).

4. Experiments

Each of the selected models was already pretrained on its own domain. Also, models were trained and evaluated with the help of BasicSR [25] – deep learning framework for solving image restoration tasks, such as denoising and SR. Each model was trained to upscale thermal images by a factor of 2.

4.1. Dataset

Since the task of thermal image SR is quite specific, the number of open-source datasets for this task is quite low.

One of the largest and most popular is the FLIR dataset [26]. This dataset offers annotated thermal and RGB images for object detection systems. The dataset contains more than 26000 annotated frames and 15 different object categories. In total, the dataset contains almost 10,000 thermal and more than 9,000 RGB images with a size of 640 by 480 pixels [26]. Figure 1 shows an example of data from FLIR dataset.

Figure 1: Example data from FLIR [26] dataset

One of the most popular datasets for solving the RGB SR problem is DIV2K [27]. This dataset contains 1000 high-resolution images (1500-2000 pixels per image side) divided into training, test and validation sets. This dataset also contains reduced images by a factor of 2, 3, 4, and 8.

In this work, both of these datasets are used. In particular, DIV2K is used to examine the performance of model trained on grayscale images of this dataset in thermal image domain.

4.2. Data Preprocessing

To use FLIR dataset for the thermal image SR task, only thermal images (single channel) were selected and a version of the dataset with thermal images reduced by a factor of 2 (320 by 240 pixels) was created.

In order to use the DIV2K dataset for the thermal image SR task, a version of the dataset with images in grayscale format was created. Thus, images have one channel with a pixel range of 0-255.

4.3. Evaluation

To evaluate trained models SSIM [28] (structural similarity index measure) and RSNR (peak signal-to-noise ratio) metrics were used. SSIM is defined as follows (1):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$ (1)

where x, y are windows of image of size $N{\times}N$;
μ_x – mean of x;
μ_y – mean of y;
σ_x – variance of x;
σ_y – variance of y;
c_1, c_2 – stabilization variables.
PSNR is defined as follows (2):

$$PSNR = 10 \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right),$$ (2)

where MAX_I is maximum pixel value of image I;
MSE – mean squared error.

4.4. Experiment Setup

Before conducting experiments, a set of hyperparameters was chosen.

Adam [29] with a learning rate 0.0002 was chosen as an optimization algorithm for all models. All models were trained for 100000 iterations. The number of images in one batch is 16 for SwinIR, ESRGAN and RRDBNet and 64 for SRGAN and SrResNet.

All models were trained not on whole images, but on square patches that were randomly selected from each image. This speeds up training and forces the neural network to pay attention to high-frequency features. Thus, for all models, a patch of 128×128 pixels was chosen, except for the classic SwinIR model (96×96 pixels).

4.5. Results

The quantitative comparison is listed in Table 1.

Table 1
Results of thermal image super-resolution

Model	SSIM	PSNR
Nearest Neighbors	0,7389	30,3350
Bilinear Interpolation	0,7557	31,1910
Bicubic Interpolation	0,7672	31,5820
SRGAN	0,6617	28,3197
SwinIRGAN	0,6751	29,6528
ESRGAN	0,6853	29,8002
SwinIR (DIV2K)	0,7642	31,4355
RRDBNet	0,7829	32,3622
SRResNet	0,7828	32,3471
SwinIR (FLIR)	**0,7833**	**32,3688**

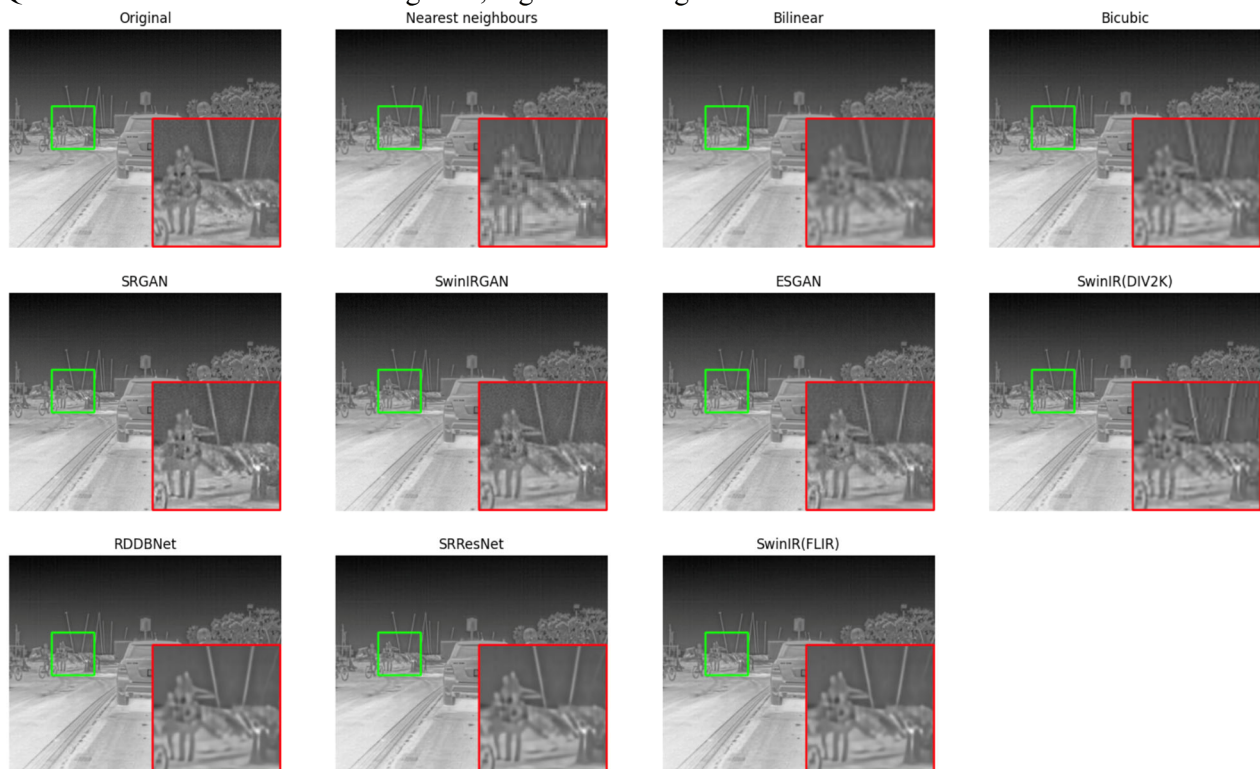Qualitative results are shown in Figure 2, Figure 3 and Figure 4.

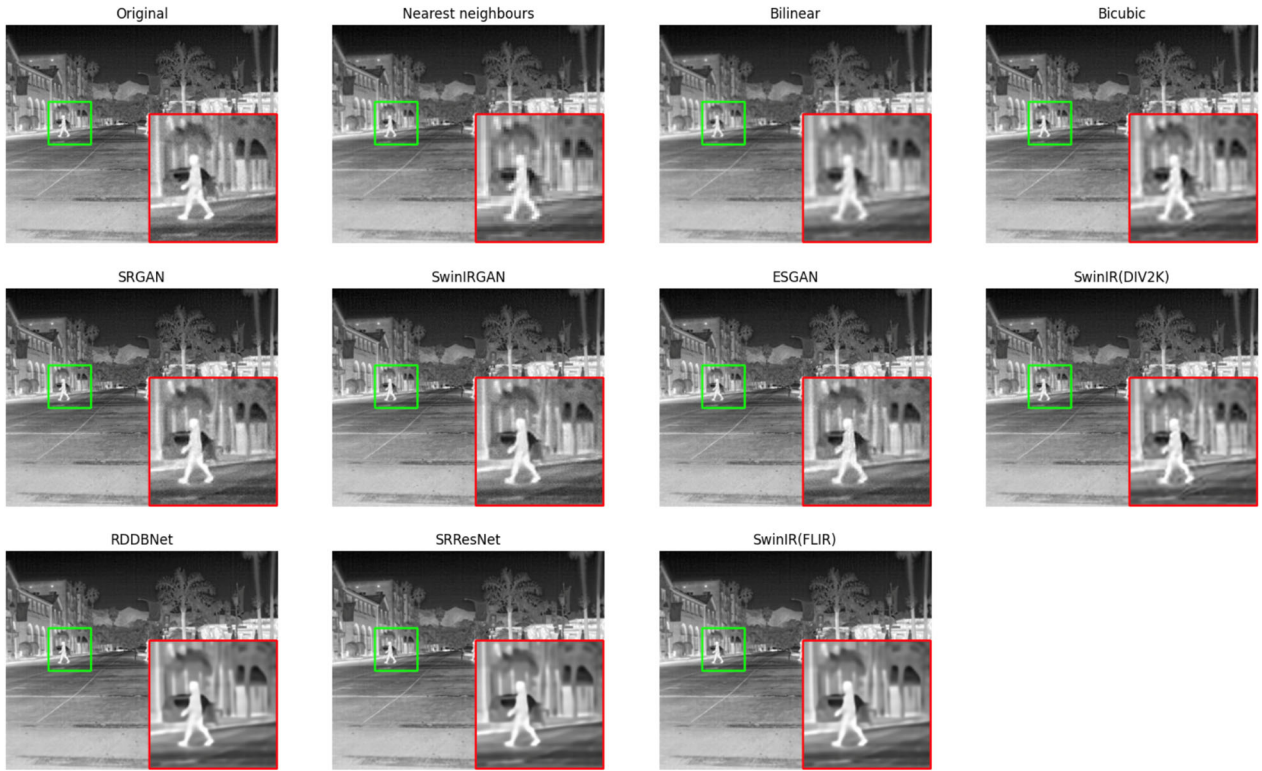**Figure 2**: Qualitative comparison on FLIR sample (small objects), scaling factor ×2

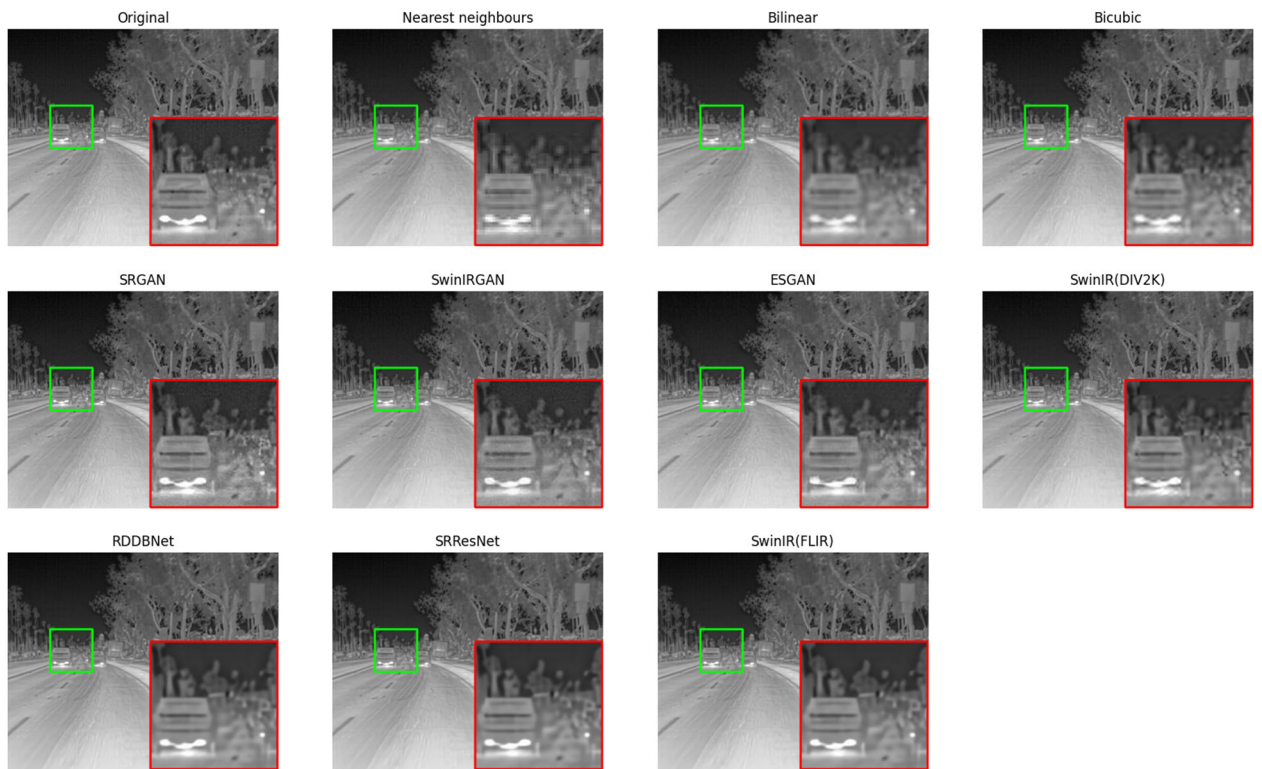**Figure 3**: Qualitative comparison on FLIR sample (pedestrian), scaling factor ×2

Figure 4: Qualitative comparison on FLIR sample (car), scaling factor ×2

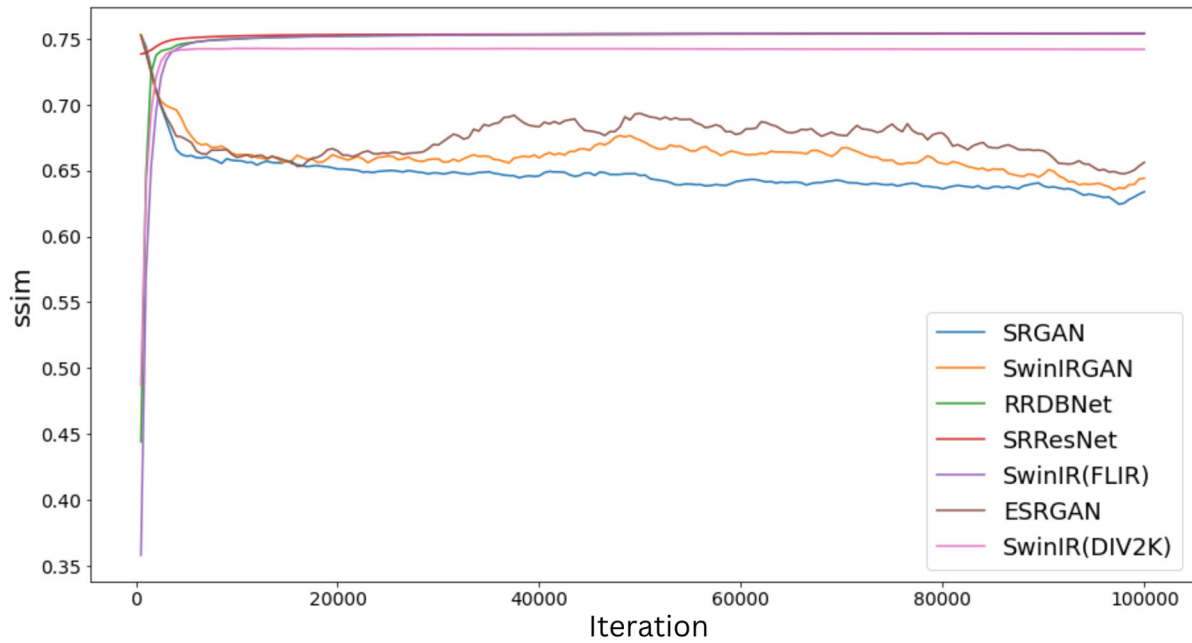Results of SSIM and PSNR metrics during training are shown in Figure 5 and Figure 6 respectively.

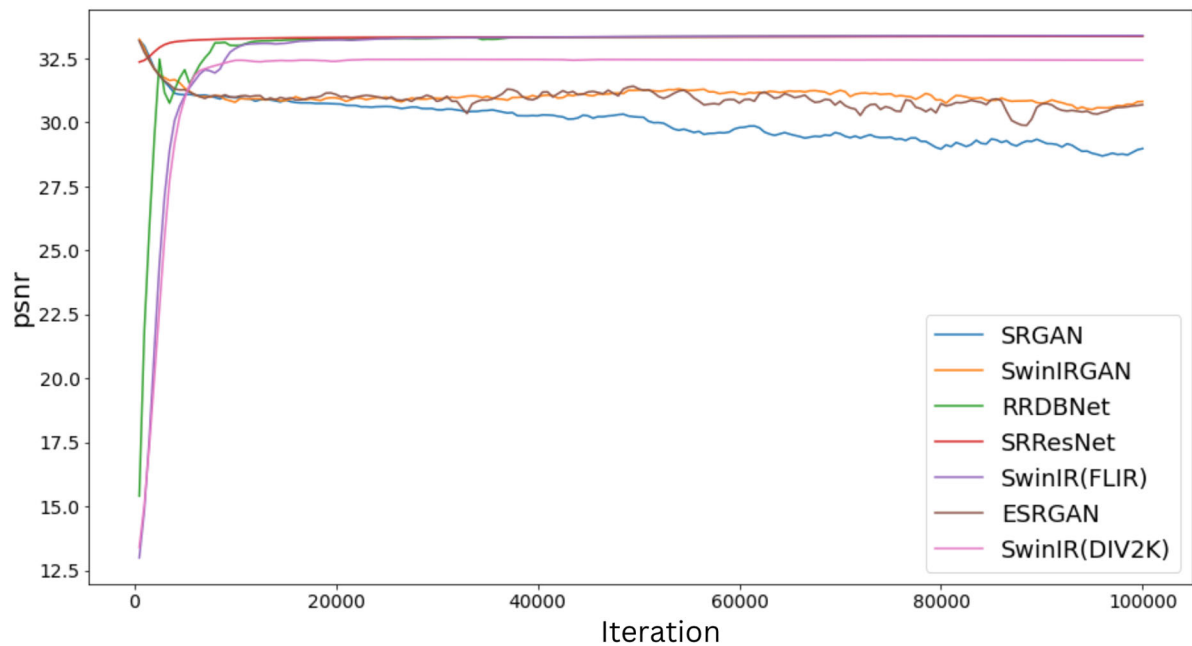**Figure 5**: SSIM metric on the validation set during training

Figure 6: PRNR metric on the validation set during training

From the obtained results, it can be concluded that SwinIR trained on FLIR dataset shows the best results among other models. Additionally, CNNs such as SrResNet and RRDBNet trained separately are better than these CNNs trained as GANs. As it can be seen from qualitative results, GANs have a lot of noise on upscaled images which causes bad quantitative results, while CNNs and Transformers almost completely denoise upscaled images. It can be explained by the fact that GANs are sensitive to the input data and selected training hyperparameters. It is also visible that SwinIR trained on FLIR dataset performs better than SwinIR trained on the grayscale DIV2K dataset.

5. Conclusion

This paper implements modern SR methods and examines their results on thermal image SR problem. The aim of this paper was to analyze the performance of popular SR methods on upscaling of thermal images.

The main problem of this task is the lack of datasets for thermal image SR. Current datasets don't have enough training data, the size of thermal images in dataset is not as big as in RGB datasets for SR tasks, and quality of thermal images is worse than quality of RGB images in training sets.

According to the obtained results of thermal image upscaling it was concluded that Transformer and CNN models can perform better than GANs and classical algorithms.

In addition, the following steps can be taken to improve the quality of SR of thermal images: increase the size of the dataset and training batch to improve the generalization ability of the models; collect dataset of thermal images with higher quality and lower percentage of noise in the images to improve the results of SR methods; tune training hyperparameters such as optimization algorithm, learning rate, loss function, etc.; improve the architecture of used models or modify them for the task of thermal image SR.

6. References

[1] J. M. Hart A practical guide to infra-red thermography for building surveys, Building Research Establishment, Garston, 1991.

[2] B.B. Lahiri, S. Bagavathiappan, T. Jayakumar, John Philip, Medical applications of infrared thermography: A review, Infrared Physics & Technology, vol. 55(4), (2012) pp. 221-235. doi: 10.1016/j.infrared.2012.03.007.

[3] C. Meola, Infrared Thermography in the Architectural Field, The Scientific World Journal, vol. 2013, (2013). doi: 10.1155/2013/323948.

[4] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun and Y. Sun, Infrared Thermal Imaging-Based Crack Detection Using Deep Learning, IEEE Access, vol. 7, (2019) pp. 182060-182077. doi: 10.1109/ACCESS.2019.2958264.

[5] M. Najafi, Y. Baleghi, S. A. Gholamian and S. Mehdi Mirimani, Fault Diagnosis of Electrical Equipment through Thermal Imaging and Interpretable Machine Learning Applied on a Newly-introduced Dataset, 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), (2020) pp. 1-7. doi: 10.1109/ICSPIS51611.2020.9349599.

[6] R. Ludovisi, F. Tauro, R. Salvati, S. Khoury, G. Scarascia Mugnozzaa, A. Harfouche, UAV-Based Thermal Imaging for High-Throughput Field Phenotyping of Black Poplar Response to Drought, Front. Plant Sci, vol. 8, (2017) doi: 10.3389/fpls.2017.01681.

[7] D. Glasner, S. Bagon, M. Irani, Super-resolution from a single image, 2009 IEEE 12th International Conference on Computer Vision, (2009) pp. 349-356. doi: 10.1109/ICCV.2009.5459271.

[8] S. Fadnavis, Image Interpolation Techniques in Digital Image Processing: An Overview, International Journal Of Engineering Research and Application, vol. 4(10), (2014) pp. 70-73

[9] 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(2), (2016) pp. 295-307. doi: 10.1109/TPAMI.2015.2439281.

[10] C. Dong, C. C. Loy, and X. Tang, Accelerating the super-resolution convolutional neural network, European Conference on Computer Vision, vol. 9906, (2016) pp. 391–407. doi: 10.1007/978-3-319-46475-6_25.

[11] X. Mao, C. Shen, and Y. Yang, Image restoration using convolutional auto-encoders with symmetric skip connections, Advances in Neural Information Processing Systems, (2016).

[12] Y. Choi, N. Kim, S. Hwang and I. S. Kweon, Thermal Image Enhancement using Convolutional Neural Network, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), (2016) pp. 223-230. doi: 10.1109/IROS.2016.7759059.

[13] K. Lee, Junhyeop Lee, Joosung Lee, S. Hwang, and S. Lee, Brightness-based convolutional neural network for thermal image enhancement, IEEE Access, vol. 5, (2017) pp. 26867–26879. doi: 10.1109/ACCESS.2017.2769687.

[14] R. E Rivadeneira, P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, Thermal image superresolution through deep convolutional neural network, International Conference on Image Analysis and Recognition, (2019) pp. 417–426. doi: 10.1007/978-3-030-27272-2_37.

[15] V. Chudasama et al., TherISuRNet - A Computationally Efficient Thermal Image Super-Resolution Network, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2020) pp. 388-397. doi: 10.1109/CVPRW50498.2020.00051.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, Advances in Neural Information Processing Systems (NIPS), vol. 3(11), (2014) pp. 2672–2680. doi: 10.1145/3422622.

[17] C. Ledig, L. Theis, F. Husz´ar, et al, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017) pp. 105-114. doi: 10.1109/CVPR.2017.19.

[18] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, European Conference on Computer Vision 2018 Workshop, vol. 11133 (2019). pp. 63-79. doi: 10.1007/978-3-030-11021-5_5.

[19] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, Thermal image super-resolution: a novel architecture and dataset, International Conference on Computer Vision Theory and Applications, (2020) pp 1–2. doi: 10.5220/0009173601110119.

[20] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, 2017 IEEE International Conference on Computer Vision (ICCV), (2017) pp. 2242-2251. doi: 10.1109/ICCV.2017.244.

[21] Vaswani, A.; Shazeer, N. M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need, Proceedings of the 31st International Conference on Neural Information Processing System, (2017) pp. 6000–6010. doi: 10.48550/arXiv.1706.03762.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, International Conference on Learning Representations, (2021). doi: 10.48550/arXiv.2010.11929.

[23] Z. Liu, Y. T. Lin, Y. Cao; H. Hu, B. N. Guo, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), (2021) pp. 9992-10002. doi: 10.1109/ICCV48922.2021.00986.

[24] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool and R. Timofte, SwinIR: Image Restoration Using Swin Transformer, 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), (2021) pp. 1833-1844. doi: 10.1109/ICCVW54120.2021.00210.

[25] Xintao Wang, Liangbin Xie, Ke Yu, Kelvin C.K. Chan, Chen Change Loy and Chao Dong. BasicSR: Open Source Image and Video Restoration Toolbox, 2022. URL: https://github.com/xinntao/BasicSR.

[26] FREE Teledyne FLIR Thermal Dataset for Algorithm Training. URL: https://www.flir.eu/oem/adas/adas-dataset-form

[27] E. Agustsson and R. Timofte, NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study, 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2017) pp. 1122-1131. doi: 10.1109/CVPRW.2017.150.

[28] Z. Wang, E. P. Simoncelli and A. C. Bovik, Multiscale structural similarity for image quality assessment, The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, vol. 2, (2003) pp. 1398-1402. doi: 10.1109/ACSSC.2003.1292216.

[29] D. P. Kingma and J. Lei Ba, Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), (2015), pp 1–13. doi: 10.48550/arXiv.1412.6980.