Multi-Image Super-Resolution for Thermal Images

Rafael E. Rivadeneira[1][a], Angel D. Sappa[1,2][b] and Boris X. Vintimilla[1][c]

[1] *Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación, CIDIS, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

[2] *Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

{*rrivaden, asappa, boris.vintimilla*}*@espol.edu.ec*

Keywords: Thermal images, multi-view, multi-frame, Super-Resolution, deep learning, attention block

Abstract: This paper proposes a novel CNN architecture for the multi-thermal image super-resolution problem. In the proposed scheme, the multi-images are synthetically generated by downsampling and slightly shifting the given image; noise is also added to each of these synthesized images. The proposed architecture uses two attention blocks paths to extract high-frequency details taking advantage of the large information extracted from multiple images of the same scene. Experimental results are provided, showing the proposed scheme has overcome the state-of-the-art approaches.

1 INTRODUCTION

Image Super-resolution (SR) is an ill-posed problem that refers to reconstructing a high-resolution (HR) image from a single or multiple low-resolution (LR) images of the same scene. HR images are often required as they provide supplementary information, making it a widely studied problem with several practical applications in domains such as: surveillance and security ((Zhang et al., 2010), (Rasti et al., 2016), (Shamsolmoali et al., 2019)), medical imaging (e.g., (Mudunuri and Biswas, 2015), (Robinson et al., 2017), (Huang et al., 2019)), object detection (e.g., (Girshick et al., 2015)), among others; in spite of the large amount of literature it is still an active research field in the computer vision community (e.g., (Han et al., 2021), (Pesavento et al., 2021), (Song et al., 2021)). In the last years, most of the SR community has focused on the single image super-resolution (SISR) problem, which estimates the HR image from a single LR input. On the contrary, multi-image super-resolution (MISR) reconstructs the original HR image using multiple LR images of the same scenes.

Deep learning techniques have shown remarkable progress with respect to conventional methods, where most state-of-the-art approaches focus on the visible domain. Long-Wave Infra-Red (LWIR) images, a.k.a. thermal images, have shown the essential applications in many fields (e.g., (Qi and Diakides, 2003), (Hermann et al., 2018)); unfortunately, the technology (thermal cameras) to acquire a higher image pixel density is usually restrictive and overpriced. Most thermal images tend to have poor resolution. Still, with effective image processing techniques, such as learning-based super-resolution methods (like those used in the visible spectral domain), it is possible to generate a high-resolution thermal image from a low-resolution.

The current work tackles the thermal image super-resolution problem in the multi-image scheme. It requires as input several LR images from the same scene. Hence, due to the lack of a benchmark of multi-thermal image datasets, a dataset with synthesized images is generated. This dataset contains several LR images of a given scene by down-sampling, adding both noise and blur, and randomly shifting (X and Y coordinates) trying to simulate being captured by a burst of input images. On the contrary to SISR baseline, the main idea of the present approach is to combine information from multiple frames to obtain a more detailed reconstruction of the HR image. Up to our humble knowledge, there is just a few approaches on the literature using a multi-image scheme to generate HR thermal images.

In summary, the contributions of this manuscript are as follows:

- It generates a synthesized dataset that simulates a RAW burst of LR images with their corresponding

[a] https://orcid.org/0000-0002-5327-2048
[b] https://orcid.org/0000-0003-2468-0031
[c] https://orcid.org/0000-0001-8904-0209

HR ground truths.

- It proposes a novel MISR architecture for thermal images, which generates a HR representation using bursts generated images.
- It proposes an attention-based module that helps to merge the input images to generate the corresponding SR representation.

The remainder of this manuscript is organized as follows: Section 2 covers the related work tackled in the current work. The proposed approach is detailed in Section 3. Experimental results and comparisons are provided in Section 4. Finally, conclusions are given in Section 5.

2 RELATED WORK

This section summarizes work-related with SR, including both SISR and MISR approaches. Section 2.1 summarizes the state-of-art on single image super-resolution, mainly approaches proposed for images from the visible spectrum. Then, Section 2.2 tackles the studies related to the multi-image super-resolution problem.

2.1 Single Image Super-Resolution

SISR techniques have been widely used in the field of image processing with a variety of proposed methods and techniques, such as interpolation, frequency domain, sparse representations, among others (e.g., (Dai et al., 2007), (Ji and Fermüller, 2008), (Yang et al., 2010), (Freeman et al., 2002)). Recently, using deep learning techniques with convolutional neural networks (CNNs) has shown a great capability to improve the quality of SR results. The first deep CNN-based approach has been proposed by (Dong et al., 2015) (SRCNN), who achieved superior results than conventional methods, training a CNN to directly map the input LR images to generate a SR image as their HR counterparts. SRCNN was followed by its faster version (FSRCNN) (Dong et al., 2016) for learning LR to HR mapping, accelerating the testing and training need in their previous work. After SRCNN, a number of different approaches have been proposed with substantial improvements using more effective network architectures (e.g., (Kim et al., 2016), (Zhang et al., 2017), (Lim et al., 2017)) and loss functions (e.g., (Ledig et al., 2017), (Wang et al., 2018)).

Most of the SISR approaches mentioned above tackle images from the visible spectrum. SR approaches have also been proposed to enhance the resolution of images from other spectral bands, such as near-infrared, hyper-spectral, thermal-infrared (e.g., (Yao et al., 2020), (Long et al., 2021), (Choi et al., 2016)). The most recent works on thermal images SR are present in the first (Rivadeneira et al., 2020b) and second (Rivadeneira et al., 2021) thermal image super-resolution challenges organized on the workshop *Perception Beyond the Visible Spectrum* of CVPR2020 and CVPR2021 conferences. Both challenges use as a novel thermal dataset acquired by (Rivadeneira et al., 2019). In these challenges, two kinds of evaluations have been proposed: $Evaluation1$ consists of down-sampling the HR thermal images by a N factor and comparing their SR results with the corresponding GT images. $Evaluation2$ obtains the $\times 2$ SR from a given LR thermal image and compares it with its corresponding semi-registered HR image. Several teams have participated in both challenges by proposing different approaches.

2.2 Multi-Image Super-Resolution

MISR aims to merge the information extracted from multiple LR inputs images of the same scene to reconstruct a HR output. MISR techniques involve different ways of degrading the GT image (burring, warping, noising, shifting, downsampling) to get several LR images. The first approach presented on MISR (Tsai, 1984) uses frequency-domain techniques, which combine the multiple LR images with their sub-pixel displacement to enhance the spatial resolution and generate a SR image. In (Peleg et al., 1987) and (Irani and Peleg, 1991) an iterative back-projection approach is introduced, which was later on extended in (Hardie et al., 1998) with an improved observation model and a regularization term. A joint multi-frame demosaicking and super-resolution approach has been presented in (Farsiu et al., 2004) . Most of MISR methods are based on sub-pixel registration between the LR images and fusion into a super-resolved image (e.g., (Milanfar et al., 2011), (Rossi and Frossard, 2018)).

Currently, the state-of-the-art in MISR is dominated by neural networks, where their architecture must be able to align the noisy LR inputs images with sub-pixel accuracy to enable the fusion. Then they should be able to fuse the information between all aligned images. MISR problem takes more interest due to the increasingly popular mobile burst photography, where images have sub-pixel shifts due to hand tremors (Wronski et al., 2019). Satellite imagery is commonly used for MISR due to the available dataset. In (Deudon et al., 2020) the HighRes-net network is proposed; it aligns each input frame, from satellite imagery dataset to a reference frame, and merges

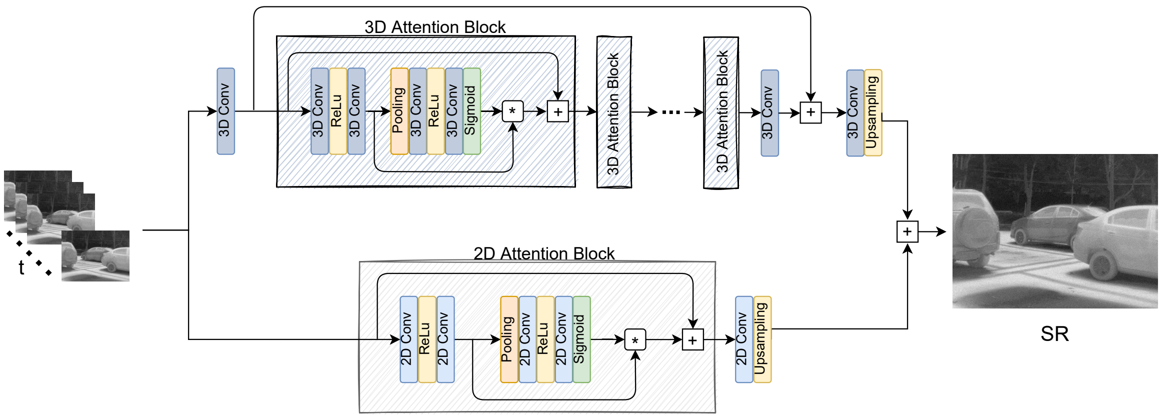

Figure 1: Proposed multi-image thermal super-resolution architecture using a 2D and 3D Attention Blocks.

them using a recursive fusion method. Similarly, DeepSUM (Molini et al., 2019) aligns each input but assumes just translation motion between frames and uses 3D convolutions for fusion. In contrast to these previous approaches, (Bhat et al., 2021) tackles the general problem of burst SR from any mobile camera. Recent works ((Salvetti et al., 2020), (Nguyen et al., 2021)) present a multi-image super-resolution architecture evaluating satellite images (PROBA-V), showing that the proposed MISR strategy overcomes state-of-the-art results.

2.3 Datasets

As SISR, most MISR approaches are focused on the visible or near-infrared spectrum. As far as we know, there are no approaches on multi-image super-resolution for thermal images. It should be noticed that in MISR it is necessary to count with multiple LR images of the same scene. In order to overcome this problem, datasets with synthesized images can be used; in this case, images can be generated following most of the degradation process applied to the GT images together with random shifts. In other words, datasets intended for SISR can be used to generate such synthesized images to be used by the MISR approaches.

Regarding SISR dataset of thermal images, (Rivadeneira et al., 2019) has presented a dataset intended for SR, which contains a total of 101 images captured with a single HR TAU2 camera from FLIR; each thermal image has a native resolution of 640×512 pixels. An extensive database has been presented in (Rivadeneira et al., 2020a); it consists of a set of 1021 thermal images acquired with three different thermal cameras at different resolutions (referred to as LR, MR, and HR camera). The cameras were mounted on a panel, trying to minimize the baseline distance between each optical axis camera. This dataset was used as a benchmark in the first and second thermal image super-resolution challenges organized on the workshop *Perception Beyond the Visible Spectrum* of CVPR2020 (Rivadeneira et al., 2020b) and CVPR2021 conferences (Rivadeneira et al., 2021). In the current work, the HR images of (Rivadeneira et al., 2020a) are used; synthesized images from this dataset are generated, simulating that the inputs were acquired in a RAW burst of LR images—multi-LR images.

3 PROPOSED APPROACH

This section presents an overview of the approach proposed for thermal multi-image SR. The neural network (as shown in Fig. 1) takes as an input a sequence of multiple noisy, RAW, LR thermal images and combines their features to generate a SR image. Inspired on (Salvetti et al., 2020), the current approach consists of two main paths, a 2D Attention Block and a 3D Attention Block. Both paths use Residual attention blocks, which are the core of the model that focuses on the images' high-frequency (HF) features. HF features have more valuable information for SR generation. For better computational performance, the up-sample operation is done at the end of each path. Finally, the results from both paths are added to generate the SR image.

The 2D Attention Path allows the network to generate a simple super-resolution solution for upsampling a set of multi-LR images. This attention path consists of: 2DConv -> ReLU -> 2DConv -> GlobalPoll -> 2DConv -> ReLU -> 2DConv -> Sigmoid, with respective skip connection, followed by 2DConv -> UpSampling.

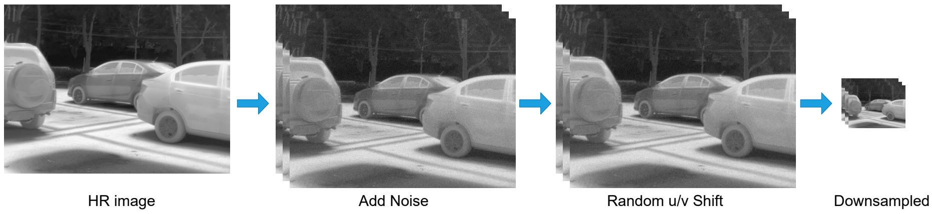

Figure 2: Illustration of the model used as degradation prepossessing.

The 3D Attention Path uses 3D convolutions residual-based blocks to extract spatial correlations from the pool of inputs LR images. This path is the main branch of the approach. First, a 3D convolution layer is applied to extract shallow features from the LR input images. After this, a cascade of N concatenates 3D Attention Blocks is applied for higher extractions of features exploiting the spatial and local, and non-local correlations. Long skip connection is used for redundant low-frequency signals and several short skip connections inside each block. Finally, the up-sample operation is done. In summary, this attention path consists of 12 times: 3DConv -> ReLU -> 3DConv -> GlobalPoll -> 3DConv -> ReLU -> 3DConv -> Sigmoid, with respective skip connection, and a long skip connection, followed by 3DConv -> UpSampling.

The multi-image super-resolution approach can be summarized as follow:

$$SR = U\left(2D_{attB}(LR_t)\right) + U\left([3D_{attB}(LR_t)]^N\right) \quad (1)$$

where U represents the up-sampling operation, 2D and 3D are each attention block paths, and N is the number of times the 3D path repeats. LR_t represents the set of multi-image from the same scene, and SR represents the generated super-resolution image.

4 EXPERIMENTAL RESULTS

This section presents the results of the MISR proposed approach, training it on a synthesized dataset and comparing its performance with state-of-the-art SISR algorithms. Section 4.1 presents information of the generated dataset; then, Section 4.2, depicts the parameters used for the training phase. Finally, Section 4.3, shows the quantitative and qualitative results obtained with the proposed approach.

4.1 Synthesized Dataset

SR reconstruction is highly dependent on the degradation model. Several factors such as relative motion (handshake), atmospheric turbulence, optical blurring, and preprocessing are used to generate a simulated burst of multi-LR thermal images. The thermal dataset used to evaluate the proposed model consists of 1021 thermal images (950 for training, 50 validating, and 20 for testing). Assuming all thermal LR images are generated under the same condition, the degradation model can be formulated as:

$$Y_t = (X + G_t) * S_t * D_t; t = 1, 2, ..., T, \quad (2)$$

where X, Y_t represent the t^{th} HR image and LR image respectively. G_t is the additional Gauss noise to Y_t. S_t and D_t represents the random u and v shift and downsampled factor by 4, respectively; ten LR images (T = 10) are generated. When random shift is done, reflect padding is performed to fill the gap of the shift. The degradation process is illustrated in Fig. 2. Random $gauss - noise$ with a value of 2 std. Random $left - right$ shift ± 4, $up - down$ shift $\pm$ 3, and $bicubic - downsampled$ method. No rotation was apply in this degradation method.

To complete the synthesized multi-image, each T-generated image is registered using the first image as a reference, which has no shift. The registration is done using an efficient sub-pixel image translation by cross-correlation to have real simulated subpixel shifts with respect to each other, as it would be generated due to, e.g., camera motion, providing different LR samplings of the underlying scene, registered patch examples are depicted in Fig. 3. Reflect padding is used to complete the dismissed pixels. The synthesized dataset is saved in npy files to be loaded during the training process. The data format of the images are in uinit8, and each image is normalized between [-1,1] at the beginning, and after passing the network, they are denormalized. No data augmentation was used.

4.2 Training

In all convolutional layers, on both paths of the network 32 filters, and a kernel size of 3×3 are set. The reduction factor in the attention blocks is set to 8. The number of times that 3D Attention Block is repeated

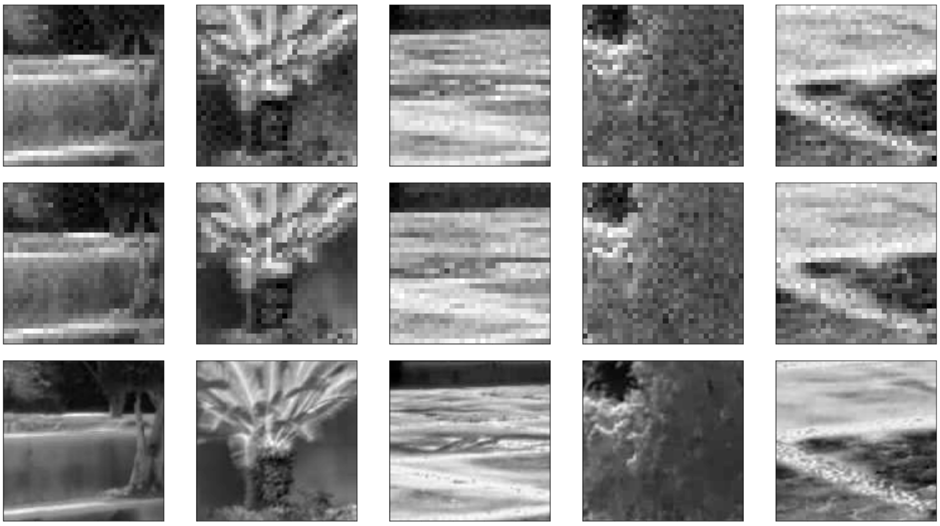

Figure 3: Examples of the patch image registration process. $(top - rows)$ represent different generated LR patches—synthesized images. $(bottom - row)$ show the corresponding HR image patch.

has been set to 12 (lower value causes a loss of performance, higher value increases the number of parameters unnecessarily). In total, the network has less than 750K parameters.

For training, patches of 32×32 pixels, with an overlap of 22 pixels, are extracted from each LR image, giving more than 23K patches for training and 1.2K for validation. An initial learning rate of 0.0005 and Adam loss function optimization is used. To learn the end-to-end mapping process, L_1 and SSIM losses are considered by minimizing their values between the generated and the ground truth images. The proposed architecture has been trained in a NVIDIA Titan X mounted in a workstation with 128GB of RAM. Python programming language and Tensorflow 2.0 library are used. The model is trained for 50 epochs, taking less than 24 hours.

4.3 Results

The standard fidelity based metrics PSNR and SSIM measures are used for testing and validating the proposed model, which consists in evaluating the SR generated from the multi noisy down-sampled image with the corresponding HR image, as shown below:

$$R = \frac{1}{N}\sum_{1}^{N} eval\,(HR, SR(LR_t)) \quad (3)$$

where *eval* is PSNR and SSIM measures metrics separately calculated, SR is the super-resolution generated image from the t multi-image LR noise inputs, and HR represents the corresponding GT image. N is the number of validation images.

Table 1: Results from the proposed multi-image SR approach, and state-of-the-art SISR approaches from PBVS 2021 Challenge (Rivadeneira et al., 2021)

Method	PSNR	SSIM
SVNIT_NTNU-1 Team	<u>30.70</u>	**0.9290**
SVNIT_NTNU-2 Team	30.69	<u>0.9288</u>
SVNIT_NTNU-3 Team	30.59	0.9254
ISESL-CSIO Team	30.39	0.8992
CVS Team	29.21	0.9032
Current work	**32.99**	0.9236

The metrics mentioned above to evaluate the results are: $i)$ Peak Signal-to-Noise Ratio (PSNR), which is commonly used to measure the reconstruction quality of lossy transformations; and $ii)$ Structural Similarity Index Metric (SSIM) (Wang et al., 2004), which is based on the independent comparisons of luminance, contrast, and structure. Due to thermal images being represented in grayscale, these metrics can also be used.

Quantitative results obtained with the proposed architecture are shown in Table 1, together with the SISR results of the state-of-the-art approaches from (Rivadeneira et al., 2021). As it can be appreciated, the proposed architecture achieves a better performance in PSNR with 32.99dB and is highly good on SSIM metrics (just 0.0054 below the best results, SVNIT_NTNU team achieves slightly better results). The SVNIT_NTNU-1 team uses an effective design of ResBlock, that preserves the HF details with fewer parameters and uses channel attention modules; using an exponential linear unit (ELU) activation function to improve learning performance at each

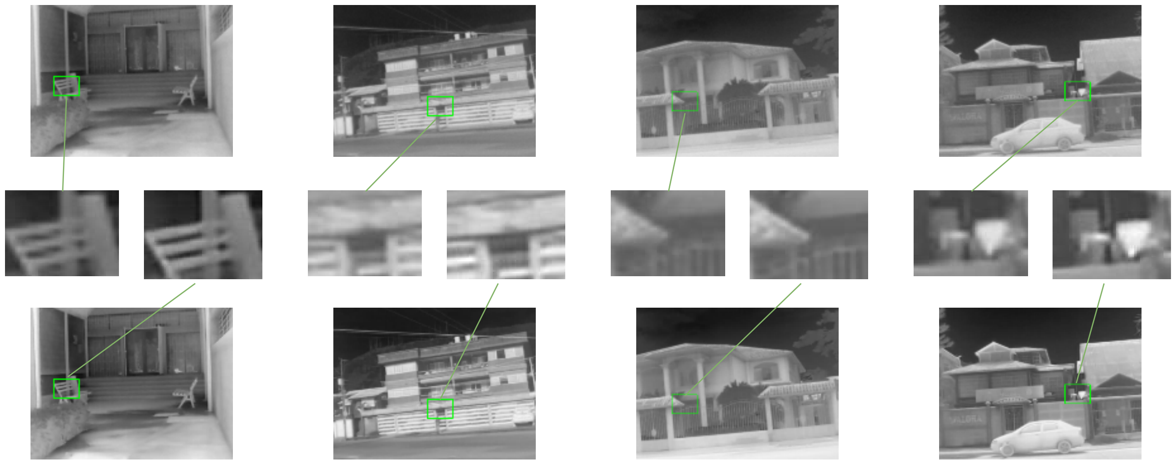

Figure 4: SR results with a $\times 4$ scale factor: $(top-row)$ results from bicubic interpolation; $(bottom-row)$ results from the proposed approach.

layer in an efficient manner. The SVNIT_NTNU-2 team uses a cascade of convolution with Layer attention which includes Residual Blocks using a self-assemble techniques to generate the SR result. Finally, the SVNIT_NTNU-3 team proposes several residual groups to learn complex and rich features from the LR observation, using subpixel convolutions in the up-sampling block, with local and long skip connections.

Qualitative comparison between bicubic interpolations and results from the proposed approach are depicted in Fig. 4. Zoomed patches are provided for a closed inspection showing that the obtained results are sharper and less noisy than bicubic interpolation. This comparison shows that using this architecture to go from a multi-LR to a HR image on the thermal spectrum is possible, even though the network is trained with synthesized images.

5 CONCLUSIONS

This paper presents a novel multi-image super-resolution architecture for thermal images, which exploits recent deep learning advancements. Two attention paths, a 2D and a 3D attentions block mechanisms, are used to train the network to perform SR at a $\times 4$ scale. To train the proposed architecture, synthesized RAW burst noise LR images are generated. As loss functions, L_1 and SSIM are considered. Results obtained with the proposed MISR approach reach the state-of-the-art SISR approaches presented in the PBVS 2021 Challenge when SSIM is considered; on the contrary, when PSNR is considered, results from the proposed approach considerably overcome results from the state-of-the-art approaches.

ACKNOWLEDGEMENTS

This work has been partially supported by the ESPOL project PhysicalDistancing (CIDIS-56-2020) and the “CERCA Programme / Generalitat de Catalunya”. The authors gratefully acknowledge the support of the CYTED Network: “Ibero-American Thematic Network on ICT Applications for Smart Cities” (REF-518RT0559) and the NVIDIA Corporation for the donation of the Titan Xp GPU used for this research. The first author has been supported by Ecuador government under a SENESCYT scholarship contract.

REFERENCES

Bhat, G., Danelljan, M., Van Gool, L., and Timofte, R. (2021). Deep burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9209–9218.

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

Dai, S., Han, M., Xu, W., Wu, Y., and Gong, Y. (2007). Soft edge smoothness prior for alpha channel super resolution. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Deudon, M., Kalaitzis, A., Goytom, I., Arefin, M. R., Lin, Z., Sankaran, K., Michalski, V., Kahou, S. E., Cornebise, J., and Bengio, Y. (2020). Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery. *arXiv preprint arXiv:2002.06460*.

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

Dong, C., Loy, C. C., and Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer.

Farsiu, S., Elad, M., and Milanfar, P. (2004). Multiframe demosaicing and super-resolution from undersampled color images. In *Computational Imaging II*, volume 5299, pages 222–233. International Society for Optics and Photonics.

Freeman, W. T., Jones, T. R., and Pasztor, E. C. (2002). Example-based super-resolution. *IEEE Computer graphics and Applications*, 22(2):56–65.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158.

Han, J., Yang, Y., Zhou, C., Xu, C., and Shi, B. (2021). Evintsr-net: Event guided multiple latent frames reconstruction and super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4882–4891.

Hardie, R. C., Barnard, K. J., Bognar, J. G., Armstrong, E. E., and Watson, E. A. (1998). High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Optical Engineering*, 37(1):247–260.

Herrmann, C., Ruf, M., and Beyerer, J. (2018). Cnn-based thermal infrared person detection by domain adaptation. In *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, volume 10643, page 1064308. International Society for Optics and Photonics.

Huang, Y., Shao, L., and Frangi, A. F. (2019). Simultaneous super-resolution and cross-modality synthesis in magnetic resonance imaging. In *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, pages 437–457. Springer.

Irani, M. and Peleg, S. (1991). Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239.

Ji, H. and Fermüller, C. (2008). Robust wavelet-based super-resolution reconstruction: theory and algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):649–660.

Kim, J., Kwon Lee, J., and Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). 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.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144.

Long, J., Peng, Y., Li, J., Zhang, L., and Xu, Y. (2021). Hyperspectral image super-resolution via subspace-based fast low tensor multi-rank regularization. *Infrared Physics & Technology*, 116:103631.

Milanfar, P., Takeda, H., and Farslu, S. (2011). Kernel regression for image processing and reconstruction. US Patent 7,889,950.

Molini, A. B., Valsesia, D., Fracastoro, G., and Magli, E. (2019). Deepsum: Deep neural network for super-resolution of unregistered multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3644–3656.

Mudunuri, S. P. and Biswas, S. (2015). Low resolution face recognition across variations in pose and illumination. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):1034–1040.

Nguyen, N. L., Anger, J., Davy, A., Arias, P., and Facciolo, G. (2021). Self-supervised multi-image super-resolution for push-frame satellite images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1121–1131.

Peleg, S., Keren, D., and Schweitzer, L. (1987). Improving image resolution using subpixel motion. *Pattern recognition letters*, 5(3):223–226.

Pesavento, M., Volino, M., and Hilton, A. (2021). Attention-based multi-reference learning for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14697–14706.

Qi, H. and Diakides, N. A. (2003). 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. IEEE.

Rasti, P., Uiboupin, T., Escalera, S., and Anbarjafari, G. (2016). Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, pages 175–184. Springer.

Rivadeneira, R. E., Sappa, A. D., and Vintimilla, B. X. (2020a). Thermal image super-resolution: A novel architecture and dataset. In *VISIGRAPP (4: VISAPP)*, pages 111–119.

Rivadeneira, R. E., Sappa, A. D., Vintimilla, B. X., Guo, L., Hou, J., Mehri, A., Behjati Ardakani, P., Patel, H., Chudasama, V., Prajapati, K., et al. (2020b). Thermal image super-resolution challenge-pbvs 2020. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 96–97.

Rivadeneira, R. E., Sappa, A. D., Vintimilla, B. X., Nathan, S., Kansal, P., Mehri, A., Ardakani, P. B., Dalal, A., Akula, A., Sharma, D., et al. (2021). Thermal image super-resolution challenge-pbvs 2021. In *Proceedings* *of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4359–4367.

Rivadeneira, R. E., Suárez, P. L., Sappa, A. D., and Vintimilla, B. X. (2019). Thermal image superresolution through deep convolutional neural network. In *International Conference on Image Analysis and Recognition*, pages 417–426. Springer.

Robinson, M. D., Chiu, S. J., Toth, C. A., Izatt, J. A., Lo, J. Y., and Farsiu, S. (2017). New applications of super-resolution in medical imaging. In *Super-Resolution Imaging*, pages 401–430. CRC Press.

Rossi, M. and Frossard, P. (2018). Geometry-consistent light field super-resolution via graph-based regularization. *IEEE Transactions on Image Processing*, 27(9):4207–4218.

Salvetti, F., Mazzia, V., Khaliq, A., and Chiaberge, M. (2020). Multi-image super resolution of remotely sensed images using residual attention deep neural networks. *Remote Sensing*, 12(14):2207.

Shamsolmoali, P., Zareapoor, M., Jain, D. K., Jain, V. K., and Yang, J. (2019). Deep convolution network for surveillance records super-resolution. *Multimedia Tools and Applications*, 78(17):23815–23829.

Song, D., Wang, Y., Chen, H., Xu, C., Xu, C., and Tao, D. (2021). Addersr: Towards energy efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15648–15657.

Tsai, R. (1984). Multiframe image restoration and registration. *Advance Computer Visual and Image Processing*, 1:317–339.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0.

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., et al. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Wronski, B., Garcia-Dorado, I., Ernst, M., Kelly, D., Krainin, M., Liang, C.-K., Levoy, M., and Milanfar, P. (2019). Handheld multi-frame super-resolution. *ACM Transactions on Graphics (TOG)*, 38(4):1–18.

Yang, J., Wright, J., Huang, T. S., and Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873.

Yao, T., Luo, Y., Hu, J., Xie, H., and Hu, Q. (2020). Infrared image super-resolution via discriminative dictionary and deep residual network. *Infrared Physics & Technology*, 107:103314.

Zhang, K., Zuo, W., Gu, S., and Zhang, L. (2017). Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938.

Zhang, L., Zhang, H., Shen, H., and Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859.