LaMAR: Laplacian Pyramid for Multimodal Adaptive Super Resolution (Student Abstract)

Aditya Kasliwal, Aryan Kamani*, Ishaan Gakhar*, Pratinav Seth, Sriya Rallabandi

Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal - 576104, India.
kasliwaladitya17@gmail.com, aryankamani@gmail.com, ishaangakhar04@gmail.com, seth.pratinav@gmail.com, sriyarallabandi@gmail.com

Abstract

Recent advances in image-to-image translation involve the integration of non-visual imagery in deep models. Non-visual sensors, although more costly, often produce low-resolution images. To combat this, methods using RGB images to enhance the resolution of these modalities have been introduced. Fusing these modalities to achieve high-resolution results demands models with millions of parameters and extended inference times. We present LaMAR, a lightweight model. It employs Laplacian image pyramids combined with a low-resolution thermal image for Guided Thermal Super Resolution. By decomposing the RGB image into a Laplacian pyramid, LaMAR preserves image details and avoids high-resolution feature map computations, ensuring efficiency. With faster inference times and fewer parameters, our model demonstrates state-of-the-art results.

Introduction

Non-visual sensors often produce low-resolution images, and high-resolution versions are expensive. Enhancing the resolution of non-visual images, such as thermal images using RGB images, is termed Guided Thermal Super Resolution. Combining data from multiple sensors, like RGB and thermal, offers comprehensive insights. However, challenges like content distortions, retaining edge information, and avoiding redundant artefacts from RGB images persist.

This paper introduces LaMAR: Laplacian Pyramid for Multimodal Adaptive Super Resolution. Our approach enhances the resolution of thermal images by using key attributes of RGB images. The model uses a lightweight neural network with cascading residual blocks, ensuring robustness and computational efficiency. The use of Laplacian pyramids is rooted in the similarity of edge maps of RGB and High-Resolution Thermal images (HR). Our model addresses Guided Thermal Super Resolution for both aligned and misaligned situations. The model's adaptability and state-of-the-art results in various settings highlight its potential in super-resolution.

The use of CNNs for Super Resolution (SR) was pioneered with SRCNN (Dong et al. 2014), which introduced the concept of end-to-end mapping. Later, RGAN (Ledig et al. 2016) incorporated adversarial training, pushing the boundaries of photo-realistic results. The potential of merging modalities in Guided Thermal Super Resolution and robustness against missing modality was demonstrated by CoReFusion (Kasliwal et al. 2023); however, the high number of parameters and complex architecture results in greater computational load. Laplacian Image Pyramids, as demonstrated in LapSRN (Lai et al. 2017), excel in preserving image edges and overall structure during progressive image reconstruction.

Proposed Method

Our proposed model, LaMAR, is an adaptive, multimodal super-resolution model which utilizes Laplacian pyramids for Guided Thermal Super Resolution on both aligned and misaligned datasets. It demonstrates high perceptual quality and faithful reconstruction with a novel loss while achieving state-of-the-art results with only 612K trainable parameters. The architecture of LaMAR (Fig: 1) consists of two branches:

- A lower branch, focused on extracting features from the low-resolution thermal image, I_{L}.
- An upper branch, responsible for extracting features from the fusion of the RGB and Thermal images.

A high-resolution, grayscaled RGB image is decomposed into a Laplacian pyramid of depth 2. The low-resolution thermal image (LR) replaces the residual of the pyramid. This image passes through the lower branch, which consists of Conv layers, InstanceNorm, 3 Residual Blocks, and leaky ReLU layers. The LR image is combined with the transformed image and passed through a 'tanh' activation. The result is upsampled by x2 and concatenated with the second-last layer of the Laplacian pyramid, forming input for the upper branch.

The upper branch includes a Conv layer serving as a realignment layer for robustness against misalignment, followed by 5 residual blocks, Conv, and Leaky ReLU layers. We calculate a mask on the low-resolution thermal image to avoid informational aberrations and progressively upsample it for other low-resolution components, multiplied with each level of the original pyramid to generate high-resolution components.

*These authors contributed equally.

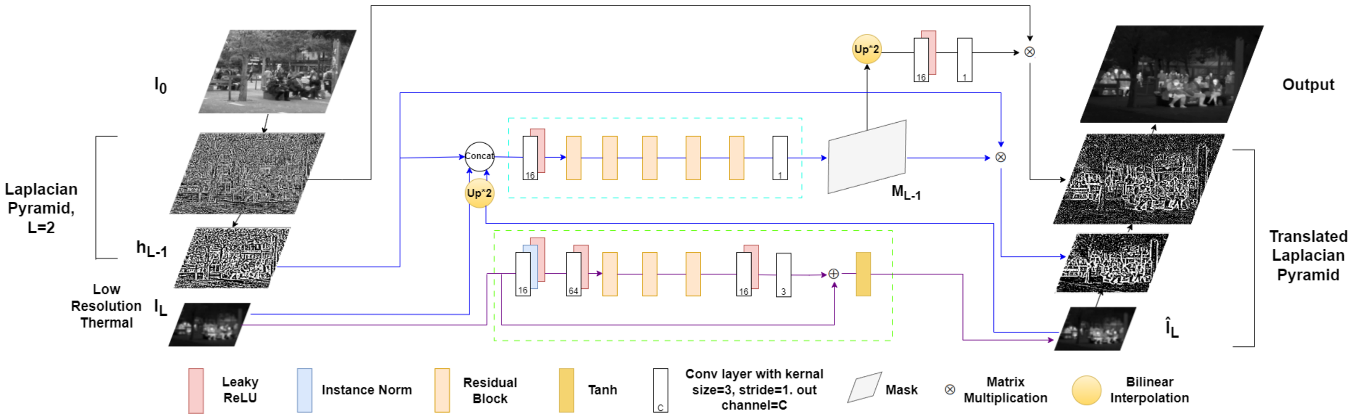

Figure 1: Proposed LaMAR model architecture. I_0 refers to the gray high-resolution RGB image. The given figure is for Pyramid Depth 2 (L=2, where L is the depth of the Laplacian Pyramid), but our model can cater to further downsampling by means of additional layers like L=3 for 8x downsampling

Our final loss function comprises a weighted addition of Mean Squares Error (MSE) Loss and (Generative Adversarial Network) GAN loss. This combination ensures high qualitative and quantitative performance across datasets. The equation for the Generator is given as:

$$\mathbb{E}_{I_0 \sim p_{\text{data}}(I_0)} \left[D(G(I_0) - 1)^2 \right] \tag{1}$$

The equation for the Discriminator is given as:

$$\mathbb{E}_{I_0 \sim p_{\text{data}}(I_0)} \left[D(G(I_0) - 1)^2 \right] + \quad \mathbb{E}_{I_0 \sim p_{\text{data}}(I_0)} \left[D(G(I_0))^2 \right] \tag{2}$$

The weighted addition is enabled by a hyperparameter λ, which optimizes the tradeoff between spatial reconstruction and perceptual quality.

$$Loss = \lambda L_{\text{MSELoss}} + L_{\text{GANLoss}} \tag{3}$$

Experimental Results and Analysis

Our Experiments are run on 2 benchmark datasets: ULB17-VT and the FLIR-ADAS dataset.

We downsample the HR Thermal image by 4x to produce the LR Thermal for the FLIR dataset. We also conducted experiments on the ULB17-VT dataset with various degrees of misalignment. We observed that 5 and 3 residual blocks in the upper and lower branches, respectively, aid in learning structural and spatial features and finer details, ensuring limited parameter count and a quick inference period.

Method	**PSNR (dB)**	**SSIM**
VTSRGAN	27.988	0.8202
VTSRCNN	27.968	0.8196
CMSR	29.928	0.882
LaMAR	**33.635**	**0.9102**

Table 1: Mean PSNR/SSIM of various models on the ULB17-VT dataset. The best results are bolded, and the second-best results are underlined. We demonstrate state-of-the-art results with significantly less trainable parameters.

A learning rate of 10^{-4} for 300 epochs and batch size 12 is used for the ULB17-VT dataset (Almasri and Debeir 2018), and 100 epochs and batch size 16 is used for the FLIR-ADAS dataset. The weight of MSELoss in the final loss, λ, is set to 2000. We use PSNR and SSIM to evaluate the performance of our model. Comparisons of our model with other models on the ULB17-VT dataset are evident in Table 1: A drawback of our model was observed in the FLIR-ADAS dataset. We observed state-of-the-art PSNR but comparable SSIM, as mentioned in Table 2. This is due to the prevalence of night imagery in the dataset. Our model relies on explicit edge guidance, missing in the Laplacian pyramids of night images, which results in lower perceptual quality of the super-resolved images. We plan to address this problem in the future by making provisions for feature extraction in low-light scenarios.

Method	**PSNR (dB)**	**SSIM**
PixTransform	24.84	0.787
Deep-ISTA	25.86	0.828
PAG-SR	29.56	0.912
LaMAR	**30.68**	**0.701**

Table 2: Mean PSNR/SSIM of various models on the FLIR-ADAS dataset.

Conclusion

We propose LaMAR, a lightweight model that uses Laplacian image pyramids to improve the resolution of thermal images. The model works by decomposing the original RGB image into a Laplacian pyramid, which preserves finer image details. This bypasses the need for heavy, high-resolution feature map computation, resulting in a compact and efficient model. LaMAR is able to produce state-of-the-art results in daylight conditions while still being quick and lightweight (612K parameters), making it suitable for real-world applications.

Acknowledgments

We would like to thank Mars Rover Manipal, an interdisciplinary student project team of MAHE, for providing the necessary resources for our research. We are grateful to our faculty advisor, Dr Ujjwal Verma, for providing the necessary guidance.

References

Almasri, F.; and Debeir, O. 2018. Multimodal Sensor Fusion In Single Thermal image Super-Resolution. *arXiv preprint arXiv:1812.09276*.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38: 295–307.

Kasliwal, A.; Seth, P.; Rallabandi, S.; and Singhal, S. 2023. CoReFusion: Contrastive Regularized Fusion for Guided Thermal Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 507–514.

Lai, W.-S.; Huang, J.-B.; Ahuja, N.; and Yang, M.-H. 2017. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5835–5843.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Aitken, A. P.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2016. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105–114.