SAR Super-Resolution Using Capella Satellite Imagery

Sewoong Ahn [1], Sohee Son [1], Dongjun Lee [1], Hyunguk Choi[1], Taegyun Jeon [1]
{anse3832, shson, djlee, hyunguk, tgjeon}@si-analytics.ai

[1] SI Analytics

Keywords: Synthetic Aperture Radar (SAR), Deep Learning, Super-Resolution, Capella Satellite, Satellite Imagery

Abstract

Electro-Optical (EO) images are limited in that they can only be captured during daylight and under clear weather conditions, which restricts their usability in certain environments. In contrast, Synthetic Aperture Radar (SAR) images have the distinct advantage of being able to capture high-quality data regardless of the time of day or weather conditions. This makes SAR images highly valuable across various fields such as national defense, remote sensing, and disaster monitoring. However, despite their advantages, SAR images often suffer from lower resolution due to artifacts like speckle noise, which can significantly degrade image quality.
To address this issue, numerous efforts have been made to enhance the resolution of SAR images through the use of SAR Super-Resolution (SR) techniques. However, research in this area is limited, primarily due to the high cost associated with acquiring real SAR data. Some existing studies in SAR SR have even resorted to using synthetic images that combine speckle noise with regular camera images, which do not accurately represent real SAR data. In response to this, our paper proposes a solution by constructing a dataset based on actual Capella SAR satellite imagery and introducing a novel SAR Super-Resolution model.
For our experiments, we collected real SAR images from the Capella satellite, used them as high-resolution references, and generated low-resolution counterparts by down-sampling. By modifying state-of-the-art image restoration models for the SR task, we demonstrate through a series of experiments that our proposed model outperforms existing SR methods in both quantitative and qualitative assessments.

1. Introduction

1.1 Differences between Electro-Optical (EO) images and SAR images

Electro-Optical (EO) imagery has a significant limitation in that it can only be captured under specific environmental conditions—specifically during daylight hours and when the weather is clear. This dependency on external lighting and atmospheric conditions restricts its usability in scenarios where continuous monitoring or data collection is required. On the other hand, Synthetic Aperture Radar (SAR) imagery offers a substantial advantage because it can be acquired regardless of the time of day or prevailing weather conditions. This capability makes SAR imaging a highly valuable tool in various applications, including national defense, remote sensing, disaster monitoring, and many other fields that require reliable and consistent image acquisition.

1.2 Problems of other SAR SR studies

Despite the numerous advantages of Synthetic Aperture Radar (SAR) imagery, it still has certain limitations, particularly in terms of resolution. Various artifacts, such as speckle noise, can degrade image quality, making it challenging to extract fine details from SAR images. To address this issue, researchers have explored techniques for enhancing SAR image resolution, commonly referred to as SAR Super-Resolution (SR). However, due to the high cost associated with acquiring real SAR data, only a limited number of studies have been conducted in this area. In fact, some existing research on SAR SR relies on artificially generated images that simulate SAR characteristics by introducing speckle noise into conventional optical images, rather than utilizing actual SAR imagery. To overcome this limitation, this paper constructs a dataset using real SAR images captured by the Capella satellite and proposes a novel SAR Super-Resolution model to enhance the resolution and quality of SAR imagery.

1.3 Contribution Points

The key contributions of our paper can be summarized as follows:
- First, unlike many previous studies that rely on synthetic SAR images, we use real SAR imagery to ensure that our model is trained and evaluated on data that accurately reflects real-world conditions, thereby enhancing the authenticity of our results.
- Second, we present a comprehensive set of both quantitative and qualitative experiments, demonstrating the effectiveness of our approach from multiple perspectives and providing a thorough evaluation of its performance.
- Lastly, we show that our proposed model outperforms existing state-of-the-art Super-Resolution (SR) models, achieving superior results in terms of both objective metrics and visual quality, thus establishing the efficacy and robustness of our method in comparison to other leading techniques in the field

1.4 Contents

- Section 2 provides an overview of various studies related to deep learning-based Super-Resolution (SR) techniques, extending to those specifically focused on SAR SR, highlighting the key advancements and methodologies in these areas.
- Section 3 introduces the baseline model, detailing its structure and performance, and also explains the modifications made to adapt it for the specific challenges of SAR SR, ensuring it is suited for high-quality SAR image enhancement.
- Section 4 presents a comprehensive analysis of the experimental results, demonstrating the superior performance of

our proposed model compared to existing approaches, with a focus on key metrics and qualitative assessments.
- Section 5 offers a conclusion that summarizes the findings of the study and outlines potential directions for future research and improvements in the field of SAR Super-Resolution.

2. Related Works

In this section, we will offer a detailed overview of the development and evolution of Super-Resolution (SR) techniques, with a particular focus on how deep learning approaches have contributed to significant advancements in the field. We will explore the key milestones in SR research, shedding light on the progression of methodologies and the innovations that have driven improvements in image resolution. Additionally, we will delve into the specific challenges and difficulties encountered when applying SR techniques to Synthetic Aperture Radar (SAR) images. These challenges include the unique characteristics of SAR data, such as noise, speckle, and resolution limitations, which make SAR SR a particularly complex task. By addressing these issues, we aim to highlight the distinctive obstacles faced in SAR Super-Resolution and the specific considerations that must be taken into account when developing models for this domain.

2.1 Deep learning based Super-Resolution

(Dong, 2014) were the first to introduce a deep learning-based approach to Super-Resolution, presenting a model known as SRCNN. This pioneering method utilizes convolutional neural networks (CNNs) to establish a direct end-to-end mapping between low-resolution (LR) and high-resolution (HR) images, effectively learning how to enhance image quality through data-driven techniques. The architecture of SRCNN is structured into three distinct stages: patch extraction, where small regions of the input image are extracted for processing; non-linear mapping, which applies deep learning techniques to transform these patches into higher-quality representations; and reconstruction, where the enhanced patches are merged to produce the final high-resolution output. Following the introduction of SRCNN, numerous subsequent studies have been inspired by this three-stage framework, leading to the development of more advanced and refined deep learning-based Super-Resolution models.
Following the introduction of SRCNN, subsequent research efforts primarily focused on increasing the size and complexity of Super-Resolution (SR) models to achieve better performance. (Kim, 2016) introduced the VDSR model, which was inspired by the VGG-Net (Simonyan, 2014) architecture originally designed for ImageNet classification. Unlike SRCNN, which utilized only 3 weight layers, VDSR significantly expanded the network depth by successfully incorporating 20 weight layers, demonstrating the effectiveness of deeper networks in SR tasks. Later, (Lim, 2017) proposed the EDSR model, which leveraged residual blocks to improve stability during the training of large-scale SR models. Their work was an extension of SRResNet (Ledig, 2017), and they further enhanced EDSR’s performance by removing batch normalization layers, which were present in SRResNet. Building on these advancements, (Zhang, 2018) introduced RCAN, which utilized residual-in-residual structures, enabling the stacking of an even greater number of deep layers. Additionally, they incorporated channel attention mechanisms to capture and exploit the interdependencies between different channels, further enhancing the model’s ability to refine image details effectively.
While some Super-Resolution (SR) studies have primarily focused on reducing pixel-wise differences between generated images and their corresponding ground-truth images, others have shifted their attention toward generating perceptually high-quality images that appear more visually realistic. (Wang, 2018) introduced the ESRGAN model, which leveraged generative adversarial networks (GAN) to enhance the perceptual quality of Super-Resolution outputs. Their work was an improvement upon SRGAN (Ledig, 2017), and they refined ESRGAN by eliminating batch normalization layers, which helped enhance the model's overall performance. Additionally, they incorporated residual-in-residual dense blocks, allowing for the construction of deeper networks capable of capturing more complex image details and textures.
Subsequently, numerous researchers began exploring the use of transformer models in Super-Resolution (SR) research. (Liang, 2021) introduced SwinIR, which leveraged the Swin Transformer—an architecture commonly utilized in object detection—to enhance SR performance. Additionally, other researchers have focused on reducing the computational cost associated with self-attention mechanisms, aiming to make transformer-based SR models more efficient and scalable.
In recent times, a number of innovative deep learning models, such as the diffusion model (Rombach, 2022) and Mamba (Cu, 2023), have been introduced, gaining significant attention in the research community due to their advanced capabilities. As a result, many researchers are increasingly adopting these cutting-edge models in an effort to achieve state-of-the-art performance in Super-Resolution (SR) tasks, pushing the boundaries of image restoration and enhancement by leveraging the strengths of these new architectures.

2.2 SAR Super-Resolution

In recent years, an increasing number of studies have investigated the application of deep learning techniques to enhance the resolution of Synthetic Aperture Radar (SAR) images, commonly referred to as SAR Super-Resolution (SAR SR). Among these studies, (Mastriani, 2016) proposed a wavelet transform-based method designed to improve both SAR super-resolution and despeckling, aiming to reduce noise while enhancing image clarity. Similarly, (Wu, 2016) employed a back-propagation neural network alongside a non-local mean filter to achieve super-resolution in SAR images, leveraging deep learning to refine image quality.
Although these research efforts have contributed valuable insights into SAR SR, a common limitation among them is their reliance on synthetic SAR images rather than real-world data.

3. Method

In this section, we present a comprehensive explanation of the proposed Super-Resolution (SR) model, detailing both its architecture and the loss function employed during the training process. To begin, we provide an in-depth overview of the baseline model, describing its fundamental structure and the underlying principles that govern its operation. Following this, we elaborate on the specific modifications introduced to tailor the model for Synthetic Aperture Radar Super-Resolution (SAR SR), ensuring its effectiveness in handling the unique characteristics of SAR images.
Furthermore, this section covers the methodology used to obtain real SAR imagery, highlighting the sources of data and the criteria for selecting high-quality SAR images. In addition to data acquisition, we describe the step-by-step process of constructing the dataset for SR, including preprocessing

techniques, the generation of low-resolution and high-resolution image pairs, and the partitioning of data into training and testing sets. By providing these detailed explanations, this section aims to offer a clear understanding of the overall workflow, from data collection to model adaptation, ensuring a solid foundation for the subsequent experimental analysis.

3.1 Model

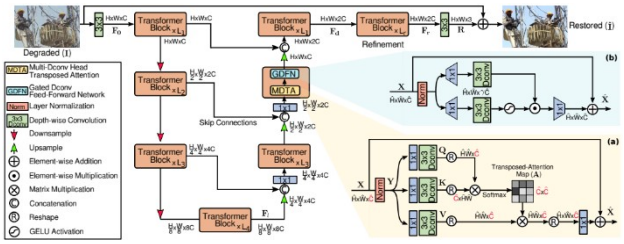


Figure 1. Architecture of Restormer. This image is originated from (Zamir, 2022)

In this study, we utilize the Restormer model (Zamir, 2022) to perform Super-Resolution (SR) on SAR images. Restormer has shown exceptional, state-of-the-art performance across a variety of image restoration tasks, making it a suitable choice for our application. The following sections will provide a detailed explanation of the Restormer model and describe how we applied it specifically to address our SR task.

1) Overall Process: Given an input image $I \in R^{H\times W\times 3}$, the model begins by applying a convolutional layer to extract low-level features. In this context, the image has a height and width denoted as H and W, respectively, and C represents the number of channels in the image. These low-level features $F_0 \in R^{H\times W\times C}$ are then passed through a series of 4-level encoder-decoder layers, which are based on the architecture of UNet (Weng, 2021). This process transforms the features into deeper, more abstract representations $F_d \in R^{H\times W\times 2C}$. Each level within the encoder-decoder structure consists of multiple transformer blocks, and the number of blocks increases progressively as we move from the higher levels of the encoder to the lower levels of the decoder.
In the encoder portion, starting from the high-resolution input, the spatial size of the image is progressively reduced, while the channel dimension is expanded to capture more complex features. In contrast, the decoder takes these low-resolution latent features and works to increase the spatial size, while simultaneously reducing the number of channels to refine the image details. For down-sampling the features in the encoder and up-sampling them in the decoder, pixel-unshuffle and pixel-shuffle operations (Shi, 2016) are applied, respectively. These operations play a key role in maintaining the integrity of the image resolution during the transformation process. More detailed explanations of each of these processes can be found in (Zamir, 2022), where the specifics of the architecture and operations are further elaborated.

2) Transformer Block: The primary computational bottleneck in transformer models arises from the intensive self-attention computation, which becomes particularly challenging as the model size increases. To address and alleviate this issue, the authors introduce a novel approach called multi-Dconv head transposed attention (MDTA), which is illustrated in Figure 1(a). The central innovation behind MDTA is the application of self-attention not along the spatial dimension, as is traditionally done, but rather across the channel dimension. This key modification allows the model to focus its attention on the relationships between different channels of the feature map, rather than on spatial interactions between pixels. By doing so, MDTA computes the cross-covariance across channels, which is then used to generate an attention map that implicitly captures the global context of the input data. This method significantly reduces the computational complexity compared to traditional self-attention mechanisms, making it more efficient while still preserving the ability to capture long-range dependencies and global context within the model.

Starting from a normalized tensor $Y \in R^{H\times W\times C}$, the MDTA mechanism first generates three key components: the query (Q), key (K) and value (V) projections. These projections are created through a two-step process that begins with the application of 1×1 convolutions, which are designed to aggregate the pixel-wise context across the different channels of the tensor. This is followed by 3×3 depth-wise convolutions, which serve to capture the spatial context along the channel dimension, thus encoding both channel-wise and spatial features within the tensor. Once these projections are generated, the next step involves reshaping the query and key projections in a specific way that allows their dot-product interaction to produce a transposed-attention map, denoted as A. Unlike traditional self-attention mechanisms, which typically produce a spatial attention map of size $R^{HW\times HW}$, the transposed-attention map A is of a different size $R^{C\times C}$, specifically designed to represent the interactions across channels rather than across spatial locations. This innovative method helps to significantly reduce computational complexity while still capturing important relationships within the data. For a more in-depth understanding of the Transformer Block and its architecture, further details can be found in (Zamir, 2022).

3) Gating mechanism: The gating mechanism in our model is designed as the element-wise product of two separate parallel paths that each consist of linear transformation layers. One of these paths is further processed by the GELU non-linearity (Hendrycks, 2016), which helps to introduce non-linearity and improve the model’s ability to capture complex patterns. This approach enables the gating mechanism to selectively control the flow of information across different stages of the model. As a result, the GDFN, as illustrated in Figure 1(b), plays a crucial role in regulating the information flow through the various hierarchical levels within our pipeline. By doing so, the GDFN ensures that each level can focus on processing specific, fine-grained details that complement the information captured by the other levels, facilitating a more effective and comprehensive representation of the input data. For a more detailed understanding of the gating mechanism and its implementation, please refer to the explanation provided in (Zamir, 2022).

4) Changes from Restormer: One of the limitations of the Restormer model is that its output dimension is identical to the input dimension, meaning that it cannot directly be applied to Super-Resolution (SR) tasks without modification. To overcome this, we made adjustments to the original Restormer model to make it suitable for SR applications. Specifically, we added a nearest neighbor up-sampling technique after the final convolutional layer of the Restormer network, which serves to increase the resolution of the input images. This nearest up-sample module, which enlarges the input images by replicating pixel values, is a common technique that has been widely used in various SR studies and has proven effective for enhancing image resolution. By incorporating this module, we enable the Restormer model to generate higher-resolution images, making it capable of handling SR tasks effectively.

3.2 Loss function

To assess the performance of the model, we use structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) as evaluation metrics. To facilitate the training of the model, we employ the L1 loss function. The L1 loss function is mathematically defined as follows:

$$L_1(\Theta) = \frac{1}{n}\sum_{i=1}^{n} |F(X_i; \Theta) - Y_i|, \quad (1)$$

where n = number of training images
X, Y = low- and high-resolution image
$F(\cdot)$ = SR model
Θ = SR model parameters

In Section 4, we will present both quantitative and qualitative results to evaluate the performance of our model using the loss function discussed earlier. The quantitative results will include numerical metrics, while the qualitative results will focus on visual assessments of the model's output. This will help provide a comprehensive evaluation of our model's effectiveness.

3.3 Dataset

In this subsection, we will provide a detailed explanation of the process for obtaining real SAR imagery, as well as the steps involved in preparing the data for use in our experiments. This includes the methods used to acquire authentic SAR images and the procedures followed to ensure that the data is properly organized and formatted for optimal use in the training and evaluation of our model.

1) SAR Satellite: For constructing our dataset, we utilize authentic Synthetic Aperture Radar (SAR) images that are captured by the Capella satellite. These real-world SAR images, obtained directly from the Capella satellite, serve as the foundation for our dataset, ensuring that the data is not artificially generated or simulated. By using actual SAR imagery, we aim to capture the true characteristics and complexities of SAR data, which allows us to create a more realistic and accurate dataset for our study. This approach provides a more reliable representation of the conditions and challenges typically encountered in SAR image processing, as opposed to relying on synthetic or overly simplified data.

2) Dataset preparation process: Initially, we gather high-resolution Capella SAR images from the Capella SAR archives. These images, which have a ground sampling distance (GSD) of 0.5 meters, serve as the high-resolution (HR) images in our dataset. To create the corresponding low-resolution (LR) images, the HR images are bicubically downsampled. Afterward, we perform patch extraction on both the LR and HR images, generating HR patches of size 1024x1024 and LR patches of size 512x512. These pairs of SAR patches are then split into training and testing datasets, with an 80:20 ratio (training: test). As a result, the final dataset consists of 22,951 pairs for training and 5,760 pairs for testing.

3.4 Implementation details

Our proposed model utilizes a 4-level encoder-decoder architecture, where each level is designed with a specific configuration to enhance the Super-Resolution process. At each level, the number of Transformer blocks varies as follows: [4, 6, 6, 8] from level 1 to level 4. Additionally, the attention heads in the Multi-Dimensional Transformer Attention (MDTA) mechanism increase progressively at each level, with values of [1, 2, 4, 8], allowing for more complex attention patterns as the network deepens. The number of channels at each level also increases as the model progresses through the layers, with the values being [64, 128, 256, 512] for each respective level.
In the refinement stage, which aims to fine-tune the output of the encoder-decoder, we include 4 specialized blocks designed to improve image quality. The Channel Expansion Factor (γ) in the Generative Deep Feature Network (GDFN) is set to 4, allowing for enhanced feature learning and image resolution improvements.
To train the model, we use the AdamW optimizer with the following parameters: β1=0.9, β2=0.999, and a weight decay of $1e^{-4}$. The training is conducted over 300,000 iterations, starting with an initial learning rate of $2e^{-4}$, which is gradually reduced to $1e^{-6}$ through cosine annealing, as described in (Loshchilov, 2016). During training, the input patch size is set to 64x64, and the batch size is 12 per GPU. For the training process, we utilize a total of 4 A6000 GPUs, while a single A6000 GPU is employed during testing.

4. Experiments

In this Section, we will present both quantitative and qualitative results that demonstrate the performance of our proposed method when evaluated using this specific loss function. These results will be analyzed to provide a comprehensive understanding of how effectively the model performs in terms of numerical metrics as well as visual quality, offering a well-rounded evaluation of its capabilities. The quantitative results will include objective performance measures, while the qualitative results will focus on the perceptual aspects of the images generated by the model, highlighting any improvements or notable features observed through visual inspection.

4.1 Quantitative performance evaluation

We utilize Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) as the primary metrics for evaluating the performance of our model. These metrics are essential for assessing the quality of the generated images in comparison to the ground-truth images. The datasets used for both training and testing are the same as those described in Section 3.3. As shown in Table 1, our proposed model achieves state-of-the-art performance in terms of both SSIM and PSNR, outperforming previous methods. In the table, the bolded values represent the best performance across all models tested. For the purpose of performance comparison, we have included several well-known models in the field, including EDSR (Lim, 2017), ESRNet (Wang, 2018), RCAN (Zhang, 2018), and SwinIR (Liang, 2021), which serve as benchmarks to demonstrate the effectiveness of our approach.

Table 1. Performance Comparison for Various Super-Resolution models

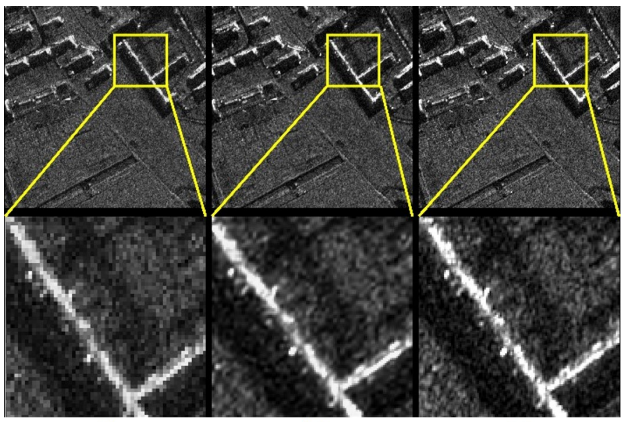
	EDSR	ESRNet	RCAN	SwinIR	Proposed
PSNR	27.07	27.00	27.09	27.08	**27.16**
SSIM	0.8888	0.8878	**0.8896**	**0.8896**	0.8891

As can be seen in Table 1, our proposed model achieves the highest score when evaluated based on Peak Signal-to-Noise Ratio (PSNR), demonstrating its superior ability to enhance the quality of the generated images in terms of pixel-level accuracy. However, when evaluating using the Structural Similarity Index (SSIM), other models such as RCAN and SwinIR outperform our proposed model by a small margin. Despite this, the difference in SSIM scores between our model and the top-

performing models is relatively minor, indicating that our approach still achieves competitive results in terms of perceptual image quality.

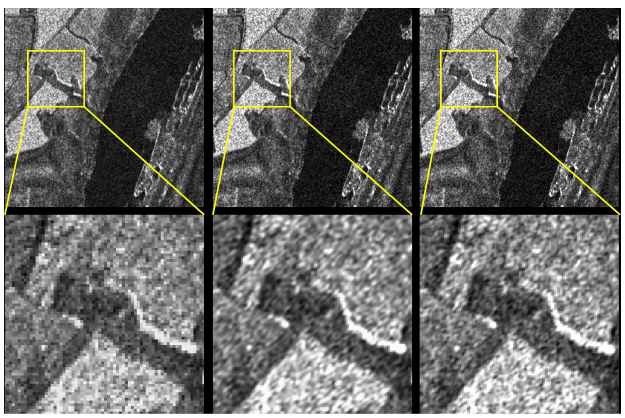
4.2 Qualitative performance evaluation

The figure provided below illustrates the qualitative performance of our proposed SAR Super-Resolution (SR) model. This visual representation highlights how well our model enhances the resolution and quality of SAR images, showcasing improvements in clarity, detail, and overall image quality. By comparing the results with those of existing methods, it is possible to observe the superior performance of our approach, particularly in terms of accurately reconstructing fine details and minimizing artifacts that are often present in lower-resolution SAR images. These qualitative results offer a clear visual validation of the effectiveness of our model in addressing the challenges inherent in SAR Super-Resolution tasks.



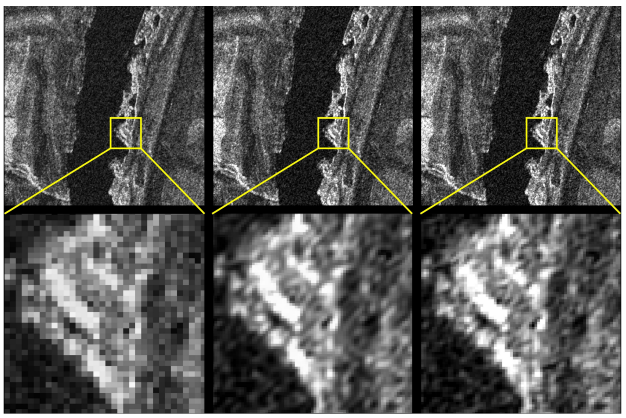
(a) (b) (c)

Figure 2. Original and zoom-in version of SAR images. (a) LR image, (b) Result of proposed SAR SR model, (c) HR image.



(a) (b) (c)

Figure 3. Original and zoom-in version of SAR images. (a) LR image, (b) Result of proposed SAR SR model, (c) HR image.



(a) (b) (c)

Figure 4. Original and zoom-in version of SAR images. (a) LR image, (b) Result of proposed SAR SR model, (c) HR image.

As illustrated in Figure 2, the images exhibit a noticeable improvement in sharpness after applying our proposed SAR Super-Resolution (SAR SR) model. This enhancement is particularly evident as the resolution of the images increases, with more intricate details being successfully reconstructed, resulting in a clearer and more detailed representation of the original scene. The same general trend is also observed in Figure 3 and 4, where the higher the resolution, the more refined the image becomes, showcasing the model's effectiveness in enhancing both the visual clarity and the structural details of the SAR images.

5. Conclusion

In this paper, we gather a diverse set of Capella SAR images to create a comprehensive SAR Super-Resolution (SAR SR) dataset, ensuring that the dataset is representative of real-world SAR imaging conditions. To achieve high performance in SAR SR, we modify a state-of-the-art image restoration model, tailoring it to effectively handle the unique challenges of the SR task. Looking ahead to future work, we aim to eliminate any synthetic processes involved in the creation of high-resolution (HR) and low-resolution (LR) image pairs. Specifically, we plan to directly collect authentic HR and LR pairs, without relying on the conventional method of using bicubic downsampling to generate LR images, thereby ensuring that the data used for training and evaluation is entirely representative of real SAR imagery.

Acknowledgements

Acknowledgements of support for the project/paper/author are welcome. Note, however, that for the paper to be submitted for review all acknowledgements must be anonymized.

References

Dong, C., Loy, C. C., He, K., & Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13* (pp. 184-199). Springer International Publishing.

Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional

networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

Lim, B., Son, S., Kim, H., Nah, S., & 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* (pp. 136-144).

Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." *Proceedings of the European conference on computer vision (ECCV) workshops*. 2018.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R., 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1833-1844).

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

Gu, A., & Dao, T., 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Mastriani, M., 2016. New wavelet-based superresolution algorithm for speckle reduction in SAR images. *arXiv preprint arXiv:1608.00270*.

Wu, Z., & Wang, H., 2016. Super-resolution reconstruction of SAR image based on non-local means denoising combined with BP neural network. *arXiv preprint arXiv:1612.04755*.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022). Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5728-5739).

Weng, W., and X. Zhu INet. "Convolutional networks for biomedical image segmentation., 2021, 9." *DOI: https://doi. org/10.1109/ACCESS* (2021): 16591-16603.

Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Hendrycks, Dan, and Kevin Gimpel. "Gaussian error linear units (gelus)." *arXiv preprint arXiv:1606.08415* (2016).

Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." *arXiv preprint arXiv:1608.03983* (2016).