Improving SWOT images by specular ringing correction with superresolution learning

Laura Manuela Castañeda Medina
Daniel Felipe Torres Robles

March 30, 2025

Abstract

SAR imagery from the SWOT mission plays a vital role in hydrological and oceanographic studies, but its utility is limited by artifacts such as specular ringing. This project investigates a hybrid approach to mitigate these artifacts while enhancing image resolution. We propose a frequency-domain filtering technique that selects and combines spectral subbands to suppress ringing in the original complex SAR images. The resulting low-resolution images were used to train a NAFNet-based superresolution model, with MERLIN-processed SWOT images serving as high-resolution targets. The pipeline demonstrates the potential of subband-based preprocessing for artifact-aware learning in SAR. Our results highlight both the promise and challenges of combining physical insight with data-driven methods for improving the visual and scientific quality of satellite radar imagery.

1 Introduction

1.1 Context and Motivation

The Surface Water and Ocean Topography (SWOT) mission, a joint effort between NASA and CNES, aims to provide unprecedented observations of Earth's surface water and ocean dynamics. By employing a novel SAR-based Ka-band radar interferometer, SWOT captures high-resolution two-dimensional measurements of water elevation, enabling detailed monitoring of rivers, lakes, reservoirs, and coastal oceans [MMB+19]. These observations are vital for understanding climate change impacts, managing freshwater resources, and improving hydrological and oceanographic models.

Synthetic Aperture Radar (SAR) plays a central role in SWOT due to its ability to acquire imagery regardless of cloud cover or solar illumination [CM91]. However, SAR images, especially in missions such as SWOT, often suffer from limitations such as low spatial resolution and artifacts such as specular ringing. These issues can significantly reduce the interpretability of the data, introducing ambiguities in the identification of hydrological structures and degrading the accuracy of scientific analyses.

Correcting artifacts in SAR images while maintaining the spatial resolution is thus a key challenge to fully exploit SWOT's scientific potential. This project address this by proposing a learning-based super-resolution approach, guided by spectral analysis, to reconstruct higher-quality images with reduced ringing artifacts.

1.2 Objectives

The main objective of this project is to mitigate specular ringing artifacts present in the original data and improve the spatial resolution and visual quality of the obtained SWOT SAR images. To achieve this, we develop a spectral-domain pre-processing strategy that filters frequency components responsible for ringing, producing improved low-resolution (LR) inputs for a learning-based super-resolution model. Specifically, we aim to:

- Select subbands in the Fourier domain from original (raw) SWOT images to suppress specular ringing while preserving essential spatial structures.
- Investigate different subband combination strategies, such as minimum energy selection, median fusion, and energy-weighted averaging, to reconstruct artifact-reduced LR inputs.
- Train a deep super-resolution model (NAFNet), using the subband-reconstructed LR images as inputs and the MERLIN-processed high-resolution (HR) images as training targets [Bre+22].
- Evaluate the effectiveness of the overall pipeline using quantitative metrics (PSNR, SSIM) and qualitative visual inspection.

2 Background

The SWOT (Surface Water and Ocean Topography) mission employs an innovative SAR-based Ka-band radar interferometer that captures fine-scale two-dimensional patterns, which are essential for monitoring water dynamics at the global scale [MMB+19]. However, like many SAR systems, SWOT imagery is prone to several limitations.

- **Specular Ringing**: Caused by sharp transitions or strong reflectors, this artifact manifests as oscillatory ripples around bright features, degrading both visual and analytical quality of the images.
- **Low Spatial Resolution**: Depending on the system configuration and observation mode, SWOT SAR images may exhibit insufficient detail, particularly in narrow or heterogeneous areas such as small rivers, reservoirs, or coastal regions.

Traditional filtering and post-processing techniques provide limited capability in addressing these issues simultaneously. Recently, learning-based restoration and super-resolution approaches have gained attention for their ability to learn mappings from low-quality to high-quality imagery using large-scale data [Zha+21].

Recent advances in deep learning, particularly in image restoration and super-resolution, have shown strong capabilities for enhancing degraded imagery in a data-driven manner [Zha+21]. Among these methods, NAFNet [Che+22] has demonstrated excellent performance in tasks such as denoising, deblurring, and super-resolution, using a nonlinear activation-free architecture that balances quality and efficiency.

In the context of the SWOT mission, SAR images are processed using the MERLIN processor [Bre+22], which performs multilook processing and denoising using a spatial adaptive Wiener filter. While MERLIN effectively suppresses speckle and thermal noise, it does not fully remove specular ringing artifacts, particularly near strong reflectors such as land-water interfaces or sharp hydrological features. These residual artifacts degrade the interpretability of the images and can interfere with downstream analysis.

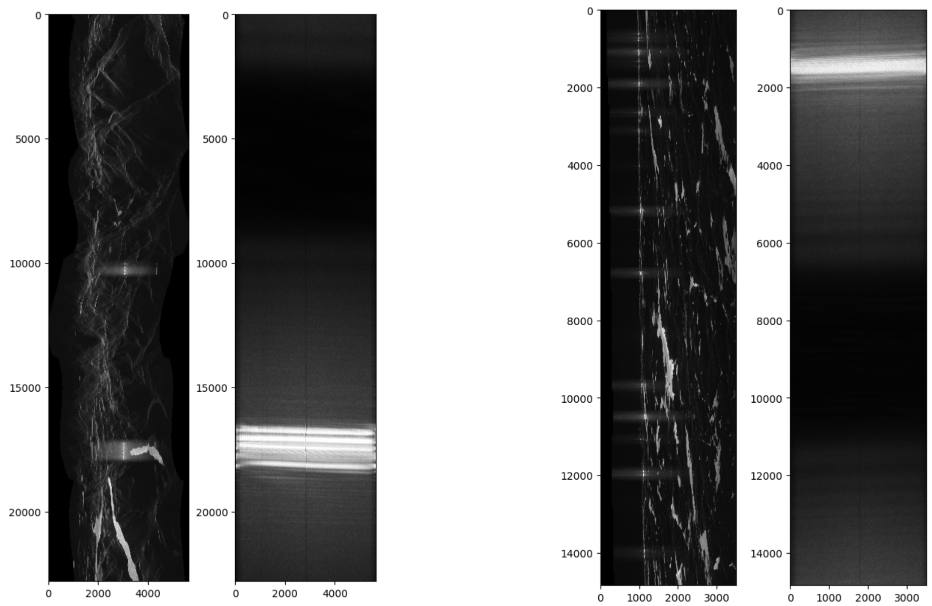
Our approach addresses this issue by identifying and selecting frequency subbands from the original SWOT images that minimize the presence of specular ringing when used to reconstruct the image. We aim to selectively filter spectral components responsible for ringing, using various strategies such as subband selection around spectral peaks, multi-subband combination, and weighted fusion. These filtered images serve as improved inputs to a learning-based super-resolution model, designed to recover fine-scale details while preserving the artifact reduction achieved by spectral filtering.

3 Method

3.1 Spectral Subband Selection

To identify relevant subbands, we first compute the energy profile along the azimuth axis and detect the index of the peak energy. Based on this peak, we select two vertical frequency bands below and two above it, forming rectangular subbands that span the outermost columns of the spectrum. These regions are chosen using configurable parameters: the relative height of the subbands (`df_width_ratio`), their vertical offset (`skip_bands_below` and `skip_bands_above`), and their width (`col_df_ratio`).

During our preliminary analysis, we observed that the location of the spectral energy peak varied across the dataset. As we only had access to 11 SWOT complex SAR images, using a fixed subband was not viable, since it would capture irrelevant frequency regions in many cases. The adaptive strategy ensures that, for each image, we extract frequency content centered around its dominant components, thus maximizing useful information while avoiding the inclusion of noisy or irrelevant bands.



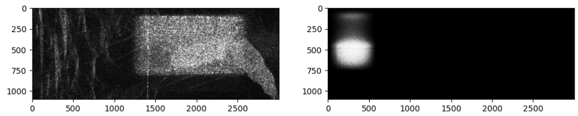
(a) `SLC_024_430_045L_minus_y_pola_V` (b) `SLC_015_001_257L_minus_y_pola_V`

Figure 1: Comparison of spectral energy peak between images of the dataset

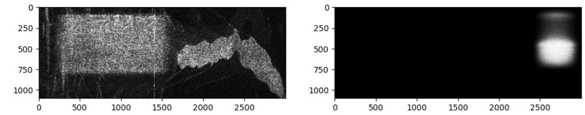
A key insight came from observing the effect of selecting only the leftmost or rightmost columns of the Fourier spectrum. When using only the left outer subband, we found that specular ringing was suppressed on the right side of the spatial-domain image, and vice versa. This suggests that the frequency components responsible for ringing are spatially directional and asymmetric. By combining the left and right subbands, we were able to reduce specular ringing on both sides of the image.

While this approach significantly reduces ringing, it comes at a cost: discarding large portions of the spectrum leads to loss of spatial resolution. To mitigate this, we select not just one subband pair but multiple subbands, both above and below the spectral peak, to retain as much information as possible without reintroducing the ringing artifacts.

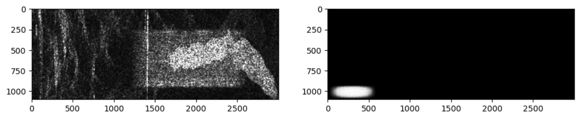
To avoid sharp cutoffs and preserve local consistency, we apply a 2D window (Hamming) to each subband before applying the inverse Fourier transform. This smoothes transitions in the frequency domain and reduces boundary artifacts in the spatial reconstruction.



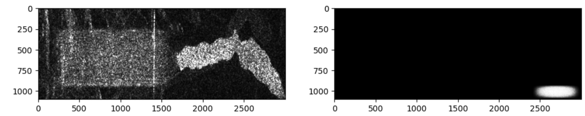
(a) Left-outer subband below the peak



(b) Right-outer subband below the peak



(c) Left-outer subband above the peak



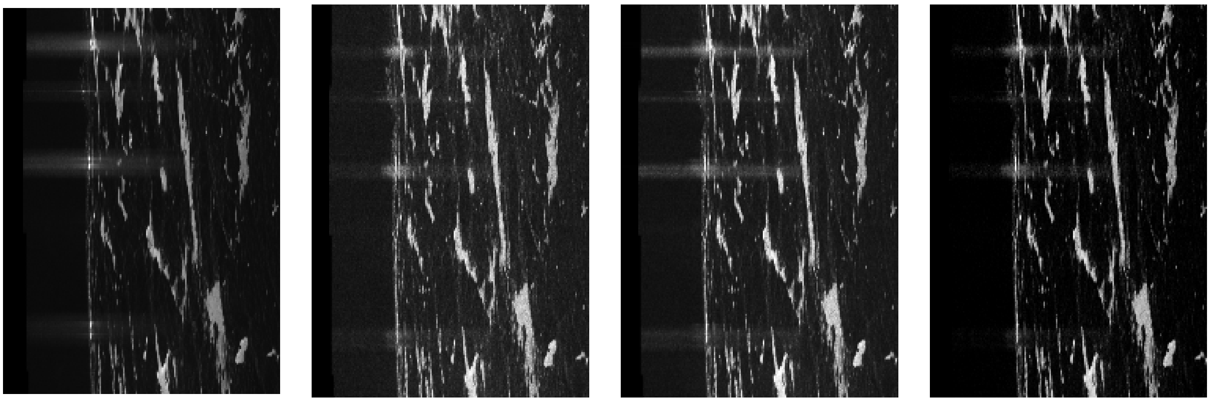
(d) Right-outer subband above the peak

Figure 2: Example of subband selection strategy: two subbands above and below the peak in the azimuth frequency axis, targeting outermost columns.

3.1.1 Subband combination strategies.

After extracting multiple subbands from each image, we tested various methods for combining them into a single low-resolution image:

- `min_energy`: Selects, at each pixel, the minimum value among all subband reconstructions. Although simple, this strategy retains many artifacts and fails to fully suppress ringing.
- **Median**: Performs pixel-wise median filtering across the subbands. This helps reduce noise and outliers, but it can smooth out important features and leave some residual artifacts.
- **AND-based fusion**: Applies a threshold to each subband and retains only the features that are consistently present across all subbands (i.e., common structures). This method proved to be the most effective: it preserves salient features while strongly suppressing specular ringing and spurious noise. It also avoids reinforcing asymmetric artifacts seen in other methods.



(a) (b) (c) (d)

Figure 3: Visual comparison of combination strategies: (a) Input with ringing; (b) min_energy; (c) median; (d) AND-based. The last one achieves the cleanest result.

The spectral subband selection and combination pipeline serves two roles: reducing specular ringing while preserving as much spatial detail as possible. The resulting images are used as low-resolution inputs for the super-resolution learning stage.

3.2 Dataset Construction

The training dataset is composed of pairs of low-resolution (LR) and high-resolution (HR) grayscale SAR images. Each pair corresponds to the same geographical region, extracted from SWOT complex SAR products.

The LR inputs are generated using the subband filtering and combination pipeline described in the previous section. This process applies spectral-domain selection and inverse transformation to the original complex-valued SAR images. The result is a real-valued image with minimized artifacts and a reduced effective resolution, which serves as the input to the super-resolution model.

The HR targets are obtained from the corresponding MERLIN-processed versions of the original images [Bre+22]. These products have undergone multilook denoising and serve as the closest available high-quality reference for training. Importantly, the LR images are aligned with these HR targets at the spatial level.

To ensure pixel-wise correspondence, all LR images are resized to match the spatial dimensions of the HR images using bicubic interpolation. Additionally, all images are normalized to the range $[0, 1]$ after clipping extreme values (e.g., infinities or NaNs) to ensure stable training and meaningful PSNR/SSIM evaluation.

3.2.1 Patch extraction and masking of ringing regions.

During early experiments, we observed that if remains of specular ringing were still present in the LR images, the model would tend to reproduce it and increase it. This suggests that, unless explicitly avoided, the network can learn the artifacts as features of the data distribution.

To prevent this, we manually identified and cropped out the regions of the LR images where specular ringing persisted. This step ensures that the network is trained exclusively on artifact-free inputs, focusing on learning true spatial detail enhancement rather than inadvertently reconstructing unwanted oscillatory patterns.

After cropping the affected regions, the remaining image area was divided into overlapping patches for training.

3.2.2 Memory constraints and patch strategy.

Training deep super-resolution networks on full-size SAR images is computationally infeasible. In addition to removing ringing regions, memory limitations of the available GPUs—particularly when training in mixed precision forced us to limit both the patch size and the batch size.

We empirically determined that patches of size 512×512 with 128 pixels of overlap provided the best trade-off between spatial coverage, learning efficiency, and GPU compatibility. Larger patches led to out-of-memory (OOM) errors, while smaller ones reduced performance due to a loss of structural context.

The overlap between patches serves to:

- Increase the number of training samples, which is critical given our limited dataset (only 11 available SWOT images).
- Reduce edge effects during patch-wise training and inference.
- Preserve continuity of spatial features such as riverbanks and coastlines.

Each training sample thus consists of a pair of aligned patches $(x_{\mathrm{LR}}, x_{\mathrm{HR}})$, where x_{LR} is the artifact-reduced, cropped low-resolution input, and x_{HR} is the MERLIN-denoised high-resolution target. These are fed to the NAFNet model in batches of size 1—the largest viable setting given the hardware constraints.

3.3 Model Architecture: NAFNet

To perform super-resolution on grayscale SAR images, we adopt the NAFNet architecture [Che+22], a recently proposed convolutional network designed for image restoration tasks such as denoising, deblurring, and super-resolution. NAFNet stands out due to its simplicity, low memory footprint, and the absence of nonlinear activation functions, which enhances stability during training and inference.

NAFNet follows an encoder–bottleneck–decoder structure. The input is first passed through a shallow convolutional layer, followed by multiple encoding stages with downsampling, a set of bottleneck residual blocks, and then decoding stages with upsampling and skip connections.

Component	Description
NAFBlock (Core Unit)	
Depthwise Conv + SimpleGate	Lightweight convolution followed by element-wise gating between channel-split halves.
Channel Attention	Global average pooling and 1×1 convolution for feature recalibration.
Feedforward Layer	Channel expansion, gating, and residual scaling via learned parameters.
Layer Normalization	Applied before each functional block to stabilize training.
NAFNet Architecture (SAR Adaptation)	
Input/Output	3×3 convolutions with residual connection to the input.
Encoder	4 stages with increasing NAFBlocks and downsampling.
Bottleneck	12 NAFBlocks at the lowest spatial resolution.
Decoder	4 upsampling stages with skip connections from encoder.
Channels	Configured for single-channel SAR input with 96 base channels.

Table 1: Overview of the NAFBlock and the NAFNet architecture used for super-resolution.

All convolutional layers use 3×3 kernels with stride 1 (except the downsampling layers, which use 2×2 convolutions). Upsampling is performed using PixelShuffle.

The network is trained in a fully supervised manner, minimizing the pixel-wise L1 loss between the predicted super-resolved image and the MERLIN-based high-resolution target. This choice of loss emphasizes fine detail reconstruction and ensures compatibility with the PSNR evaluation metric.

To reduce GPU memory usage and accelerate training, automatic mixed precision (AMP) is enabled through PyTorch's `torch.amp` module. This allows the network to perform computations in lower precision (e.g., float16) while maintaining numerical stability via gradient scaling.

As discussed previously, each image is split into overlapping patches, and due to GPU memory constraints, training is conducted with batch size 1. During each iteration, a batch contains all the patches extracted from a single LR–HR image pair, and the model updates weights patch by patch.

This approach allows the model to learn from local variations while preserving spatial context, and is necessary when using high-resolution SAR images that would otherwise exceed memory limits if processed as a whole.

To avoid overfitting, early stopping is implemented based on the average epoch loss. If no improvement is observed after 10 consecutive epochs, training is halted. The best-performing model (with the lowest training loss) is saved to disk and used for all evaluations.

4 Results and Results Analysis

In this section, we evaluate the performance of the proposed pipeline using both qualitative and quantitative methods. The trained the NAFNet model was tested using the best-performing subband combination strategy (`and`) as determined through visual inspection. While we explored various strategies for subband combination, the majority of evaluations were conducted visually, as specular ringing artifacts are easily detectable by eye.

4.1 Quantitative Evaluation

We computed PSNR and SSIM metrics between the predicted super-resolved (SR) images and the MERLIN-denoised ground truth for all test scenes using the final model trained on AND-filtered LR inputs.

The quantitative metrics reveal a consistent drop in reconstruction quality after super-resolution. On average, PSNR dropped from 51.51 dB (LR) to 45.66 dB (SR), and SSIM dropped from 0.1147 to 0.0322. These results indicate that the model tends to degrade the fidelity of the input images in most cases, with only 23.6% of patches showing PSNR improvement, and just 1.08% showing SSIM improvement.

Interestingly, the SAM metric showed a moderate gain, with 44.75% of patches scoring better in SR than LR. This suggests that while the model may reduce fine-grained accuracy and perceptual quality, it sometimes manages to preserve or even improve directional or angular consistency in the reconstructed signal.

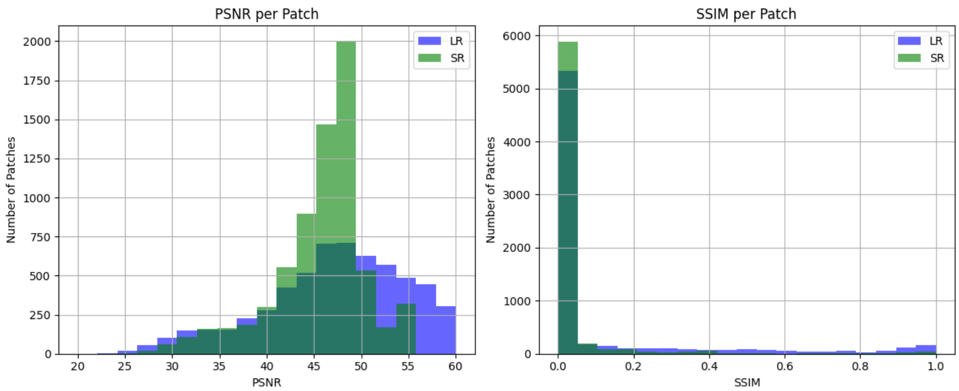


Figure 4: Histograms of PSNR and SSIM scores across all test images using the final subband filtering and super-resolution pipeline.

Overall, the model's performance appears limited by the size and variability of the training dataset. A larger and more diverse dataset, along with explicit artifact masking or a better target reference, might be necessary to achieve meaningful improvements in all metrics.

4.2 Qualitative Evaluation

Figure 5 shows examples of LR inputs and their corresponding SR outputs. In many cases, the model successfully enhances image sharpness and spatial resolution, preserving key structural features.

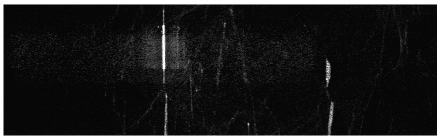
(a) HR target: SLC_021_119_246L_plus

(b) LR input: SLC_021_119_246L_plus

(c) HR target: SLC_019_387_209L_plus

(d) LR input: SLC_019_387_209L_plus

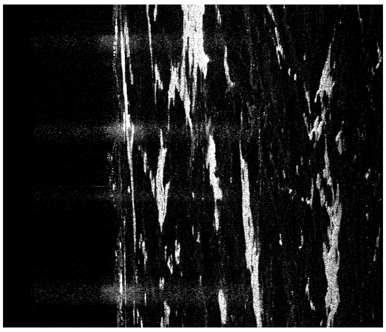
(e) HR target:
SLC_016_447_265R_minus_y_pola_H



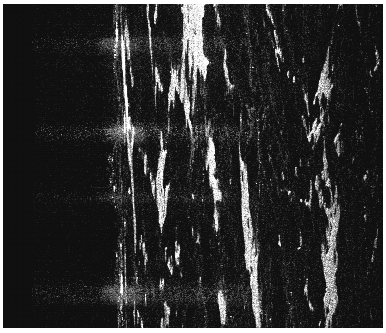
(f) LR input:
SLC_016_447_265R_minus_y_pola_H

Figure 5: Visual comparison of LR input and SR output (NAFNet).

However, some failure cases remain. In Figure 6, we show patches where specular ringing is still reconstructed by the model, despite being largely absent from the LR input. This suggests the model has learned to "hallucinate" the artifacts, likely due to their presence in the ground truth or insufficient training diversity.



(a) LR target:
SLC_015_001_257L_minus_y_pola_V



(b) SR output:
SLC_015_001_257L_minus_y_pola_V

(c) LR input:
SLC_015_550_099L_plus_y_pola_V

(d) SR output:
SLC_015_550_099L_plus_y_pola_V

Figure 6: Examples of failure cases where the model reintroduces specular ringing.

4.3 Subband Combination and Final Performance

During our experiments, we explored multiple subband fusion strategies to reduce specular ringing. These methods aimed to reconstruct the cleanest possible low-resolution (LR) version of the image by selecting and combining frequency subbands in the spectral domain.

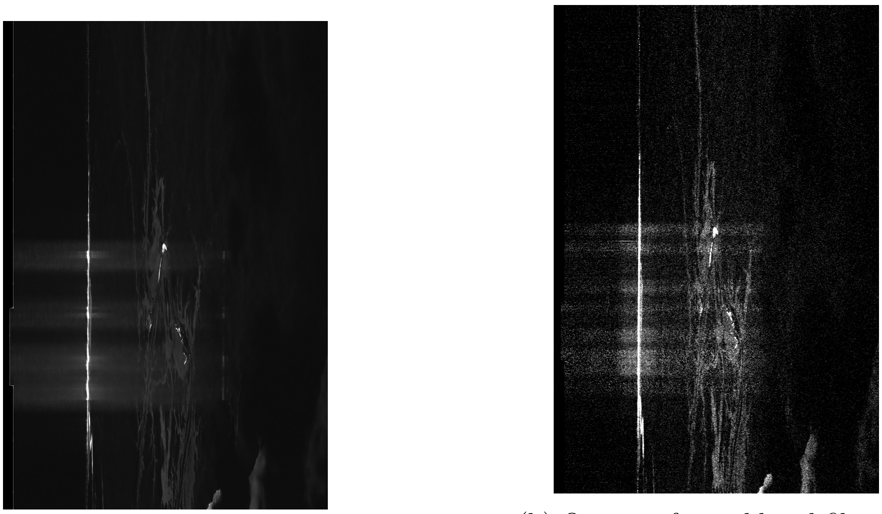
However, not all approaches worked as intended:

- `min_energy` removed many spatial features along with the ringing, resulting in over-smoothed images.
- `median` was moderately successful, but often left residual ringing near transitions.
- `weighted` introduced instability in low-SNR regions, and could even enhance unwanted components.

Only the `and`-based fusion consistently produced cleaner results, suppressing ringing while preserving detail. This method worked well in most cases and was therefore selected for building the final training dataset for the NAFNet model.

Still, one important limitation emerged: **in some images, subband filtering alone was not sufficient to eliminate specular ringing**, regardless of the combination method or the tuning of spectral parameters.

A concrete example of this is shown in Figure 7, where the original image exhibits strong ringing near bright transitions. Despite extensive fine-tuning of the parameters `df_width_ratio`, `col_df_ratio`, `skip_bands_below`, and `skip_bands_above`, as well as testing all available fusion strategies, the artifacts persist in the filtered image.



(a) Original image with strong ringing

(b) Output after subband filtering (AND-based)

Figure 7: Example of a scene where subband filtering failed to suppress ringing, even with extensive parameter tuning.

This failure case illustrates the limits of our current spectral selection pipeline. In such regions, the ringing may span broader frequency bands than expected, or be mixed with meaningful signal components that are difficult to isolate.

Despite these limitations, the subband filtering strategy was overall effective in preparing a cleaner dataset for training. But this example also highlights the need for complementary techniques, either in the frequency or spatial domai, to address particularly challenging cases.

Finally, even with improved inputs, the NAFNet model occasionally reintroduced artifacts during reconstruction. This is likely due to overfitting to residual ringing present in the HR targets (MERLIN), reinforcing the need for cleaner supervision and loss functions that explicitly penalize artifact generation.

5 Discussion

Our experiments show that filtering subbands in the frequency domain is a promising approach for specular ringing removal. However, the results also exposed several challenges that limit the robustness of this pipeline.

First, while the AND-based fusion strategy produced strong results in most scenes, it did not generalize perfectly. In some cases, no combination of parameters successfully removed ringing without degrading important structures. This suggests that ringing is not always confined to isolated subbands and may overlap with signal-dense regions of the spectrum. More adaptive or data-driven subband selection strategies may be required to handle such variability.

Second, the use of MERLIN images as HR targets allowed us to train a model in a weakly supervised fashion. But because these images still contain some residual ringing, the network occasionally learned to reconstruct the very artifacts we aimed to remove. This was especially visible in regions with high contrast or strong signal discontinuities.

Another factor is the limited dataset size. With only 11 images, the training set was not diverse enough to expose the network to the full range of ringing patterns. Despite using overlapping patches and normalization to increase data diversity, we observed clear signs of overfitting and poor generalization.

Finally, while NAFNet offered a solid baseline for restoration, its inductive bias may be too limited for SAR-specific features. Architectures with stronger global reasoning, such as SwinIR or attention-based models, could better handle subtle, structured artifacts like ringing — especially if coupled with perceptual or adversarial losses.

Overall, our results highlight that both preprocessing and supervision quality are essential. Better fusion, better targets, and better models must come together for truly effective artifact-aware restoration.

6 Conclusion

This work proposed a hybrid pipeline for enhancing SWOT SAR images by combining spectral subband filtering and deep learning-based superresolution. We targeted specular ringing artifacts by analyzing the frequency spectrum of complex SAR images and isolating the subbands that contributed to their formation.

Through extensive experimentation, we found that subband filtering alone can significantly reduce artifacts, particularly when using an AND-based combination strategy. This provided a practical and interpretable way to build clean low-resolution inputs from unprocessed SWOT data.

We then trained a NAFNet model using MERLIN-processed images as high-resolution targets. In scenes where artifacts were successfully removed by preprocessing, the model enhanced spatial detail and edge structure. However, in other cases, residual ringing in the HR targets led to undesired reconstructions, limiting the overall performance.

Despite these mixed results, the pipeline offers valuable insights and a strong foundation for future work. These directions include:

- Developing learnable or adaptive subband selection mechanisms that can tailor filtering to each image.
- Acquiring larger and cleaner datasets for training, or building synthetic datasets with controllable artifacts.
- Exploring advanced models (e.g., SwinIR, NAFSSR) and losses that explicitly penalize ringing or hallucination.

This project demonstrates both the promise and complexity of artifact-aware learning in SAR imaging. Even when results fall short, the failure cases point the way to more informed designs in future iterations.

References

[CM91] John C Curlander and Robert N McDonough. "Synthetic aperture radar: systems and signal processing". In: (1991).

[MMB+19] Alexis Mouche, François Mercier, Sylvain Biancamaria, et al. "The SWOT mission: A new era of space altimetry for hydrology". In: *Surveys in Geophysics* 40.4 (2019), pp. 881–901.

[Zha+21] Kai Zhang et al. "Plug-and-Play Image Restoration with Deep Denoiser Prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[Bre+22] Maxime Breil et al. "MERLIN: The SWOT SAR processor". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15.

[Che+22] Liangyu Chen et al. "NAFNet: Nonlinear Activation Free Network for Image Restoration". In: *ECCV* (2022).