**Enhanced Ship Detection through Deep-Learning-Based Super-Resolution Embedding**

Akshit Sharma Satvik Pandey

**Abstract**

In this project, we work on the principles of small object detection through super-resolution techniques to the maritime domain, specifically for ship detection in satellite and aerial imagery. Recognizing the unique challenges posed by the aquatic environment, such as variable lighting, sea clutter, and the diverse sizes of ships, our model integrates deep learning-based super-resolution methods to enhance image quality and facilitate finer detection. Our methodology integrates the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) for image super-resolution, which significantly improves the resolution and clarity of low-resolution maritime images. This enhanced image quality is crucial for detecting smaller and distant ships, which are often challenging to identify in lower resolution. Following the super-resolution process, we employ YOLOv5, a state-of-the-art object detection model, to accurately identify and localize ships within these high-resolution images. YOLOv5's efficient architecture allows for rapid detection while maintaining high accuracy, making it well-suited for real-time maritime monitoring applications, especially can contribute to improved maritime safety and traffic management.

1. **Introduction**

Detecting small objects, such as ships in vast maritime environments, presents unique challenges, especially when working with satellite and aerial imagery. The primary limitation in this context is the minimal pixel representation of these objects. In large water bodies, ships, varying greatly in size and shape, occupy a relatively small portion of the image, making them hard to distinguish from the surrounding sea clutter. This difficulty is exacerbated in low-resolution images where ships may appear minuscule or indistinct, leading to inaccurate

detections that can significantly impact applications requiring high precision, such as maritime traffic monitoring, illegal fishing detection, and border security.

* 1. **Background and related work**

***Small Object Detection in Satellite Imagery***

The challenge of small object detection in satellite imagery is a widely recognized problem in the AI community. This difficulty arises from several key limitations inherent to satellite images:

* Low Resolution: The relatively low resolution of satellite images makes it hard to discern small objects such as ships [1].
* Occlusion: Small objects may be obscured by larger objects, weather phenomena, or sea clutter [2].
* Scale Variability: Ships vary significantly in size, complicating detection algorithms' ability to consistently recognize them across different scales [3].
* Background Clutter: The presence of waves, clouds, and other environmental elements adds complexity to identifying small, discrete objects like ships [1].

Addressing these challenges requires innovative approaches that enhance image quality and refine detection algorithms.

***Super-Resolution Networks and Their Evolution***

The advancement of super-resolution techniques, particularly through Generative Adversarial Networks (GANs)[5], has significantly contributed to improving small object detection. Models like SRCNN, EDSR, SRGAN, and ESRGAN have each progressively enhanced the ability to convert low-resolution (LR) images to high-resolution (HR) images. Specifically, ESRGAN [6], an advancement over SRGAN, addresses the issue of artifacts in super-resolved images. Its key features, such as the Residual-in-Residual Dense Block (RRDB) and the elimination of batch normalization, have streamlined training processes and improved visual quality, particularly in preserving sharp edges and realistic textures.

***Attention Mechanisms in Super-Resolution***

Attention mechanisms have become integral in refining GANs for super-resolution [7]. The implementation of channel and spatial attention, as discussed in studies like "Remote Sensing Image Scene Classification Based on an Enhanced Attention Module," [8] has proven effective in focusing on critical aspects of images. In super-resolution, these mechanisms direct the model to prioritize important image features like edges and textures, enhancing detail preservation in HR outputs. This approach is especially pertinent in maritime environments, where attention to detail is crucial for distinguishing ships from the surrounding sea clutter.

***Advancements in Small-Object Detection***

Research, such as "Small-Object Detection with Super Resolution Embedding," demonstrates the efficacy of combining super-resolution models like ESRGAN with object detection algorithms to improve small object detection. This method, primarily tested on ground vehicles in air traffic control surveillance, is directly applicable to maritime surveillance for ship detection. Enhancements in image resolution facilitate better performance of object detection algorithms on small, distant objects.

* 1. **Summary of proposed work**

Our innovative approach for ship detection in satellite imagery involves the integration of Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) with advanced attention mechanisms, coupled with the utilization of YOLOv5 for object detection. This methodology specifically targets the challenges associated with detecting small ships in high-resolution satellite images.

**Enhanced Super-Resolution with Attention Mechanisms**: We employ ESRGAN, augmented with both channel and spatial attention mechanisms. This enhancement is crucial for emphasizing and preserving vital features of small ships in the up-scaled high-resolution images. The channel attention focuses on important feature channels, improving feature representation, while the spatial attention pinpoints relevant spatial locations in the image, ensuring that critical ship features are not lost during the super-resolution process.

**Adapted YOLOv5 for Maritime Environments**: YOLOv5, known for its efficiency in real-time object detection, is adapted for maritime surveillance after being fine-tuned using the Ship Satellite Imagery dataset. This dataset provides a more relevant and challenging context for maritime object detection, with various ship sizes and complex oceanic backgrounds.

**Real-Time Processing and Enhanced Detection Accuracy**: The combination of YOLOv5’s real-time processing capabilities with the high-resolution images generated by our enhanced ESRGAN allows for rapid and accurate detection of ships. This setup is particularly effective in dynamic oceanic environments, where real-time data processing is essential.

**Technical Advancements**: Our approach adds technical depth to the field of small object detection in satellite imagery. By merging ESRGAN's advanced super-resolution capabilities with the precision of YOLOv5 in object detection, and refining this synergy with attention mechanisms, we address key challenges like scale, resolution, and complex maritime scene interpretation.

In essence, our approach represents a significant leap in satellite-based maritime surveillance technology, offering a sophisticated, real-time solution for ship detection in challenging oceanic environments.

1. **Proposed Work**

The proposed approach for Enhanced Ship Detection via Deep-Learning-Based Super-Resolution Embedding commences with the preprocessing of images to adapt them for the model. Images are decoded, potential alpha channels are removed to retain only RGB components, and the dimensions are adjusted to multiples of four for uniformity. The images are then downscaled to a lower resolution, reducing the original size by a factor of four using bicubic interpolation, a process that maintains the integrity of the image details as much as possible.

The core of the super-resolution method is a convolutional neural network constructed from a series of blocks designed to iteratively enhance the image quality. The network starts with an initial convolution layer that extracts fundamental features from the low-resolution input. It then employs a series of residual blocks, each composed of convolution layers followed by batch normalization and Parametric Rectified Linear Units (PReLU), which introduce non-linearity and allow for learning of complex features. The output of these blocks is a high-level feature map that retains the input's content while adding details absent from the lower-resolution version.

Subsequently, the model incorporates upscale blocks that apply convolutions to increase the number of filters, followed by upsampling operations to enlarge the feature map's spatial dimensions, thus bringing the image closer to the desired high resolution. The upscale process is repeated to progressively refine the image details.

A diagram of a computer

Description automatically generated

**Figure 1 Proposed Model**

Attention mechanisms are integrated into the model to further refine the super-resolution results. These mechanisms consist of spatial and channel attention modules. Spatial attention focuses on the localization of relevant features in the image, employing convolutions with varying dilation rates to capture information at different scales. The channel attention identifies and emphasizes informative feature channels while suppressing the less relevant ones through global pooling operations.

A diagram of a schematic diagram of a schematic diagram

Description automatically generated

**Figure 2 Spatial Attention Module**

A diagram of a channel

Description automatically generated

**Figure 3 Channel Attention Module**

The discriminator component of the model is a deep convolutional network that classifies images as either high-resolution or generated by the super-resolution model. It consists of a series of discriminator blocks that apply convolutions with increasing filters and strides, interspersed with leaky ReLU activations and batch normalization. This structure enables the model to effectively discern the nuances between the super-resolved images and their original high-resolution counterparts.

The model's performance is further boosted by using features extracted by a VGG network, which captures high-level perceptual and texture information from images. This feature extraction allows the model to orient its generation process towards producing images that are not only high in resolution but also rich in details that align closely with those present in natural high-resolution images.

The entire model is trained end-to-end, with the generator and discriminator updating their weights iteratively. The generator strives to produce images that are indistinguishable from true high-resolution images, while the discriminator improves its ability to detect differences between the two. This adversarial training process is key to achieving a state-of-the-art performance in super-resolution tasks.

* 1. **Training and YOLO**

**References**

[1] Wu, H., Zhang, L., & Ma, J. (2022). Remote Sensing Image Super-Resolution via Saliency-Guided Feedback GANs. (IEEE Transactions on Geoscience and Remote Sensing).

[2] Biswas, D., & Tešić, J. (2022). Small Object Difficulty (SOD) Modeling for Objects Detection in Satellite Images. (14th International Conference on Computational Intelligence and Communication Networks).

[3] Biswas, D., & Tevsi'c, J. (2022). Progressive Domain Adaptation with Contrastive Learning for Object Detection in Satellite Imagery. (ArXiv)

[4] Song, Z., Sui, H., & Hua, L. (2021). A hierarchical object detection method in large-scale optical remote sensing satellite imagery using saliency detection and CNN. (International Journal of Remote Sensing)

[5] Tian, C., Zhang, X., Lin, C., Zuo, W., & Zhang, Y. (2022). Generative Adversarial Networks for Image Super-Resolution: A Survey.

[6] Vonderfecht, J., & Liu, F. (2022). Fingerprints of Super Resolution Networks.

[7] Zhu, H., Xie, C., Fei, Y., & Tao, H. (2021). Attention Mechanisms in CNN-Based Single Image Super-Resolution: A Brief Review and a New Perspective.

[8] Li, G., Wei, Y., Li, T., & Li, M. (2022). Image Super-resolution Algorithm Based on Dual Regression Networks with Cross-scale Non-local Attention Mechanisms. (2022 International Conference on Machine Learning and Intelligent Systems Engineering)