# Learning Satellite Image Recovery Through Turbulence

Kimmy Chang, Justin Fletcher

Space Systems Command (A&AS)

#### ABSTRACT

This paper presents a study of deep learning approaches to image recovery using spatially-extended sequential observations of near-Earth satellites. Image recovery is often a prerequisite for use of ground-based extended imagery in space domain awareness (SDA) due to aberrations induced by atmospheric turbulence along the path from satellite to sensor. Traditional deconvolution-based image recovery methods are sensitive to factors such as observation sequence length and estimates of the point spread function (PSF), which has motivated recent interest in autoencoders and other learned approaches. However, no previous study has applied general state-of-the-art image restoration models to the space domain data. In this work, we evaluate the effectiveness of recent deep learning methods, specifically Generative Adversarial Networks (GANs) and Vision Transformers, for image restoration of satellites. We analyze the tradeoffs between restoration quality, time, and computational complexity of each method. We experimentally demonstrate that deep learning models provide high-quality image restoration with less data than traditional deconvolution methods. We further optimize the most successful state-of-the-art model and demonstrate its efficacy in image restoration at a previously unseen degradation level (SNIIRS=2.5). Our deep learning models are trained on simulated data from the SILO dataset and require no training on real data, yet they restore the most severely degraded real satellite imagery with state-of-the-art performance of 27.0 dB PSNR and 0.95 SSIM on the SILO dataset, as well as better visual results on the real satellite images.

### 1. INTRODUCTION

Spatially-resolved imagery of satellites in low-Earth orbit reveals otherwise inaccessible satellited health and status information and underpins many space domain awareness (SDA) characterization tasks [11, 9, 18]. Unfortunately, this information resides in the object plane of the target satellite which we may access only imperfectly due to optical and, in the case of ground-based observations, atmospheric aberrations. The task of correcting for these aberrations, traditionally called image recovery in astronmical contexts, has been extensively studied. In computer vision the closely related task of image restoration deals with the general problem of retrieving information from a degraded version of an image. In this work, we leverage recent advances in learned image restoration to advance the state of the art image recovery for ground-based observation of satellites.

# 2. RELATED WORKS

#### 2.1 Image Restoration

In recent years, CNNs have become a popular solution to super-resolution tasks due in part to their high task performance, comparatively reliable training dynamics, and general applicability. The Super-Resolution Convolutional Neural Network (SRCNN) [5] demarcates the beginning of this trend. Following that milestone, several CNN-based models have been proposed that have advanced the state-of-the-art in super-resolution performance while sometimes reducing the computational costs of training and inference. For instance, the Enhanced Deep Residual Network for Single Image Super-Resolution (EDSR) [8] simplified the residual block and created a multi-scale model with fewer parameters.

The Cascading Residual Network (CARN) [10] addressed the heavy computation seen in prior deep learning methods by proposing an architecture that implemented a cascading mechanism on a residual network. The cascading connections between intermediary layers increased the flow of gradient information. This resulted in fewer parameters

and operations whilst maintaining state-of-the-art performance. CARN-M used a group size of four in the efficient residual block, which reduced the number of parameters by five and number of operations by four times with minimal PSNR loss.

The Information Multi-distillation Network (IMDN) [20] constructed cascaded information distillation blocks that extracted hierarchical features step-by-step and aggregated them according to importance determined by the contrast-aware channel attention mechanism. IMDN touts competitive results with a modest number of parameters and offered a guideline for lightweight network design. Despite the success of these models, none of them have been applied to the space domain for super-resolution tasks.

In the area of blind super-resolution tasks, General Adversarial Networks (GANs) have recently gained attention as an approach to restore low-resolution images that suffer from unknown and complex degradations [19, 1, 15]. Real-ESRGAN [16] proposes a high-order degradation process to model realistic degradations. Although this model is trained with synthetic data, it can restore real-world data and is more practical than other GANs for image restoration.

The success of the Transformer architecture in natural language processing (NLP) tasks [14] [3] [4] [21] has sparked interest in the architecture within the computer vision community. Among the state-of-the-art methods for image restoration and denoising, SwinIR [7] has emerged as a top performer while also reducing the number of parameters by up to 67%. SwinIR is based on the Swin Transformer [?], which combines the strengths of the CNN and Transformer architectures. This allows for the processing of large images through the use of a local attention mechanism, while also modeling long-range dependencies using the shifted window scheme.

### 2.2 Space Domain Awareness

The task of learned image reconstruction for satellited imagery was pioneered by Kyono et al. in [?]. This work demonstrated the efficacy of learned and generative approaches to satellite image restoration, but did not extend these methods to real data.

In the field of image recovery for space applications, the canonical approach is the multi-frame blind deconvolution method, as explored by several studies, including [6] and [12]. Schmidt et al. developed a real-time restoration process for ground-based images of satellites at video frame rates, utilizing multiple frames to jointly estimate the object and point spread function (PSF) while also implementing windowing, registration, and enhancement techniques [12]. However, the effectiveness of this method is limited by the need for a good initial guess of the PSF for a fast convergence, and the possibility of introducing edge blur and feature size artifacts during post-processing. Another study, [6], employed a physics-based model for the PSF and imposed strong constraints on its pupil plane and Fourier plane properties to extend the dynamic range and spectral region. Despite relatively good performance across various image modalities, this method requires prior information on the PSFs and is highly dependent on the length of the observation period.

Recently, attention-based U-nets have been explored as learned methods for deblurring atmospheric turbulence [2] [13]. Chen et al. [2] proposed a deep-stacked autoencoder neural network model for image reconstruction. This model comprises an encoder network that extracts high-level features from the input blurred image and a decoder network that reconstructs the clean image from the extracted features. On the other hand, Shu et al. [13] proposed a Noise Suppression-based Restoration Network (NSRN) for turbulence-degraded images that employs a U-net architecture and utilizes a curriculum learning strategy. Both studies showcase experimental results that demonstrate the subjectively superior performance of learned methods compared to traditional methods in restoring atmospheric turbulence-blurred images, in terms of both objective and subjective image quality metrics. However, the efficacy of these learned methods on real-world turbulence-degraded data has been primarily evaluated on relatively low levels of atmospheric turbulence.

# 3. APPROACH

### 3.1 Dataset

In this paper, we present experiments conducted on the Scored Images of LEO Objects (SILO) dataset [17]. The SILO dataset was generated using a Time-domain Analysis Simulation for Advanced Tracking (TASAT) tool configured for early morning observations of 38 unique satellites from the 3.6m Advanced Electro-Optical System (AEOS) telescope located on Mt. Haleakala. The dataset includes 19 subsets of atmospheric degradation levels, ranging from 2.5 to 7.0



Fig. 1: Example simulated images of Almaz at SNIIRS 2.5-7.0 with target pristine from Scored Images of LEO Objects (SILO) dataset.

with increments of 0.25 SNIIRS. Unless otherwise specified, we downsampled the target (pristine) data to a resolution of (256, 256) and input (degraded) data to (64, 64) for our models, which aim to upsample by a factor of four. A set of example images is presented in Fig. 1.

### 3.2 Evaluating Models

The efficacy of our proposed approach will be assessed through an extensive evaluation that involves both qualitative and quantitative methods. The quality of the upsampled images produced by our method will be visually inspected for qualitative evaluation. To perform quantitative evaluation, we will utilize two standard performance metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

PSNR (Eq 1) measures the average amount of noise present in the reconstructed images relative to the original images. Specifically, it is given by the equation below, where L represents the number of maximum possible intensity levels in an image, and MSE is the mean squared error.

$$10\log_{10}\left(\frac{(L-1)^2}{\text{MSE}}\right) \tag{1}$$

SSIM (Eq 2), on the other hand, evaluates the similarity between the reconstructed and original images by considering their luminance, contrast, and structure. It is computed using the equation below, where  $\mu_x, \mu_y$  and  $\sigma_x^2, \sigma_y^2$  are the averages and variances of *x*, *y*,  $\sigma_{xy}$  is the covariance of *x* and *y*, and  $c_1$  and  $c_2$  are variables that stabilize the division with weak denominator.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$
(2)

It is worth noting that SSIM is designed to better reflect human visual perception than PSNR. Therefore, both metrics will be employed to quantitatively evaluate the performance of our approach. PSNR is particularly useful for assessing the quality of noisy images, while SSIM provides a more comprehensive evaluation by taking into account various aspects of image quality.

#### 3.3 Performance on Pretrained Models

To evaluate the performance of the pretrained super-resolution models on atmospherically turbulent satellite images, we randomly select twenty images from each SNIIRS level, resulting in a test set of 380 images. The models under



Fig. 2: SNIIRS scores v. PSNR and SSIM by Model. Smaller models like SRCNN, EDSR, CARN-M, and IMDN display abnormal drops in performance across SNIIRS scores. And while models like SwinIR and Real-ESRGAN consistently outperform other models based on the SSIM metric, the PSNR metric for these two models are on average the lowest.

consideration include SRCNN [5], EDSR [8], CARN-M [10], IMDN [20], SwinIR [7], and Real-ESRGAN [16].

We examine the relationship between PSNR, SSIM, and the number of model parameters for each of these models using provided pretrained models and provide the results in Table 1. This analysis enables us to compare the performance and computational efficiency of each model and select the most suitable approach for our use case.

Model	Number of Parameters	PSNR	SSIM
SRCNN [5]	57K	18.6522	0.5263
CARN-M [10]	414K	18.6684	0.6068
IMDN [20]	715K	20.5112	0.6067
EDSR [8]	1.5M	20.5074	0.5934
SwinIR [7]	11.8M	19.5349	0.7324
Real-ESRGAN [16]	16.7M	19.2718	0.7281

Table 1: Quantitative comparison (average PSNR/SSIM) with pretrained state-of-the-art methods. Best and second best performance are in red and blue, respectively.

After analyzing the PSNR and SSIM values presented in Table 1, we further investigate the performance of the aforementioned models for image restoration under specific levels of atmospheric turbulence. In addition to numerical analysis, visual inspection of the reconstructed satellite images is also considered. The plots presented in Fig. 2 demonstrate that the pretrained Real-ESRGAN and SwinIR models exhibit the highest and the least variability in their performance. Moreover, upon visual inspection of the differences between the prediction outputs of the two models (see Fig. 3), it can be observed that the Real-ESRGAN and SwinIR model produce the best image reconstructions amidst their large number of parameters. Thus, for image reconstruction using pretrained models, we recommend using pretrained Real-ESRGAN or SwinIR models.



Fig. 3: Example upsampling results of Seasat for pretrained models.

# 4. EXPERIMENTS

# 4.1 Experimental Setup

A subset of the SILO dataset [17] consisting of 1000 images from each individual SNIIRS subset is used to train six deep learning models. This subset is divided into 800 training, 100 validation, and 100 test images. The resulting training set comprises of 15,200 images, while the validation and test sets each consist of 1,900 images. Model training is conducted on an NVIDIA DGX system with 4 GPUs for 100 epochs.

# 4.2 Results on Image Super Resolution

We conduct a comparative analysis of several models for image restoration, including Enhanced Deep Super-Resolution (EDSR) [8], Cascading Residual Network with Multi-scale Feature Fusion (CARN-M) [10], Iterative Multi-Domain Network (IMDN) [20], SwinIR [7], and Real-ESRGAN [16]. Our analysis is based on an extensive set of experiments, the results of which are presented in Fig. 4.

Our findings show that Real-ESRGAN consistently outperforms IMDN and SwinIR, while CARN-M and EDSR exhibit slightly inferior performance. Moreover, we observe that Real-ESRGAN achieves faster and better convergence than any other method. Consequently, the subsequent sections of this manuscript are devoted to an in-depth investigation of the effectiveness of Real-ESRGAN for image restoration.

# 4.3 Ablation Study on Real-ESRGAN

# **Robustness to Varying Levels of Degradation**

In this study, we explore various training approaches to determine the pipeline for restoring atmospherically turbulent images. The three approaches examined in our experiments include: (1) a set of 19 individual models, each designed to handle a specific SNIIRS class input; (2) a group of 3 models, designed for severe (SNIIRS 2.5-3.5), moderate (SNIIRS 3.75-4.75), and low degradation (SNIIRS 5.0-7.0) inputs; and (3) a single model capable of handling degradation across the entire SNIIRS range from 2.5 to 7.0.

We plot average PSNR and SSIM values across SNIIRS levels for the groupings to better understand image performance metrics and dataset selection in Fig. 5. Our results show that there is no significant benefit to training nineteen models that specialize in restoring a specific value of SNIIRS over a general model that is trained on SNIIRS that range from severe to low degradation. Furthermore, the quality of restoration does not vary significantly over the SNIIRS range, staying within the range of 5.0dB PSNR and 0.05 SSIM. These findings highlight the robustness of ESRGAN in restoring images degraded by a large range of atmospheric turbulence.



Fig. 4: Quantitative comparison (average PSNR/SSIM) of state-of-the-art methods.



Fig. 5: Ablation Study on Robustness of Real-ESRGAN.



Fig. 6: Ablation Study on Number of Observations. Blue line represents PSNR and orange line represents SSIM.

### **Optimizer and Upsampling Factor**

Table 2 presents the results of an experiment designed to assess the performance of Real-ESRGAN for image restoration using different optimizers and x2 upsampling. The optimizers used in the study are Adam, Adam-W, RMSprop, Adagrad, and Lion, and the upsampling is performed from an input image resolution of (256, 256) to an output resolution of (512, 512). The table shows that Adam performs the best among the optimizers, with improvements of 0.28 dB in validation PSNR and 0.0033 in SSIM. Lion comes a close second with slightly higher training PSNR and SSIM than Adam. The PSNR and SSIM for x2 upsampling were 25.074dB and 0.9522, respectively. Thus, upsampling by a factor of 2 does not generally improve image restoration. This could be attributed to the benefits of downsampling in the x4 case, which mitigates the effects of atmospheric turbulence. It is important to note that we only present the best learning rate and weight decay hyperparameter combinations for each optimizer in the table.

	Adam	Adagrad	RMSprop	AdamW	Lion
PSNR	27.0363	24.6793	13.0152	26.5103	26.7541
SSIM	0.9461	0.9366	0.0113	0.9426	0.9428

Table 2: Ablation Study on Optimizers. Best and second best performance are in red and blue, respectively.

### Number of Observations

The present study examines the relationship between the number of observations and the performance of Real-ESRGAN. As illustrated in Fig. 6, our findings demonstrate that the algorithm's performance improves markedly as the number of observations increases. Specifically, once we obtain more than 250 observations from each SNIIRS class, the algorithm's performance dramatically improves. These results indicate that Real-ESRGAN can achieve high performance with a relatively small number of observations from each degradation level, implying that the algorithm is well-suited for use in real-world settings where access to data may be limited.

### 4.4 Real-ESRGAN: Results on Unseen Satellites

To evaluate the generalization capability of Real-ESRGAN for satellite image restoration, we conducted experiments on a modified version of our original SILO dataset subset (as described in section 4.1). Specifically, we removed all instances of a single satellite from the training set and designated them as a validation subset. We repeated this process for ten different satellites. The performance of Real-ESRGAN on this modified dataset is presented in Table 3, and we provide corresponding qualitative results in Fig. 7. It is worth noting that the modified dataset does not have the same train/validation splits. Thereby we provide the exact number of training images and testing images in the table. Despite the comparatively low values of average PSNR and SSIM when the training data lacks examples of unseen satellites, it is observed that Real-ESRGAN performs well in restoring images even in cases of the lowest average



Fig. 7: Visual results of unseen SILO satellites at SNIIRS=2.5. Rows 1, 2, and 3 show the image restorations of the Explorer, NOAA, and Argos satellites.

PSNR and SSIM, as demonstrated by the Ekran and HST examples depicted in Fig.s 8 and 9, respectively. These results suggest that Real-ESRGAN is capable of effectively restoring images of satellites that were not present in the training data.

Satellite Name	<b>Train/Test Number of Images</b>	PSNR	SSIM
Argos	18088/912	21.2211	0.9325
Atlas	18591/409	19.9626	0.9255
Ekran	18518/482	16.0243	0.8776
Explorer	18305/695	31.4276	0.9856
Geosat	18624/376	21.9698	0.9425
HST	18622/378	17.6035	0.7544
Molniya	18693/307	19.1247	0.9212
NOAA	18704/296	16.7814	0.8782
Seasat	17842/1158	18.2409	0.8132

Table 3: Quantitative comparison (average PSNR and SSIM) for different unseen satellites.

### 4.5 Real-ESRGAN: Results on Real-World Data

The ultimate goal of satellite image recovery is for real-world applications. We test our trained model on linear pass satellite images analyzed obtained from a 3.6m Advanced Electro-Optical System (AEOS) telescope located at the Air Force Maui Optical and Supercomputing (AMOS) Observatory. As shown in Fig. 10, Real-ESRGAN produces visually pleasing images with clear and sharp edges in contrast to other methods. Notably, we observed unexpectedly poor performance from SwinIR. We believe a potential explanation for this phenomenon is the inherent emphasis of the transformer architecture on sequence-to-sequence mapping, as opposed to the complex statistical relationships that are essential for generating realistic images from degraded image data. We present examples of the effectiveness of our approach on a total of fifteen different satellites in Fig. 11. In conclusion, we establish the validity of our method for preserving the geometric shape of a satellite while accommodating for angular rotations in the presence of diverse input images of the same satellite. This is illustrated in Fig. 12, which provides examples of the behavior of the method on SPOT 2 data.



Fig. 8: Visual results of "unseen" Ekran satellite. Rows 1, 2, and 3 show the image restorations of the satellite at SNIIRS = 2.5, 4.25, 6.75.



Fig. 9: Visual results of "unseen" HST satellite. Rows 1, 2, and 3 show the image restorations of the satellite at SNIIRS = 2.5, 4.25, 6.75.



(a) Input

(b) Richardson-Lucy Deconvolution



(c) EDSR



(d) CARN-M



(e) IMDN

(f) SwinIR



(g) Real-ESRGAN DISTRIBUTION A. Approved for public release: distribution is unlimited. Fig. 10: Visual result Public Affairs celease approval #AffaIa72023384 for Super-Resolution.



Fig. 11: Visual results of custom Real-ESRGAN model on real-world data. Rows 1, 2, 3, and 4 show the image restorations of Cosmos 1803, Cosmos 2084, Meteor-M 2, and Starlink-3558.



Fig. 12: Visual results of custom Real-ESRGAN model on real-world data. Rows 1, 2, 3, and 4 show the image restorations of SPOT 2.

#### 5. CONCLUSION

This research proposes a new standard for image restoration of atmospherically turbulent satellite images [6, 12, 2, 13]. Our method achieves subjectively superior performance on both the Synthetic Image Library for Optical Imagery (SILO) dataset and real-world data acquired from a 3.6m telescope, as measured by peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Through a systematic evaluation of recent deep learning methods for satellite image restoration, we find that General Adversarial Networks (GANs) are effective in this context, while remaining vulnerable to hallucination in the low-SNIIRS setting.

We investigated the robustness of our custom GAN model to atmospheric turbulence within the range of SNIIRS 2.5-7.0 and demonstrate its ability to restore heavily degraded images without relying on prior point spread function (PSF) knowledge or observation period length. Our model can also recover images of various satellites not present in the training set. Notably, we show that the GAN model trained on only synthetic images can restore real-world satellite images of corresponding synthetic satellites not present in the training data.

Our work provides a foundation upon which future space domain awareness applications may be built. By restoring heavily degraded satellite images, our approach enables improved object tracking, pose estimation, and 3D reconstruction. We present our method as a baseline for future research in this area, and we anticipate that our learned solutions will become increasingly valuable as the demand for astronomical survey missions and orbital asset monitoring continues to grow.

#### 6. ACKNOWLEDGEMENTS

The authors would like to thank the 15th Space Surveillance Squadron for access to AEOS data. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government.

#### REFERENCES

- Kelvin C.K. Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. Glean: Generative latent bank for largefactor image super-resolution. arXiv preprint arXiv:2012.00739, 2021.
- [2] Gongping Chen, Zhisheng Gao, Qiaolu Wang, and Qingqing Luo. U-net like deep autoencoders for deblurring atmospheric turbulence. J. Electron. Imaging, 2021.
- [3] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2019.
- [5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. arXiv preprint arXiv:1501.00092, July 2015.
- [6] Stuart M. Jefferies. Next generation image restoration for space situational awareness. 2009.
- [7] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. arXiv preprint arXiv:2108.10257, 2021.
- [8] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. arXiv preprint arXiv:1707.02921, July 2017.
- [9] Jacob Lucas, Trent Kyono, Julia Yang, and Justin Fletcher. Discovering 3-d structure of leo objects. AMOS, 2021.
- [10] Ahn Namhyuk, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. *arXiv preprint arXiv:1803.08664*, October 2018.
- [11] Klaus Okkelberg, Jacob Lucas, Trent Kyono, Michael Abercrombi, Justin Fletcher, and Matthew Phelps. Self-supervised auxiliary task learning for estimating satellite orientation. *AMOS*, 2021.
- [12] Jason D. Schmidt, Kevin Jackovitz, Joseph T. Riley, and Jason A. Tellez. Real-time image restoration for space-object imaging. *Applied Optics, Optica Publishing Group*, 2019.
- [13] Jie Shu, Chunzhi Xie, and Zhisheng Gao. Blind restoration of atmospheric turbulence-degraded images based on curriculum learning. *Remote Sensing*, 2022.
- [14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017.
- [15] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. *arXiv preprint arXiv:2101.04061*, 2021.

- [16] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. arXiv preprint arXiv:2107.10833, 2021.
- [17] Michael Werth, Jacob Lucas, Trent Kyono, Ian McQuaid, and Justin Fletcher. Silo: A machine learning dataset of synthetic ground-based observations of leo satellites. *IEEE Aerospace*, 2020.
- [18] Julia Yang, Jacob Lucas, Trent Kyono, Michael Abercrombie, Justin Fletcher, and Ian McQuaid. Semantic segmentation of low earth object satellites using convolutional neural networks. AMOS, 2021.
- [19] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-restore: Learning network path selection for image restoration. arXiv preprint arXiv:1904.10343, 2019.
- [20] Hui Zheng, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multidistillation network. arXiv preprint arXiv:1909.11856, September 2019.
- [21] Chen Zhu, Yu Cheng, Zhe Gan, Siqi Sun, Tom Goldstein, and Jingjing Liu. Freelb: Enhanced adversarial training for natural language understanding. *arXiv preprint arXiv:1909.11764*, 2020.