

Towards Real-Time Image Reconstruction Using Deep Neural Networks

Trent Kyono, Jacob Lucas, Michael Werth

The Boeing Company

Justin Fletcher

Odyssey Systems Consulting

Ian McQuaid

AFRL

Abstract

Imaging space objects from ground-based stations results in images perturbed with atmospheric turbulence. To reconstruct the original image from an observed perturbed image, Multi-frame Blind Deconvolution (MFBD) is a state-of-the-art maximum-likelihood algorithm that in practice is often computationally expensive in terms of time and space. It may take on the order of hours to generate reconstructions and typically requires hours of compute time on a super computer. Even GPU based approaches currently on the order of minutes to converge which is not fast enough for real-time situational awareness. We present preliminary results on neural network reconstructions for solving this inverse imaging problem on a large synthetic dataset.

1. Introduction

The problem of blind deconvolution is a well-studied problem applicable to several imaging domains including astronomical ground-based imaging (through a turbulent atmosphere) and medical imaging techniques (ultrasound and MRI). In this work we address the problem of blind deconvolution for astronomical reconstruction of images from a ground-based sensor that have been degraded and perturbed by atmospheric turbulence. The amount of apparent blurring of astronomical objects are due to the turbulent mixing in the Earth's atmosphere, which causes variations of the optical refractive index. While large telescopes have theoretically milli-arcsecond resolution, the observed images are limited to the average point spread function at the time of observation. State-of-the-art techniques for reconstructing satellite images, such as multi-frame blind deconvolution (MFBD) [1], requires iteratively searching through a combinatorial search space of Point Spread Functions (PSFs) convolved with a pristine image of the target to arrive at the optimal answer. This process alone may take on the order of minutes to hours to reconstruct an image making real-time decisions impossible.

Existing works using convolutional neural networks (CNNs) have shown state-of-the-art performance on many super-resolution and image denoising tasks [cite](#). Motivated by this, we present a method to remove atmospheric turbulence, i.e, denoise, from ground-based images of satellites using a CNN. We train and evaluate our models on a synthetic dataset comprised of 11 million satellite image pairs. We provide an analysis of existing image recovery methods in terms of Peak Signal to Noise Ratio (PSNR), Structural

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

Similarity (SSIM), as well as a scoring neural network used for estimating Space-object National Imagery Interpretability Rating Scale (SNIIRS), a variant of the National Imagery Interpretability Scale (NIIRS) [2].

We provide a discussion of related works in Section 2. Section 3 details a formalization of our problem and approach. Section 4 provides a discussion of our dataset, training architecture, hyperparameters, and our experimental results. We conclude with brief remarks and a discussion of limitations in Section 5.

2. Related Works

Several related works have explored the applications of CNNs and deep learning to astronomy. For noise reduction, [3] recently presented a proof-of-concept neural network for denoising the bispectrum for astronomical image recovery on synthetic data. For classification, [4, 5, 6] investigated the application of object classification using neural networks on photometric light curves and showed promising results. Using Generative Adversarial Networks (GANs), [7] recovered features from artificially degraded images with worse seeing and higher noise than the original with a performance that far exceeded the capabilities of simple deconvolution. Additionally, [8] used a GAN to generate more realistic images of galaxies than existing state of the art. [9] used machine learning to automatically segment and label galaxies in astronomical images. [10] showed promising results using an autoencoder for real-time MFBD of solar images. Additionally there have been probes into image scoring with deep learning; [11] applies a CNN to images and yields a ‘human opinion’ quality score, and [12] uses a relatively shallow network to give a quality score to distorted images. Later [13] applied CNNs to predict SNIIRS for ground-based images of satellites under various atmospheric perturbations. Lastly, [14, 15] attempt to determine the likelihood of whether a perturbed image will result in an coherent MFBD reconstruction. The success of these approaches motivates our application of neural networks to atmospheric removal.

The problem of atmospheric removal (or MFBD function learning) presented in this work is closely related to CNNs for image denoising and super-resolution. These fields are vast and span hundreds of publications within the year. Without going in to too much detail we point the reader to deep learning survey papers for image denoising [16] and super-resolution [17]. A majority of the methods are based on variations of fully convolutional neural networks (FCNNs) with no dense layering. These variations may or may not have skip layers, such as in the popular U-NET [18]. Recent approaches are centered around using GANs for improving the photorealism of reconstructions [16, 17], which has been the motivation for this work.

3. Problem Formulation

We consider MFBD an optimal reconstruction, defined as R^* , that performs the mapping

$$R^* : \mathcal{X} \rightarrow \mathcal{Y}, \quad (1)$$

where \mathcal{X} is an input tensor in $\mathbb{R}^{\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}}$ and $\hat{\mathcal{Y}}$ is an output matrix in $\mathbb{R}^{\mathbb{Z} \times \mathbb{Z}}$. The input tensor, \mathcal{X} , comprises the sequence $\{\mathcal{X}_n, \mathcal{X}_{n+1}, \dots, \mathcal{X}_{n+T}\}$, in which each element \mathcal{X}_i is a single frame speckle image transformation of a pristine image \mathcal{X}_i . We impose the constraint that $\mathcal{Y} = \mathcal{X}_{n+T}$. Thus, R^* exactly recovers the final pristine image from a length T sequence of degraded images. We may also define a reconstruction estimator, R , as

$$R : \mathcal{X} \rightarrow \hat{\mathcal{Y}}, \quad (2)$$

which has the same form as R^* , but has no domain constraints. So, $\hat{\mathcal{Y}}$ may assume any value in $\mathbb{R}^{\mathbb{Z} \times \mathbb{Z}}$. For any reconstruction estimator R , $R \rightarrow R^*$ as $\hat{\mathcal{Y}} \rightarrow \mathcal{Y}$; equivalently, an optimal estimator minimizes $d(\hat{\mathcal{Y}}, \mathcal{Y})$

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

where d is any metric function on $\mathbb{R}^{\mathbb{Z} \times \mathbb{Z}}$. In this work, we seek an instantiation of R which approximates R^* as near as possible within a finite computational budget. Formally, we define this minimization as.

$$\min_{\mathcal{X}_i} \ell(d(R(\mathcal{X}_i), R(\mathcal{X}_i))) \quad (3)$$

To solve this non-convex optimization we leverage a deep neural network (DNN). A DNN may be formalized as a composition, f , of L parameterized functions

$$f(\mathbf{x}) = f^{(L)}(f^{(L-1)}(\dots f^{(2)}(f^{(1)}(\mathbf{x}))), \quad (4)$$

where $f^{(i)}$ is the i -th layer of the DNN and \mathbf{x} is an input tensor. Each layer computes a feature map, $\mathbf{h}^{(i)}$, of the preceding layer given by

$$\mathbf{h}^{(i)} = g^{(i)}(\mathbf{W}^{(i)\top} \mathbf{h}^{(i-1)} + \mathbf{b}^{(i)}), \quad (5)$$

where $\mathbf{W}^{(i)}$ and $\mathbf{b}^{(i)}$ are trainable parameters comprising $\Theta^{(i)}$, $g^{(i)}$ is an activation function, and $\mathbf{h}^{(0)} = \mathbf{x}$. The feature representation computed by the outermost function, $\mathbf{h}^{(L)}$, is the output of the DNN, and is denoted \hat{y} .

In this work we use conditional generative adversarial networks (CGAN) [19, 20] which is an extension of the original generative adversarial networks (GAN) by [21]. A GAN is comprised of two adversarial models with disparate objectives: (1) a generative model G that captures the data distribution and (2) a discriminative model D that estimates the probability that a sample belongs to the training data rather than a generated sample from G . The components G and D are represented by any non-linear model. To learn the generator distribution p_g over some data x , G builds a function from a prior noise distribution $p_z(z)$ to data space $G(z; \theta_g)$; $D(x; \theta_d)$ outputs the probability that a sample x is from p_g generated by G or not. G and D are trained simultaneously and objectively play a two-player min-max game with a value function $V(G, D)$ as:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (6)$$

We can extend this GAN objective to be conditioned on some labeled information y . In our case, this is the pristine satellite images, which is fed into D , and the paired degraded image, which is fed into G . We can rewrite $V(G, D)$ for conditioning on this information as the following.

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|x)))]. \quad (7)$$

This objective can be satisfied according to the following loss:

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))]. \quad (8)$$

As shown in prior works, such as [20], it is useful to include an auxiliary loss term to G , such as L1, L2, or SSIM. In this work, we use structural similarity (SSIM) which we define later. Therefore we rewrite our final objective function as:

$$G^* = \arg \min_G \max_D \mathcal{L}_{CGAN}(G, D) + \lambda \mathcal{L}_{SSIM}(G). \quad (9)$$

We present an illustrative schematic for CGAN in Figure 1.

4. Experiments

In this section, we highlight our synthetic dataset details. We introduce our benchmark methods as well as discuss our evaluation metrics. We present results using our method on synthetic data.

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

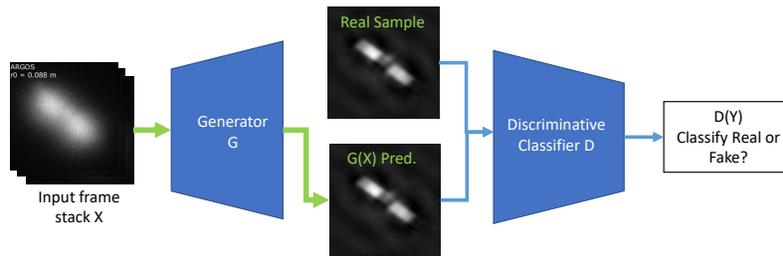


Figure 1: Schematic of C-GAN.

4.1. Dataset details

The SNIIRS score of a space object is a measure of the smallest resolvable feature. It is a \log_2 based score, with a larger value indicating that smaller features are resolvable. An analyst is required to determine what features are resolvable in a given image, and to determine the size of the smallest resolvable feature. Possible scores range from 0 (no resolvable features) to 12 (features smaller than 5mm can be resolved). In practice many of the higher scores are not currently attainable for ground based observations. The usable range of this metric for ground based systems can be viewed in terms of r_0 and object range. A reasonable span of r_0 values might be from 5cm (poor seeing) to 100cm (adaptive optics corrected). At a typical LEO range of 600km, these r_0 values equate to resolutions of approximately 10m (SNIIRS 3) and .5m (SNIIRS 7) respectively in the I-band. For this study a significant database of analyst-scored ground-based LEO observations was made available, however this data contained a lack of diversity that even with class weighting was difficult if not impossible to overcome, with several classes having no representation in the data at all. The solution posed was to construct a simulated basis set for initial training, and then for future work fold in actual sensor data.

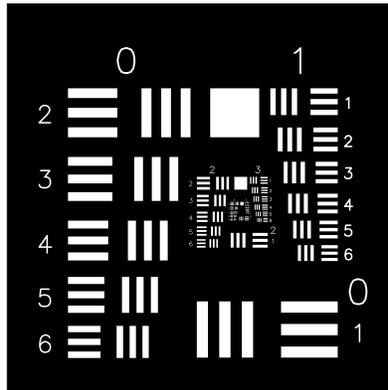


Figure 2: 3-bar target used to calibrate the simulation SNIIRS scores.

To create the simulated dataset a metric was constructed using a validated atmospheric simulation code [22] to apply a broad range of atmospheric turbulence characterized by Fried Parameter r_0 to a scaled 3-bar target render with cascading target sizes (Fig. 2). By setting the simulated range and instantaneous field of view (IFOV), we designated the simulated physical size of the 3-bar targets. Measuring the smallest resolvable 3-bar target for a simulated r_0 and then fitting in the style of a General Image Quality Equation (GIQE) [23]

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

produced a mapping from Δr_0 to Δ SNIIRS. Simulating a reasonable set of r_0 values allowed us to degrade an image by multiple SNIIRS with a resolution of approximately 1/4 of a SNIIRS. In the simulation used r_0 was a dependent variable and could not be directly set, only measured. This resulted in range of SNIIRS scores with less recognizable but still entirely valid values, as can be viewed in Figure 3. We constrained our scores to a range from 3 to 7, corresponding to r_0 of 5cm and 100cm as described in the previous paragraph.

An image set containing 5000 satellite renders with 38 discrete satellites in multiple poses and configurations was used as the basis for the simulated set. The SNIIRS score of each render was established by a trained analyst. This is referred to as the absolute score. The absolute score is the SNIIRS of the target as viewed from a diffraction-limited optical system with a circular aperture, i.e. the upper limit for the simulated SNIIRS is capped at the absolute score on the basis of resolvable features discernible in the diffraction limited image. The basis set contains a range of initial SNIIRS values, which after degradation provides a diverse data set of greater than 90,000 scenarios each containing 125 simulated images, resulting in over 11 million unique images generated by 500,000 CPU hours on a supercomputer. From here on this simulated data will be referred to as SILO (Simulated Images of LEO Objects) [24].

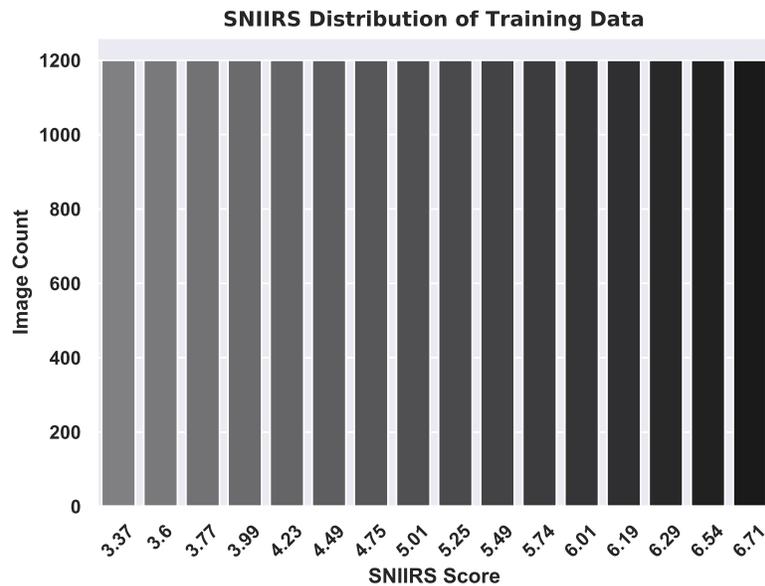


Figure 3: Histogram illustrating the distribution of SNIIRS scores in the simulated training data.

A subset of SILO was used for network training. This subset was evenly sampled across SNIIRS 3-7 as shown in Fig. 3. To prevent contamination of the validation data, the subset was segregated by target render. Of the 38 discrete targets, the training set contained 30, and the remaining 8 were reserved for validation and testing (discrete sets). This ensured that images in the validation data were dissimilar to the training images. 80% of data was used for training, 20% for validation and testing.

4.2. Benchmark methods

We benchmark our method against two common atmospheric removal techniques: Bispectrum [25] and GPU-based MFBD [1] (LUCID) [26]. For both bispectrum and LUCID we have a trained image analyst select the optimal tuning parameters resulting in best quality image. For both bispectrum and LUCID we use a standard ensemble duration of 0.25 seconds which is equivalent to 32 frames in our neural network.

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

4.3. Evaluation metrics

We evaluate methods using the following metrics: PSNR, SSIM [27], and SNIIRS-NN. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation and is often used to measure image quality between two images [28]. PSNR between two images x and y is often defined in terms of MSE as follows

$$PSNR(x, y) = 10 \log_{10} \left(\frac{MAX(x)^2}{MSE(x, y)} \right), \quad (10)$$

where $MAX(X)$ is the maximum pixel value of x , and $MSE(x, y)$ is the mean squared error between the pixels of x and y . SSIM considers image degradation as a perceived change in structural information, while also incorporating important perceptual phenomena, including luminance masking and contrast masking terms. SSIM is designed to improve upon traditional methods such as PSNR and MSE. For two images x and y , SSIM is given as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (11)$$

where μ_i is the average of i , σ_i^2 is the variance of i , and σ_{xy} is the covariance of x and y . C_1 and C_2 are used to stabilize the division with a weak denominator and is a function of the dynamic range of the pixel values [27]. In addition to these common metrics used, we also report a scoring method in terms of SNIIRS. Image quality is often not easily quantifiable by a statistical metric such as PSNR and SSIM. To go beyond this, often a human is necessary for determining true utility of an image. However, having a trained analyst examine and rate each image is prohibitively costly, and may be subject to human biases. In the prior work of [13] a neural network was used to score atmospheric images with accuracy better than a trained analyst. We reproduce the network they used and report the predicted score as a metric in our experiments. We refer to this metric as SNIIRS-NN.

4.4. Experimental Results

We trained our CGAN network with SSIM loss according to equation 9. We use the hyperparameters specified in [20] and we performed a grid search for the hyperparameter $\lambda \in \{1, 0.1, 0.01\}$. We use 32 consecutive frames for each benchmark method, which we consider one input sequence or ensemble. When training CGAN, for each input image sequence we image register all frames to the first frame in the sequence. In Figure 4 we a sample comparison of CGAN to our existing benchmarks as well as the ground-truth renders over various SNIIRS rating on the SILO dataset. Upon visual inspection, CGAN has sharper edges and higher contrast than the existing benchmark methods. It is also important to note that Bispectrum seems to be the best in this case for SNIIRS 3-4, however qualitatively over the test dataset, Bispectrum appears to be the worst under strong turbulence. Furthermore, it appears that CGAN does really well at reconstructing the general outline of the image suggesting semantic segmentation as a promising research direction. Figure 5 shows our performance in terms of various metrics of each benchmark method on a held-out test set of SILO. CGAN has superior performance in comparison to the benchmark methods in terms of PSNR and SSIM. In terms of SNIIRS-NN, CGAN has improved performance for lower SNIIRS values and degraded performance for higher values suggesting that the reconstructions made by CGAN are out-of-distribution for the SNIIRS-NN. Note that our network has never seen these test satellites or PSF perturbation instances before.

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

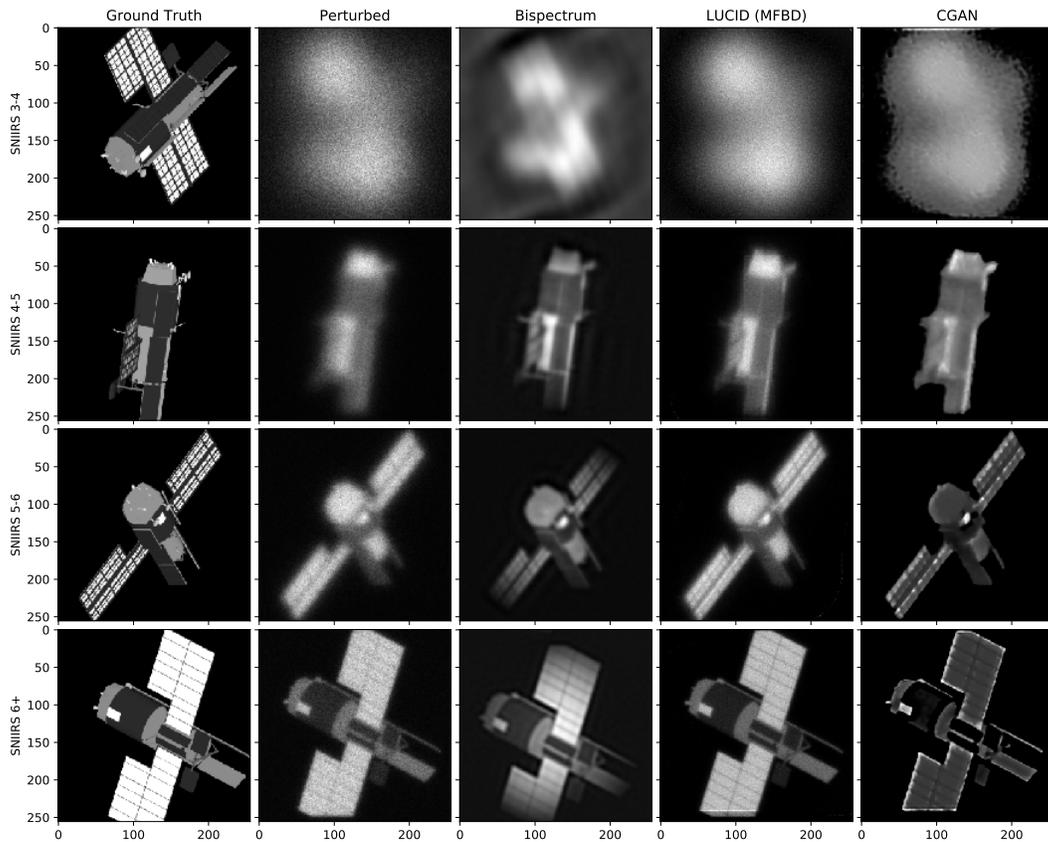


Figure 4: Example simulated images of benchmark methods vs CGAN over various atmospheric turbulence strengths.

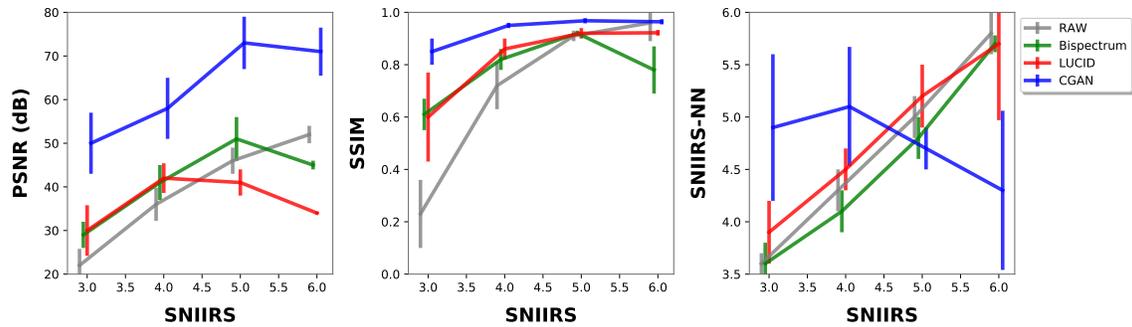


Figure 5: Comparison of benchmark methods in terms of PSNE, SSIM and SNIIRS-NN.

5. Conclusion

In this work, we have provided a neural network approach for removing atmospheric turbulence. To the best of our knowledge, we are the first to do so for ground-based observations of resolved LEOs. We have proposed a CGAN approach on a large synthetic dataset that demonstrated improvements over existing atmo-

**DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049**

spheric denoising approaches on synthetic data. For future work, we wish to validate and transfer our method to real data, investigate joint recovery of PSFs for improving performance, and increase the input ensemble size. We hope this work will inspire other data-driven approaches for atmospheric denoising.

DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049

References

- [1] Timothy J. Schulz. Multiframe blind deconvolution of astronomical images. *J. Opt. Soc. Am. A*, 10(5):1064–1073, May 1993.
- [2] John M. Irvine. National imagery interpretability rating scales (niirs): overview and methodology, 1997.
- [3] Jacob Lucas, Brandoch Calef, and Trent Kyono. Recovering astronomical images with deep neural network supported bispectrum processing. In *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*. 2018.
- [4] B. Jia, K. D. Pham, E. Blasch, Z. Wang, D. Shen, and G. Chen. Space object classification using deep neural networks. In *2018 IEEE Aerospace Conference*, pages 1–8, March 2018.
- [5] Richard Linares. Space object classification using deep convolutional neural networks. 07 2016.
- [6] Robert Furfaro, Richard Linares, and Vishnu Reddy. Space objects classification via light-curve measurements: Deep convolutional neural networks and model-based transfer learning. 09 2018.
- [7] Kevin Schawinski, Ce Zhang, Hantian Zhang, Lucas Fowler, and Gokula Krishnan Santhanam. Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Monthly Notices of the Royal Astronomical Society: Letters*, 467(1):L110–L114, 01 2017.
- [8] Levi Fussell and Ben Moews. Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial networks. *Monthly Notices of the Royal Astronomical Society*, 485(3):3203–3214, 03 2019.
- [9] Alex Hocking, James E. Geach, Yi Sun, and Neil Davey. An automatic taxonomy of galaxy morphology using unsupervised machine learning. *Monthly Notices of the Royal Astronomical Society*, 473(1):1108–1129, 09 2017.
- [10] A. Asensio Ramos, J. de la Cruz Rodriguez, and A Pastor Yabar. Real-time multiframe blind deconvolution of solar images. 06 2018.
- [11] Hossein Talebi Esfandarani and Peyman Milanfar. NIMA: neural image assessment. *CoRR*, abs/1709.05424, 2017.
- [12] Le Kang, Peng Ye, Yi Ci Li, and David S. Doermann. Convolutional neural networks for no-reference image quality assessment. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, 2014.
- [13] Jacob Lucas, Michael Werth, Trent Kyono, Ian McQuaid, and Justin Fletcher. Automated interpretability scoring of ground-based observations of leo objects with deep learning. In *IEEE Aerospace*. 2020.
- [14] Trent Kyono, Jacob Lucas, Michael Werth, Ian McQuaid, and Justin Fletcher. Determining multi-frame blind deconvolution resolvability using deep learning. In *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*. 2019.
- [15] Trent Kyono, Jacob Lucas, Michael Werth, Brandoch Calef, Ian McQuaid, and Justin Fletcher. Machine learning for quality assessment of ground-based optical images of satellites. *Optical Engineering*, 59(5):1 – 15, 2020.
- [16] Chunwei Tian, Yong Xu, Lunke Fei, and Ke Yan. Deep learning for image denoising: A survey. In *ICGEC*, 2018.

DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049

- [17] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [19] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. 11 2014.
- [20] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [21] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14*, page 2672–2680, Cambridge, MA, USA, 2014. MIT Press.
- [22] V. S. Rao Gudimetla, Richard B. Holmes, Thomas C. Farrell, and Jacob Lucas. Phase screen simulations of laser propagation through non-kolmogorov atmospheric turbulence. In *Proc. SPIE, Atmospheric Propagation VIII*, volume 8038, 2011.
- [23] Jon C. Leachtenauer, William Malila, John Irvine, Linda Colburn, and Nanette Salvaggio. General image-quality equation: Giqe. *Appl. Opt.*, 36(32):8322–8328, Nov 1997.
- [24] Michael Werth, Jacob Lucas, Trent Kyono, Ian McQuaid, and Justin Fletcher. Silo: A machine learning dataset of synthetic ground-based observations of leo satellites. In *IEEE Aerospace*. 2020.
- [25] P. Negrete-Regagnon. Practical aspects of image recovery by means of the bispectrum. *Journal of the Optical Society of America A*, 13:1557–1576, July 1996.
- [26] Michael Werth and Kevin Roe. Lucid: High-resolution ground-based observations of leo satellites with multi-frame blind deconvolution. In *NVIDIA GPU Technology Conference*, 2020.
- [27] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [28] A. Horé and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

DISTRIBUTION A. Approved for public release: distribution is unlimited.
Public Affairs release approval #AFRL-2020-0049