

SUPER-RESOLUTION OF GOES IMAGERY FOR NEAR REAL TIME HIGH RESOLUTION IMAGERY

James Mullen
Virginia Tech

mullenj@vt.edu

Brian Lattimer
Virginia Tech

lattimer@vt.edu

Abstract

Remote sensing continues to gain relevance in modern society in part due to the development of new image processing methods and imagery platforms. The rise of convolutional neural networks (CNNs) for processing imagery has coincided with an increase in publicly available remote sensing imagery, expanding the envelope of possibility when processing satellite imagery. For many use cases, it is desirable to have remote sensing imagery with high temporal and spatial resolution. The temporal resolution of GOES imagery at 15 minutes is unmatched in other publicly available products, but GOES's spatial resolution of 2 km per pixel is too coarse for many use cases. Conversely, VIIRS-I imagery has the high resolution needed for many applications, but the temporal resolution limits time-sensitive applications.

Our contribution combines two disparate areas of research by applying state-of-the-art CNN-based super-resolution techniques to GOES imagery, while utilizing high-resolution VIIRS-I imagery as ground truth, bringing its effective spatial resolution down to sub-kilometer pixels. We attempt this super-resolution using both an autoencoder-based architecture of our own creation and the Very Deep Super-Resolution network [6] (VDSR). Both our model and the VDSR provide an improvement in the peak signal-to-noise ratio over the original imagery of 1.37 and 1.20 dB, respectively.

1. Introduction

For remote sensing imaging applications, imagery with both high temporal-resolution and high spatial-resolution unlocks many use cases such as tracking and forecasting natural disasters in real time, such as wildland fires. Low spatial resolution imagery prevents users from accurately deploying resources and understanding the true shape or extent of the phenomenon of interest. Low temporal resolution imagery prevents users from adapting to fast changing situations. For example, in a wildfire response, fires change rapidly with the weather and the situation from hour to hour can be completely different. Sending first responders out into the field with up-to-date, accurate, information would increase firefighting effectiveness and decrease danger.

With the current array of publicly available remote sensing imagery, users must choose between data with satisfactory temporal resolution or satisfactory spatial resolution, with no satellite platform having both. The recent launch of NOAA's Geostationary Operational Environmental Satellites (GOES) has brought forth a new public platform with a temporal resolution of 15 minutes, unmatched by competitors. However, the imagery from GOES has a temporal resolution of only 2 km per pixel, which is overly coarse for many resolution sensitive applications. Another publicly available remote sensing imagery platform is the Visible Infrared Imaging Radiometer Suite (VI-

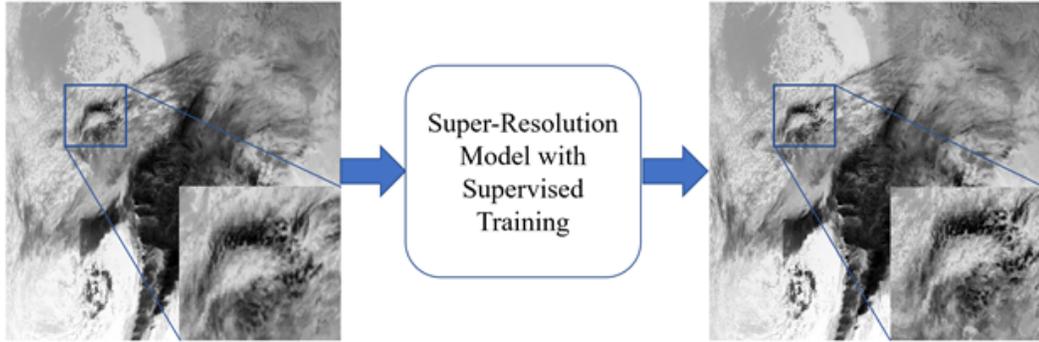


Figure 1: An overview of our system: taking GOES-17 band 7 imagery, training a super-resolution model with VIIRS-I band 4 ground truth, to produce images more closely matching the VIIRS-I imagery spatial resolution at the high temporal resolution of GOES.

IRS) on the NOAA-20 and Suomi NPP satellites. VIIRS has superior spatial resolution at 350 m per pixel but a temporal resolution of 12 hours makes it impossible to use for time sensitive applications. A solution that combines the strengths of these two platforms is needed is for time sensitive, spatially resolved applications such as wildland fires.

Super-resolution is an image processing technique that retrieves a high spatial resolution image from a single or series of low spatial resolution images [23]. There are two approaches typically utilized to ascertain the high-resolution information: i) produce new pixel values by estimating them from the local area on the image or ii) through a learned model such as a neural network. Convolutional neural networks (CNNs) have made breakthroughs in many image processing tasks, including super-resolution (SR). This is an active area of research in the computer vision community [6, 7, 17, 19, 20]. Some of these works have proven successful in recovering the high-resolution information.

While many of these neural networks are trained on images or videos taken by people, there is a growing body of research attempting to apply these state-of-the-art SR techniques to satellite imagery [1, 8, 9, 10]. Research has shown that performing a super resolution on satellite imagery can improve its results in an object detection or

segmentation task [16]. Super-resolving an image by a factor of 4 would require 15 new pixels be generated per pixel, effectively quartering the ground sampling distance. For example, if the original resolution was 100 meters per pixel, the super-resolved image would have a resolution of 25 meters per pixel. This additional detail could help analysts or detection algorithms better understand a scene and its semantic makeup [22].

The impacts of effective super-resolution of overhead imagery would be significant. Numerous remote sensing applications would materially benefit from this technology including natural disaster response, automatic target recognition, crop and deforestation monitoring, and reconnaissance. These methods can also be applied to many imagery modalities including electro-optical, infrared, and hyperspectral.

Most super-resolution research focuses on first down sampling an image to be used as a model input and using the original image as truth. In this work, we have instead investigated the possibility of utilizing a low-resolution satellite image as an input and a high-resolution satellite image of the same area at the same time as the ground truth. Specifically, we utilized the 7th band of the GOES-17 satellite as an input and the 4th band of the VIIRS-I imaging platform as ground truth in our supervised training regime. In theory, this will

allow us to get super-resolved, high spatial resolution GOES images at high temporal frequency for use in time sensitive applications. Figure 1 shows an example of the flow of our system.

1.1. Contributions

In this paper, we describe the integration of super-resolution models with overhead imagery from two different satellite imagery platforms. Our main contributions can be summarized as follows:

- Investigated the application of existing state-of-the-art super-resolution models to overhead imagery of two publicly available platforms, one with a high temporal resolution and low spatial resolution, and one with a low temporal resolution and high spatial resolution, using one as an input and one as training truth.
- Proposed a network architecture based upon the autoencoder or fully convolutional network for this task.
- Evaluated several factors in the super-resolution process to understand the limits of the applications of these techniques on satellite imagery from two disparate imagery platforms.

2. Methods

The methods utilized for data collection and processing and setting up our neural network architectures are outlined below.

2.1. Data Collection and Processing

There are several public overhead imagery datasets including *SpaceNet Challenge* [15], the *IARPA Multi-View Stereo Satellite Challenge* [2], the *Vehicle Detection in Aerial Imagery dataset (VEDAI)* [14], and the *Overhead Imagery Research Dataset (OIRDS)* [18]. None of these included the data required for our investigation,

namely images of the same place and time from two imagery platforms of differing spatial resolution, requiring us to create our own dataset. Our dataset consists of VIIRS-I imagery and GOES-17 imagery. Specifically, we focused on the 4th VIIRS imagery band which operates at a wavelength from 3.550 - 3.930 μm and the 7th GOES-17 band which operates at a wavelength from 3.80 - 3.99 μm . In theory, this should make the images similar and increases the likelihood of success in training our model.

As the GOES 17 data covers the western portion of the United States, we started by retrieving VIIRS-I granules that intersect the western continental United States using the NASA VIIRS Atmosphere SIPS API [12]. These files were then downloaded before using the timestamp associated with each file to find the most closely connected GOES-17 imagery. As GOES-17 imagery is produced every 15 minutes, as opposed to every 12 hours for VIIRS, the maximum differential in time for each image pair is 7.5 minutes. The corresponding GOES 17 images were then downloaded from the AWS S3 server using the *rclone* program [3]. After collecting the raw data files for each satellite, we preprocessed them to be in the same orientation and orthorectified. We utilized the *SatPy* python package to complete this; opening each file as a “Scene” and resampling the GOES image onto the location of the VIIRS image. This injected some amount of noise into our dataset as the *SatPy* resample process is not flawless. With our data preprocessed, we had to break each paired image into reasonably sized image chips for our model. Using a sliding window technique, we produced 128, 128-pixel image chips (using a step size of 128). This left us with 2,118,371 total image pairs for training our model. We split these images into a training set of 80% of the images and a testing set of 20% of the images.

Figure 2 shows a sampling of our created dataset with some failure modes. The leftmost image pair is a good example of most of the images

in our dataset. As can be seen, the GOES-17 image is much lower resolution than the VIIRS-I image. In the middle of the GOES-17 image, you can see the artifacts that *SatPy* leaves on some images when resampling through the horizontal lines in the middle of the image and duplicated, not smooth graphics. This could have a negative effect on the training and results of our model. The third image pair shows a significant failure in our dataset where some image chips seem to have wildly different information encoded in them. We believe this is likely caused by subtle differences in the imagery platforms and the band operating wavelengths. This is not very common in the dataset but could be injected enough to confuse our models during training or lower our evaluation metrics.

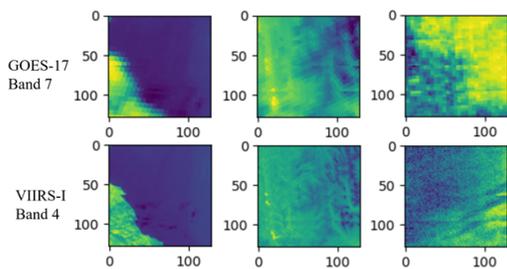


Figure 2: Three sample sets of GOES-17 to VIIRS-I image training pairs after preprocessing.

2.2. Network Architectures

As we are primarily evaluating the efficacy of super-resolving remote sensing imagery from one platform onto another platform, the models utilized were strongly influenced by the literature. The first model utilized borrowed heavily from autoencoder techniques. Autoencoders are known to learn a representation of a set of data very effectively, mainly by training a network to ignore signal noise [4, 5, 11, 13]. Autoencoders consist of a reduction side, or encoder, and a reconstruction side, or decoder. In theory these two portions of the network will work together to effectively dis-

till the relevant information from an input and reproduce it as an output. In our implementation, the autoencoder was tasked with distilling the relevant portions of the input GOES-17 imagery, and reproducing a clearer, higher resolution output, similar to the VIIRS-I imagery.

The encoder portion of our model consisted of five two-dimensional convolution layers with kernel size of three and a padding of one. Following each convolutional layer was a ReLu activation and after the second and fourth convolutions, a max pool. For the decoder, the network architecture had two blocks of one transpose convolution, and two normal convolutions, all with a kernel size of 3. After each of these blocks was a final convolutional layer. Again, ReLu activation was utilized after each convolutional layer. A graphic of this architecture and its implementation data flow can be seen in Figure 3. This model was trained with a mean squared error loss comparing the decoder output to the high-resolution VIIRS-I imagery. A learning rate of 1.0×10^{-4} and a batch size of 32 images was utilized over five epochs with the Adam optimizer.

In addition to the custom autoencoder implementation utilized, we also trained and tested a commonly used method in super-resolution literature, the Very Deep Super Resolution network, or VDSR. VDSR is a single image super-resolution (SISR) network originally proposed by Kim et al. in 2016 [6]. Its main contribution was increasing the depth of a network and finding a significant improvement in accuracy. VDSR is a commonly used baseline in the literature due to its ubiquity and large jump in performance versus earlier models [17].

Our VDSR model was trained using MSE loss and a learning rate of 0.1 with a standard gradient descent optimizer. The details of the implementation were taken from the original paper and a Pytorch implementation found at [21]. Small modifications to the model needed to be added as the resolution change was not a clean 4x increase between 2 km pixels and 350 m pixels. The same

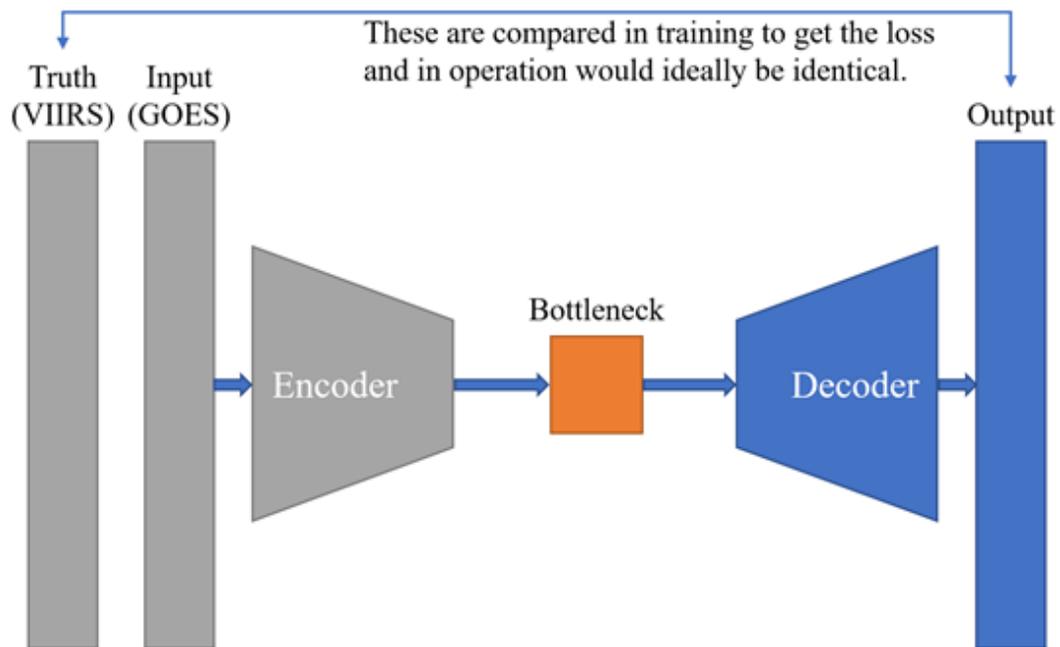


Figure 3: Overview of autoencoder architecture and the data flow through the model. The output is trained against the VIIRS-I ground truth while utilizing the GOES-17 input. In theory, the encoder should distill information important to the reconstruction of high resolution VIIRS imagery from the GOES input.

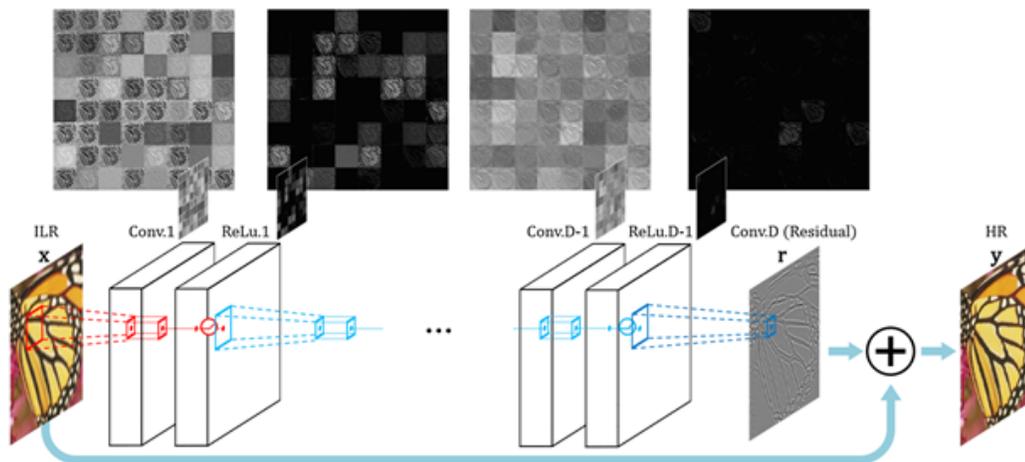


Figure 4: The VDSR network architecture from [1]. We utilized the VDSR due to its standing in the literature as a standard state-of-the-art model in super-resolution.

testing and validation sets were utilized for both networks.

3. Results

After conducting our experiments as described in the previous sections, results are reported using PSNR on the validation set of image chips. Table 1 summarizes a comparison of our two approaches on the data with the PSNR of the original low resolution GOES images, utilizing the VIIRS-I images as the target. The comparison between the GOES input image and the VIIRS-I reference image is referred to as *Unresolved* while our autoencoder based method is referred to as *RS-AE*, remote sensing autoencoder, and the Very Deep Super Resolution Network [6] is referred to as *VDSR*.

Algorithm	Quality (dBs vs. VIIRS-I Imagery)
Unresolved	70.73
VDSR	72.04
RS-AE	72.11

Table 1: Comparison between no super-resolution, the state-of-the-art VDSR network, and our autoencoder network for our task of super-resolving low-resolution GOES-17 Imagery to high-resolution VIIRS-I Imagery.

In addition to PSNR as a quantitative metric, we also collected a set of qualitative images for reference. In Figure 5, we show the results of both of our models against the input GOES imagery and the goal VIIRS imagery.

4. Discussion

With many remote sensing applications for high resolution, high frequency imagery, our goal is to explore the possibility of creating this imagery through super-resolving imagery from the high frequency but low-resolution GOES-17 platform. We have designed a system to leverage existing

super-resolution approaches and high-resolution but low frequency VIIRS-I imagery to produce the desired high frequency, high-resolution imagery.

With our results, we have shown that this type of super-resolution has its own unique challenges with the methods used to pre-process the imagery, differences in the imaging platforms themselves, and misaligned spectral ranges of the imagery. This can be seen clearly in Figure 5, where the color of the GOES-17 imagery is different than that in the VIIRS-I despite showing the same overarching features. Specifically, in the second image from the left, you can see the dark blues and bright yellows are more pronounced in the GOES imagery than for the same features in the VIIRS-I imagery. It is difficult to filter many of these issues out of the training set as there is no known tools to convert the imagery of each platform to be closer to the other and prevent the creation of added noise in the image preprocessing stage.

Comparing the results shown in Table 1 of the two super-resolution architectures explored with the unresolved imagery, we can see that our methods represent a significant improvement of over 1.3 dBs in the PSNR metric. Additionally, our custom AutoEncoder based architecture produced marginally improved results, 0.07 dBs, relative to the VDSR network. This is likely due to the autoencoder structure being better suited to the unique nature of our super-resolution task. For the autoencoder architecture, no improvement was observed by altering the activation functions, adding or removing convolutions, or adding more dense blocks.

Our qualitative results in Figure 5 show that both of our models seemed to approach the problem in a similar way, softening the edges of the coarse GOES-17 image input and trying to convert the imagery to be more like the VIIRS-I imagery in terms of the measured intensity represented by color. This can especially be seen in the right-most image in Figure 5 where the VDSR and AutoEncoder both decrease the intensity of the bright yellow part of the GOES-17 image signifying that it

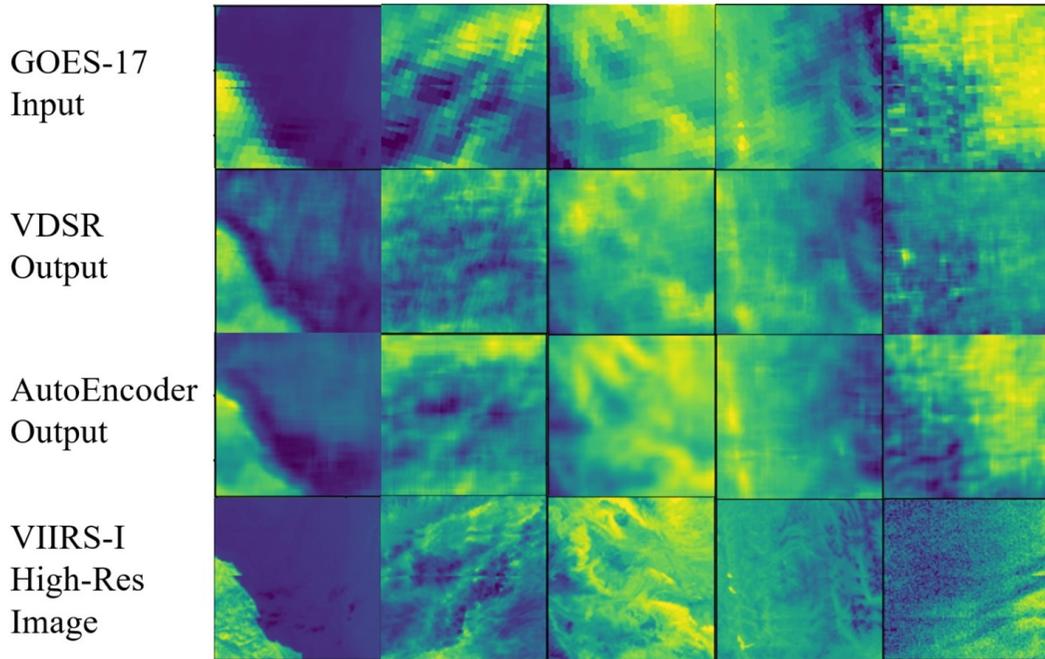


Figure 5: Qualitative results on our unique GOES-17 to VIIRS-I Dataset with our AutoEncoder network and the state-of-the-art VDSR network.

has learned that the VIIRS-I images typically have a lower intensity.

While our results are an improvement on the original GOES-17 imagery, the models were unable to recover detail that would be important in object recognition tasks. This likely decreases our methods efficacy for real-world use cases that require such data. Despite this, we feel that these results reveal the potential of these methods and we are encouraged that the network can recover some details and improve upon the input images provided. As we enter a period where remote sensing is increasing in prevalence alongside the maturation of super-resolution computer vision methods, we believe that there is clear challenge in optimizing both super-resolution and image preprocessing methods to produce higher quality data for users.

5. Conclusion

Remote sensing is one of many fields that can leverage recent advances in super-resolution pow-

ered by deep learning. In this exploratory work, we developed a unique dataset to approach the super-resolution problem utilizing multiple satellite imagery platforms, trained multiple super-resolution models on this unique training scenario, and evaluated their successes on this problem.

Our results have shown that current methods can produce an improvement on original GOES-17 input images, when trained using VIIRS-I imagery as the ground truth image, of over 1.3 dBs in the PSNR metric. However, these improvements might not be significant enough to power many of the intended use cases that require high-resolution, high frequency remote sensing imagery.

We have also shown that current methods for preprocessing remote sensing imagery do not meet the needs of current state-of-the-art super-resolution models when using different imagery platforms as the input and expected truth. New methods need to be created that inject less noise or artifacts and compensate between differences in the original sensors the images were taken on.

References

- [1] M. Bosch, C. M. Gifford, and P. A. Rodriguez. Super-resolution for overhead imagery using densenets and adversarial learning. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1414–1422. IEEE, 2018. 2
- [2] M. Bosch, Z. Kurtz, S. Hagstrom, and M. Brown. A multiple view stereo benchmark for satellite imagery. In *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–9, 2016. 3
- [3] N. Craig-Wood. 3
- [4] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang. Single sample face recognition via learning deep supervised autoencoders. *IEEE transactions on information forensics and security*, 10(10):2108–2118, 2015. 4
- [5] K. Ghasedi Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang. Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization. In *Proceedings of the IEEE international conference on computer vision*, pages 5736–5745, 2017. 4
- [6] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1, 2, 4, 6
- [7] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2
- [8] S. Lei, Z. Shi, and Z. Zou. Super-resolution for remote sensing images via local–global combined network. *IEEE Geoscience and Remote Sensing Letters*, 14(8):1243–1247, 2017. 2
- [9] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017. 2
- [10] Y. Luo, L. Zhou, S. Wang, and Z. Wang. Video satellite imagery super resolution via convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2398–2402, 2017. 2
- [11] A. Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011. 4
- [12] U. of Wisconsin Madison. 3
- [13] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin. Variational autoencoder for deep learning of images, labels and captions. *arXiv preprint arXiv:1609.08976*, 2016. 4
- [14] S. Razakarivony and F. Jurie. Vehicle detection in aerial imagery : A small target detection benchmark. *Journal of Visual Communication and Image Representation*, 34:187–203, 2016. 3
- [15] J. Shermeyer, D. Hogan, J. Brown, A. Van Etten, N. Weir, F. Pacifici, R. Hansch, A. Bastidas, S. Soenen, T. Bacastow, and R. Lewis. Spacenet 6: Multi-sensor all weather mapping dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 3
- [16] J. Shermeyer and A. Van Etten. The effects of super-resolution on object detection performance in satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2
- [17] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017. 2, 4
- [18] F. Tanner, B. Colder, C. Pullen, D. Heagy, M. Eppolito, V. Carlan, C. Oertel, and P. Sallee. Overhead imagery research data set — an annotated data library tools to aid in the development of computer vision algorithms. In *2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009)*, pages 1–8, 2009. 3
- [19] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on*

Computer Vision (ECCV) Workshops, pages 0–0, 2018. 2

- [20] Z. Wang, J. Chen, and S. C. Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2
- [21] J. Xu. 4
- [22] D. Yang, Z. Li, Y. Xia, and Z. Chen. Remote sensing image super-resolution: Challenges and approaches. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 196–200, 2015. 2
- [23] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128:389–408, 2016. 2