DISCLAIMER

ORNL/TM-2025/3836

Illuminating the Night: A Survey of Super-Resolution Methods for Night-time Light Images

Tony Allen
Haley Duba-Sullivan
Emma J. Reid

April 2025

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-2025/3836

Cyber Resilience and Intelligence Division

ILLUMINATING THE NIGHT: A SURVEY OF SUPER-RESOLUTION METHODS FOR NIGHTTIME LIGHT IMAGES

Tony Allen
Haley Duba-Sullivan
Emma J. Reid

April 2025

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AWEI	Automated Water Extraction Index
AWGN	Additive White Gaussian Noise
CMDM	Conditional Multiscale Downscaling Model
CNN	Convolutional Neural Network
DMSP-OLS	Defense Meteorological Satellite Program's Operational Line-scan System
DN	Digital Number
DNB	Day/Night Band
FSIM	Feature Similarity Index Measure
GWR	Geographically Weighted Regression
HR	High-Resolution
INR	Implicit Neural Representation
ISS	International Space Station
JPSS	Joint Polar Satellite System
LAADS DAAC	Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center
LR	Low-Resolution
LST	Land Surface Temperature
LUCC	Land Use & Cover Change
MGWR	Multiscale Geographically Weighted Regression
MLP	Multilayer Perceptron
MMSR	Multimodal Super-Resolution
MSE	Mean Squared Error
MSI	Multispectral Image
MSI-SR	Multispectral Image Super-Resolution
NASA	National Aeronautics and Space Administration
NDBI	Normalized Built-Up Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIQE	Naturalness Image Quality Evaluator
NIR	Near-Infrared
NRMSE	Normalized Root Mean Square Error
NTL	Nighttime Lights
OSM	OpenStreetMap
POI	Point of Interest
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
S-NPP	Suomi National Polar-orbiting Partnership
SISR	Single Image Super-Resolution
SSIM	Structural Similarity Index Measure
VIIRS	Visible Infrared Imaging Radiometer Suite
VNCI	Vegetation Nighttime Condition Index

FOREWORD

ACKNOWLEDGEMENTS

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (https://energy.gov/doe-public-access-plan). This research used resources from the ORNL Research Cloud Infrastructure at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

ABSTRACT

Nighttime Light (NTL) images provide critical insights into urbanization, disaster response, and energy consumption. The VIIRS Day/Night Band (DNB) sensor offers high-quality NTL imagery with daily revisit rates, but the available spatial resolution hinders fine-grained accurate analysis. Super-resolution techniques aim to increase the resolution of NTL images, enabling more detailed assessments of infrastructure, light pollution, economic activity, and power outages. However, existing state-of-the-art super-resolution methods designed for natural images struggle with the unique characteristics of NTL data. This work provides a comprehensive review of super-resolution methods across multiple image modalities, evaluates their effectiveness on VIIRS DNB data, and proposes a multi-modal super-resolution approach tailored to NTL imagery. The proposed approach integrates VIIRS DNB data with road networks and land use information to improve reconstruction accuracy and spatial detail. Code is available for this project at https://code.ornl.gov/viirs-sr/sr-demos.

1. INTRODUCTION

Nighttime Lights (NTL) images are satellite images that capture artificial light emissions on Earth at night [1]. These images primarily show city lights, roads, industrial areas, fishing fleets, and even some natural events like wildfires or auroras. NTL images have a wide variety of use cases, including economics, environmental science, and urban planning [2]. There are several NTL sensors mounted on satellites with publicly available data, including the Defense Meteorological Satellite Program's Operational Line-scan System (DMSP-OLS), the Visible Infrared Imaging Radiometer Suite (VIIRS) on-board the Suomi National Polar-orbiting Partnership (S-NPP) and Joint Polar Satellite System (JPSS) satellites, and photographs taken by astronauts on the International Space Station (ISS) [2]. However, in this work, we will focus on the Day/Night Band (DNB) sensors on the VIIRS due to their high resolution and wide use.

The VIIRS DNB sensors on board the S-NPP and JPSS satellites capture global daily measurements of visible and Near-Infrared (NIR) light emissions [3]. These sensors capture 22-band images at spatial resolutions of 375m and 750m with a swath width of 3000 km. The DNB has a spectral range of 0.5-0.9μm and has three gain settings which allow for dynamic range of $3 \times 10^{-9} - 2 \times 10^{-2}$ W·cm^{-2}·sr^{-1} with a noise floor of about 5×10^{-11} W·cm^{-2}·sr^{-1} [2]. More details on the data specifications and post-processing of VIIRS DNB images are available in the user guide [3]. VIIRS DNB images have emerged as a valuable source of information for a variety of applications, including urbanization monitoring, disaster risk assessment and response monitoring, regional conflict monitoring, light pollution studies, and power outage detection [4, 5, 6, 7, 8]. Despite the current application of VIIRS DNB images to critical applications, there is continued discussion of opportunities for more effective use of NTL data, implying that NTL images will continue to be at the widely used [9, 10].

While NTL images offer valuable insight into human activity, urbanization, and energy consumption, localized and accurate analysis of these acquired images is often limited by their spatial resolution [11]. Super-resolution methods for NTL images attempt to recover a High-Resolution (HR) NTL image from its Low-Resolution (LR) NTL counterpart, enabling finer-grained analysis and more accurate assessments of infrastructure development, disaster impacts, and light pollution [12]. Figure 1 shows an example of NTL super-resolution, with a LR NTL image on the left and a HR NTL image on the right.

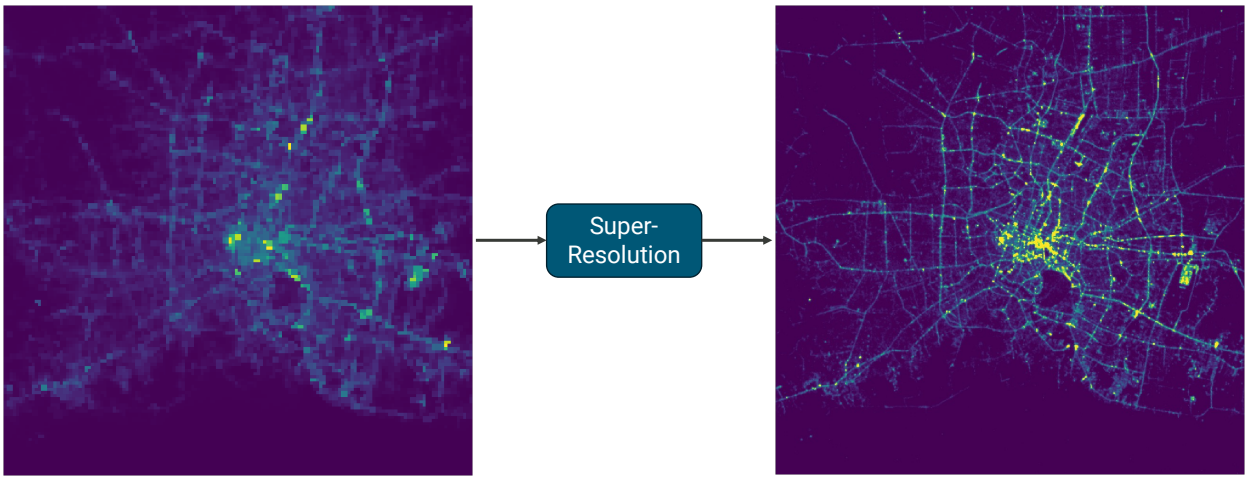


Figure 1. Example of nighttime light super-resolution. A low-resolution nighttime light image on the left is input into a super-resolution method, which generates a high-resolution nighttime light image on the right.

Super-resolution is a commonly researched topic in the image processing and computer vision community, and has demonstrated impressive performance on natural images, which tend to have complex textures, colors, and fine lines (i.e. images that depict real-world scenes such as landscapes, animals, human faces, etc.). However, there is a significant difference between the statistics of these natural images and NTL images, since NTL images generally consist of low-contrast dark regions and a high dynamic range of brightness values. Due to these differences, the performance of state-of-the-art natural image super-resolution methods significantly deteriorates when applied directly to NTL images.

In this work, we first provide a thorough literature review of existing super-resolution methods designed for a variety of image modalities, including natural images, multispectral satellite images, and NTL images. Then, we select a comprehensive set of methods to test on VIIRS DNB data in order to analyze the effectiveness of various super-resolution approaches for NTL imagery. Additionally, we construct a multi-modal super-resolution network designed specifically for VIIRS DNB images which exploits correlation with auxiliary modalities. To the best of our knowledge, this is the first comprehensive review of super-resolution methods for NTL images, which will help to guide future research.

2. PROBLEM FORMULATION

We denote the rasterized unknown HR NTL image that we wish to recover as $x \in \mathbb{R}^{MN}$, where $M, N \in \mathbb{N}$ are the number of rows and columns respectively. Then, we assume that the LR NTL image is given by

$$y = Ax + \epsilon, \tag{1}$$

where $y \in \mathbb{R}^{\frac{MN}{L^2}}$, A is an operator that reduces the resolution of the HR NTL image by a factor of L, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is independent Additive White Gaussian Noise (AWGN).

The operator A is modeled as the composition of a blurring operator and a downsampling operator. The result of applying A to an image is as follows. First, the image is blurred by convolving it with a kernel k. Then, the image is downsampled, in which the image is divided into non-overlapping $L \times L$ blocks, with each block being replaced by a single pixel related to the pixel values in the corresponding $L \times L$ block. In this work, we assume that downsampling is performed using a process known as block averaging, in which the pixel values are averaged over each $L \times L$ block.

The kernel k is related to the Point Spread Function (PSF) of the imaging system, which describes how a point source of light is spread out in the image. The PSF is determined by various factors such as the optics, detector, and atmospheric conditions. While the PSF for satellite imagery can be spatially and temporally variant due to atmospheric conditions, motion of the satellite, or oblique viewing angles, we will assume that the PSF is spatially and temporally invariant for practical purposes. For many super-resolution methods, the PSF is assumed to be known. In these methods, an accurate estimation of the PSF is crucial. In the literature, there are several approaches to estimating the PSF of NTL images. Zheng, Weng, and Wang estimate the PSF by analyzing known point sources of light surrounded by relative darkness, such as gas flares in remote oil extraction sites or ships at sea [13, 14]. Similarly, Xue et al and Sanchez de Miguel et al analyze the spread over known high-contrast edges such as the Ross Ice Shelf in Antarctica, or the Lake Michigan shoreline [14, 15]. On the other hand, Tziokas et al assume k to be a Gaussian kernel based on case studies that they performed [16]. Namely, in two case studies, one of Los Angeles, California, and the other of Delhi, India, the authors found through statistical analysis that a Gaussian kernel with a standard deviation of 572 and 676 meters, respectively, was the best fit. Since we are working with data from the same satellite platform, this is the approach we take in this work. Specifically, we assume that the PSF is a Gaussian kernel with a standard deviation of 625 meters.

The goal of super-resolution is to recover the high-resolution image x from the low-resolution image y using the relationship established in Equation (1). However, the super-resolution problem is severely ill-posed due to the information loss during the blurring and downsampling process. Notably, the solution space is infinite, i.e., there are infinitely many high-resolution images that could have produced the low-resolution image. Moreover, the noise term ϵ further complicates the problem — small perturbations in the low-resolution image can lead to large changes in the reconstructed high-resolution image which are not visually plausible. Thus, a robust inversion method is needed to recover the high-resolution image. Instead of directly inverting the forward model in Equation (1), super-resolution methods seek to recover some map $f : \mathbb{R}^{\frac{MN}{L^2}} \to \mathbb{R}^{MN}$ such that $x = f(y)$. This is done assuming some prior knowledge about the high-resolution image, such as smoothness, sparsity, or some other property. This assumed information can then be incorporated explicitly through regularization techniques, or implicitly through a data-driven approach.

However, these approaches are not without caveats, such as bias, overfitting, and generalization issues.

In the case of multimodal super-resolution, these problems are further eased by incorporating additional sources of data. Namely, in addition to the low-resolution image, we are given some set $\{a_i\}$ of auxiliary modalities (e.g. depth maps, street maps, environmental indices) which are assumed to have some correlation with the input image. Multimodal super-resolution methods seek to leverage this information when recovering the map $f : \mathbb{R}^{\frac{MN}{L^2}} \to \mathbb{R}^{MN}$ such that $x = f(y, \{a_i\})$.

In the following section, we survey existing super-resolution techniques for a variety of image modalities, outlining the approaches they take for dealing with the ill-posed nature of super-resolution.

3. LITERATURE REVIEW OF SUPER-RESOLUTION METHODS

In this section, we outline various super-resolution methods that could be applied to NTL super-resolution. Table 1 gives a high-level overview of each of these methods. We split them into three categories: Single Image Super-Resolution (SISR), Multispectral Image Super-Resolution (MSI-SR), and Multimodal Super-Resolution (MMSR).

3.1 SINGLE IMAGE SUPER-RESOLUTION (SISR)

SISR methods seek to enhance the resolution given only a single low-resolution image. These methods can be classified as unsupervised, semi-supervised, and supervised methods.

Unsupervised SISR methods do not require any training, enabling application when training data is unavailable. In general, these methods either take a zero-shot deep learning-based approach or a model-based approach. Zero-shot deep learning-based approaches rely on self-similarity of images to generate super-resolved images using a deep learning network. On the other hand, model-based unsupervised SISR methods solve an optimization problem using a model of the imaging system along with prior knowledge of the high-resolution distribution.

Supervised SISR methods rely on a training set of paired low- and high-resolution images to learn a mapping from a low-resolution image to the corresponding high-resolution image. They learn this mapping by minimizing a loss function between the predicted super-resolved image and the high-resolution ground truth image. Common loss functions include pixel-based loss such as mean squared error or mean absolute error (L1 loss), total variation loss [47], perceptual loss [48], and adversarial loss [49]. Supervised SISR methods are particularly successful when large amounts of high-quality training data is available. However, it is impractical to collect this training data in many applications, limiting the applicability of supervised SISR methods. Additionally, performance of these methods can deteriorate if the training data is not representative of the testing data (i.e. the testing data is out-of-distribution).

Semi-supervised SISR methods attempt to mitigate the limitations of supervised and unsupervised methods by using unpaired high-resolution images to learn a prior distribution paired with a data consistency term. Generally, these methods includes the incorporation of a deep learning model within a model-based approach.

3.1.1 Unsupervised SISR Methods

Unsupervised SISR methods do not require any training before being applied to a test image. Instead, these methods generally use either a zero-shot deep learning-based approach or a model-based approach. Model-based approaches depend on modeling of the imaging system along with some prior knowledge of the high-resolution distribution. Various priors have been proposed for SISR, including edge sharpening [20], total variation [19], and quadratic regularization [21]. Frequency-based approaches have also been taken using either the discrete Fourier transform [17] or wavelet-based approaches [18, 50].

Zero-shot deep learning-based methods use a single test image to leverage internal self-similarities for learning parameters of a deep learning network, thus exploiting the success of deep learning architectures without requiring paired low- and high-resolution training data. For example, ZSSR [51] exploits internal redundancy in an image to train a small image-specific Convolutional Neural Network (CNN) at inference time. More specifically, ZSSR uses downscaled versions of the

Method	Reference	Method Type	Input Type	Training Data	Auxiliary Inputs	Evaluation Metrics	Code
DFT-based method	[17]	Unsupervised SISR	RGB	–	–	PSNR	–
Wavelet-based method	[18]	Unsupervised SISR	RGB	–	–	–	–
Total Variation Prior	[19]	Unsupervised SISR	RGB	–	–	PSNR	–
Edge Sharpening Prior	[20]	Unsupervised SISR	RGB	–	–	–	–
Quadratic Regularization	[21]	Unsupervised SISR	RGB	–	–	RMSE, PSNR, ISNR, MSSIM	Unofficial
ZSSR	[22]	CNN, Unsupervised SISR	RGB	–	–	PSNR, SSIM	Official
SinGAN	[23]	GAN, Unsupervised SISR	RGB	–	–	RMSE, NIOE	Official
SRCNN	[24]	CNN, Supervised SISR	RGB	ImageNet, Set5, Set14	–	PSNR, SSIM, IFC, NQM, WPSNR, MSSSIM	Unofficial
VDSR	[25]	CNN, Supervised SISR	RGB	Set5, Set14, BSD500	–	PSNR, SSIM	Unofficial
ESPCN	[26]	CNN, Supervised SISR	RGB	ImageNet	–	PSNR	Unofficial
EDSR	[27]	CNN, Supervised SISR	RGB	DIV2K	–	PSNR, SSIM	Official
SRResNet+	[28]	CNN, Supervised SISR	RGB	DIV2K	Noise map	PSNR	Official
SRGAN	[29]	GAN, Supervised SISR	RGB	ImageNet	–	PSNR, SSIM, Human Evaluation (MOS)	Unofficial
ESRGAN	[30]	GAN, Supervised SISR	RGB	Div2K, Flickr2K, OurdoorScene-Training	–	PSNR, Perceptual Index	Official
SR3	[31]	Diffusion, Supervised SISR	RGB	Flickr-Faces-HQ, ImageNet 1K	–	PSNR, SSIM, LR MSE, Human Evaluation	Unofficial
USRNet	[32]	Supervised SISR	RGB	DIV2K, Flickr2K	–	PSNR	Official
SwinIR	[33]	Vision transformer, Supervised SISR	RGB	DIV2K, Flicker2K	–	PSNR, SSIM	Official
Swin2SR	[34]	Vision transformer, Supervised SISR	RGB	DIV2K	–	PSNR, SSIM	Official
Swin2-MoSE	[35]	Vision transformer, Supervised SISR	Remote Sensing Images	Sen2Venμs, OLI2MSI	–	PSNR, SSIM	Official
SupReME	[36]	Unsupervised MSI-SR	MSI	–	–	SRE, SAM	Official
SMUSH	[37]	Unsupervised MSI-SR	MSI	–	–	SNR	–
S2Sharp	[38]	Unsupervised MSI-SR	MSI	–	–	SRE, SAM, RMSE, SSIM	Official
DSen2	[39]	CNN, Supervised MSI-SR	MSI	Reduced Resolution Sentinel-2	–	RMSE, SRE, UIQ	Official
ResSR	[40]	Unsupervised MSI-SR	MSI	–	–	PSNR, SSIM	Official
GWR-based method	[41]	Unsupervised MMSR	NTL	–	NDVI, LST, AWEI, population density	R^2, RMSE	–
BlackMarbleHD	[42]	Unsupervised MMSR	NTL	–	NDWI, NDVI, road maps	–	–
MGWR-based method	[43]	Unsupervised MMSR	NTL	–	NDVI, NDBI, LST, LUCC, road density, and POI density	R^2, RMSE	–
VNCI-based method	[44]	Unsupervised MMSR	NTL	Annual NPP DNB, Luojia1-01	NDVI	RMSE, R^2	–
DeepLight	[45]	Supervised MMSR	NTL	PANDA, Luojia 1-01	MSI, elevation maps, impervious surface	PSNR, SSIM, SAM, CC, PIQE, UIQI	–
NTL-CMDM	[46]	Supervised MMSR	NTL	Black Marble, Luojia 1-01	Road density, impervious surface, highlight features	RMSE, R^2 CV, AD	–

Table 1. Comparison of reviewed super-resolution methods.

test image to learn cross-scale relationships within the image, which are then applied to generate the super-resolved image. SinGAN [23] extends this idea from CNNs to GANs by using a pyramid of convolutional GANs to learn cross-scale relationships within the testing image.

More recently, zero-shot deep learning-based SISR methods using an Implicit Neural Representation (INR) of the image have gained attention in the computer vision community. These methods attempt to learn a continuous representation of a discrete image using a coordinate-based neural network. After learning a continuous representation of a low-resolution image, the INR can then be used to super-resolve the image by inputting a finer pixel grid to the network. Most INRs use a Multilayer Perceptron (MLP) for the neural network. The first proposed INRs used a simple ReLU MLP [52]. However, Rahaman et al [53] showed that these networks prioritize learning low frequencies, a phenomenon called spectral bias. There are generally two approaches for dealing with the spectral bias of ReLU MLPs – encoding the coordinates in a higher-dimensional space or changing the activation function. By encoding the coordinates in a higher-dimensional space, the network is able to express higher frequencies on the manifold of the new space using lower frequencies with respect to the original coordinate space [53]. There are three encoders generally used in literature: basic encoding, positional encoding [54, 55, 56], and random Fourier features [57]. Swapping ReLU with a different activation function can also mitigate the spectral bias of ReLU MLPs. The first proposed alternative activation function was a sinusoidal activation function [58] which enables better representation of higher-order derivatives and higher-frequency detail. Since then, other periodic activation functions have been used, such as Gabor wavelet activation functions [59], HOSC functions [60], and modified sinusoidal activation functions [61]. Non-periodic activation functions have also been proposed, such as Gaussian activation functions [62] and the Sinc activation function [63]. Other INR-based unsupervised SISR methods update the training process based on the super-resolution problem being solved, such as LINR [64] More specifically, they use similar architectures to general INR methods with a modified loss function which incorporates the super-resolution forward model.

3.1.2 Semi-supervised SISR Methods

Semi-supervised SISR methods use unpaired high-resolution images to learn high-frequency detail without requiring access to a paired training set. Generally, these methods use a plug-and-play approach [65] which incorporates a deep learning model as a prior along with a data-fitting term. MDF [66] showed success in super-resolving microscopy imagery using PnP-ADMM with DnCNN [67] trained on high-resolution imagery as the prior. By training the prior on high-resolution images, the algorithm is able to effectively fuse high-frequency information into the super-resolved images. DPSR [28] is a similar method that makes small adjustments to the degradation operator so that a super-resolver can be used as the prior instead of a denoiser. This procedure does require access to paired high- and low-resolution data for training the prior, but is not an end-to-end super-resolution method, which is why we categorize it as a semi-supervised SISR method.

3.1.3 Supervised SISR Methods

The first supervised SISR method used machine learning techniques to learn relationships between low- and high-resolution images without the use of a deep learning network. Common examples use neighbor-based methods [68], sparse representations [69], and example-based approaches [70]. However, with the rise of deep learning, supervised SISR methods have primarily shifted to deep learning-based methods.

The first supervised deep learning-based SISR method was proposed in 2014 [24]. This network, called SRCNN, is a fully convolutional network that learns an end-to-end mapping between in-

terpolated low- and high-resolution images. It takes an interpolated low-resolution image as input and consists of three convolutional layers which extract a set of feature maps, map the feature maps non-linearly to high-resolution patch representations, and combine the predictions within a spatial neighborhood to produce the final high-resolution image. The proposed architecture is guided by reformulating conventional sparse coding-based super-resolution methods into a convolutional network. For training, they use mean squared error as the loss function and minimize the loss using stochastic gradient descent with the standard back-propagation. Their results show improved performance over sparse coding-based methods. Kim et al improved upon SRCNN in their proposed VDSR [25], which uses contextual information across large image regions, exploits residual-learning and gradient clipping for faster training, and allows for multi-scale super-resolution with a single network.

One limitation of SRCNN and VDSR is that the input image is interpolated before being propagated through the network, which greatly increases the computational complexity. To address this, Shi et al proposed an efficient sub-pixel convolutional neural network (ESPCN) which takes the measured low-resolution image and learns upsampling filters for each feature map that are applied at the last layer [26]. This adjustment allows the majority of the computation to be performed in low-resolution space, significantly reducing the computational complexity compared to SRCNN and VDSR. Results show that ESPCN can significantly increase the speed and reduce the complexity compared to other super-resolution methods, enabling super-resolution of HD videos in real time on a single GPU.

Residual learning, originally proposed in ResNet [71], has been very successful in high-level computer vision problems like classification and detection. In 2017, Ledig et al applied residual learning to super-resolution in their proposed SRResNet [29]. However, Lim et al noted that modifications must be made to this architecture to optimize the network for super-resolution. They propose an enhanced deep super-resolution network (EDSR) which removes unnecessary modules from the SRResNet structure [27]. Namely, they remove batch normalization layers, which increases performance and decreases GPU memory usage. They also propose a multi-scale model which uses scale-specific pre-processing and upsampling modules. Results on public benchmarks datasets demonstrate better performance than SRCNN, VDSR and SRResNet.

Ledig et al harnessed the promise of generative adversarial networks (GANs) in their super-resolution GAN, called SRGAN [29]. They use a deep residual network, SRResNet, for the generator with a novel perceptual loss rather than the traditional MSE loss (which tends to create blurry images). More specifically, the perceptual loss is calculated as a weighted sum of a content loss and an adversarial loss, where the content loss is calculated using high-level feature maps of a VGG classification network [72]. Results demonstrated state-of-the-art performance (as of 2017) on benchmark datasets for $4\times$ super-resolution in terms of PSNR, SSIM, and mean opinion score.

Wang et al proposed an improvement on the original SRGAN called Enhanced SRGAN (ESRGAN) [30]. They improve the original model in three specific ways – the network structure, the discriminator, and the perceptual loss. For the network structure, they introduce the Residual-in-Residual Dense Block (which combines multi-level residual network and dense connections), remove Batch Normalization layers (as in [27]), and use residual scaling and smaller initialization. For the discriminator, they use a relativistic average GAN (RaGAN) [73] which helps the generator recover more realistic texture detail. Finally, they improve the perceptual loss by using the VGG features before activation instead of after activation. ESRGAN is shown to consistently achieve better perceptual quality than previous SR methods, including SRGAN. The method won the first place in the PIRM-SR Challenge in terms of the perceptual index.

SwinIR [33] is a transformer-based image restoration method, which produces state-of-the-art results on super-resolution, denoising, and JPEG compression artifact reduction tasks. It is built on the Swin Transformer [74], which uses a shifted window mechanism to produce hierarchical representations of images, thus reducing complexity of the method and allowing learning at various scales. Swin2SR [34] extends SwinIR using SwinV2 [75], which is a modified version of the Swin transformer that better scales model capacity and window resolution. Swin2SR produces results comparable to SwinIR with 100k fewer training iterations. Swin2-MoSE [35] extends Swin2SR by replacing the Feed-Forward inside all transformer blocks with an enhanced Mixture-of-Experts (MoE), using positional encoding with a combination of per-head and per-channel bias, and training with a combination of Normalized-Cross-Correlation and SSIM losses. The presented results show slight improvements over Swin2SR and SwinIR.

USRNet [32] is a deep unrolled super-resolution network that bridges the gap between model-based and learning-based methods. While learning-based methods have been proven to be better than model-based methods in effectiveness and efficiency, model-based methods are more generally applicable since they can directly handle different scale factors, blur kernels and noise levels. USRNet combines the advantages of both approaches by using a deep unrolled network to learn the mapping from low-resolution to high-resolution images, while incorporating the problem-specific scale factor, blur kernel, and noise level. Algorithms such as plug-and-play and their unrolled variants can be limited by a number of hyperparameters, which may require tuning for each specific problem. To mitigate this, USRNet includes a hyperparameter tuning network which learns the optimal hyperparameters for the specific problem at hand during the course of training. USRNet reaches state-of-the-art performance on benchmark datasets while remaining flexible like model-based methods. However, the method requires a known and fixed degradation model, which may not be available in all applications.

HCFlow [76] is a unified framework for image super-resolution and image rescaling that learns a bijective map between HR and LR image pairs by modeling the LR image and the high-frequency component distributions of the HR image simultaneously. It combines the methods from two previous papers: SRFlow [77] and IRN [78]. First, by normalizing flow, HCFlow learns a distribution of HR images from a LR image. Next, by conditioning the high-frequency components on the LR image in a hierarchical manner, it enhances reconstruction quality. Beyond leveraging normalizing flow, HCFlow improves training by additionally using L1 loss, perceptual loss, and adversarial loss. However, as with most generative models, HCFlow suffers from hallucination artifacts and bias from the training data, which are inadmissible in critical applications.

Super-Resolution via Repeated Refinement (SR3) [31] is a conditional generative diffusion model that uses a series of invertible transformations to map Gaussian noise to a high-resolution image. Using a U-net architecture, the model learns to transform a standard normal distribution into an empirical data distribution through a sequence of refinement steps, resembling Langevin dynamics. SR3 works on a range of magnification factors and input resolutions, and can be easily cascaded to achieve higher magnification to obtain impressive 16x super-resolution results. However, this method is prone to bias; experiments on real-world data show that the model does not reliably generate textures which often are the most important features to resolve.

In imaging scenarios where the blur kernel is unknown, or changes over time, blind super-resolution methods are necessary. Blind methods work without knowing the forward model beforehand – either by training on a variety of degradations or by estimating the degradation along with the HR image.

BSRGAN [79] is a blind generative method for image super-resolution trained on a suite of realistic SISR degradations. First, they randomly shuffle the degradation sequence (i.e. blur, downsampling, noise, JPEG compression artifacts, etc). Then they randomly sample from the implemented downsampling methods as well as various parameters like kernel width, rotation angle, and noise standard deviation. From this, they generate a number of different LR candidates for each given HR image. This variety of LR images ensures that the trained model is robust and more broadly applicable to real degradation models.

Iterative Kernel Correction (IKC) [80] is a blind super-resolution method that estimates the blur kernel and the high-resolution image simultaneously. When incorrect kernels are used, regular and predictable artifacts, such as over-sharpening or over-smoothing, are introduced into the super-resolved image. This information can be leveraged to better estimate the blur kernel and the high-resolution image. IKC is composed of three networks: a kernel estimation network, a high-resolution image estimation network, and a kernel correction network. These networks are applied iteratively eight times to refine the kernel and the high-resolution image. It is then trained in an end-to-end manner to minimize the error between the estimated high-resolution image and the ground truth. The training is done using known Gaussian blur kernels with varying sizes and standard deviations, while other blur kernels which are more common in real-world scenarios, such as motion blur, are said to be the subject of future work.

3.2 MULTISPECTRAL IMAGE SUPER-RESOLUTION (MSI-SR)

MSI-SR methods enhance the lower spatial resolution spectral bands of a Multispectral Image (MSI) by leveraging the high spatial resolution spectral bands of the same MSI, exploiting the correlation between the LR and HR spectral bands. The VIIRS satellite captures the panchromatic NTL image as well as 5 high-resolution spectral bands and 16 low-resolution spectral bands. In theory, MSI-SR methods can be used to super-resolve the NTL image by leveraging the 5 high-spatial resolution bands. Similar to SISR methods, these can be split into unsupervised and supervised methods.

3.2.1 Unsupervised MSI-SR Methods

Unsupervised MSI-SR methods generally take a model-based approach, which solve for the super-resolved image using a Bayesian framework. Specifically, they minimize a cost function that consists of a data-fitting term, which models the physics of the imaging system, and a regularization term, which models the prior distribution of expected high-resolution images. Since the spectral bands are highly correlated, some model-based MSI methods use singular value decomposition to reduce the dimensionality. The resulting optimization problem solves for the representation coefficients for the new subspace representation. One such example is SupReME [36], which uses an L2 norm of spatial derivatives for regularization and minimizes the cost function iteratively using C-SALSA [81]. SMUSH [37] extends upon SupReME by hierarchically super-resolving the various resolutions and including a BM3D plug-and-play-based regularizer [82]. S2Sharp [38] also extends on SupReME by minimizing both the subspace representation coefficients and the basis vectors for the subspace. Other methods don't explicitly reduce the dimensionality, but rather use a learned sparse representation or enforce a low-rank solution [83, 85, 86, 87, 88, 89, 90, 91, 92, 84]. While these iterative methods can attain high-quality super-resolution results, they are often computationally expensive, limiting their scalability. Recently, a computationally efficient model-based method, ResSR [40], was proposed. By using only pixel-wise regularization and approximating the downsampling operator, ResSR significantly reduces the computation complexity compared to alternative methods, enabling scalable processing of large data sets without loss of reconstruction quality.

3.2.2 Supervised MSI-SR Methods

Supervised MSI-SR methods generally use a neural network to learn the relationship between low-resolution and high-resolution bands of an MSI. Various network architectures have been proposed for this problem, ranging from CNNs to GANs. DSen2 [39] is a supervised CNN trained on reduced resolution Sentinel-2 data, which consists of separate networks for different super-resolution factors. RS-ESRGAN [93] is a similar supervised method that uses the ESRGAN architecture [30] instead of a CNN. MoG-DCN [94] is an unrolled network that attempts to bridge the gap between model-based and deep learning-based MSI-SR methods by unrolling each iteration of the proximal gradient method into a layer in a network.

3.3 MULTIMODAL SUPER-RESOLUTION (MMSR)

MMSR refers to the process of enhancing the resolution of an image by integrating information from multiple sources of information, such as different imaging modalities. These methods are particularly useful when multiple, highly-correlated data sources are available. In this section, we will restrict our discussion to MMSR methods for NTL super-resolution. In this context, the goal is to enhance the spatial resolution of a NTL image by leveraging images from higher spatial resolution auxiliary data, such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Land Surface Temperature (LST), and road networks. Similar to SISR and MSI-SR methods, MMSR methods can be split into two categories: unsupervised and supervised methods.

3.3.1 Unsupervised MMSR Methods

Unsupervised methods generally use spatial regression to determine a spatially-varying relationship between the NTL image and the downsampled auxiliary images. Then they apply this relationship using the HR auxiliary images to reconstruct the super-resolved NTL image, assuming that this relationship is scale-invariant [95, 96]. The National Aeronautics and Space Administration (NASA) has developed a method, called BlackMarbleHD, which combines NTL images, upsampled by cubic convolution interpolation, with NDVI, NDWI, and road maps, to produce a HR NTL image [42]. Similarly, Ye et al propose an unsupervised MMSR NTL super-resolution method based on Geographically Weighted Regression (GWR) [41]. GWR extends standard regression by allowing the relationships to vary over space. This method uses the NDVI, LST, Automated Water Extraction Index (AWEI), and population density data. Liu et al [43] extends this method by using Multiscale Geographically Weighted Regression (MGWR) instead of GWR, which allows the use of an spatially-adaptive kernel. This method uses NDVI, Normalized Built-Up Index (NDBI), LST, Land Use & Cover Change (LUCC), road density, and Point of Interest (POI) density data. In a similar vein, [44] constructs a joint distribution of the NTL and NDVI images, called Vegetation Nighttime Condition Index (VNCI), and uses this distribution to estimate a HR NTL image from given LR NTL and HR NDVI images.

3.3.2 Supervised MMSR Methods

Unlike unsupervised methods, supervised approaches learn the relationship between the LR NTL image, the auxiliary images, and a corresponding HR NTL image. In [46], the authors propose a deep learning-based MMSR method called NTL-Conditional Multiscale Downscaling Model (CMDM) that employs CNN blocks to extract features from the LR NTL image and road networks, impervious surfaces, and a custom highlight feature map. Additional CNN blocks then model the relationships between these features, producing a super-resolved NTL image as the final output. Similarly, DeepLight [45] utilizes MSI, elevation maps, and impervious surface data

to super-resolve NTL images. DeepLight also introduces a learned alignment module that learns to align the input LR NTL image with the auxiliary images. The alignment module consists of a learned downsampling operator to reduce the resolution of the auxiliary images, and a transformed-based alignment network to align the LR NTL image with the downsampled auxiliary images. The features of each image are then extracted via a series of parallel residual networks with cross-modality connections, and then fused together to produce the super-resolved NTL image.

4. PROPOSED MMSR METHOD

While there are existing MMSR methods for NTL images (as outlined in Section 3.3), none of them have public code or datasets, making their results very difficult to replicate. Instead, we use the foundational ideas from existing supervised MMSR methods to create our own MMSR model and dataset consisting of LR VIIRS DNB images, HR Luojia 1-01 images, HR OpenStreetMap (OSM) images, and HR land use data from DynamicWorld. In this section, we first outline the data that we use for our model. Then, we present an overview of the model architecture and implementation details.

4.1 DATA

At inference time, our MMSR model requires a LR NTL image, along with two registered HR auxiliary images. For training, our model also requires a paired HR NTL image that is used as ground truth for computing the loss. Our LR NTL dataset consists of images from the Black Marble product suite and our HR NTL dataset consists of images from the Luojia 1-01 satellite. Additionally, our two HR auxiliary datasets consist of road networks from OSM and land use data from DynamicWorld. The entire dataset contains 418 unique location-time pairs, each consisting of a HR NTL image, a LR NTL image, a road network image, and a land use image. We split this dataset into a training, validation, and testing dataset each with 281, 67, and 70 images respectively.

While the image sizes vary, the LR NTL images are approximately 600×600 pixels, whereas the HR NTL and auxiliary images are roughly 2400×2400 pixels. Figure 2 shows an example of each image modality incorporated in our dataset. The top row of shows the auxiliary images from OSM and DynamicWorld, while the bottom row shows the LR and HR NTL images from Black Marble and Luojia 1-01 respectively. The full-sized image is shown on the left, while the right side shows a zoomed-in view of the region in the red box. The remainder of this section outlines the data specifications and preprocessing used for each of these image modalities.

4.1.1 LR NTL Dataset

For our LR NTL dataset, we use NASA's Black Marble Nighttime Lights product [3], downloaded from the NASA Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center (LAADS DAAC)[1] using the Python library blackmarblepy [97].

The Black Marble product suite provides reliable and accurate global NTL images at 500 m resolution from January 2012 to the present. It addresses several issues with using raw DNB data from VIIRS, such as the need to correct for atmospheric effects, light from the moon, light from snow cover, and other sources of aberrations. Moreover, it fills missing data gaps by interpolating between nearest available data points in time. NASA provides several versions of the Black Marble product, including daily, monthly, and yearly composites. Since some applications of this work require high temporal resolution, we use the VNP46A2 product, which is a daily composite of the corrected NTL radiance values in units of $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$.

In order to ensure that the LR NTL images are suitable for training, we preprocess the images in three steps. First, we remove the background noise by setting all values below 1.5 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$

[1] https://www.earthdata.nasa.gov/centers/laads-daac

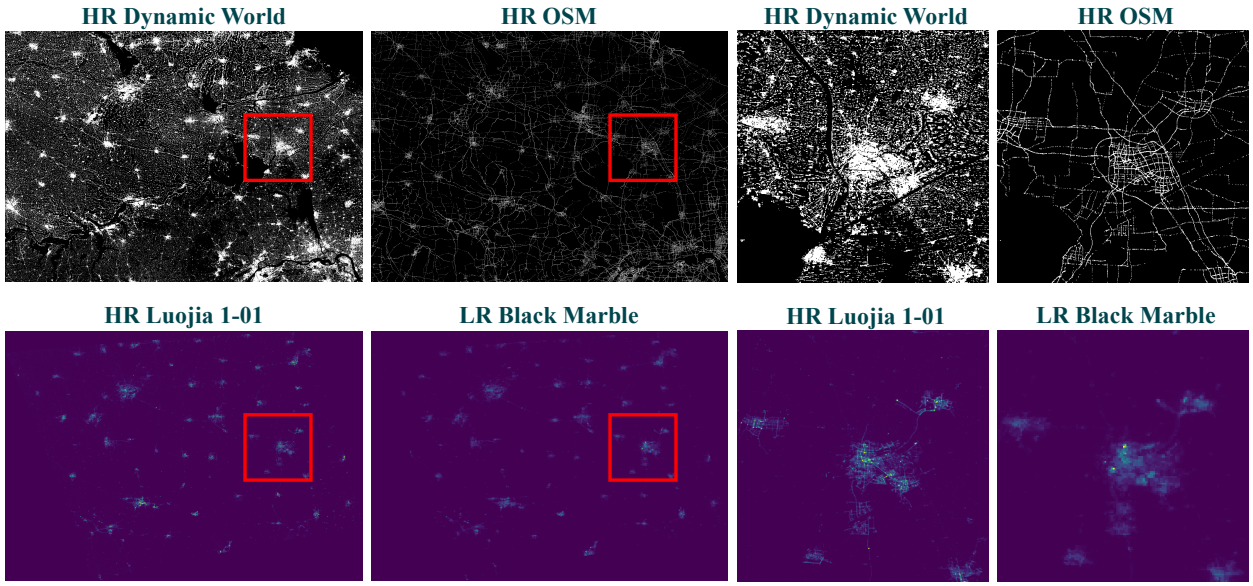


Figure 2. Example images of the four modalities used in our dataset. The top row shows the road network and land use images, while the bottom row shows the LR and HR NTL images. The left side shows the full-sized image, while the right side shows a zoomed-in view of the region in the red box.

to zero. This threshold was determined by examining the distribution of radiance values in regions with no artificial light sources. Second, we clip the maximum value of the images to 300 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$ to prevent outliers from skewing the data, as done in [44]. Finally, we normalize the images by the mean and standard deviation radiance of the entire training set, which were found to be 1.15 and 6.85 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$ respectively.

4.1.2 HR NTL Dataset

For our HR NTL dataset, we use the Luojia 1-01 NTL dataset from the HR Earth Observation System Hubei Data and Application Network [98]. This satellite was launched in 2018 and was operational until mid-2019; it orbited Earth at an altitude of 645 kilometers and had a revisit time of 15 days. This dataset consists of global NTL images at 130 m GSD captured on a 2048 $\times$ 2048 CMOS sensor, covering a spectral range from 480 nm to 800 nm at 15-bit radiometric resolution.

The data is stored as a Digital Number (DN) image, which is a raw image that has not been radiometrically calibrated. To convert the DN image to radiance in units of $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$, we use the following formula [99]:

$$\text{radiance} = \text{DN}^{3/2} \times 10^{-10} \times 0.52, \tag{2}$$

where the factor of 10^{-10} is a conversion factor from DN to radiance, and 0.52 is the bandwidth of the sensor in micrometers.

Originally, the Luojia 1-01 NTL dataset contains 8671 images. However, a significant portion of these images have partial cloud cover or other aberrations that make them unsuitable for training. In order to select a suitable subset of training images, we manually inspected a random sample of 1000 images. Using 800 of these, we trained a logistic regression classifier on feature vectors extracted from ResNet50 [100] to predict whether an image is suitable for training based on the

image's features. The trained classifier was 86% accurate on the remaining 200 images. We then used this classifier to predict the suitability of the remaining images in the dataset. Of those images the classifier predicted to be suitable, we manually selected a subset of 458 images to use for training and testing. The location and time of these images are used as the base for extracting the corresponding LR NTL images and auxiliary datasets.

Similar to the LR NTL dataset, we preprocess the HR NTL images by removing background noise, clipping the maximum value, and normalizing the images by the mean and standard deviation of the training set. In this case, the noise threshold was set to 5 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$, as was done in [46], the maximum value was clipped to 500 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$, and the mean and standard deviation were found to be 1.11 and 10.81 $\text{nW}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$ respectively.

4.1.3 HR Auxiliary Datasets

In addition to the HR NTL dataset, we use two HR auxiliary datasets – road networks and land use data. These datasets serve as auxiliary inputs to our MMSR model, enhancing the quality of the super-resolved images by providing additional regional information. They were specifically chosen because they are publicly available, have global coverage, and are highly correlated with NTL data.

For the road network dataset, we use OSM [101], a comprehensive open-source dataset containing roads, buildings, and other geographic features. OSM is structured as a network of nodes, which contain the geographic coordinates of the feature, and edges which define connections between nodes. Additionally, the data contains information about the type of feature, such as the type of road, building, or land use. Since roads are the primary features visible in NTL imagery, we filter the OSM data to only include road nodes classified as motorways, trunks, primary, and secondary roads. For each region in the HR NTL dataset, we extract the OSM data for that region and convert it to an image format. Each pixel in the image represents the number of road nodes within that pixel's area. We then create a binary image where each pixel is 1 if the number of road nodes is greater than 0, and 0 otherwise.

For land use data, we use the DynamicWorld dataset [102], which classifies land cover into 10 categories, including urban areas, forests, and water bodies. Derived from Sentinel-2 satellite imagery, DynamicWorld provides land use classifications at a 10 m resolution. Using Google Earth Engine, we extract the land use data for each region in the HR NTL dataset at the same resolution as the HR NTL images (125 m). The 'built' land use class strongly correlates with NTL, while other classes tend to correlate negatively. To leverage this information, we generate a binary image where pixels are set to 1 if the predominant land use classification is 'built' and 0 otherwise.

4.1.4 Image Alignment

The four images (LR NTL, HR NTL, road network, and land use) are not perfectly aligned due to differences in the sensors, resolutions, and geolocation errors. To address this, we use SimpleITK [103] to perform image registration by computing the affine transformation that minimizes misalignment between the LR NTL images, HR NTL images, and auxiliary images. We use the HR NTL image as the reference image and align the other images to it. The registration process minimizes the mean squared error between the images with respect to the transformation parameters by gradient descent, and stops when the norm of the gradient is less than 10^{-6}, ensuring numerical convergence. We validated the alignment by visually inspecting the images. This alignment process is essential for MMSR, as it ensures that corresponding features across the images are correctly mapped, allowing the model to leverage auxiliary information effectively.

4.2 MODEL ARCHITECTURE

Figure 3 shows a pipeline of our proposed MMSR method. Our method is a multiscale super-resolution method, which does 4× super-resolution by first super-resolving by a factor of 2 and then super-resolving by a factor of 2 again. When compared to directly doing 4× super-resolution, this approach has several advantages including a reduction in computational complexity, the ability to learn finer details, and more stable training. For each 2× super-resolution step, the LR NTL image is first upsampled by a factor of 2 using bicubic interpolation. Then, the HR auxiliary images are rescaled to match the LR NTL resolution using bicubic downsampling. Finally, all three of these images are input to the "2× super-resolution module," which extracts features from each image, fuses the features, and performs artifact reduction to generate the 2× super-resolved image.

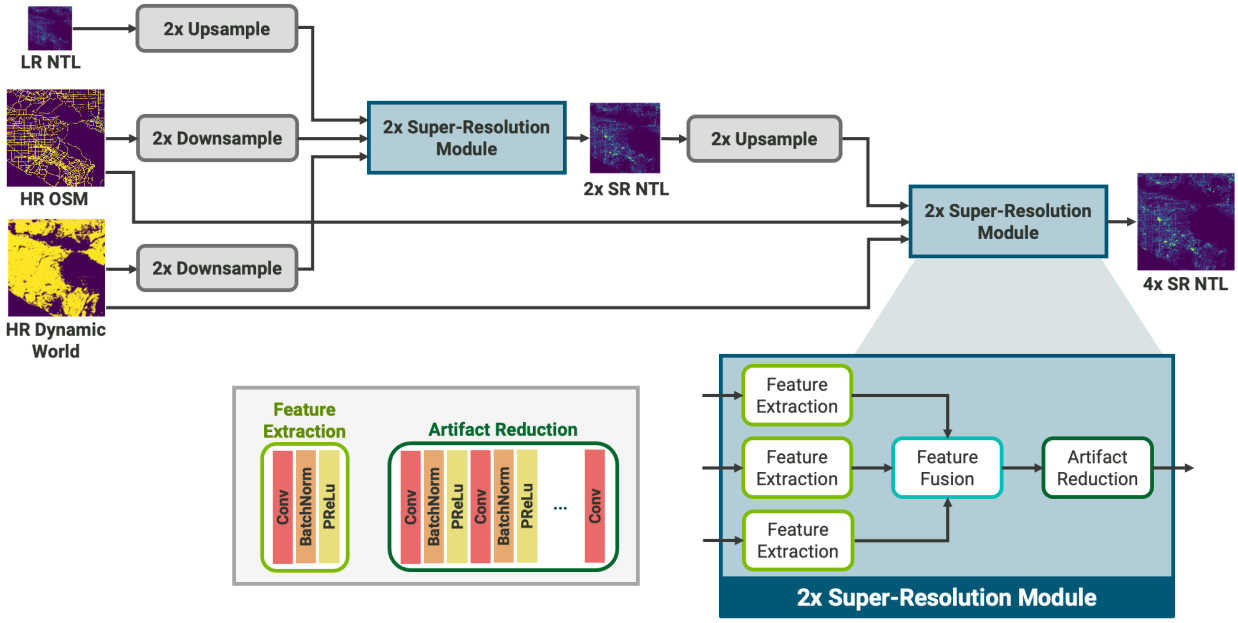


Figure 3. Pipeline of proposed MMSR method, which uses a multi-scale architecture with feature extraction, feature fusion, and artifact reduction. The feature extraction module consists of one convolutional layer with batch normalization and a parametric ReLU activation function. The feature fusion module is concatenation. The artifact reduction module consists of 8 convolutional layers with batch normalization and a parametric ReLU activation function.

There are various approaches for the architecture used to do feature extraction, feature fusion, and artifact reduction. In our MMSR model, the feature extraction module has one convolutional layer with batch normalization and a parametric ReLU activation function. The convolutional layer consists of 64 feature maps with 3×3 convolutional kernels. The feature fusion module concatenates the features of the three input images. Finally, the artifact reduction module consists of eight convolutional layers, where all but the last have batch normalization and parametric ReLU activation functions. The convolutional layers consist of 64 feature maps with 3×3 convolutional kernels.

To train the model, we compute a multiscale loss which includes the loss for 2× and 4× super-resolution. Namely, we compute the 2× loss after the first 2× super-resolution module by comparing the 2× super-resolved LR NTL image with the 2× downsampled HR NTL image. Additionally, we compute the loss after the second 2× super-resolution module by comparing the 4× super-resolved LR NTL with the HR NTL image. Then the final loss is a weighted sum of the two losses:

$$\text{loss} = \gamma \text{loss}_{2\times} + \text{loss}_{4\times}. \tag{3}$$

Experimentally, we use $\gamma = 1.0$, weighting the 2× loss equally with the 4× loss. The two losses, $\text{loss}_{2\times}$ and $\text{loss}_{4\times}$, are each computed as a weighted sum of the L2 and TV loss at the corresponding scale. Formally, if $\hat{x}$ and x are the super-resolved image and ground truth image respectively, then

$$\text{loss}(\hat{x}, x) = \alpha_{L2} L2(\hat{x}, x) + \alpha_{TV} TV(\hat{x}, x), \tag{4}$$

where $L2(\hat{x}, x), TV(\hat{x}, x)$ are the L2 and TV losses and α_{L2}, α_{TV} are weights corresponding to the L2 and TV loss. For our experiments, we set $\alpha_{L2} = 1.0$ and $\alpha_{TV} = 0.1$. The L2 loss is defined as:

$$L2(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - x_i)^2, \tag{5}$$

where N is the number of pixels in the batch and $\hat{x}_i$ and x_i are the pixel values of the super-resolved and ground truth images respectively. The TV loss is defined as:

$$TV(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^{N} \sqrt{|\hat{x}_{i+1,j} - \hat{x}_{i,j}|^2 + |\hat{x}_{i,j+1} - \hat{x}_{i,j}|^2 + \varepsilon}, \tag{6}$$

where i and j are the pixel indices in the horizontal and vertical directions respectively, and $\varepsilon = 10^{-6}$ ensures numerical stability. The TV loss is used to encourage spatial smoothness in the super-resolved image, which is important for reducing artifacts and improving the overall quality of the image.

We extract 100 patches of size 100×100 from each image in the training and validation datasets introduced in Section 4.1. The patches are randomly selected from the images, where the probability of selecting a patch center is proportional to the average radiance in a small region around the patch center. This ensures that the patches are selected from regions with a high density of artificial light sources, which are the primary features of interest in NTL images. Using these training and validation patches, we run a training loop for 250 epochs with an initial learning rate of 1×10^{-3} that is gradually decreased using exponential scheduling with decay 0.99. At each epoch, we backpropagate the loss computed over the training dataset and save the average Structural Similarity Index Measure (SSIM) over the validation patches. After all 250 training epochs, we save the model with the best validation SSIM to use as our final model for testing.

5. EXPERIMENTAL RESULTS

We chose a variety of super-resolution methods with open-source code from Section 3 to compare performance, as outlined in Table 2. We do not include any MSI-SR methods because they depend on high correlation between the low-resolution nighttime light image and the high-resolution bands, which we have observed is not the case, causing severe deterioration in results. Additionally, we do not include any unsupervised MMSR methods because they are generally not as effective as supervised methods.

5.1 IMPLEMENTATION DETAILS

We trained all of the methods using the same paired Black Marble and Luojia 1-01 images from the dataset in Section 4.1 for 50 epochs. Our training was performed on one NVIDIA A100 GPU with 80GB of memory with resources provided by ORNL Research Cloud Infrastructure at the Oak Ridge National Laboratory. For all methods, we extract 100 patches of low-resolution size 128×128 (except for ESRGAN which uses patches of size $64\times$ 64) selected with the same distribution as our MMSR method. Unless specified otherwise, the training optimizer is an ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, and no weight decay. The rest of this subsection outlines the method-specific parameters that we used.

The plug-and-play method runs for 50 iterations with hyperparameter $\mu = 50$. The prior that we use is a BM3D denoiser for AWGN with standard deviation 0.01.

MDF also runs for 50 iterations with hyperparameters $\mu = 50$ and $\sigma = 0.01$. The prior is a DRUNet model trained to remove AWGN with $\sigma = 0.1$ using the architecture proposed in [106]. The model is trained for 50 epochs with an L1 loss using a learning rate of 1×10^{-4}.

DPSR runs for 50 iterations with hyperparameters $\mu = 50$ and $\sigma = 0.005$. The prior is the SRResNet+ model, with implementation details described below.

The architecture of SRCNN is the same as what was proposed in [24], including kernel size, number of filers, and activation, with an Mean Squared Error (MSE) loss. We use a learning rate of 1×10^{-5} for the entire training process.

We use the proposed EDSR architecture from [27] with an MSE loss. We use an initial learning rate of 1×10^{-4}, which decreases by a factor of $\frac{1}{2}$ at epochs 3, 6, 13, and 20.

Method	Reference	Method Type
Plug-and-Play	[104]	Model-based, Unsupervised SISR
MDF	[66]	Model-based, Semi-supervised SISR
DPSR	[28]	Model-based, Supervised SISR
SRCNN	[24]	CNN, Supervised SISR
EDSR	[27]	CNN, Supervised SISR
ESRGAN	[30]	GAN, Supervised SISR
SRResNet+	[105]	Supervised SISR
MMSR	–	Supervised MMSR

Table 2. Super-resolution methods used for testing

Our ESRGAN network has the same architecture as originally proposed in [30]. The loss consists of an L1 loss, an perceptual loss and an adversarial loss, weighted as 0.01, 1, and 0.005 respectively. For the discriminator, we use a learning rate of 1×10^{-5}. For the generator, we use an initial learning rate of 1×10^{-4}, which decreases by a factor of $\frac{1}{2}$ at every 10 epochs.

Most of the SRResNet+ architecture is the same as what was proposed in [30], but it additionally includes a noise map as input to include denoising in the network as done in [28]. The noise map consists of AWGN with standard deviation randomly chosen in the range $[0, 0.1]$. The loss is MSE. We use an initial learning rate of 1×10^{-4}, which decreases by a factor of $\frac{1}{2}$ at epoch 40.

5.2 METRICS

We compute a variety of metrics to assess the quality of a super-resolved image $\hat{x}$ based on the input Black Marble image y and the ground truth Luojia 1-01 image x. Namely, we compute the Normalized Root Mean Square Error (NRMSE), LR consistency, SSIM, and Feature Similarity Index Measure (FSIM). The NRMSE is given by

$$\text{NRMSE}(x, \hat{x}) = \frac{\|\hat{x} - x\|}{\|x\|}. \tag{7}$$

Ideally, this value should be close to zero, indicating that the super-resolved image is close to the ground truth image in terms of pixel intensity. The LR consistency is given by

$$\text{LR Consistency}(y, \hat{x}) = \frac{\|A\hat{x} - y\|}{\|y\|}. \tag{8}$$

The LR consistency metric measures how closely the super-resolved image $\hat{x}$ produces the input image y when downsampled by the matrix A as defined in Section 2. A low value of this metric indicates that the super-resolved image is consistent with the data collected. The SSIM metric measures the perceptual similarity between two images based on luminance, contrast, and structure. The SSIM is given by

$$\text{SSIM}(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \tag{9}$$

where μ_x and $\mu_{\hat{x}}$ are the mean intensities of x and $\hat{x}$ respectively, σ_x and $\sigma_{\hat{x}}$ are the standard deviations of x and $\hat{x}$ respectively, $\sigma_{x\hat{x}}$ is the covariance between x and $\hat{x}$, and C_1 and C_2 are small constants which prevent division by zero. We use the scikit-image implementation[2]. Similarly, the FSIM is another perceptual metric that measures the similarity between two images based on the low-level features [107]. We calculate the FSIM using the provided implementation[3]. For both the SSIM and FSIM, a value close to 1 indicates that the super-resolved image is similar to the ground truth image.

5.3 SUPER-RESOLUTION RESULTS

In this section, we show the results of the super-resolution methods from Table 2 on 7 test images: 00069, 00118, 00144, 00277, 00421, 00423, and 00451. These images were randomly selected from the test set of 70 images. In order to show the details of the super-resolved images, we show only a small patch of the full image.

[2]https://scikit-image.org/docs/stable/api/skimage.metrics.html
[3]https://github.com/nekhtiari/image-similarity-measures

Figures 4–10 show the input auxiliary images (land-use and street maps), the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image, as well as the super-resolved images from each of the methods shown in Table 2. The super-resolved images presented in these figures showcase the tendencies of the different methods. The Plug-and-Play and MDF methods tend to produce images similar to the input Black Marble image, with very little improved quality, while the DPSR and ESRGAN methods produce images with patterns which do not resemble natural nighttime light images. The SRCNN and EDSR methods produce images with some improved quality, and even begin to extract some features of the HR NTL images, such as roads and waterways, but they still have some artifacts and do not resemble the ground truth images. This is particularly clear in Figure 5, 9 and 10, where the road-like structures are pronounced in the super-resolved images, but they do not match the HR Luojia 1-01 images. The SRResNet+ method produces images generally sharper than the SRCNN and EDSR methods, with improved quality. However, this method struggles to produce fine details in dense urban areas, as evident in Figure 6 and 7. Our MMSR method produces the highest quality images, with the most natural appearance and the least artifacts. Our MMSR method is able use the contextual information from the land-use and street maps to improve the super-resolved images — learning the relationship between the auxiliary images and the nighttime light images. This allows the network to produce images with more features that resemble the HR Luojia 1-01 image, while also maintaining the overall structure and appearance of the input nighttime light image. Importantly, the network generally does not hallucinate features in the auxiliary images that are not present in the NTL images. For example, in Figure 6, the auxiliary images show a large region of built-up area with roads, but both NTL images show no light in this region. The super-resolved image from the MMSR method does not produce any light in this region.

In Figure 11, we show scatter plots of the pixel intensity of the Black Marble and super-resolved images against the ground truth Luojia 1-01 image for the test images 00069, 00118, 00144, 00277, 00421, 00423, and 00451. For each test image, we show the scatter plot of the pixel intensity of the Black Marble image (left) and the super-resolved image using our MMSR method (right) against the ground truth Luojia 1-01 image. The x-axis is the pixel intensity of the Black Marble or super-resolved image and the y-axis is the pixel intensity of the ground truth Luojia 1-01 image. The color indicates the number of pixels in each bin. We also show a line of best fit (in red) for each scatter plot, which is computed as the least squares regression line. The black dashed line indicates a perfect prediction (i.e. the predicted pixel intensity is equal to the ground truth pixel intensity). The scatter plots and their line of best fit show that the super-resolved images using our MMSR method are more closely aligned with the ground truth than the Black Marble images, indicating that our method is able to produce high-quality super-resolved images that are more consistent with the ground truth.

In Table 3, we show the image quality metrics for each of the test images. We show the NRMSE, LR consistency, SSIM, and FSIM for each of the considered methods. In Table 4, we show the average image quality metrics for all of the test images. Note that the images in Table 3 are computed over the small patches that are displayed in figures, so they contain less background area than the full-size testing images. The NRMSE and LR consistency metrics are known to favor blurrier images, a phenomena known as the variance-bias tradeoff. This explains the fact that EDSR, SRCNN, and SRResNet+ are achieving lower values for these metrics compared to MMSR, which generates sharper images. On the other hand, SSIM and FSIM metrics are higher for images that have similar structure to the ground truth, which is consistent with the qualitative SRResNet+ and MMSR results we have observed. It is important to note that while these metrics are commonly used for evaluating natural image super-resolution techniques, they may not ade-

Test Image	Metric	PnP	MDF	DPSR	SRCNN	EDSR	ESRGAN	SRResNet+	MMSR
00069	NRMSE (↓)	0.856	0.856	0.914	0.863	**0.853**	0.896	0.873	0.856
	LR Cons. (↓)	0.224	0.223	0.465	**0.118**	0.121	0.235	0.178	0.136
	SSIM (↑)	0.901	0.902	0.905	**0.915**	0.906	0.895	0.917	0.901
	FSIM (↑)	0.622	0.624	0.565	**0.643**	0.614	0.611	0.630	0.536
00118	NRMSE (↓)	0.921	0.921	1.003	0.738	0.734	1.042	0.712	**0.688**
	LR Cons. (↓)	0.220	0.220	0.553	**0.116**	0.124	0.205	0.251	0.175
	SSIM (↑)	0.780	0.790	0.831	0.849	0.835	0.810	0.865	**0.881**
	FSIM (↑)	0.539	0.542	0.491	0.596	0.581	0.556	0.597	**0.606**
00144	NRMSE (↓)	0.920	0.920	0.957	0.929	0.929	0.945	0.937	0.937
	LR Cons. (↓)	0.238	0.239	0.369	**0.068**	0.072	0.218	0.102	0.099
	SSIM (↑)	0.759	0.769	0.787	0.811	0.797	0.779	0.806	**0.824**
	FSIM (↑)	0.481	0.495	0.452	0.534	0.518	0.503	0.519	**0.537**
00277	NRMSE (↓)	0.857	0.857	0.999	0.805	0.797	0.947	0.782	**0.781**
	LR Cons. (↓)	0.239	0.239	0.650	**0.112**	0.121	0.260	0.185	0.177
	SSIM (↑)	0.433	0.437	0.396	0.572	0.590	0.451	0.582	**0.663**
	FSIM (↑)	0.400	0.402	0.290	0.440	0.449	0.414	**0.461**	0.402
00421	NRMSE (↓)	0.894	0.894	0.959	**0.850**	0.851	0.958	0.855	0.854
	LR Cons. (↓)	0.237	0.237	0.589	**0.133**	**0.133**	0.282	0.188	0.161
	SSIM (↑)	0.800	0.806	0.803	**0.871**	0.859	0.797	0.855	0.869
	FSIM (↑)	0.497	0.503	0.428	**0.554**	0.533	0.478	0.551	0.539
00423	NRMSE (↓)	1.087	1.088	1.250	0.759	0.766	1.281	**0.644**	0.738
	LR Cons. (↓)	0.227	0.227	0.742	**0.142**	0.136	0.299	0.255	0.250
	SSIM (↑)	0.668	0.676	0.719	0.778	0.782	0.703	0.836	**0.857**
	FSIM (↑)	0.481	0.481	0.382	0.539	0.526	0.435	0.570	**0.599**
00451	NRMSE (↓)	0.606	0.606	0.914	0.567	**0.539**	0.818	0.492	0.652
	LR Cons. (↓)	0.245	0.246	0.647	**0.093**	0.105	0.281	0.152	0.576
	SSIM (↑)	0.622	0.622	0.467	0.661	0.671	0.544	0.700	**0.710**
	FSIM (↑)	0.441	0.441	0.328	0.490	0.498	0.436	0.527	**0.556**

Table 3. Image quality metrics for test images 00069, 00118, 00144, 00277, 00421, 00423, and 00451.

quately capture the quality of NTL images due to the unique properties of NTL images, such as low contrast in dark regions and a wide dynamic range of brightness values. We discuss this further in Section 7.

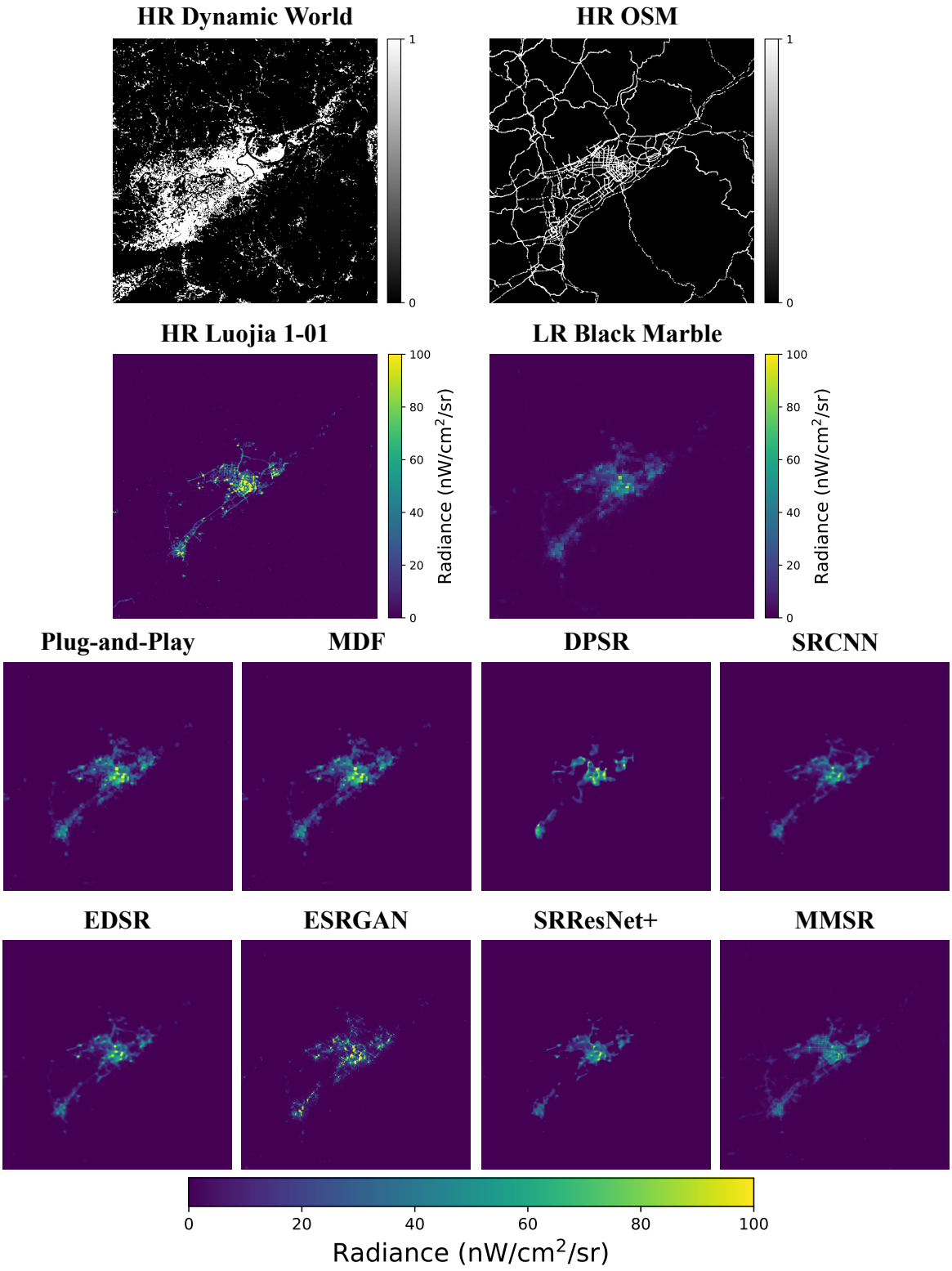


Figure 4. Comparison of input, ground truth, and super-resolved images for test image 00069. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

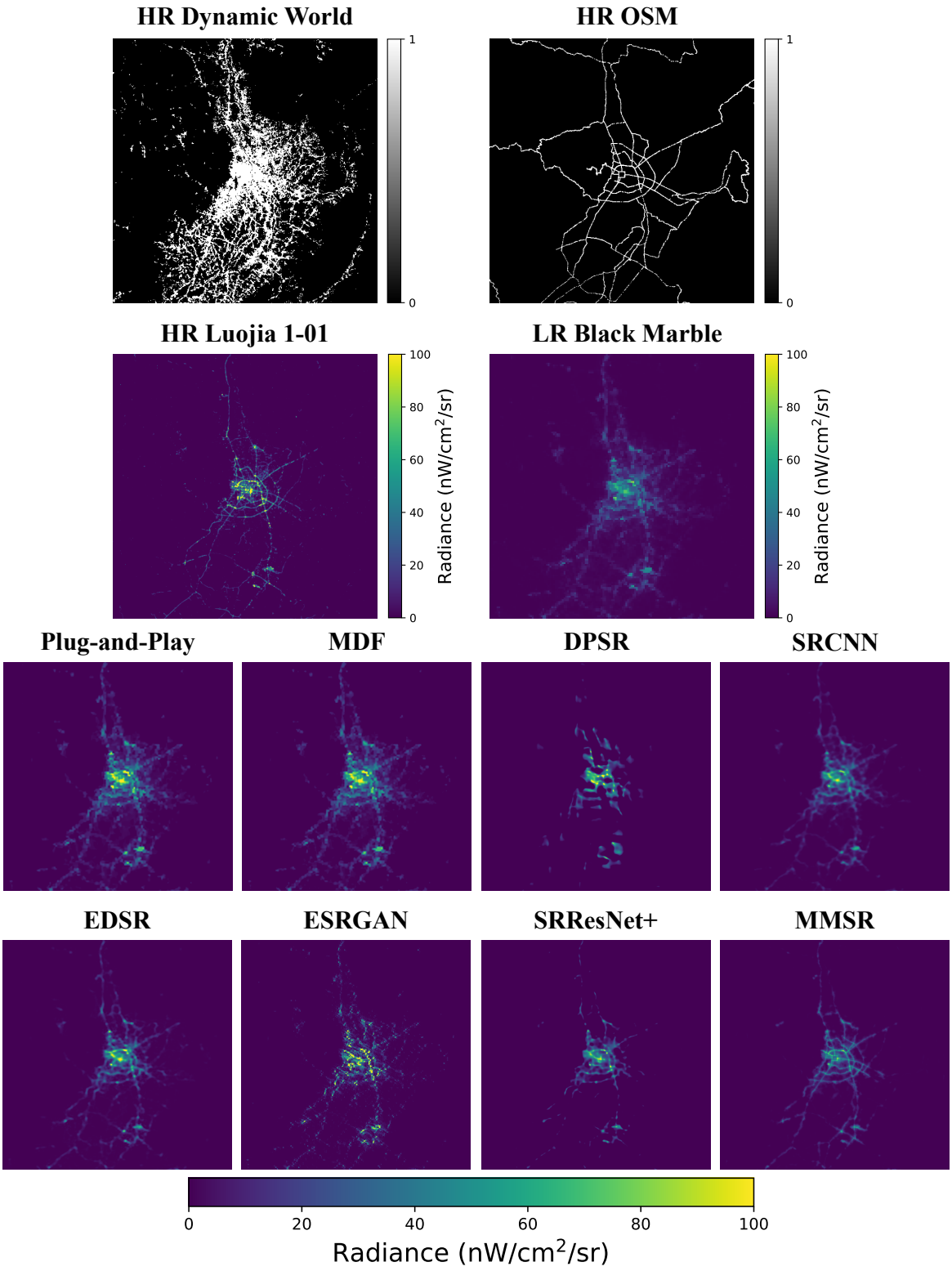


Figure 5. Comparison of input, ground truth, and super-resolved images for test image 00118. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

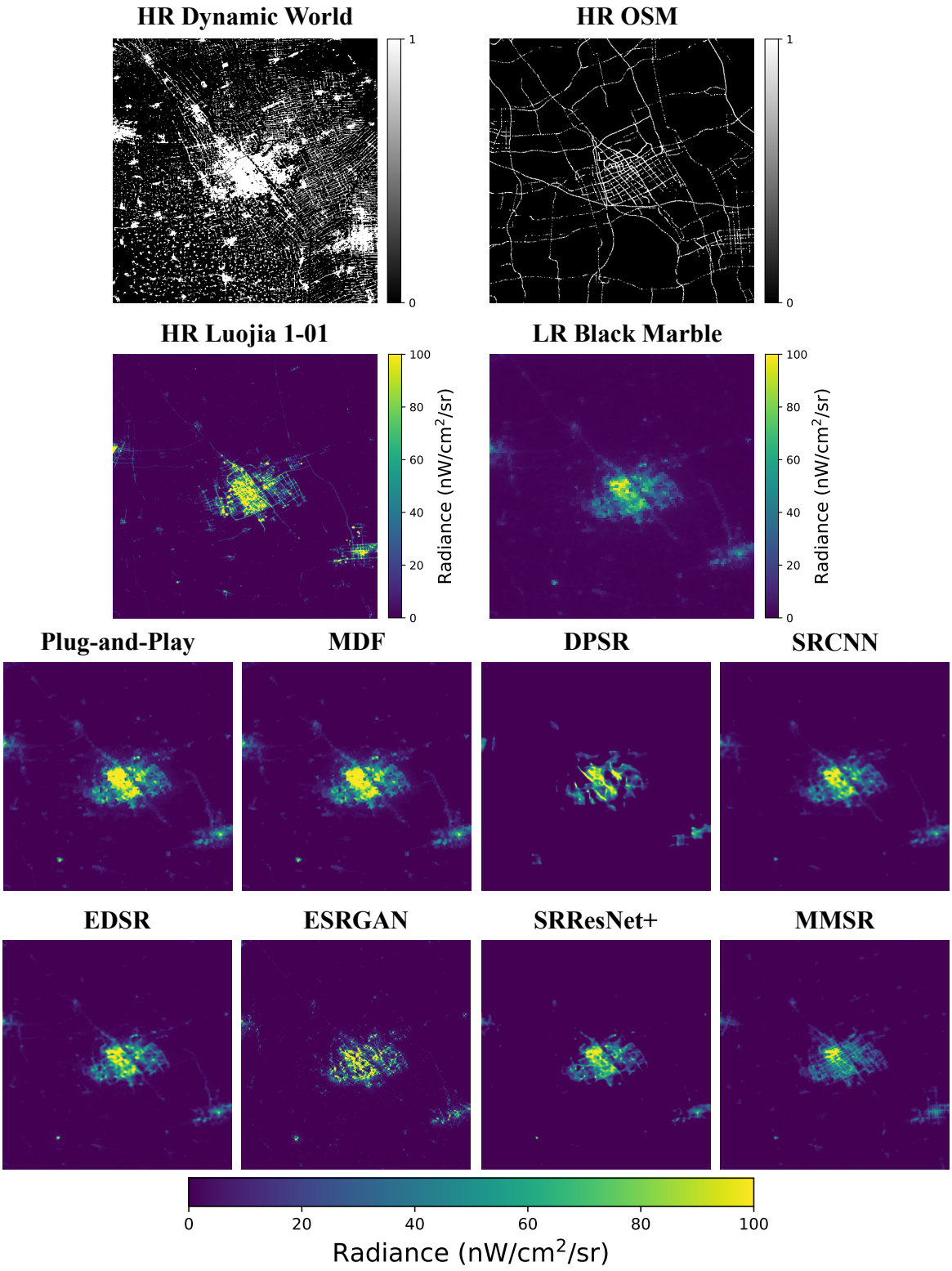


Figure 6. Comparison of input, ground truth, and super-resolved images for test image 00144. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

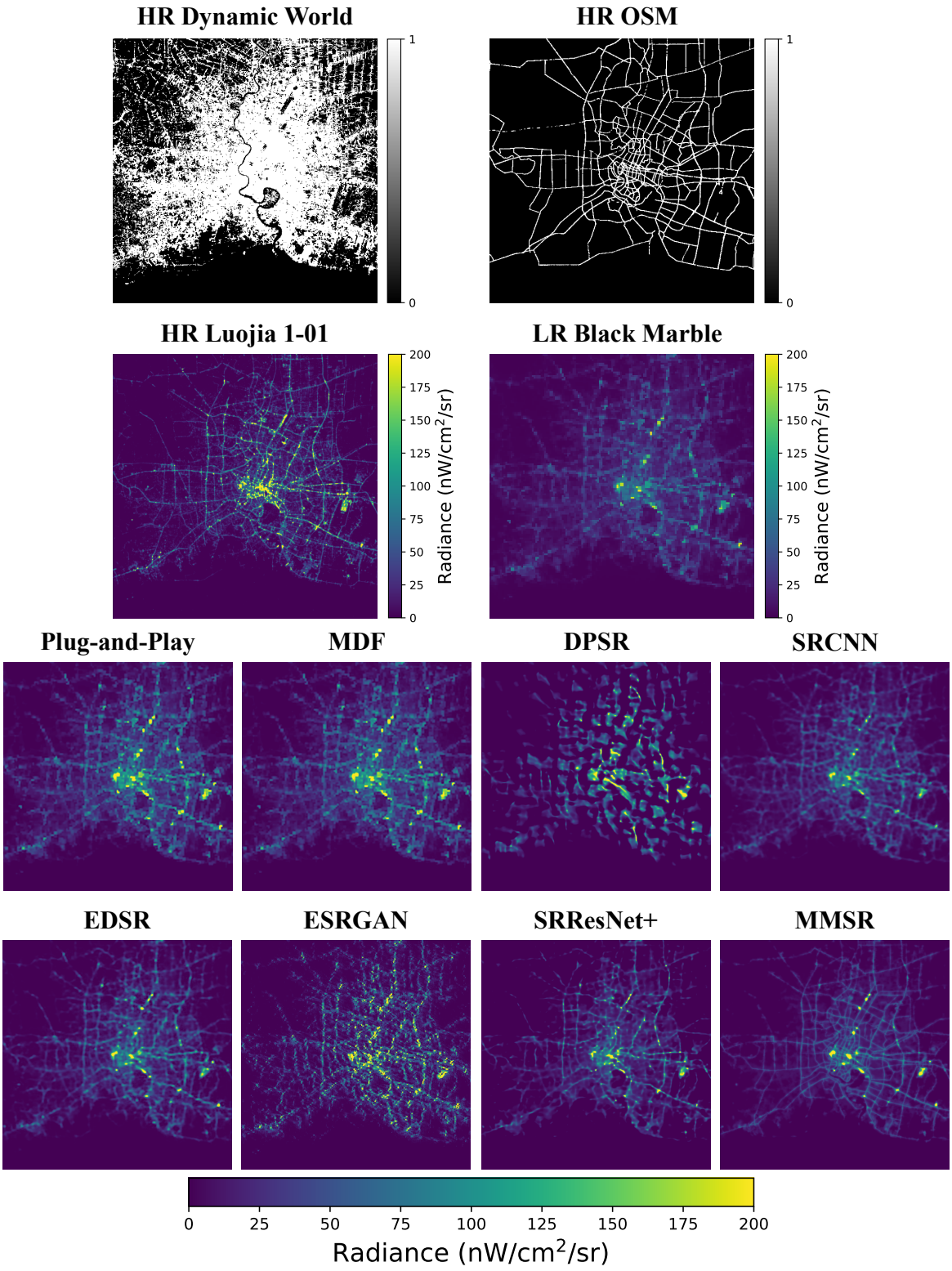


Figure 7. Comparison of input, ground truth, and super-resolved images for test image 00277. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

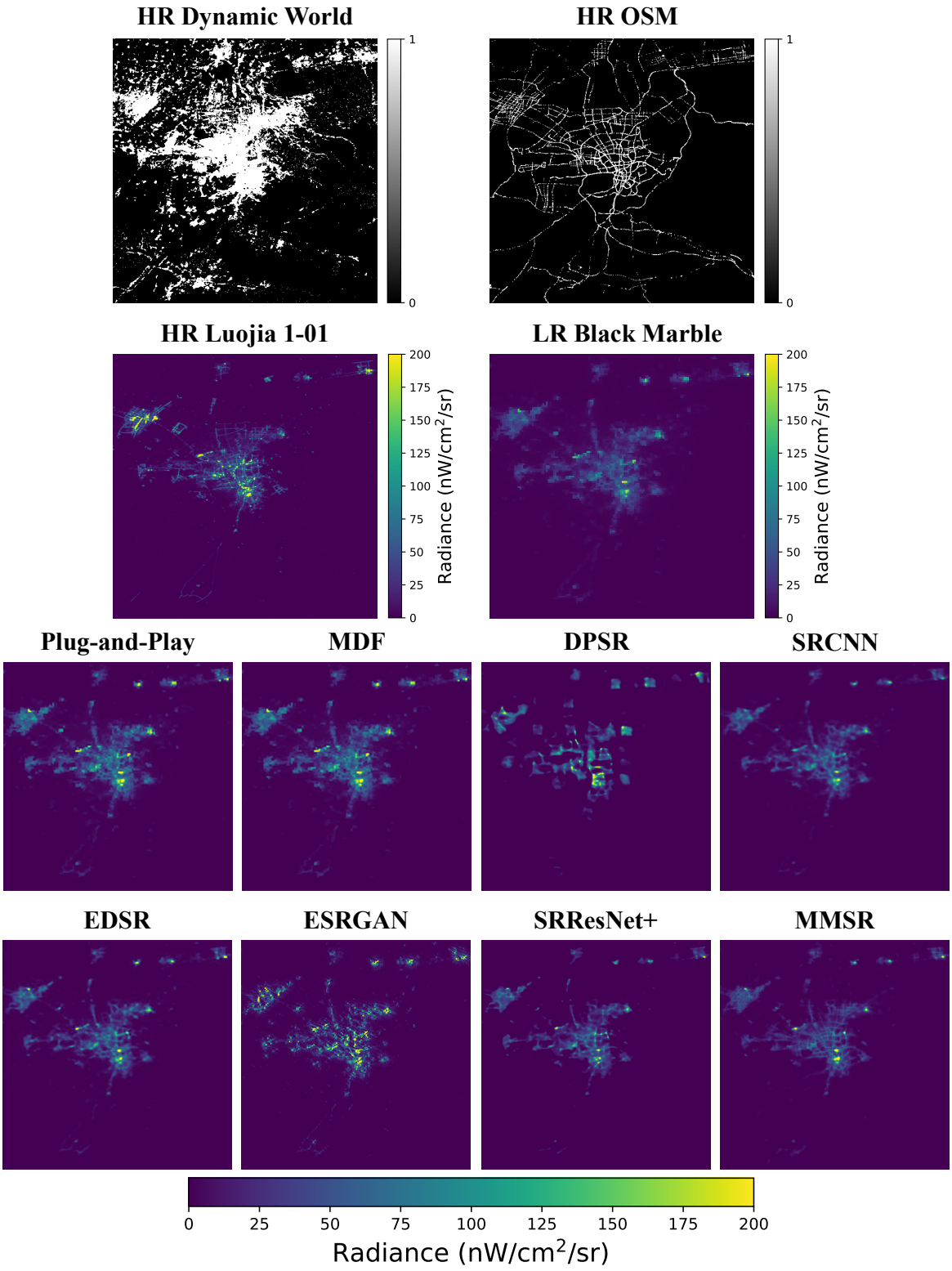


Figure 8. Comparison of input, ground truth, and super-resolved images for test image 00421. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

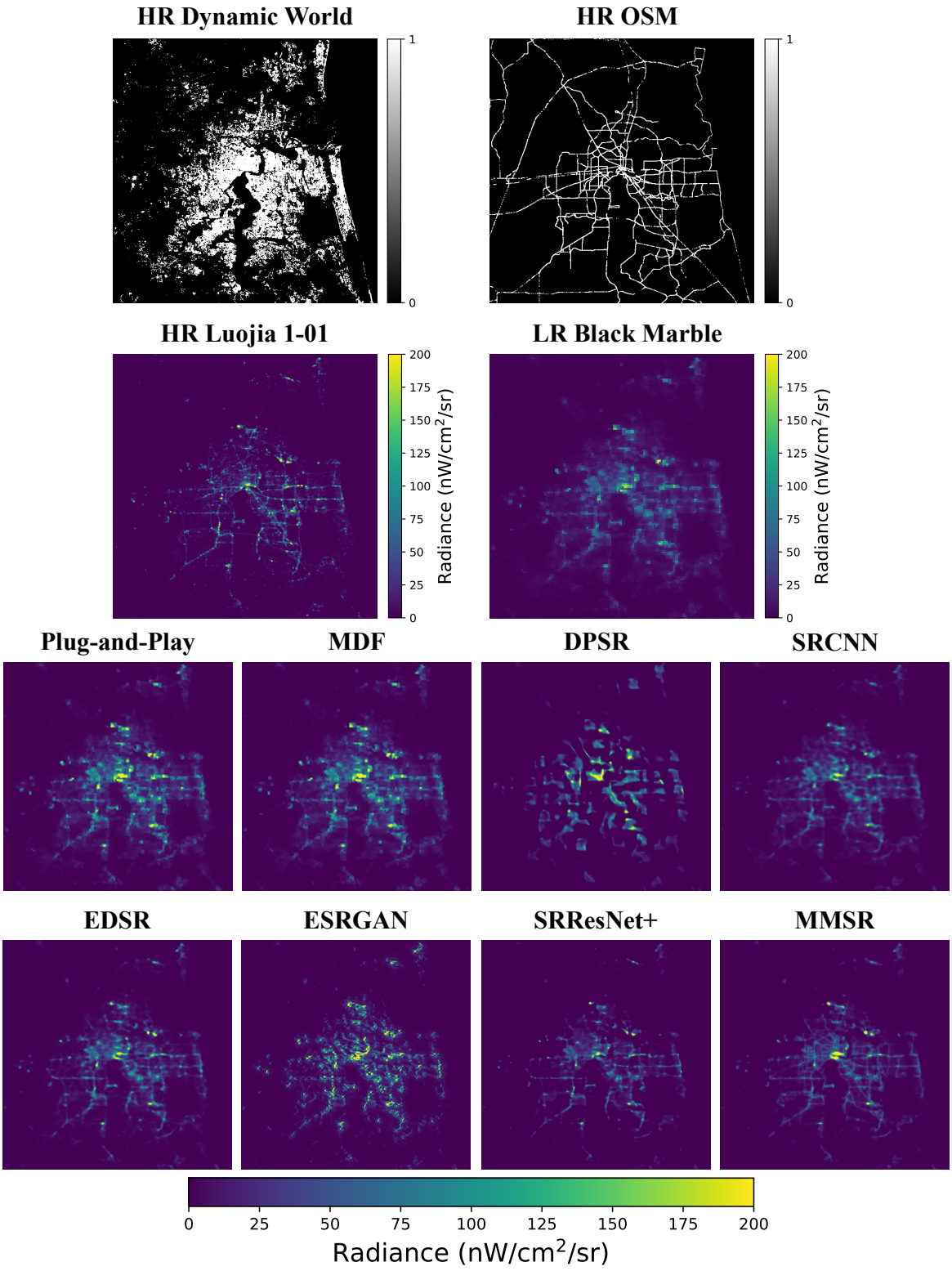


Figure 9. Comparison of input, ground truth, and super-resolved images for test image 00423. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

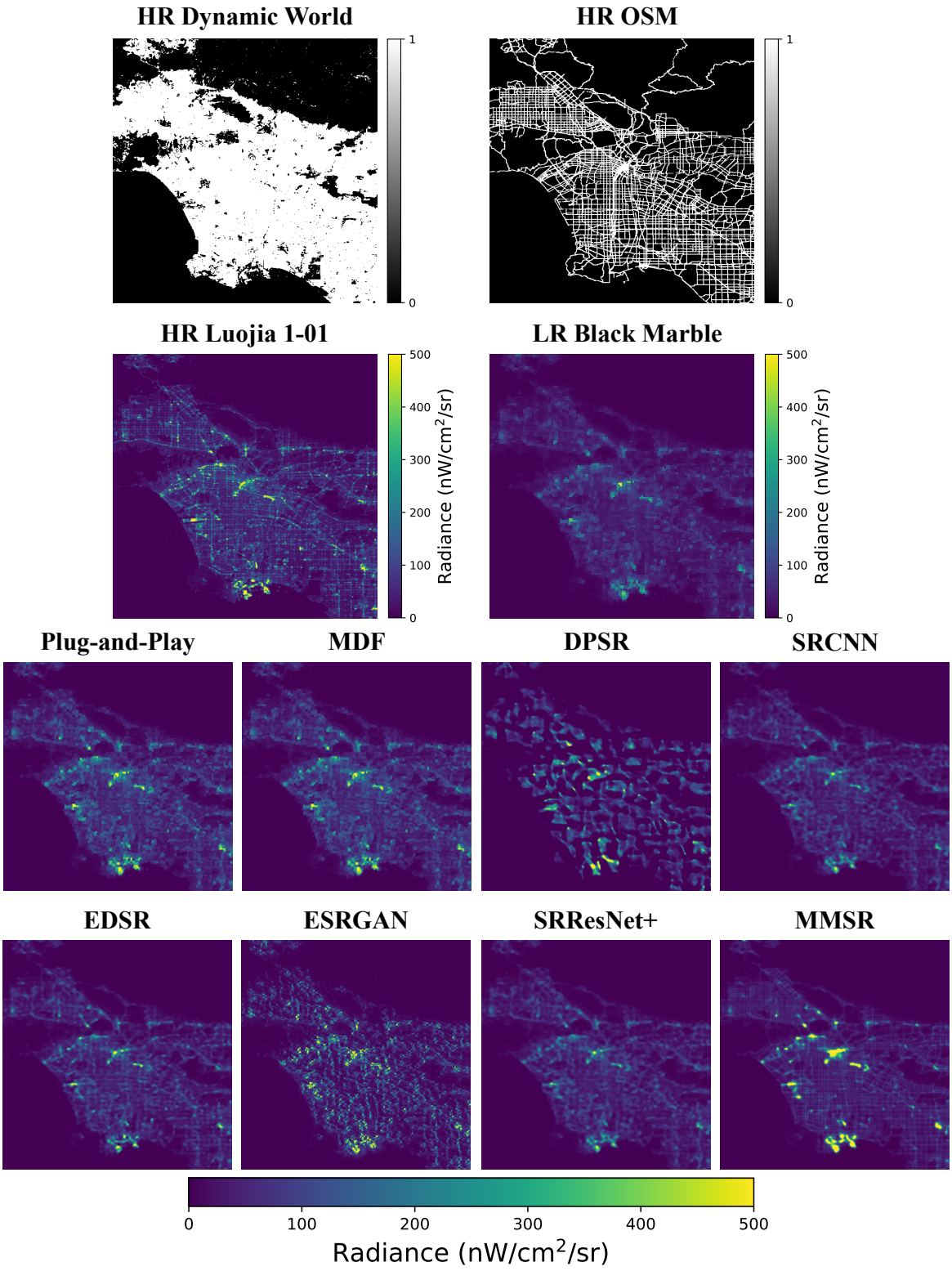
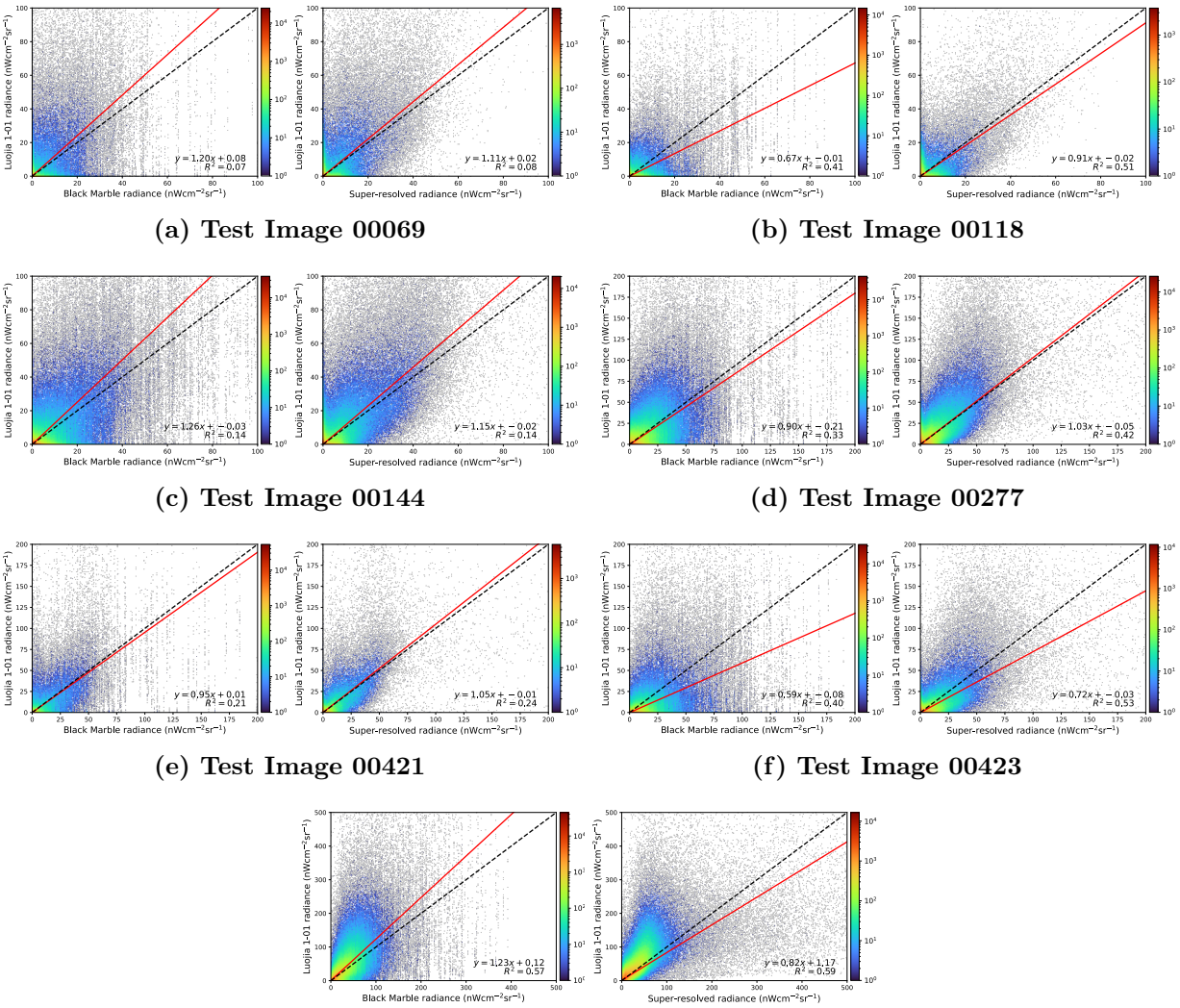


Figure 10. Comparison of input, ground truth, and super-resolved images for test image 00451. (Top) Input land-use and street maps, the HR Luojia 1-01 NTL image, and the LR Black Marble NTL image. (Bottom) Super-resolved images from each of the methods shown in Table 2.

Method	NRMSE (↓)	LR Consistency (↓)	SSIM (↑)	FSIM (↑)
PnP	0.946	0.235	0.768	0.499
MDF	0.946	0.235	0.764	0.505
DPSR	1.039	0.585	0.752	0.437
SRCNN	0.805	0.587	0.756	0.504
EDSR	0.815	**0.141**	0.818	0.535
ESRGAN	1.088	0.309	0.760	0.495
SRResNet+	**0.797**	0.225	0.828	**0.555**
MMSR	0.953	0.495	**0.840**	0.552

Table 4. Image quality metrics averaged over all test images.



(a) Test Image 00069 **(b) Test Image 00118**

(c) Test Image 00144 **(d) Test Image 00277**

(e) Test Image 00421 **(f) Test Image 00423**

(g) Test Image 00451

Figure 11. Scatter plots of Black Marble (left) and the MMSR super-resolved image (right) against the ground truth Luojia 1-01 image for the test images 00069, 00118, 00144, 00277, 00421, 00423, and 00451.

6. A CASE STUDY: 2025 NEW YEAR'S POWER OUTAGE IN PUERTO RICO

In this section, we present a preliminary case study of the 2025 New Year's wide-area power outage in Puerto Rico, which was caused by a line failure in the transmission system. The outage occurred on December 31, 2024. The outage was reported to have affected 1.2 million customers, with the majority of the island experiencing a blackout. To analyze the outage, we use super-resolved NTL images of Puerto Rico before and after the outage from our proposed MMSR method described in Section 3.3. We hope to demonstrate the utility of the MMSR method for analyzing the spatial extent of the outage and assess the subsequent recovery of the power supply.

To construct a consistent temporal dataset, we downloaded NTL data from Black Marble over Puerto Rico for each night between December 1, 2024, and January 31, 2025. This timeframe allows us to capture pre-outage baseline conditions, the immediate effects of the blackout, and the gradual restoration of electricity. To apply the MMSR method, we also extracted the OSM and DynamicWorld data for each night. Since DynamicWorld is a near real-time product of Sentinel-2, there is some temporal variance between this data and the NTL data, which may affect the results of the MMSR network.

To analyze the outage, we compare the NTL images of Puerto Rico before and after the power outage. We compare the outage occurrence on the night of January 1, 2025 with a base composite image, which is obtained by averaging the NTL images from December 1st, 2024 to December 31st, 2024. In Figure 12, we show the NTL images of Puerto Rico before and after the power outage. The left column shows the base composite NTL image, while the right column shows the NTL image on the night of the power outage. The top row shows the LR NTL image, while the bottom row shows the HR NTL output by the proposed MMSR method.

Additionally, we compute outage and recovery images, which are measures of the percentage of light that was lost during the outage and restored after the outage respectively. We calculate these images as follows:

$$\text{Outage} = \frac{\text{NTL}_{\text{Pre}} - \text{NTL}_{\text{Jan. 1st}}}{\text{NTL}_{\text{Pre}}} \times 100\%, \tag{10}$$

and

$$\text{Recovery} = \frac{\text{NTL}_{\text{Post}} - \text{NTL}_{\text{Jan. 1st}}}{\text{NTL}_{\text{Post}}} \times 100\%, \tag{11}$$

where NTL_{Pre} is a composite image computed by averaging the NTL images prior to the outage, $\text{NTL}_{\text{Jan. 1st}}$ is the NTL image on the night of the outage, and NTL_{Post} is a composite image computed by averaging the NTL images after the outage. These measurements are shown in Figure 13 for both the Black Marble and MMSR images. Here red and blue indicate light lost and gained respectively, with darker colors indicating a larger percentage of light lost or gained.

A common method for analyzing the instability of electric power is to compute the standard deviation of NTL images over time [108]. Regions with high standard deviation, particularly when compared to the mean, may indicate an unreliable power supply, areas of high human activity, or areas with seasonal activities. In this case study, we examine the temporal instability of NTL images for both the Black Marble and MMSR images of Puerto Rico during the two months of December 2024 and January 2025. This is illustrated in Figure 14.

In Figure 15, we a show a scatter plot of the mean and standard deviation of the NTL images over time for both the Black Marble and MMSR images. Each point in the scatter plot represents

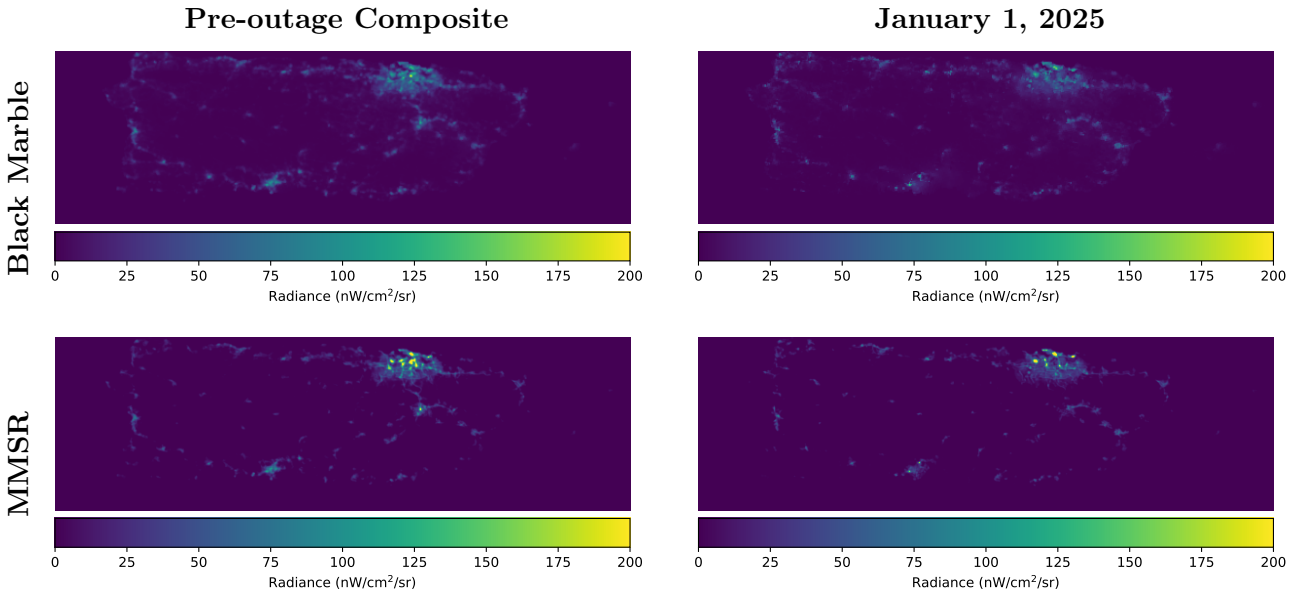
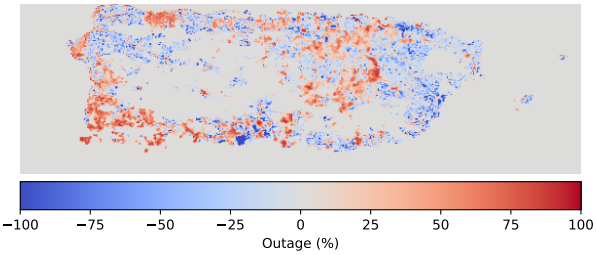


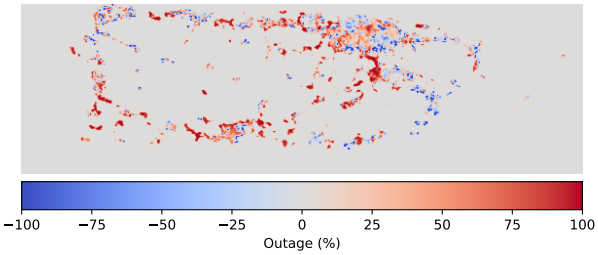
Figure 12. Comparison of the NTL images of Puerto Rico before and after the power outage on January 1, 2025. The change in radiance is evident in both Black Marble and MMSR, but the MMSR reconstruction is sharper. This sharpness allows for discernment of point sources of light, allowing for a fine-grained outage analysis.

a pixel in the image, with the x-axis representing the mean value of the pixel over time and the y-axis representing the standard deviation of the pixel over time. Of special concern are regions with a high standard deviation relative to the mean; the red line indicates when the standard deviation exceeds twice the mean. This line helps identify areas that are have unstable NTL values over time. In Figure 16, we plot the points which are above the red line in the scatter plot on the mean NTL image. The increased resolution provided by the MMSR method allows us to identify regions of instability more precisely.

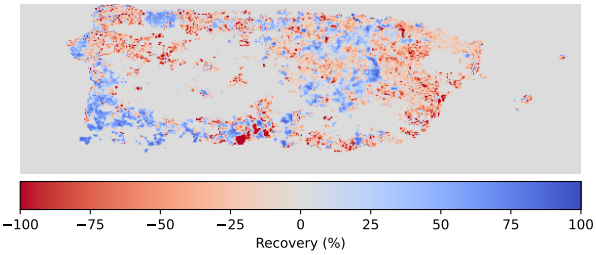
In all three of these outage metrics, MMSR offers more localized detections as compared to Black Marble. However, since we do not have access to HR NTL images for this data, it's difficult to determine correctness of these detections. In future work, we would like to do a more thorough case study which compares our findings with reported outage and recovery data to verify that accuracy of this localization.



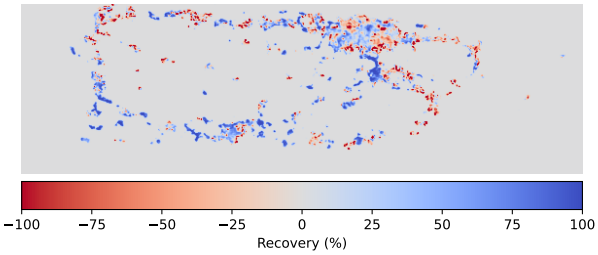
(a) Black Marble Outage



(b) MMSR Outage



(c) Black Marble Recovery



(d) MMSR Recovery

Figure 13. Comparison of outage computed from Black Marble and MMSR images. The outage index is a measure of the percentage of light that was lost during the outage, and the recovery index is a percentage of light restored. In both outage and recovery, MMSR precisely indicates location and severity more accurately than Black Marble.

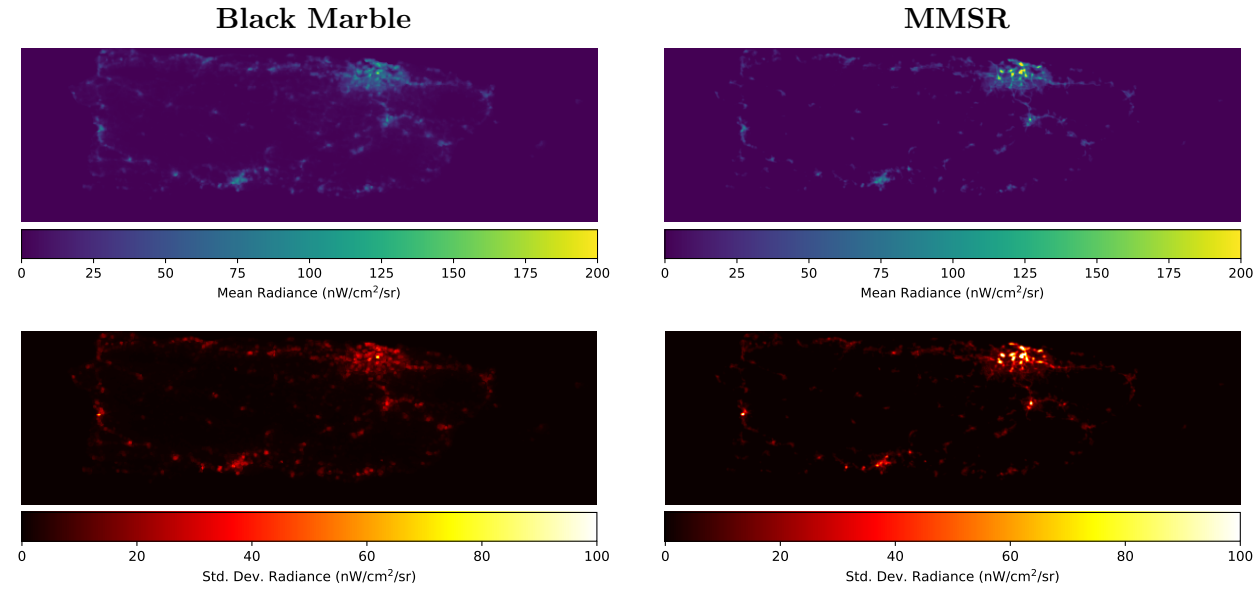
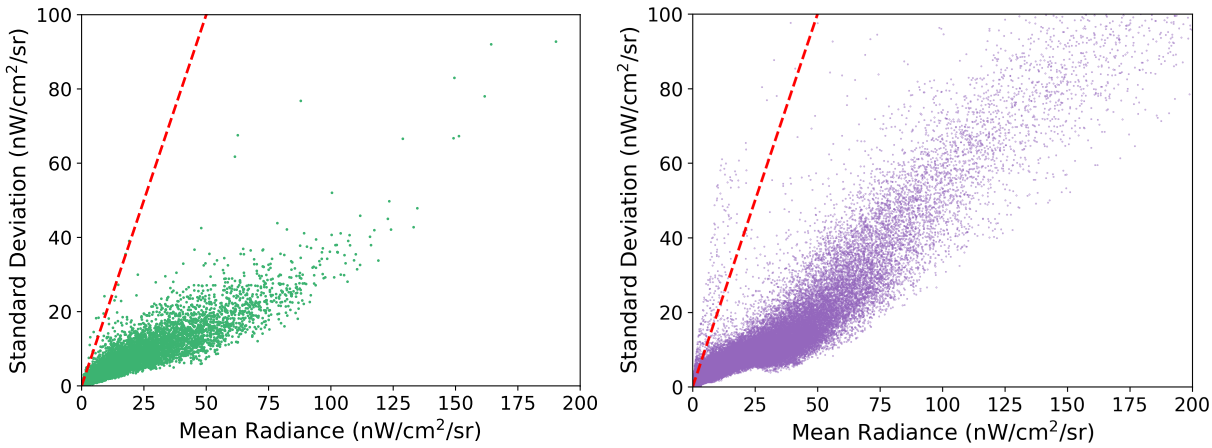
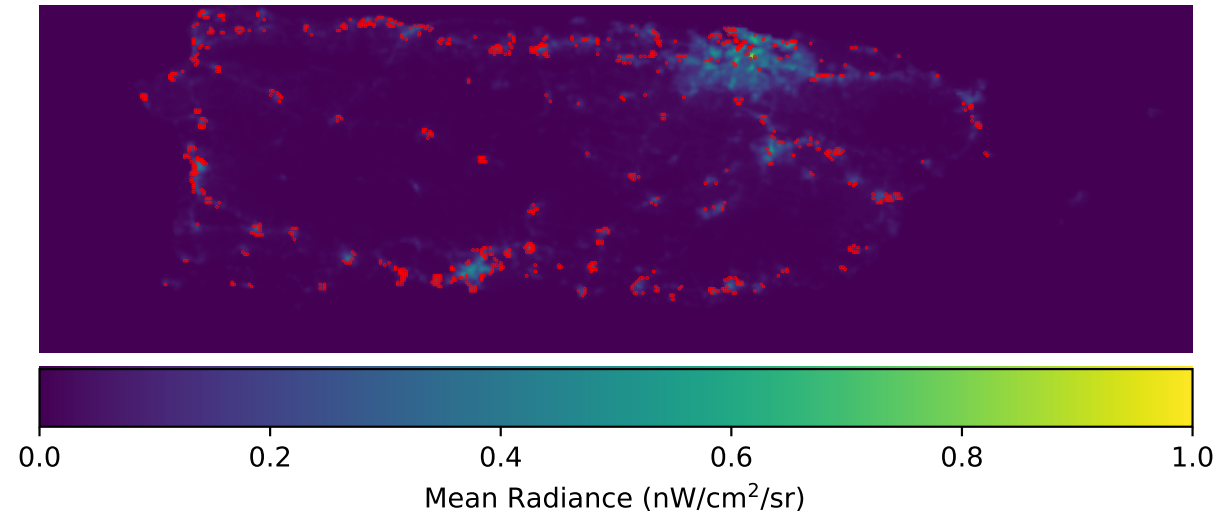


Figure 14. Mean and standard deviation of the NTL images over time for both the Black Marble and MMSR images of Puerto Rico over the months of December 2024 and January 2025. The top row shows the mean image, while the bottom row shows the standard deviation image.

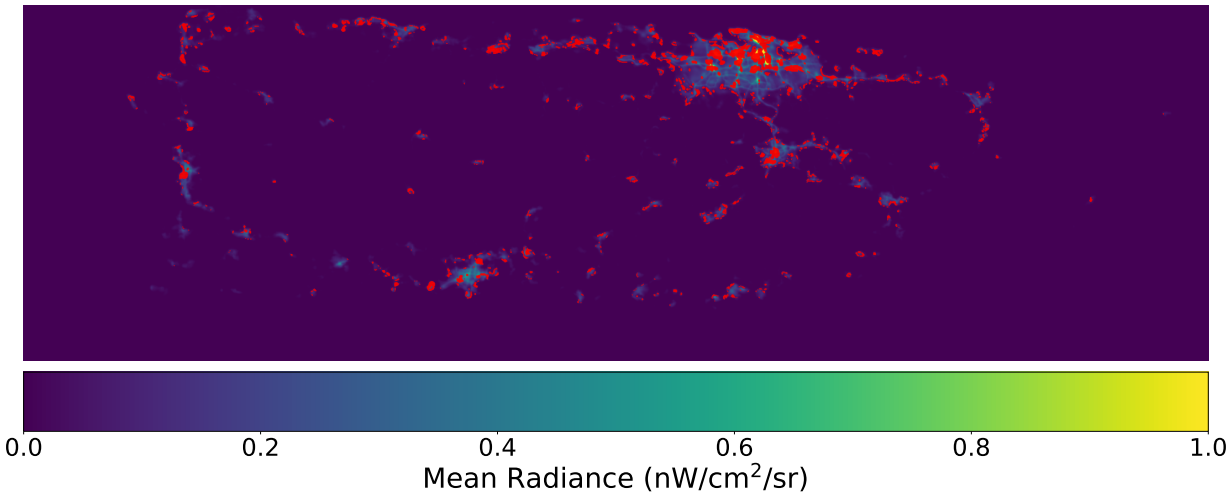


(a) Black Marble (b) MMSR

Figure 15. Mean and standard deviation scatter plot for the (a) Black Marble and (b) MMSR images of Puerto Rico over the months of December 2024 and January 2025. The red line indicates the threshold for instability, where the standard deviation is greater than twice the mean. Points above this line are plotted in Figure 16.



(a) Black Marble



(b) MMSR

Figure 16. Instability of the NTL images over time for both the (a) Black Marble and (b) MMSR images of Puerto Rico over the months of December 2024 and January 2025. The red dots indicate regions determined to be unstable by the line shown in Figure 15. MMSR is able to localize areas of high variance better than Black Marble.

7. PROPOSED FUTURE WORK

Our experimental results focused mainly on investigating the super-resolution quality; however, the preliminary results in our case study imply that super-resolution can help to detect and localize power outages. In future work, we plan to do a more thorough investigation of the impact of super-resolution power outage detection, using existing power outage prediction methods from NTL [8, 109, 110].

Additionally, the proposed MMSR method can be extended in several ways. The first is to incorporate more auxiliary inputs, such as the NDVI or other bands from the VIIRS sensor. This would allow for a more comprehensive analysis of the relationship between the LR and HR images, potentially leading to improved super-resolution results. Second, the MMSR architecture can be modified to include more advanced techniques, such as residual networks, transformers, or masked autoencoders. These architectures have shown promise in various image processing tasks and could enhance the performance of the MMSR method. Third, the MMSR architecture could incorporate a learned image registration or deformation process. This would allow the model to learn the optimal transformation between the LR, HR images, and auxiliary inputs, potentially improving the super-resolution results [45, 111].

A consistent theme in the literature of NTL image super-resolution is the need for a metric that accurately reflects the quality of the generated HR images. Current metrics, such as the ones used in this work, Peak Signal-to-Noise Ratio (PSNR), SSIM, and Naturalness Image Quality Evaluator (NIQE), are not well-suited due to the variation in radiance values recorded by different sensors. Moreover, current perceptual metrics, such as SSIM and NIQE, are designed for natural images and do not capture the unique characteristics of NTL images. Developing a perceptual metric specifically for NTL images would be a valuable contribution to the field and could significantly improve the evaluation of super-resolution methods.

Finally, the proposed MMSR method could be combined with temporal super-resolution in the HR domain. In [112], the authors use HR images at different times to guide the super-resolution of the LR image at a time in which the HR image is not available. This approach could be extended to the multimodal super-resolution framework by incorporating the auxiliary inputs into the temporal super-resolution process. Due to its limited time in orbit, Luojia 1-01 is not suitable for this approach. However, there are other satellites that could be used for this purpose such as the SDGSAT-1 or EROS-B satellites.

8. CONCLUSION

This report demonstrates the capabilities of super-resolution for enhancing the spatial detail and analysis of NTL imagery. Through a comprehensive literature review and evaluation of existing super-resolution methods, we identified the shortcomings of traditional super-resolution techniques when applied to NTL images, which exhibit distinct statistical properties compared to natural images. Furthermore, our proposed multimodal super-resolution approach demonstrates that integrating contextual information such as road networks and land use data with NTL imagery can improve the super-resolution quality. These advancements provide a proof of concept for the role of super-resolution in more nuanced assessments of urbanization, infrastructure, and energy dynamics.

To the best of our knowledge, this research serves as the first comprehensive exploration of super-resolution methods tailored for NTL imagery, offering valuable insights for future studies in this domain. By making our code and data available to other researchers at ORNL, we aim to foster reproducibility and encourage further advancements in NTL image processing and analysis.

REFERENCES

[1] Thomas A Croft. "Nighttime images of the earth from space". In: *Scientific American* 239.1 (1978), pp. 86–101.

[2] Min Zhao et al. "Applications of satellite remote sensing of nighttime light observations: Advances, challenges, and perspectives". In: *Remote Sensing* 11.17 (2019), p. 1971.

[3] Miguel O Román et al. "NASA's Black Marble nighttime lights product suite". In: *Remote Sensing of Environment* 210 (2018), pp. 113–143.

[4] Gary E Machlis, Miguel O Román, and Steward TA Pickett. "A framework for research on recurrent acute disasters". In: *Science Advances* 8.10 (2022), eabk2458.

[5] Srija Chakraborty and Eleanor C Stokes. "Adaptive modeling of satellite-derived nighttime lights time-series for tracking urban change processes using machine learning". In: *Remote Sensing of Environment* 298 (2023), p. 113818.

[6] Noam Levin. "The impact of seasonal changes on observed nighttime brightness from 2014 to 2015 monthly VIIRS DNB composites". In: *Remote Sensing of Environment* 193 (2017), pp. 150–164.

[7] Eleanor C Stokes et al. "Urban applications of nasa's black marble product suite". In: *2019 Joint Urban Remote Sensing Event (JURSE)*. IEEE. 2019, pp. 1–4.

[8] Bandana Kar et al. "Spatiotemporal Tracking of Wide Area Power Outage from Night-Time Light Imagery". In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 2021, pp. 566–569. DOI: 10.1109/IGARSS47720.2021.9554293.

[9] Eleanor C Stokes et al. "Retired satellites: A chance to shed light". In: *Science* 373.6562 (2021), pp. 1451–1452.

[10] Morgan R Edwards et al. "Satellite data applications for sustainable energy transitions". In: *Frontiers in Sustainability* 3 (2022), p. 910924.

[11] Noam Levin and Yishai Duke. "High spatial resolution night-time light images for demographic and socio-economic studies". In: *Remote Sensing of Environment* 119 (2012), pp. 1–10.

[12] Nikolaos Tziokas et al. "Downscaling satellite night-time lights imagery to support within-city applications using a spatially non-stationary model". In: *International Journal of Applied Earth Observation and Geoinformation* 122 (2023), p. 103395.

[13] Qiming Zheng, Qihao Weng, and Ke Wang. "Correcting the Pixel Blooming Effect (PiBE) of DMSP-OLS nighttime light imagery". In: *Remote Sensing of Environment* 240 (2020), p. 111707. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2020.111707. URL: https://www.sciencedirect.com/science/article/pii/S0034425720300766.

[14] Chengcheng Xue et al. "Automatic boat detection based on diffusion and radiation characterization of boat lights during night for VIIRS DNB imaging data". In: *Opt. Express* 30.8 (Apr. 2022), pp. 13024–13038. DOI: 10.1364/OE.455555. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-8-13024.

[15] Alejandro Sanchez de Miguel et al. "The nature of the diffuse light near cities detected in nighttime satellite imagery". In: *Scientific Reports* 10.1 (2020), p. 7829.

[16] Nikolaos Tziokas et al. "Downscaling Satellite Night-Time Light Imagery While Addressing the Blooming Effect". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), pp. 13678–13693. DOI: 10.1109/JSTARS.2024.3429244.

[17] Seunghyeon Rhee and Moon Gi Kang. "DCT-based regularized algorithm for high-resolution image reconstruction". In: *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*. Vol. 3. IEEE. 1999, pp. 184–187.

[18] Nhat Nguyen and Peyman Milanfar. "A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution)". In: *Circuits, Systems and Signal Processing* 19 (2000), pp. 321–338.

[19] Hussein A Aly and Eric Dubois. "Image up-sampling using total-variation regularization with a new observation model". In: *IEEE Transactions on Image Processing* 14.10 (2005), pp. 1647–1659.

[20] Shengyang Dai et al. "Soft edge smoothness prior for alpha channel super resolution". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2007, pp. 1–8.

[21] Ningning Zhao et al. "Fast Single Image Super-Resolution Using a New Analytical Solution for $\ell_2 - \ell_2$ Problems". In: *IEEE Transactions on Image Processing* 25.8 (2016), pp. 3683–3697.

[22] Assaf Shocher, Nadav Cohen, and Michal Irani. ""zero-shot" super-resolution using deep internal learning". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3118–3126.

[23] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. "Singan: Learning a generative model from a single natural image". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 4570–4580.

[24] Chao Dong et al. "Image Super-Resolution Using Deep Convolutional Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (Dec. 2014). DOI: 10.1109/TPAMI.2015.2439281.

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[26] Wenzhe Shi et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1874–1883.

[27] Bee Lim et al. "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. July 2017.

[28] Kai Zhang, Wangmeng Zuo, and Lei Zhang. "Deep plug-and-play super-resolution for arbitrary blur kernels". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 1671–1681.

[29] Christian Ledig 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 (CVPR)*. July 2017.

[30] Xintao Wang et al. "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Sept. 2018.

[31] Chitwan Saharia et al. "Image Super-Resolution via Iterative Refinement". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.4 (2023), pp. 4713–4726. DOI: 10.1109/TPAMI.2022.3204461.

[32] Kai Zhang, Luc Van Gool, and Radu Timofte. "Deep Unfolding Network for Image Super-Resolution". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 3214–3223. DOI: 10.1109/CVPR42600.2020.00328.

[33] Jingyun Liang et al. "Swinir: Image restoration using swin transformer". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 1833–1844.

[34] Marcos V Conde et al. “Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration”. In: *European Conference on Computer Vision*. Springer. 2022, pp. 669–687.

[35] Leonardo Rossi et al. “Swin2-MoSE: A New Single Image Super-Resolution Model for Remote Sensing”. In: *arXiv preprint arXiv:2404.18924* (2024).

[36] Charis Lanaras et al. “Super-resolution of multispectral multiresolution images from a single sensor”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017, pp. 20–28.

[37] Claudia Paris, José Bioucas-Dias, and Lorenzo Bruzzone. “A hierarchical approach to superresolution of multispectral images with different spatial resolutions”. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2017, pp. 2589–2592.

[38] Magnus O Ulfarsson et al. “Sentinel-2 sharpening using a reduced-rank method”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6408–6420.

[39] Charis Lanaras et al. “Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (2018), pp. 305–319.

[40] Haley Duba-Sullivan et al. “ResSR: A Residual Approach to Super-Resolving Multispectral Images”. In: *arXiv preprint arXiv:2408.13225* (2024).

[41] Yang Ye et al. “A feasible framework to downscale NPP-VIIRS nighttime light imagery using multi-source spatial variables and geographically weighted regression”. In: *International Journal of Applied Earth Observation and Geoinformation* 104 (2021), p. 102513.

[42] Miguel O Román et al. “Satellite-based assessment of electricity restoration efforts in Puerto Rico after Hurricane Maria”. In: *PloS one* 14.6 (2019), e0218883.

[43] Shangqin Liu et al. “Spatial downscaling of NPP-VIIRS nighttime light data using multiscale geographically weighted regression and multi-source variables”. In: *Remote Sensing* 14.24 (2022), p. 6400.

[44] Bin Wu, Yu Wang, and Hailan Huang. “Downscaling NPP–VIIRS Nighttime Light Data Using Vegetation Nighttime Condition Index”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), pp. 18291–18302. DOI: 10.1109/JSTARS.2024.3476191.

[45] Lixian Zhang et al. “DeepLight: Reconstructing High-Resolution Observations of Nighttime Light With Multi-Modal Remote Sensing”. In: *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence AI for Good*. 2024, pp. 7563–7571.

[46] Weixing Xu et al. “Spatial Downscaling of NPP/VIIRS DNB Nighttime Light Data Based on Deep Learning”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), pp. 16787–16798. DOI: 10.1109/JSTARS.2024.3454093.

[47] Leonid I Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[48] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. “Perceptual losses for real-time style transfer and super-resolution”. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer. 2016, pp. 694–711.

[49] Ian Goodfellow et al. “Generative adversarial networks”. In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

[50] Hui Ji and Cornelia Fermüller. “Robust wavelet-based super-resolution reconstruction: theory and algorithm”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.4 (2008), pp. 649–660.

[51] Assaf Shocher, Nadav Cohen, and Michal Irani. ""zero-shot" super-resolution using deep internal learning". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3118–3126.

[52] Vincent Sitzmann, Michael Zollhoefer, and Gordon Wetzstein. "Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019. (Visited on 02/15/2024).

[53] Nasim Rahaman et al. "On the spectral bias of neural networks". In: *International conference on machine learning*. PMLR. 2019, pp. 5301–5310.

[54] Ben Mildenhall et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[55] A Vaswani. "Attention is all you need". In: *Advances in Neural Information Processing Systems* (2017).

[56] Ellen D Zhong et al. "Reconstructing continuous distributions of 3D protein structure from cryo-EM images". In: *arXiv preprint arXiv:1909.05215* (2019).

[57] Matthew Tancik et al. "Fourier features let networks learn high frequency functions in low dimensional domains". In: *Advances in neural information processing systems* 33 (2020), pp. 7537–7547.

[58] Vincent Sitzmann et al. "Implicit neural representations with periodic activation functions". In: *Advances in neural information processing systems* 33 (2020), pp. 7462–7473.

[59] Vishwanath Saragadam et al. "Wire: Wavelet implicit neural representations". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 18507–18516.

[60] Danzel Serrano, Jakub Szymkowiak, and Przemyslaw Musialski. "HOSC: A Periodic Activation Function for Preserving Sharp Features in Implicit Neural Representations". In: *arXiv preprint arXiv:2401.10967* (2024).

[61] Zhen Liu et al. "FINER: Flexible spectral-bias tuning in Implicit NEural Representation by Variable-periodic Activation Functions". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 2713–2722.

[62] Sameera Ramasinghe and Simon Lucey. "Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps". In: *European Conference on Computer Vision*. Springer. 2022, pp. 142–158.

[63] Hemanth Saratchandran et al. "A sampling theory perspective on activations for implicit neural representations". In: *arXiv preprint arXiv:2402.05427* (2024).

[64] Wentian Xu and Jianbo Jiao. "Revisiting implicit neural representations in low-level vision". In: *arXiv preprint arXiv:2304.10250* (2023).

[65] Suhas Sreehari et al. "Plug-and-play priors for bright field electron tomography and sparse interpolation". In: *IEEE Transactions on Computational Imaging* 2.4 (2016), pp. 408–423.

[66] Emma J Reid et al. "Multi-resolution data fusion for super resolution imaging". In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 81–95.

[67] Kai Zhang et al. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising". In: *IEEE transactions on image processing* 26.7 (2017), pp. 3142–3155.

[68] Xinbo Gao et al. "Image super-resolution with sparse neighbor embedding". In: *IEEE Transactions on Image Processing* 21.7 (2012), pp. 3194–3205.

[69] Jianchao Yang et al. "Image super-resolution via sparse representation". In: *IEEE transactions on image processing* 19.11 (2010), pp. 2861–2873.

[70] William T Freeman, Thouis R Jones, and Egon C Pasztor. "Example-based super-resolution". In: *IEEE Computer graphics and Applications* 22.2 (2002), pp. 56–65.

[71] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
[72] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
[73] Alexia Jolicoeur-Martineau. “The relativistic discriminator: a key element missing from standard GAN”. In: *arXiv preprint arXiv:1807.00734* (2018).
[74] Ze Liu et al. “Swin transformer: Hierarchical vision transformer using shifted windows”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 10012–10022.
[75] Ze Liu et al. “Swin transformer v2: Scaling up capacity and resolution”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 12009–12019.
[76] Jingyun Liang et al. “Hierarchical Conditional Flow: A Unified Framework for Image Super-Resolution and Image Rescaling”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 4056–4065. DOI: 10.1109/ICCV48922.2021.00404.
[77] Andreas Lugmayr et al. “Srflow: Learning the super-resolution space with normalizing flow”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer. 2020, pp. 715–732.
[78] Mingqing Xiao et al. “Invertible image rescaling”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer. 2020, pp. 126–144.
[79] Kai Zhang et al. “Designing a Practical Degradation Model for Deep Blind Image Super-Resolution”. In: *IEEE International Conference on Computer Vision*. 2021, pp. 4791–4800.
[80] Jinjin Gu et al. “Blind Super-Resolution With Iterative Kernel Correction”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.
[81] Manya V. Afonso, José M. Bioucas-Dias, and Mário A. T. Figueiredo. “Fast Image Recovery Using Variable Splitting and Constrained Optimization”. In: *IEEE Transactions on Image Processing* 19.9 (2010), pp. 2345–2356. DOI: 10.1109/TIP.2010.2047910.
[82] Kostadin Dabov et al. “Image denoising by sparse 3-D transform-domain collaborative filtering”. In: *IEEE Transactions on image processing* 16.8 (2007), pp. 2080–2095.
[83] Na Liu et al. “Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9 (2021), pp. 7817–7830.
[84] Weisheng Dong et al. “Hyperspectral image super-resolution via non-negative structured sparse representation”. In: *IEEE Transactions on Image Processing* 25.5 (2016), pp. 2337–2352.
[85] Xuan-Qi Wang and Teng-Yu Ji. “NSTMR: Super resolution of Sentinel-2 images using nonlocal nonconvex surrogate of tensor multirank”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 5694–5706.
[86] Ting Xu et al. “An Iterative Regularization Method Based on Tensor Subspace Representation for Hyperspectral Image Super-Resolution”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–16. ISSN: 1558-0644. DOI: 10.1109/TGRS.2022.3176266.
[87] Jun Zhang, Zhaoyang Liu, and Mingxi Ma. “Hyperspectral image fusion with a new hybrid regularization”. en. In: *Computational and Applied Mathematics* 41.6 (July 2022), p. 241. ISSN: 1807-0302. DOI: 10.1007/s40314-022-01950-y. (Visited on 09/20/2023).
[88] Yifan Zhang, Steve De Backer, and Paul Scheunders. “Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images”. In: *IEEE Transactions on Geoscience*

and Remote Sensing 47.11 (Nov. 2009), pp. 3834–3843. ISSN: 1558-0644. DOI: 10.1109/TGRS.2009.2017737.

[89] Shutao Li et al. "Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization". In: *IEEE Transactions on Image Processing* 27.8 (Aug. 2018), pp. 4118–4130. ISSN: 1941-0042. DOI: 10.1109/TIP.2018.2836307.

[90] Meng Xu et al. "Hyperspectral and Multispectral Image Fusion via Regularization on Nonlocal Structure Tensor Total Variation". en. In: *Proceedings of the International Conference on Aerospace System Science and Engineering 2021*. Ed. by Zhongliang Jing and Dmitry Strelets. Lecture Notes in Electrical Engineering. Singapore: Springer Nature, 2023, pp. 225–238. ISBN: 9789811681547. DOI: 10.1007/978-981-16-8154-7_18.

[91] Renwei Dian and Shutao Li. "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization". In: *IEEE Transactions on Image Processing* 28.10 (2019), pp. 5135–5146.

[92] Lei Zhang et al. "Exploiting clustering manifold structure for hyperspectral imagery super-resolution". In: *IEEE Transactions on Image Processing* 27.12 (2018), pp. 5969–5982.

[93] Luis Salgueiro Romero, Javier Marcello, and Verónica Vilaplana. "Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks". en. In: *Remote Sensing* 12.15 (Jan. 2020), p. 2424. ISSN: 2072-4292. DOI: 10.3390/rs12152424. (Visited on 10/02/2023).

[94] Weisheng Dong et al. "Model-Guided Deep Hyperspectral Image Super-Resolution". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 5754–5768. ISSN: 1941-0042. DOI: 10.1109/TIP.2021.3078058.

[95] Ian Dowman and Hannes I Reuter. "Global geospatial data from Earth observation: status and issues". In: *International Journal of Digital Earth* 10.4 (2017), pp. 328–341.

[96] Guo Ni and Wang Xiaoping. "Advances and developing opportunities in remote sensing of drought". In: *Journal of Arid Meteorology* 33.1 (2015), p. 1.

[97] Gabriel Stefanini Vicente and Robert Marty. *BlackMarblePy: Georeferenced Rasters and Statistics of Nighttime Lights from NASA Black Marble*. https://worldbank.github.io/blackmarblepy. 2023. DOI: 10.5281/zenodo.10667907. URL: https://worldbank.github.io/blackmarblepy.

[98] Guoqing Zhang et al., eds. *The Design, Data Processing and Applications of Luojia 1-01 Satellite*. 2019. URL: https://www.mdpi.com/journal/sensors/special_issues/RS_Luojia_01.

[99] Xi Li et al. "A preliminary investigation of Luojia-1 night-time light imagery". In: *Remote Sensing Letters* 10.6 (2019), pp. 526–535. DOI: 10.1080/2150704X.2019.1577573.

[100] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[101] OpenStreetMap. 2024. URL: https://www.openstreetmap.org.

[102] Christopher F. Brown et al. "Dynamic World, Near real-time global 10m land use land cover mapping". In: *Scientific Data* 9.1 (2022), p. 251. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01307-4. URL: https://doi.org/10.1038/s41597-022-01307-4.

[103] Richard Beare, Bradley Lowekamp, and Ziv Yaniv. "Image Segmentation, Registration and Characterization in R with SimpleITK". In: *Journal of Statistical Software* 86 (Aug. 2018), p. 8. ISSN: 1548-7660. DOI: 10.18637/jss.v086.i08. pmid: 30288153.

[104] Gregery T Buzzard et al. "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium". In: *SIAM Journal on Imaging Sciences* 11.3 (2018), pp. 2001–2020.

[105] Jaimyun Jung et al. "Super-resolving material microstructure image via deep learning for microstructure characterization and mechanical behavior analysis". In: *npj Computational Mathematics* 7, 96 (Jan. 2021), p. 96. DOI: 10.1038/s41524-021-00568-8.

[106] Kai Zhang et al. "Plug-and-play image restoration with deep denoiser prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2021), pp. 6360–6376.

[107] Lin Zhang et al. "FSIM: A feature similarity index for image quality assessment". In: *IEEE transactions on Image Processing* 20.8 (2011), pp. 2378–2386.

[108] Christopher D. Elvidge et al. "Indicators of Electric Power Instability from Satellite Observed Nighttime Lights". In: *Remote Sensing* 12.19 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12193194. URL: https://www.mdpi.com/2072-4292/12/19/3194.

[109] Tony A. Cole et al. "Synergistic Use of Nighttime Satellite Data, Electric Utility Infrastructure, and Ambient Population to Improve Power Outage Detections in Urban Areas". In: *Remote Sensing* 9.3 (2017). ISSN: 2072-4292. DOI: 10.3390/rs9030286. URL: https://www.mdpi.com/2072-4292/9/3/286.

[110] Juan P. Montoya-Rincon et al. "On the Use of Satellite Nightlights for Power Outages Prediction". In: *IEEE Access* 10 (2022), pp. 16729–16739. DOI: 10.1109/ACCESS.2022.3149485.

[111] Guy Shacht et al. "Single Pair Cross-Modality Super Resolution". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 6374–6383. DOI: 10.1109/CVPR46437.2021.00631.

[112] Jinhu Bian et al. "A new spatiotemporal fusion model for integrating VIIRS and SDGSAT-1 Nighttime light data to generate daily SDGSAT-1 like observations". In: *International Journal of Digital Earth* 18.1 (2025), p. 2472912. DOI: 10.1080/17538947.2025.2472912.