DeepLight: Reconstructing High-Resolution Observations of Nighttime Light With Multi-Modal Remote Sensing Data

Lixian Zhang[1,5] , **Runmin Dong**[2,5,*] , **Shuai Yuan**[3] , **Jinxiao Zhang**[2] , **Mengxuan Chen**[2] , **Juepeng Zheng**[4] , **Haohuan Fu**[2,5,*]

[1]National supercomputing center in Shenzhen
[2]Tsinghua University
[3]The University of Hong Kong
[4]Sun-Yat San University
[5]Tsinghua University - Xi'an Institute of Surveying and Mapping Joint Research Center

Abstract

Nighttime light (NTL) remote sensing observation serves as a unique proxy for quantitatively assessing progress toward meeting a series of Sustainable Development Goals (SDGs), such as poverty estimation, urban sustainable development, and carbon emission. However, existing NTL observations often suffer from pervasive degradation and inconsistency, limiting their utility for computing the indicators defined by the SDGs. In this study, we propose a novel approach to reconstruct high-resolution NTL images using multi-modal remote sensing data. To support this research endeavor, we introduce DeepLightMD, a comprehensive dataset comprising data from five heterogeneous sensors, offering fine spatial resolution and rich spectral information at a national scale. Additionally, we present DeepLightSR, a calibration-aware method for building bridges between spatially heterogeneous modality data in the multi-modality super-resolution. DeepLightSR integrates calibration-aware alignment, an auxiliary-to-main multi-modality fusion, and an auxiliary-embedded refinement to effectively address spatial heterogeneity, fuse diversely representative features, and enhance performance in $8\times$ super-resolution (SR) tasks. Extensive experiments demonstrate the superiority of DeepLightSR over 8 competing methods, as evidenced by improvements in PSNR (2.01 dB $\sim$ 13.25 dB) and PIQE (0.49 $\sim$ 9.32). Our findings underscore the practical significance of our proposed dataset and model in reconstructing high-resolution NTL data, supporting efficiently and quantitatively assessing the SDG progress. The code and data will be available at https://github.com/xian1234/DeepLight.

1 Introduction

The pursuit of the Sustainable Development Goals (SDGs) [UnitedNations, 2015] encounters various hurdles such as data insufficiency and lack of quantitative progress assessment [Xi *et al.*, 2023]. Earth observation data from satellites emerges as a promising solution by providing large-scale quantitative attributes [Zhang *et al.*, 2024b] and capturing detailed spatial information essential for understanding Earth's ecosystem, thus being recognized as a highly relevant tool in accomplishing SDGs [Guo *et al.*, 2023].

Observations of nighttime light (NTL) over several decades effectively characterize both the extent and intensity of human activities, providing unique insight for quantitatively assessing the progress toward the SDGs [Maso *et al.*, 2023], such as investigating poverty and inequality [McCallum *et al.*, 2022], monitoring urban sustainable developments [Elvidge *et al.*, 1997], and estimating carbon emissions [Huang *et al.*, 2014; Levin *et al.*, 2020]. Due to the lack of onboard calibration, sensor degradation, and satellite orbit drift, the existing historical record of low-resolution NTL suffers from severe quality degradation like blurring, blooming effect, and over-saturation phenomenon [Abrahams *et al.*, 2018], highlighting the urgent need for improving the historical NTL data collection. To alleviate these problems, a cost-effective way is to introduce super-resolution (SR) techniques, which reconstruct high-resolution (HR) NTL images from low-resolution (LR) NTL images.

However, as illustrated in Fig. 1, the degradations in historical LR NTL observations are composed of spatially heterogeneous inconsistency, coarse resolution, sensor blurring, atmospheric scattering, etc. The complexity of degradation types makes it a crucial task to improve the quality of LR NTL data for over two decades [Falchi *et al.*, 2016]. Therefore, We present the task of reconstructing HR NTL with multiple modalities in this study. Given an LR NTL image, the algorithm is expected to fuse multi-modality features to guide the reconstruction of calibrated HR NTL image. Nevertheless, the LR images in existing SR datasets are usually manual-simulated with common degradations [Dong *et al.*, 2023]. Considering the generalization of the model in addressing various degradations relies on the representativeness of SR datasets, the existing SR datasets are inadequate to support the training for this task.

To address the aforementioned challenge, we build a pioneering large-scale multi-modality dataset, DeepLightMD,
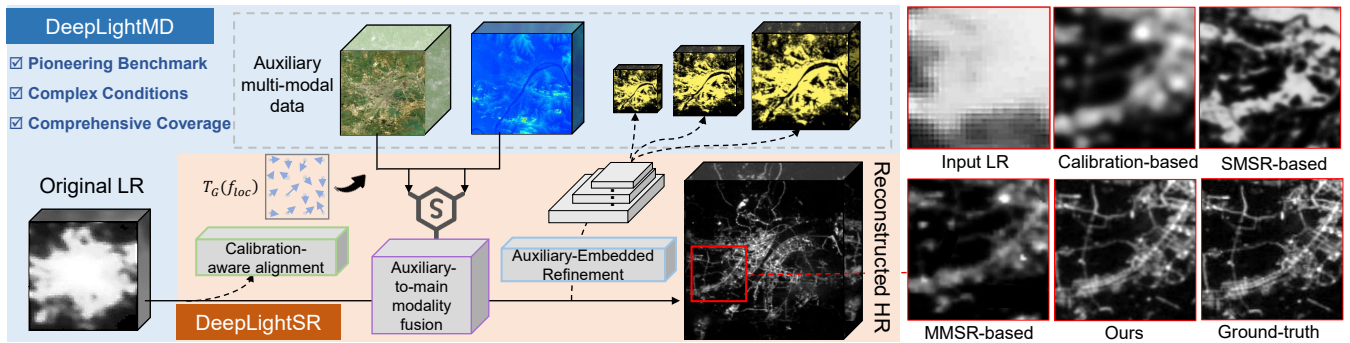
Figure 1: An intuitive example of reconstructing HR NTL image in DeepLightMD using DeepLightSR. The modalities of Daytime Multispectral Observations (DMO), Digital Elevation Model (DEM), and Impervious Surface Product (ISP) are jointly utilized to boost the reconstruction of HR NTL images from LR NTL images.

including LR NTL data, HR NTL data, daytime multispectral observations (DMO), digital elevation maps (DEM), and impervious surface products (ISP). DeepLightMD is built to address misalignment, blur, and noise in the $8\times$ reconstruction of HR NTL by incorporating multi-modality data. The involvement of auxiliary modalities has different purposes, with DMO providing finer spatial information and richer spectral information, DEM for spatial consistency, and ISP for importance guidance. A composition of pre-processing methods is introduced to remove noises and geometrical errors in modalities, achieving a geometrically aligned multi-modality dataset. The spatially heterogeneous misalignment and complex inconsistency make DeepLightMD a challenging benchmark dataset for multi-modality super-resolution under complicated degradations.

To cope with the reconstruction of HR NTL with multi-modality, we propose DeepLightSR, including three modules (*i.e.*, calibration-aware alignment, auxiliary-to-main multi-modality fusion, and auxiliary-embedded refinement) to calibrate modalities under complex inconsistent degradation, address feature fusion of different representative modes, and conduct large-ratio super-resolution tasks. First, the calibration-aware alignment calibrates the main modality and implicitly aligns auxiliary modalities under different representative modes. Second, an auxiliary-to-main multi-modality fusion module extract and distinguish useful low-level features from auxiliary modalities, and then fuse the representative multi-modality features in a stepwise manner to achieve harmonious and aligned features. Third, auxiliary-embedded refinement is proposed to cooperate with auxiliary supervision for detail reconstruction improvement, and multiscale outputs for multi-scale structural information recovering in $8\times$ super-resolution tasks.

We summarize the contributions of this paper as follows: 1) We propose the DeepLightMD, a pioneering national-scale dataset incorporating multi-modal remote sensing data. The DeepLightMD brings a promising way of reconstructing HR NTL images and also a challenge for developing new SR methods capable of addressing complicated degradations in misaligned multi-modality data. 2) We propose DeepLightSR with a calibration-aware alignment, an auxiliary-to-main multi-modality fusion, and an auxiliary-embedded re-

finement, to build bridges between spatially heterogeneous modality data in reconstucting $\times 8$ SR NTL task. 3) Experimental results demonstrate that our method achieves the best results compared with 8 competing methods, highlighting the effectiveness of generating large-scale high-resolution NTL data with our DeepLight SR, supporting efficiently and quantitatively assessing the SDG progress.

2 Related Work

Considering that this research is the very first attempt to reconstruct HR NTL data using LR NTL with complex degradation and inconsistency in $\times 8$ SR tasks, we focus on investigating the efforts that have been paid (although with a smaller SR scaling factor) or can potentially be applied to achieve improved NTL datasets based on historical NTL records of Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) from both remote sensing (RS) and computer vision (CV) sides.

2.1 Calibration-Based Methods in Remote Sensing

Assuming that there is a place where ground truth NTL intensity seldom changed over the past several decades, the main theory of calibration-based methods is to calibrate any pairs of cross-sensor NTL images from one to another using this invariant target. Lots of efforts have been paid to reconstructing calibrated NTL dataset with a single invariant target [Elvidge *et al.*, 2009], several invariant targets [Wu *et al.*, 2013], and invariant regions [Wang *et al.*, 2022]. Apart from expert knowledge, another way relies on determining invariant pixels using machine learning and deep-learning algorithms. For example, A ridge sampling and regression method is proposed to calibrate the NTL time series over the entire globe [Zhang *et al.*, 2016]. An embedded convolutional neural network (CNN) framework is proposed to calibrate NTL images [Chen *et al.*, 2021]. Modified U-Nets were introduced to achieve NTL images from different remote sensing modalities [Zhang *et al.*, 2021b]. However, the existing calibration-based methods are proposed to calibrate NTL images with $\times 2$ scaling factor and find it challenging to address most degradation issues in DMSP-OLS NTL data, resulting in limitations of practical and extended applications.

2.2 Super-Resolution -Based Methods in Computer Vision

Quantities of deep learning methods have been proposed in effectively conducting super-resolution (SR) tasks in CV [Zhang et al., 2021a; Dong et al., 2024], such as SRCNN [Dong et al., 2016], SRResNet [Ledig et al., 2017], and RCAN [Haut et al., 2019]. However, most existing SR methods are designed based on the assumption that LR input and HR target lie in the same domain, which is not valid in this task. We summarize cross-modality SR methods which are or can potentially be applied in applications where LR and HR belong to different domains. The first category is the single-modality SR methods, which do not involve auxiliary modality in the reconstruction of SR results, such as unpaired LR-HR under different domains [Maeda, 2020]. Besides, several efforts have been paid to different SR targets, such as magnetic resonance (MR) images [Fang et al., 2022], depth images [Sun et al., 2021], and thermal images [Gupta and Mitra, 2022]. In order to improve the reconstruction performance, auxiliary modalities are introduced to provide useful information to guide the reconstruction. For example, HR RGB visual modality is introduced in the reconstruction of non-visual data [Shacht et al., 2021]. JMDL model is proposed to solve the general SR problem with one auxiliary modality [Deng and Dragotti, 2020]. However, the existing multi-modality SR methods are designed to alleviate simple degradation and weak misalignment problem in conducting SR tasks, which are difficult to address the complicated degradation and spatially heterogeneous inconsistency of multi-modality NTL SR tasks.

3 DeepLightMD

3.1 Data Collection

LR NTL: Considering that the officially released DMSP-OLS NTL data suffers from interannual inconsistency [Abrahams et al., 2018], the Prolonged Artificial Nighttime-light DAtaset (PANDA) [Zhang et al., 2024a] is selected as LR NTL input, which is a long-term DMSP-like NTL dataset of China ranging from 1984-2020. The PANDA can be openly available with longer temporal coverage, better temporal consistency, and improved quality compared to the official DMSP annual composited NTL. The LR NTL data serves as the main modality input, providing information on coarse NTL intensity.

HR NTL: The composed Luojia NTL data [Li et al., 2019] is selected as HR NTL in DeepLightMD. The Luojia-01 provides global NTL images at 130 m with 15 days as revisit time. Compared to other commonly used NTL datasets such as VIIRS, the Luojia-01 provides NTL observations with higher spatial resolution, more detailed information, and a wider radiometric range. Besides, the overpass time gap between DMSP and Luojia-01 is much smaller than any other publicly available NTL pairs, reducing the challenges posed by overpass time misalignment [Li et al., 2020].

DMO: Landsat-8 Operational Land Imager (OLI) is selected as DMO, providing relatively high-resolution (30 m) multi-spectral images since 2013. The OLI collects data with 8 spectral bands and one panchromatic band at $\sim$ 10 am on each

pass with higher spatial resolution, more spectral bands, and a wider radiometric range than NTL data, resulting in much richer features at both global and local scales. We retrieve 7 bands (Band 1-7) of DMO with 30 m resolution of Landsat-8 Top of Atmosphere (TOA) Reflectance imageries.

DEM: We select 90 m DEM product releases by Shuttle Radar Topography Mission (SRTM) V4.1 [Jarvis, 2008] in DeepLightMD. The SRTM dataset was collected over an 11-day mission in February 2000, aboard the Space Shuttle Endeavor, and is available for ~80% of the globe (up to $60°$ north and south). The DEM records the elevation information, which is rarely changed over time. The temporal stability and spatial consistency of DEM make it beneficial for improving the spatiotemporal consistency in the reconstruction of NTL data.

ISP: We select GAIA [Gong et al., 2020] as ISP in this study. The GAIA is retrieved from Landsat series collections using a machine learning algorithm at 30 m resolution. The impervious surface usually indicates the region with constructions, making it an important indicator of human activities. Considering the stable NTL is usually relevant to the impervious surface area, the ISP is selected as auxiliary supervision information to improve the performance of the human settlement region. The impervious surface product of the year 2018 is selected in this research to meet up with the temporal availability of the main modalities.

All multi-modality remote sensing datasets are processed with normalization, spatial cropping, and random splitting, generating training, testing, and validating materials. Please refer to the supplementary material for details of the data-specific preprocessing.

3.2 Dataset Exploration

Overview: As illustrated in Fig. 2, DeepLightMD is a national-scale multi-modality SR dataset, consisting of training, validating, and testing materials. The training material contains 18,032 pairs of clips, and the validating material contains 2,254 pairs of samples. To demonstrate evaluation at both local and regional scales, the testing material contains 2,254 pairs of clips and three province-level samples, covering Anhui, Beijing, and Taiwan. Each pair of clips contains LR NTL images (256×256 in size), HR NTL images ($2,048 \times 2,048$ in size), DMO ($2,048 \times 2,048$ in size, 7 bands), DEM ($2,048 \times 2,048$ in size), and ISP ($2,048 \times 2,048$ in size).

Statistics: Different from existing multi-modality SR datasets, the modalities in DeepLightMD have huge gaps regarding distributions, radiometric ranges, and spatial resolutions, increasing the difficulties and complexities in model convergence. For example, the DMO data has 7 spectral bands, much more than the others (1 band). The DEM data is in long-tail distribution with a maximum of over 8,000 m, but the majority of pixel values are lower than 1,000. The ISP data is a binary product in which the number of pixels with a value of 0 is 112 times that of pixels with a value of 1. The original LR NTL data is with a radiometric range of 6 bits, smaller than DMO (14 bits) and HR NTL data (12 bits). Considering that the resolution of PANDA is 1 km and the multi-modality super-resolution task is conducted in an $\times 8$
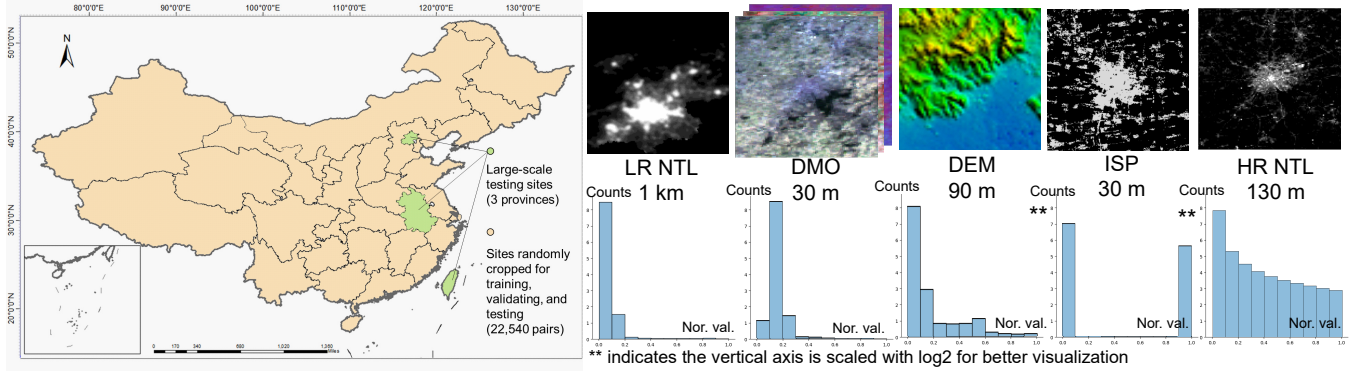
Figure 2: An exploration of the spatial coverage and normalized pixel value distribution histograms of modalities in DeepLightMD.

super-resolution task, we aim at generating SR NTL images with 125 m resolution in this task.

Highlights: We summarize the highlights of DeepLightMD as follows: 1) Pioneering Benchmark: DeepLightMD introduces a pioneering benchmark dataset for NTL multimodality SR tasks, facilitating robust reconstruction of high-resolution NTL imagery and advancing research in both RS and CV domains. 2) Complex Conditions: DeepLightMD presents challenges with complex degradation factors, including spatial misalignment and diverse distributions, demanding models to generalize fusion, alignment, and deblurring techniques for effective performance. As a result, DeepLightMD is more difficult than most existing multi-modality SR datasets. 3) Comprehensive Coverage: DeepLightMD encompasses diverse NTL types across China, spanning various landcover, energy consumption, socioeconomic, and cultural attributes. It features five modalities, from daytime to nighttime observations, capturing natural and human-related measures for comprehensive analysis.

4 DeepLightSR

4.1 Overall Architecture

As illustrated in Fig. 3, the DeepLightSR framework partitions multi-modality data into four distinct categories, including main modality input (LR NTL), auxiliary modality inputs (DMO and DEM), main modality supervision (HR NTL), and auxiliary modality supervision (ISP). First, the inputs from both primary and auxiliary modalities undergo calibration-aware alignment to attain aligned features for each modality (i.e., F_{NTL}, F_{DMO}, and F_{DEM}). Second, the aligned features of auxiliary modalities, F_{DMO} and F_{DEM}, undergo fusion and combination via a cross-modality fusion module to yield the fused auxiliary feature, F_{aux}^f. This fused feature, F_{aux}^f, is then integrated with the calibrated F_{NTL} to produce the primary fused feature, F_{main}^f. Third, the final SR NTL output is attained under refinement incorporating with multi-scale and auxiliary supervision information in the upsampling process.

The DeepLightSR is designed in an end-to-end scheme, consisting of two functions $\mathcal{F}_1$ and $\mathcal{F}_2$. Given the LR NTL image $N_L \in \mathbb{R}^{h \times w}$ and the multi-modality input set

$M = M^{DMO} \in \mathbb{R}^{rh \times rw}, M^{DEM} \in \mathbb{R}^{rh \times rw}$, $\mathcal{F}1$ generates multi-scale SR NTL sets $NG = NG, 1, \cdots, NG, m$. Subsequently, $\mathcal{F}2$ utilizes the obtained NG, m as input to reconstruct the auxiliary SR ISP image $G_{ISP} \in 0, 1^{rh \times rw}$.

Overall, DeepLightSR can be optimized with supervision from the HR NTL image $N_H \in \mathbb{R}^{rh \times rw}$ and auxiliary supervision from the ISP M_{ISP}. Here, r, h, and w denote the ratio in the SR task, the height, and the width of N_L, respectively. The comprehensive objective function for training DeepLightSR is formulated as follows:

$$\Theta = \underset{\Theta}{\arg\min}[\alpha\mathcal{F}_1 || \sum_{j=1}^{m} \beta_j (\sum_{i=1}^{n} (N_L^i, M^i; \Theta)_j - N_{H,j}^i||_1) +$$
$$(1-\alpha)\mathcal{F}_2 \sum_{i=1}^{n} (M_{ISP}^i \cdot log(N_{G,m}^i; \Theta) +$$
$$(1 - M_{ISP}^i) \cdot log(1 - (N_{G,m}^i; \Theta))],$$
(1)

where Θ is the parameters of DeepLightSR, the n and m denote the number of training samples and multiscale outputs, respectively. The $N_{G,m} \in \mathbb{R}^{rh \times rw}$ represents the obtained SR NTL image in $\mathcal{F}_1$. The α and β denote the weights of two functions and multiscale tasks.

4.2 The Calibration-Aware Alignment (CAA)

In DeepLightMD, the complexity of inconsistency and misalignment across different modality inputs varies significantly, underscoring the necessity of calibrating and aligning each input modality before feature fusion. To this end, the CAA leverages the learned warp parameters to align the severely misaligned primary modality and implicitly guide the alignment of auxiliary modalities. Compared to independently aligning each modality, the proposed calibration-aware strategy optimally utilizes the learned warp for efficient alignment of weakly misaligned modalities and enhances overall model convergence.

The CAA contains three key submodules. First, the auxiliary modalities (i.e., DMO and DEM) M are rescaled to $M \downarrow \in \mathbb{R}^{h \times w}$ via learneable downsampling block. The downsampling block comprises several CNN layers with a stride of 2, incorporating learnable parameters to retain essential information while minimizing redundant details during
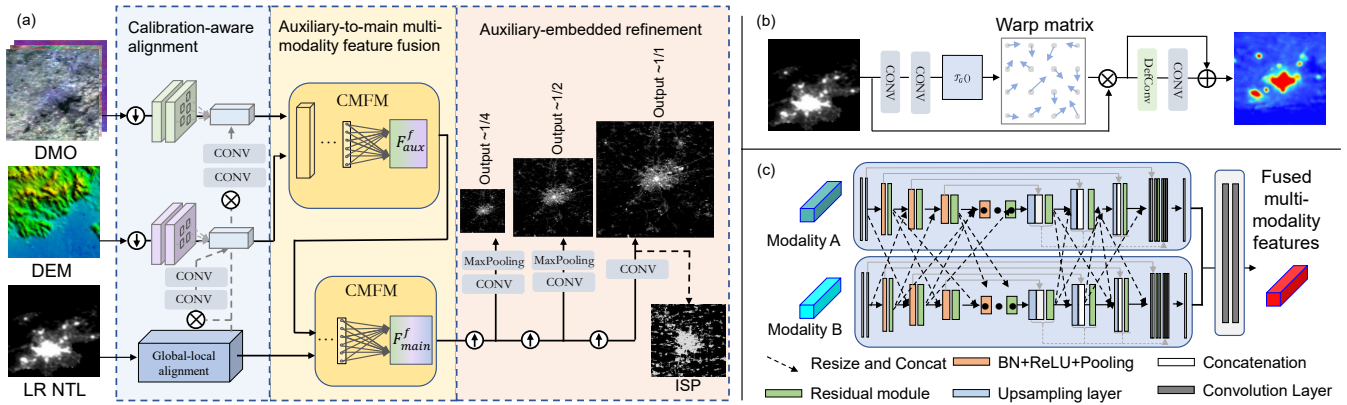
Figure 3: Overview of the proposed DeepLightSR. (a) The overall architecture of DeepLightSR. (b) The proposed global-local alignment in CAA. (c) The proposed cross-modality fusion module in AMFF.

the downscaling process. Second, for the LR NTL input, we propose a global-local alignment (GLA) submodule to generate calibrated NTL feature maps. As illustrated in Fig. 3 (b), the GLA contains two components: spatial transformer networks (STN) [Jaderberg *et al.*, 2015] for global alignment and deformable convolutional neural alignment networks (DefCAN) for local alignment. The STN layer employs a localization network f_{loc} and a parameterized grid generator $\mathcal{T}_G$ to capture the global warp matrix ω.

$$\omega = \mathcal{T}_G(f_{loc}(\boldsymbol{N}_L)), \qquad (2)$$

where the localization network f_{loc} comprises two CNN layers for regression. The obtained pixel-wise warp matrix ω is utilized as a global approximation to align the input NTL $\boldsymbol{N}_L$. To address the local irregular positions, DefCAN employs a deformable CNN layer (DefConv) [Dai *et al.*, 2017] and a reconstruction CNN layer to generate the aligned NTL features F_{NTL}. Third, the DEM and DMO are calibrated with the assistance of the obtained $\boldsymbol{N}_L$, leveraging geometric dependencies between modalities and high spatial consistency in DEM data to simplify auxiliary modality alignment. Two additional CNN layers are employed to surpass remaining fractional misalignments between the NTL feature and each auxiliary modality feature. By cooperating with the obtained warp matrix ω, the DMO implicitly "awares" the geometrical deformation information into the calibration process, resulting in coherent and aligned features.

4.3 Auxiliary-to-main Multi-Modality Feature Fusion (AMFF)

Significant representativeness gaps exist between different input modality features within DeepLightMD. DMO features (F_{DMO}) encapsulate multispectral daytime observations with rich texture details, while DEM features (F_{DEM}) offer precise elevation information alongside high-frequency details. On the other hand, LR NTL features (F_{NTL}) primarily convey artificial light intensity information in the low-frequency domain. To mitigate cross-modality feature inconsistencies, we adopt a stepwise fusion approach, enabling flexible extraction and differentiation of low-level features while effec-

tively fusing and synthesizing high-level features, thereby enhancing the overall performance in generating fused features (F_{main}^f).

Consequently, the fusion of three modalities is simplified into a two-step fusion task. To facilitate efficient fusion of two modality features, we propose the cross-modality fusion module (CMFM) for both auxiliary modality feature fusion and main modality feature fusion tasks. As illustrated in Fig. 3 (c), CMFM comprises two parallel branches, each incorporating several repeated residual blocks to progressively extract and fuse features from both modality A and the projected feature from modality B in shallower layers. For deeper layer fusion, the extracted feature from a shallower layer is resized and concatenated into the deeper layer to aid in fusion. This resize-and-concatenate strategy enhances the preservation of global information by providing larger receptive fields, thereby facilitating fusion between modalities with differing representative characteristics and efficient generation of high-level fused features. Subsequently, the extracted modality features are concatenated and further fused using two convolutional blocks to produce the fused multi-modality features.

4.4 Auxiliary-Embedded Refinement (AER)

Achieving SR results for large upsampling factors usually suffers from high uncertainty and incompleteness of the missing details and high-frequency information. To boost the reconstruction performance in the $8\times$ SR task, as illustrated in Fig. 3 (a), the AER incorporates multi-scale supervision alongside auxiliary supervision to furnish ample guidance for the model in reconstructing intricate textures while mitigating artifact generation during the upsampling process. For multiscale predictions, we employ three upsampling blocks, each comprising a 1×1 point-wise CNN layer followed by a pixel shuffle layer [Shi *et al.*, 2016], facilitating the reconstruction of feature maps into sizes enlarged by a factor of $2\times$. To refine NTL reconstruction across different scales, we compute losses between the output of each upsampling block and the ground truth, which is interpolated to match the same size.

Moreover, as the task entails long-tail distribution regression, reconstructing HR NTL images with L1 loss encounters significant overall underestimation owing to extremely im-

balanced distributions (e.g., more than 95% of pixels exhibit zero NTL intensity in China). Given that NTL applications primarily focus on areas with high human activity and NTL values, an additional CNN layer is introduced to derive ISP G_{ISP} from the obtained $N_{G,m}$. The auxiliary supervision of ISP furnishes crucial guidance to prioritize attention towards manually constructed areas, addressing this imbalance and enhancing model performance.

5 Experiments

5.1 Experimental Setups

Evaluation metrics: We select 6 metrics for the comprehensive evaluation, including PSNR, SSIM, spectral angle mapper (SAM) [Kruse *et al.*, 1993], universal image quality index (UIQI) [Wang and Bovik, 2002], correlation coefficient (CC), and perception-based image quality evaluator (PIQE) [N *et al.*, 2015]. Apart from the commonly used PSNR, SSIM, and CC, the SAM, UIQI, and PIQE are utilized to evaluate spectral fidelity, global correlation, and perceptual quality, respectively.

Competing methods: We compare the proposed method with 7 representative methods, which can be divided into three categories: calibration-based methods (denoted by "Calibration"): calibration-based auto-encoder CNN (CACNN) [Chen *et al.*, 2021], single modality super-resolution methods (denoted by "SMSR"): SMSR: SR-ResNet [Ledig *et al.*, 2017], TTSR [Yang *et al.*, 2020], UPSR [Maeda, 2020], and UGSR [Gupta and Mitra, 2022], and multi-modality super-resolution methods(denoted by "MMSR"): BDT[Deng *et al.*, 2023], CMSR [Shacht *et al.*, 2021] and MNSR [Xie *et al.*, 2022]. Note that the SRResNet and TTSR are trained using pairs of simulated LR NTL downscaled from HR NTL.

5.2 Experimental Results on DeepLightMD

As depicted in Tab. **??**, our proposed approach consistently outperforms competing methods across all six evaluation metrics, showcasing significant enhancements, particularly in SAM and UIQI metrics. These improvements indicate the ability of our method to preserve both spectral and geometric fidelity in reconstructing SR NTL images. Specifically, while the Calibration method (*i.e.*, CACNN) excels in CC scores but demonstrates lower performance in pixel-wise evaluation metrics, other Single-Modality SR (SMSR) methods exhibit superior pixel-wise correctness. On the other hand, Multi-Modality SR (MMSR) methods generally achieve superior reconstruction performance, especially in scenarios with severe misalignment. However, both Calibration and SMSR methods fall short in correcting misalignment or achieving spatially precise reconstructions, as illustrated in Fig. 4. By leveraging multi-modality data sources, DeepLightSR achieves substantial improvements in pixel-wise performance compared to competing MMSR methods, demonstrating superior overall performance in addressing inconsistencies and enhancing NTL image quality.

To further validate the capabilities of DeepLightSR in large-scale NTL SR tasks, we present regional-scale SR results for three province-level cases. Fig. 5 showcases visual

evaluations for Anhui province, located in central China. We provide representative examples of zoomed-in areas to illustrate the effectiveness of our method. Notably, the outer shape and intensity of light in areas outside the dashed boxes are aligned with HR NTL images. However, within the dashed boxes, the intensity of light appears excessively high due to over-saturation and blooming effects. Despite these challenges, our super-resolution results demonstrate the method's capability to effectively utilize multi-modality information to correct various reconstruction errors. Absolute errors between retrieved SR NTL images and ground truth further underscore the potential of our method in large-scale multi-modality NTL SR tasks. For additional evaluation results, please refer to the supplementary material.

5.3 Ablation Study

Contributions of Multi-Modality Inputs: We conduct experiments to assess the contribution of each multi-modality input in NTL image reconstruction by training the model with only one or two inputs of multi-modality. We simulate the reduction of specific modalities by replacing the extracted modality features in the cross-modality fusion module with other remaining features. Results, reported in Tab. **??**, indicate a significant decrease in all evaluation metrics, particularly in CC, when excluding DMO modality, highlighting the valuable insights provided by daytime observations in distinguishing human activity areas from natural ones. Similarly, reducing DEM information as input results in a notable drop in spectral alignment metrics like SAM, emphasizing the crucial role of DEM in maintaining spatial alignment and calibrating other modalities.

Contributions of Auxiliary Supervision Information: We further investigate the contributions of auxiliary supervision information by conducting experiments without auxiliary supervision (*i.e.*, ISP) during model training. Tab. **??** reveals performance degradation, especially in perceptual evaluator PIQE, indicating that on-the-fly auxiliary supervision guides the model to focus on reconstruction with improved realistic perception, enhancing its applicability in real-world scenarios. Additionally, since ISP products are retrieved from Landsat series observations, incorporating ISP as a modality input introduces redundancy, leading to decreased overall performance. Hence, ISP serves primarily as auxiliary supervision to enhance the recovery performance of structural information.

Effectiveness of Each Component of the Proposed Method: Additional experiments are conducted to demonstrate the effectiveness of each component in DeepLightSR. Specifically, replacing CAA removes the cross-modality alignment network, while the setups of other components remain unchanged. Similarly, replacing AMFF removes the connection between input modality networks within the fusion module, and replacing AER eliminates the multiscale output module and corresponding losses. Results in Tab. **??** show that the calibration-alignment component endows the network with the ability to align spatially irregular inaccuracies for calibrated features. The feature fusion component addresses variance between fused features across network stages, enhancing the model's overall fusion of multi-
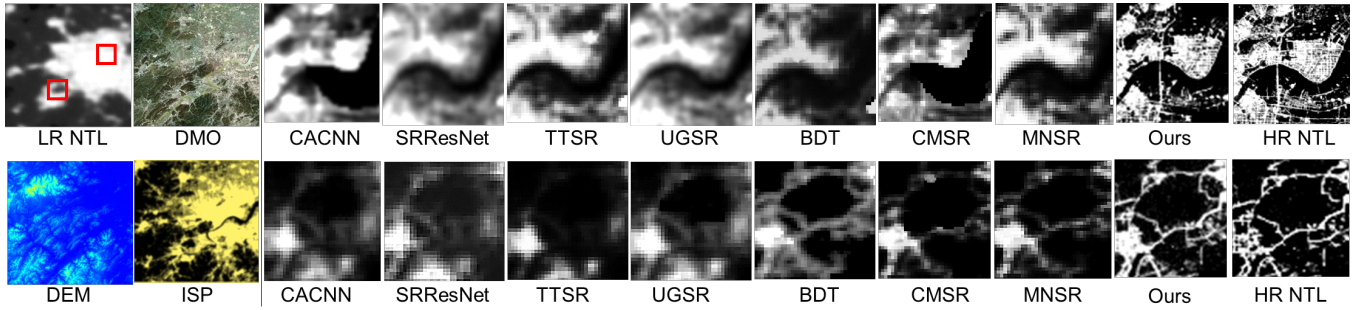
Figure 4: Visual examples for comparison in this study. Five modalities (*i.e.*, LR NTL data, DMO data, DEM data, ISP data, and HR NTL data), eight reconstruction results of competing methods, and the absolute difference between the results of DeepLightSR and HR NTL are presented.

Method	Category	Used modality(s)	PSNR ↑	SSIM ↑	SAM ↓	UIQI ↑	CC ↑	PIQE ↓
CACNN [Chen *et al.*, 2021]	Calibration	LR NTL + HR NTL	19.14	0.7542	0.1892	0.9105	0.9544	16.8792
SRResNet [Ledig *et al.*, 2017]		HR NTL	19.63	0.7617	0.2045	0.8826	0.9036	14.9643
TTSR [Yang *et al.*, 2020]	SMSR	HR NTL	21.86	0.8129	0.1957	0.9022	0.9163	13.1045
UPSR [Maeda, 2020]		LR NTL + HR NTL	23.67	0.8892	0.1451	0.9189	0.9570	9.8691
UGSR [Gupta and Mitra, 2022]		LR NTL + HR NTL	25.68	0.9038	0.1268	0.9269	0.9634	9.3519
BDT [Deng *et al.*, 2023]		LR NTL + HR NTL + DMO + DEM + ISP	29.31	0.9202	0.1034	0.9545	0.9793	9.1688
CMSR [Shacht *et al.*, 2021]	MMSR	LR NTL + HR NTL + DMO + DEM + ISP	28.68	0.9144	0.0994	0.9577	0.9711	8.3894
MNSR [Xie *et al.*, 2022]		LR NTL + HR NTL + DMO + DEM + ISP	30.38	0.9187	0.1083	0.9639	0.9775	8.0422
DeepLightSR (ours)		LR NTL + HR NTL + DMO + DEM + ISP	**32.39**	**0.9317**	**0.0849**	**0.9717**	**0.9846**	**7.5547**

Table 1: Quantitatively comparison on DeepLightMD regarding 6 evaluation metrics.
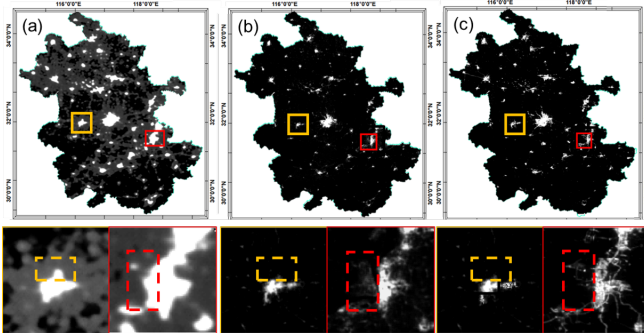

Figure 5: An example of large-scale visual evaluation. (a)-(c) are LR NTL image, SR NTL result of DeepLightSR, and HR NTL image. Dashed boxes are illustrated to better distinguish the effectiveness of DeepLightSR.

Ablation setups	PSNR ↑	SSIM ↑	SAM ↓	UIQI ↑	CC ↑	PIQE ↓
w/o LR NTL	22.67	0.7930	0.1739	0.9174	0.9133	12.7153
w/o DMO	23.17	0.8004	0.1705	0.9025	0.9086	13.1425
w/o DEM	27.57	0.8455	0.1461	0.9382	0.9529	10.6529
w/o ISP	31.86	0.9273	0.0975	0.9633	0.9825	9.9297
w/o CAA	26.56	0.8861	0.1472	0.9586	0.9477	11.592
w/o AMFF	28.62	0.8929	0.1644	0.9475	0.9208	13.5697
w/o AER	30.65	0.9238	0.0919	0.9641	0.9515	10.6388
Ours full	**32.39**	**0.9317**	**0.0849**	**0.9797**	**0.9846**	**8.5547**

Table 2: Experimental results of ablation study.

modality features with different representative modes. Lastly, AER enables the model to discern and align necessary details, improving its capability to address the blooming effect and provide reasonable estimations of NTL intensity.

Please refer to supplementary material for efficiency comparison and more ablation studies, such as multi-scale supervisions and hyper-parameter setups.

6 Conclusion and Future Work

In this paper, we introduce DeepLightMD, a pioneering national-scale benchmark dataset encompassing data from five modalities. Building upon this dataset, we propose DeepLightSR, a calibration-aware approach designed to re-construct HR NTL data. Both qualitative and quantitive results on DeepLightMD demonstrate the effectiveness of the proposed DeepLightSR in generating large-scale, high-resolution NTL data. Furthermore, we identify several avenues for future research with potential societal impact aligned with SDGs: 1) Consistent long-term NTL data reconstruction. There is a growing demand to extend our methodology to reconstruct consistent long-term NTL datasets, addressing temporal variability while ensuring alignment. 2) Incorporating more modalities in the reconstruction of NTL images. Incorporating additional modalities, such as Synthetic Aperture Radar (SAR) and social media data, holds promise in enhancing performance by capturing more nuanced aspects related to terrain and socioeconomic factors.

Acknowledgements

References

[Abrahams *et al.*, 2018] Alexei Abrahams, Christopher Oram, and Nancy Lozano-Gracia. Deblurring dmsp nighttime lights: A new method using gaussian filters and frequencies of illumination. *Remote Sensing of Environment*, 210:242–258, 2018.

[Chen *et al.*, 2021] Zuoqi Chen, Bailang Yu, Chengshu Yang, Yuyu Zhou, Shenjun Yao, Xingjian Qian, Congxiao Wang, Bin Wu, and Jianping Wu. An extended time series (2000-2018) of global npp-viirs-like nighttime light data from a cross-sensor calibration. *Earth System Science Data*, 13(3):889–906, 2021.

[Dai *et al.*, 2017] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE International Conference on Computer Vision, pages 764–773, 2017.

[Deng and Dragotti, 2020] Xin Deng and Pier Luigi Dragotti. Deep coupled ista network for multi-modal image super-resolution. *IEEE Transactions on Image Processing*, 29:1683–1698, 2020.

[Deng *et al.*, 2023] Shangqi Deng, Liang-Jian Deng, Xiao Wu, Ran Ran, and Rui Wen. Bidirectional dilation transformer for multispectral and hyperspectral image fusion. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023.

[Dong *et al.*, 2016] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.

[Dong *et al.*, 2023] Runmin Dong, Lixian Zhang, Weijia Li, Shuai Yuan, Lin Gan, Juepeng Zheng, Haohuan Fu, Lichao Mou, and Xiao Xiang Zhu. An adaptive image fusion method for sentinel-2 images and high-resolution images with long-time intervals. *International Journal of Applied Earth Observation and Geoinformation*, 121:103381, 2023.

[Dong *et al.*, 2024] Runmin Dong, Shuai Yuan, Bin Luo, Mengxuan Chen, Jinxiao Zhang, Lixian Zhang, Weijia Li, Juepeng Zheng, and Haohuan Fu. Building bridges across spatial and temporal resolutions: Reference-based super-resolution via change priors and conditional diffusion model. *arXiv preprint arXiv:2403.17460*, 2024.

[Elvidge *et al.*, 1997] Christopher D Elvidge, Kimberly E Baugh, Eric A Kihn, Herbert W Kroehl, and Ethan R Davis. Mapping city lights with nighttime data from the dmsp operational linescan system. *Photogrammetric Engineering and Remote Sensing*, 63(6):727–734, 1997.

[Elvidge *et al.*, 2009] Christopher D. Elvidge, Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3):595–622, 2009.

[Falchi *et al.*, 2016] Fabio Falchi, Pierantonio Cinzano, Dan Duriscoe, Christopher C. M. Kyba, Christopher D. Elvidge, Kimberly Baugh, Boris A. Portnov, Nataliya A. Rybnikova, and Riccardo Furgoni. The new world atlas of artificial night sky brightness. *Science Advances*, 2(6):e1600377, 2016.

[Fang *et al.*, 2022] Chaowei Fang, Dingwen Zhang, Liang Wang, Yulun Zhang, Lechao Cheng, and Junwei Han. Cross-modality high-frequency transformer for mr image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, pages 1584–92. Association for Computing Machinery, 2022.

[Gong *et al.*, 2020] Peng Gong, Xuecao Li, Jie Wang, Yuqi Bai, Bin Chen, Tengyun Hu, Xiaoping Liu, Bing Xu, Jun Yang, Wei Zhang, and Yuyu Zhou. Annual maps of global artificial impervious area (gaia) between 1985 and 2018. *Remote Sensing of Environment*, 236:111510, 2020.

[Guo *et al.*, 2023] Huadong Guo, Changyong Dou, Hongyu Chen, Jianbo Liu, Bihong Fu, Xiaoming Li, Ziming Zou, and Dong Liang. Sdgsat-1: the world's first scientific satellite for sustainable development goals. *Sci. Bull*, 68(1):34–38, 2023.

[Gupta and Mitra, 2022] Honey Gupta and Kaushik Mitra. Toward unaligned guided thermal super-resolution. *IEEE Transactions on Image Processing*, 31:433–445, 2022.

[Haut *et al.*, 2019] Juan Mario Haut, Ruben Fernandez-Beltran, Mercedes E Paoletti, Javier Plaza, and Antonio Plaza. Remote sensing image superresolution using deep residual channel attention. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9277–9289, 2019.

[Huang *et al.*, 2014] Qingxu Huang, Xi Yang, Bin Gao, Yang Yang, and Yuanyuan Zhao. Application of dmsp/ols nighttime light images: A meta-analysis and a systematic literature review. *Remote Sensing*, 6(8):6844–6866, 2014.

[Jaderberg *et al.*, 2015] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15. MIT Press, 2015.

[Jarvis, 2008] H.I. Reuter A. Nelson E. Guevara Jarvis, A. Hole-filled srtm for the globe version 4. Available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org), 2008.

[Kruse *et al.*, 1993] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2-3):145–163, 1993.

[Ledig *et al.*, 2017] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017.

[Levin *et al.*, 2020] Noam Levin, Christopher CM Kyba, Qingling Zhang, Alejandro Sánchez de Miguel, Miguel O Román, Xi Li, Boris A Portnov, Andrew L Molthan, Andreas Jechow, and Steven D Miller. Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment*, 237:111443, 2020.

[Li *et al.*, 2019] Xin Li, Taoyang Wang, Guo Zhang, Boyang Jiang, Peng Jia, Zhuxi Zhang, and Yuan Zhao. Planar block adjustment for china's land regions with luojia1-01 nighttime light imagery. *Remote Sensing*, 11(18), 2019.

[Li *et al.*, 2020] Xi Li, Noam Levin, Jinlong Xie, and Deren Li. Monitoring hourly night-time light by an unmanned aerial vehicle and its implications to satellite remote sensing. *Remote Sensing of Environment*, 247:111942, 2020.

[Maeda, 2020] Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 288–297, 2020.

[Maso *et al.*, 2023] Joan Maso, Alaitz Zabala, and Ivette Serral. Earth observations for sustainable development goals, 2023.

[McCallum *et al.*, 2022] Ian McCallum, Christopher Conrad Maximillian Kyba, Juan Carlos Laso Bayas, Elena Moltchanova, Matt Cooper, Jesus Crespo Cuaresma, Shonali Pachauri, Linda See, Olga Danylo, Inian Moorthy, et al. Estimating global economic well-being with unlit settlements. *Nature Communications*, 13(1):2459, 2022.

[N *et al.*, 2015] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Sumohana S. Channappayya, and Swarup S. Medasani. Blind image quality evaluation using perception based features. In *2015 Twenty First National Conference on Communications (NCC)*, 2015.

[Shacht *et al.*, 2021] Guy Shacht, Dov Danon, Sharon Fogel, and Daniel Cohen-Or. Single pair cross-modality super resolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6374–6383, 2021.

[Shi *et al.*, 2016] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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*, pages 1874–1883, 2016.

[Sun *et al.*, 2021] Baoli Sun, Xinchen Ye, Baopu Li, Haojie Li, Zhihui Wang, and Rui Xu. Learning scene structure guidance via cross-task knowledge transfer for single depth super-resolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7788–7797, 2021.

[UnitedNations, 2015] UnitedNations. Transforming our world: the 2030 agenda for sustainable development. *United Nations: New York, NY, USA*, 2015.

[Wang and Bovik, 2002] Zhou Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

[Wang *et al.*, 2022] Mingyue Wang, Chunhui Feng, Bifeng Hu, Nan Wang, Jintao Xu, Ziqiang Ma, Jie Peng, and Zhou Shi. A new framework for reconstructing time series dmsp-ols nighttime light data using the improved stepwise calibration (isc) method. *Remote Sensing*, 14(17), 2022.

[Wu *et al.*, 2013] Jiansheng Wu, Shengbin He, Jian Peng, Weifeng Li, and Xiaohong Zhong. Intercalibration of dmsp-ols night-time light data by the invariant region method. *International Journal of Remote Sensing*, 34(20):7356–7368, 2013.

[Xi *et al.*, 2023] Yanxin Xi, Yu Liu, Tong Li, Jintao Ding, Yunke Zhang, Sasu Tarkoma, Yong Li, and Pan Hui. A satellite imagery dataset for long-term sustainable development in united states cities. *Scientific data*, 10(1):866, 2023.

[Xie *et al.*, 2022] Wuyuan Xie, Tengcong Huang, and Miaohui Wang. Mnsrnet: Multimodal transformer network for 3d surface super-resolution. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12693–12702, 2022.

[Yang *et al.*, 2020] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. Learning texture transformer network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 5791–5800, 2020.

[Zhang *et al.*, 2016] Qingling Zhang, Bhartendu Pandey, and Karen C Seto. A robust method to generate a consistent time series from dmsp/ols nighttime light data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):5821–5831, 2016.

[Zhang *et al.*, 2021a] Lixian Zhang, Runmin Dong, Shuai Yuan, Weijia Li, Juepeng Zheng, and Haohuan Fu. Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction. *Remote Sensing*, 13(15):2872, 2021.

[Zhang *et al.*, 2021b] Lixian Zhang, Zhehao Ren, Runmin Dong, Bing Xu, and Haohuan Fu. Monitoring daily nighttime light based on modis and deep learning: A belgium case study. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 5032–5035, 2021.

[Zhang *et al.*, 2024a] Lixian Zhang, Zhehao Ren, Bin Chen, Peng Gong, Bing Xu, and Haohuan Fu. A prolonged artificial nighttime-light dataset of china (1984-2020). *Scientific Data*, 11(1):414, 2024.

[Zhang *et al.*, 2024b] Lixian Zhang, Shuai Yuan, Runmin Dong, Juepeng Zheng, Bin Gan, Dengmao Fang, Yang Liu, and Haohuan Fu. Swcare: Switchable learning and connectivity-aware refinement method for multi-city and diverse-scenario road mapping using remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 127:103665, 2024.