

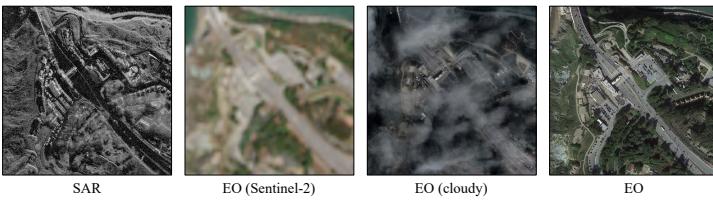
000 001 SAR2EARTH: A SAR-TO-EO TRANSLATION 002 DATASET FOR REMOTE SENSING APPLICATIONS 003 004

005 **Anonymous authors**

006 Paper under double-blind review

007 008 ABSTRACT 009

011 Electro-optical (EO) images are essential to a wide range of remote sensing ap-
012 plications. With the advent of data-driven models, the efficiency of EO image
013 analysis has significantly improved, enabling faster and more effective outcomes
014 in these applications. However, EO images have inherent limitations—they can-
015 not penetrate cloud cover and are unable to capture imagery at night. To overcome
016 these challenges, synthetic aperture radar (SAR) images are employed, as they can
017 operate effectively regardless of weather conditions or time of day. Despite this
018 advantage, SAR images come with their own difficulties: they are affected by
019 speckle noise, complicating analysis, and existing algorithms developed for EO
020 imagery are not directly transferable to SAR data. To address these issues, we
021 introduce SAR2Earth, a benchmark dataset specifically designed for SAR-to-EO
022 translation. By translating SAR images into EO-like representations, SAR2Earth
023 allows the extensive range of algorithms developed for EO imagery to be applied
024 effectively to SAR data. The dataset consists of 18 spatially aligned pairs of SAR
025 and EO images, collected from 8 distinct regions encompassing both urban and
026 rural. We provide comprehensive evaluations, detailed model analyses, and exten-
027 sive experimental results. All codes and datasets will be made publicly available
028 at <https://sar2earth.github.io>.



035 Figure 1: Examples based on various sensors and resolutions, captured around a similar time. The
036 SAR imagery is sourced from the Capella Space, Sentinel-2 data from Copernicus, and other EO
037 images from Google Earth.

038 1 INTRODUCTION 039

041 Remote sensing images provide the capability to observe the Earth on a large scale, making them
042 invaluable for analysis in various applications such as transportation (Ball et al., 2017), defense (Xu
043 et al., 2024), natural resource management (Kumar et al., 2015), disaster response (AlAli & Al-
044 abady, 2022), and environmental monitoring (Burke et al., 2021; Teng et al., 2023). However, the
045 vast amount of data generated poses significant challenges for manual analysis due to the time and
046 expertise required. The advent of data-driven models (Wang et al., 2021; Oh et al., 2023; Mall
047 et al., 2024; Kuckreja et al., 2024) has enabled more efficient and effective analysis of these images.
048 Electro-optical (EO) imagery has been the primary modality for remote sensing applications due to
049 its intuitive representation of the Earth. However, EO imagery has significant limitations: it cannot
050 penetrate cloud cover and is unable to capture images at night, restricting its utility in many scenar-
051 os (Seo et al., 2023; Low et al., 2023b). For instance, during natural disasters like floods—which
052 are often accompanied by heavy cloud cover—EO imagery becomes ineffective for timely disaster
053 assessment and response. To overcome these limitations, synthetic aperture radar (SAR) imagery is
employed. SAR sensors can operate independently of daylight and weather conditions, providing
consistent imaging capabilities, as illustrated in Figure 1. However, SAR images suffer from speckle

Table 1: Comparison of SAR2Earth dataset with existing SAR and EO multi-modality datasets. P, MS, and PS represent Panchromatic (P), Multi-Spectral (MS), and Pan-Sharpened (PS) image types. Tasks (DF, BE, I2I) denote Data Fusion, Building Extraction, and Image-to-Image translation.

| Modality | Dataset | Year | Sensors | Resolution (m) | Types | Task | Domain | Regions | Temporal |
|----------|------------------------------------|------|---------------|----------------|----------------|------|---------|-----------|---------------------|
| SAR | SEN12 (Schmitt et al., 2018) | 2018 | Sentinel-1 | 10 | VV | DF | Both | ≥ 80 | Dec 2016 – Nov 2017 |
| | SARpical (Wang & Zhu, 2018) | 2018 | TerraSAR-X | 1 | Unknown | DF | Urban | 1 | Jan 2009 – Dec 2013 |
| | SpaceNet6 (Shermeyer et al., 2020) | 2020 | Capella | 0.5 | HH, HV, VH, VV | BE | Urban | 1 | Aug 2019 |
| | MAVIC-T (Low et al., 2023a) | 2023 | GOTCHA | Unknown | Unknown | I2I | Unknown | 1 | Aug 2008 |
| | MAGIC (Low et al., 2024) | 2024 | GOTCHA, Umbra | Unknown | Unknown | I2I | Unknown | 4 | Unknown |
| | SAR2Earth (ours) | 2024 | Capella | 0.3 – 0.6 | HH | I2I | Both | 8 | Feb 2021 – Sep 2024 |
| EO | SEN12 (Schmitt et al., 2018) | 2018 | Sentinel-2 | 10 | RGB | DF | Both | ≥ 80 | Dec 2016 – Nov 2017 |
| | SARpical (Wang & Zhu, 2018) | 2018 | Aerial | 0.2 | RGB | DF | Urban | 1 | Jan 2009 – Dec 2013 |
| | SpaceNet6 (Shermeyer et al., 2020) | 2020 | WorldView-2 | 0.5 | P, MS, PS | BE | Urban | 1 | Aug 2019 |
| | MAVIC-T (Low et al., 2023a) | 2023 | Aerial | Unknown | RGB | I2I | Unknown | 1 | Aug 2008 |
| | MAGIC (Low et al., 2024) | 2024 | Aerial | Unknown | RGB | I2I | Unknown | 4 | Unknown |
| | SAR2Earth (ours) | 2024 | Google Earth | 0.15 – 0.6 | RGB | I2I | Both | 8 | Nov 2016 – Apr 2024 |

noise due to the coherent nature of radar signal processing, which introduces granular interference patterns. This speckle noise makes SAR images challenging to interpret (Spigai et al., 2011; Zhang et al., 2015), especially for non-experts, and complicates the application of algorithms developed for EO imagery. To bridge this gap, SAR-to-EO translation methods (Fuentes Reyes et al., 2019; Wang et al., 2022; Yang et al., 2022; Lee et al., 2023) have been proposed, aiming to translate SAR images into EO-like images that are more accessible for analysis using existing EO-based algorithms.

Despite these efforts, there has been a lack of comprehensive analysis of these methods, and they often remain isolated applications without standardized benchmarks. Existing SAR and EO multi-modal datasets (Schmitt et al., 2018; Wang & Zhu, 2018; Shermeyer et al., 2020; Low et al., 2023a; 2024) are limited in both quantity and diversity, often being captured within specific regions. This lack of diversity restricts the ability to generalize the performance of SAR-to-EO translation models across varying geographical contexts. Additionally, many of these datasets feature only one-day temporal differences between SAR and EO image pairs, which fails to reflect real-world data collection scenarios where temporal discrepancies can be substantial. Such discrepancies can arise from various factors, including satellite revisit intervals, cloud cover in EO imagery, and nighttime acquisition conditions.

As shown in Table 1, we present a summary of the characteristics of existing SAR and EO multi-modality datasets. These datasets are either not publicly available (Low et al., 2023a; 2024), limited in the number of regions they cover (Wang & Zhu, 2018; Shermeyer et al., 2020; Low et al., 2023a), have very low resolutions (Schmitt et al., 2018) that hinder generalization to objects like buildings, or do not consider real-world environments (Shermeyer et al., 2020; Low et al., 2023a; 2024).

To address these challenges, we introduce SAR2Earth, a comprehensive benchmark dataset for SAR-to-EO translation. SAR2Earth consists of spatially aligned SAR and EO images collected from 8 regions, encompassing both urban and rural environments. The dataset accounts for varying temporal differences between image pairs, reflecting realistic conditions encountered in practical applications. All codes and datasets are being made publicly available to support future research in this domain.

2 RELATED WORK

2.1 APPLICATIONS OF SAR IMAGERY

Numerous applications have been proposed to leverage SAR images across various domains. For instance, Li et al. (2024) collected and labeled 100,000 SAR images to perform object detection directly on SAR data. Similarly, Rambour et al. (2020) utilized spatially aligned SAR and EO images for multi-modal segmentation tasks, such as analyzing disasters like floods. Additionally, Low et al. (2023b) focused on the classification of objects such as cars and buses within SAR imagery. Despite these efforts, SAR datasets face significant limitations. SAR data collection is costly and technically complex due to advanced radar technology, making SAR sensors more expensive and challenging than EO sensors. Processing SAR data is difficult due to speckle noise and other artifacts, requiring specialized expertise. These challenges hinder researchers and limit the creation of public datasets. Additionally, the lack of standardized datasets complicates widespread use, as SAR data varies in format and resolution depending on the provider, unlike standardized EO images.

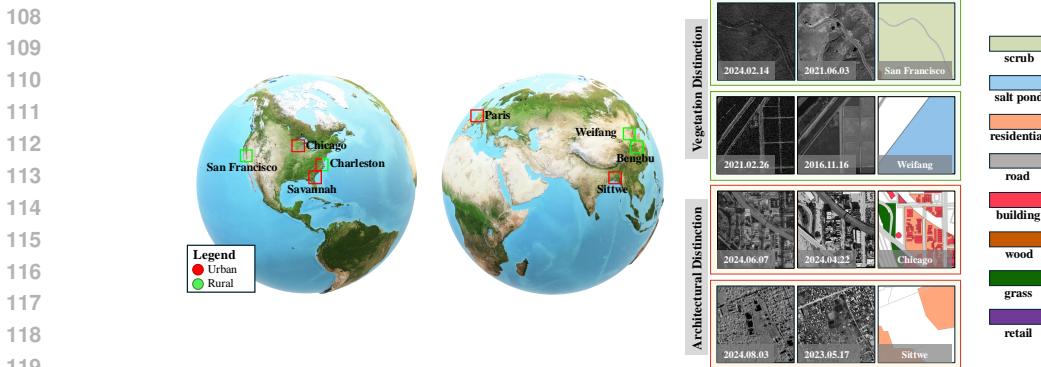


Figure 2: Geographic overview of the SAR2Earth dataset. This dataset covers eight distinct regions—Chicago, San Francisco, Charleston, Savannah, Paris, Bengbu, Weifang, and Sittwe—spanning both urban and rural areas across North America, Europe, and Asia. The consecutive columns on the right represent SAR imagery, EO imagery, and OSM-based label maps.

2.2 SAR-TO-EO TRANSLATIONS

To overcome the limitations of SAR datasets, SAR-to-EO translation techniques have been proposed. For instance, Low et al. (2023a) introduced a method to utilize SAR images by translating them into EO images. To enhance the performance of SAR-to-EO translation, models such as Pix2Pix (Isola et al., 2017), Pix2PixHD (Wang et al., 2018), and CycleGAN (Zhu et al., 2017) have been employed.

In applications such as Amazon deforestation monitoring (Cha et al., 2023), diffusion-based approaches (Rombach et al., 2022; Li et al., 2023) and generative adversarial networks (Isola et al., 2017; Wang et al., 2018) have been widely used for SAR-to-EO translation.

Despite the numerous SAR-to-EO translation methods proposed, there has not been a rigorous comparison among paired methods, unpaired methods, and diffusion-based approaches. Furthermore, because the pre-processing and post-processing pipelines differ across studies, accurate analysis and benchmarking have been lacking.

2.3 REMOTE SENSING APPLICATIONS

Recent advancements in large foundation models and generalization models have brought significant benefits to satellite image analysis. GeoChat (Kuckreja et al., 2024) has demonstrated an EO (Electro-Optical) image-based language model by efficiently fine-tuning large language models. Segment Anything (Kirillov et al., 2023) introduced a segmentation model that can be utilized across any domain by training on billion-scale general vision datasets. These technologies have also been applied in the remote sensing domain, being used in various tasks such as change detection (Oh et al., 2023; Ding et al., 2024) and building segmentation (Osco et al., 2023). However, as revealed in the study (Yan et al., 2023), models based on Segment Anything and large language models like GeoChat do not perform effectively on SAR images due to their training on EO images, which have significantly different characteristics. Consequently, in the context of SAR imagery, the benefits of advancements in large foundation models and generalization models have not been fully harnessed.

3 SAR2EARTH DATASET

In this section, we provide a detailed description of the SAR2Earth dataset. The SAR2Earth dataset has the following key characteristics:

- **Global Data Collection for Generalization:** To evaluate generalization performance, the SAR2Earth dataset includes data collected from 8 regions across North America, Europe, and Asia.

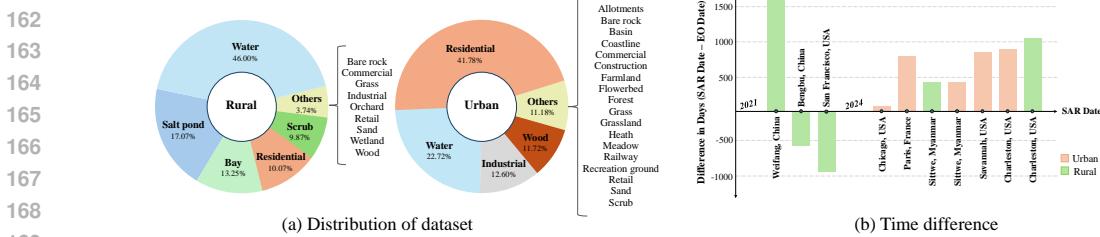


Figure 3: Statistics for the topological distribution and temporal differences in the dataset. (a) Distribution of urban and rural areas by topological elements. (b) Time differences between SAR and EO image captures across regions, indicating the satellite revisit cycles.

- **High Resolution Imagery:** The dataset consists of high resolution images, ranging from 0.15m to 0.6m, offering a diverse mix of spatial resolutions.
- **Consideration of Temporal Shifts:** The dataset accounts for a variety of temporal shifts, ranging from as close as a 1-month difference to as far as a 5-year gap, providing a wide spectrum of temporal scenarios.
- **Structural Diversity:** To address structural shifts, the data is divided into urban and rural categories. The classification is based on the ratio of buildings, amenities, and other structural elements, ensuring a balanced representation of diverse environments.

For sample images and detailed statistics of the dataset, please refer to Figure 2 and Figure 3.

3.1 DATASET DESIGN

Data acquisition SAR imagery is sourced from the Capella Space Open Data Program, with a resolution ranging from 0.3 to 0.6 meters per pixel. Its capability to capture detailed information irrespective of weather, cloud cover, or lighting makes it reliable for continuous monitoring.

EO imagery is obtained from Google Earth, with resolutions between 0.15 and 0.6 meters per pixel and only cloud-free imagery is collected.

SAR Pre-processing SAR images require significant pre-processing to address noise (such as speckle), geometric distortions, and the wide dynamic range of pixel values. One of the critical steps is translating the raw amplitude or intensity values into decibels (dB), which enhances interpretability by compressing the dynamic range and providing a logarithmic representation suitable for further analysis. The conversion to decibels is performed using the following equation:

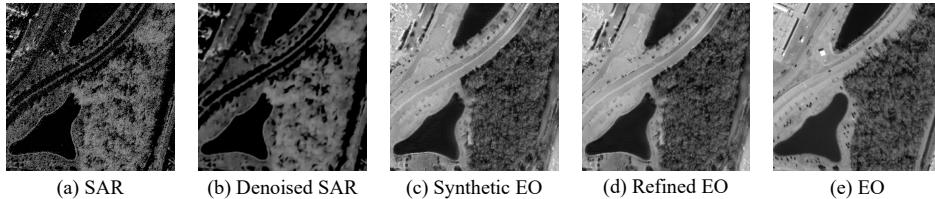
$$\sigma_{\text{dB}}^0 = 10 \log_{10}(S \cdot D^2) \quad (1)$$

where σ_{dB}^0 is the backscatter coefficient in decibels, S is a scaling factor specific to the sensor, and D is the calibrated digital number (DN) values in geocoded format. Note that D is typically the square root of the intensity value, as SAR data is often represented in amplitude.

This conversion provides several benefits: it compresses the dynamic range for enhanced visualization, reduces the influence of extreme pixel values, and improves overall data interpretability, which are crucial for subsequent analysis steps.

Co-registration of SAR and EO A significant challenge in SAR-to-EO translation is achieving precise co-registration between the two image modalities due to inherent differences in both spatial resolution and coordinate systems. Accurate spatial alignment is essential to ensure that corresponding features in both modalities are matched correctly. To address the georeferencing discrepancies, both SAR and EO data are reprojected to a unified coordinate system, specifically the World Geodetic System 1984 (WGS84), which is the most widely adopted geodetic reference framework in remote sensing and geospatial applications. This reprojection guarantees spatial consistency, enabling accurate overlay and comprehensive analysis across both data types.

The co-registration process is performed using QGIS, a robust geographic information system platform. By leveraging the longitude and latitude coordinates inherent to WGS84, we executed image



216
217
218
219
220
221
222
223 Figure 4: The results of SAR-to-EO translation at each step. (a) the original SAR image, (b)
224
225 denoised SAR, (c) the SAR-to-EO translation result, (d) the output from the refinement model, and
(e) the EO image.

226 spatial alignment to achieve pixel-level precision. This procedure facilitated the precise synchro-
227 nization of spatial features across SAR and EO imagery, thus enabling more effective translation
228 and interpretation between the two data sources.
229

230 3.2 DATASET STATISTICS

231 To obtain detailed topological information, we utilized OpenStreetMap (OSM), classifying a total
232 of 25 distinct land cover classes across all regions. The entire dataset covers a combined area of
233 1444.91 km². The regions are classified as urban if residential areas cover at least 25% of the total
234 area. Additionally, if non-residential human-made areas, such as commercial, industrial, or retail
235 spaces, occupy at least 5% of the total area, the region is also categorized as urban (Pesaresi et al.,
236 2013; Esch et al., 2017; Wang et al., 2021).

237 As shown in Figure 3-(a), this classification provides an overview of the topological distribution of
238 urban and rural areas. Specifically, rural areas predominantly consist of natural landscapes, such
239 as vegetation and bodies of water, while urban areas are marked by the presence of human-made
240 structures, including residential, commercial, and industrial buildings.
241

242 To assess the temporal diversity of our dataset, Figure 3-(b) illustrates the temporal differences
243 between SAR and EO imagery acquisition across various regions. These temporal gaps vary sig-
244 nificantly between regions, offering a wide range of temporal shifts. To the best of our knowledge,
245 this makes our dataset the first to incorporate such diverse temporal differences across a broad set
246 of geographic locations. Acquiring temporally aligned SAR-EO pairs without time discrepancies
247 is particularly challenging in real-world settings, making this diversity crucial for practical applica-
248 tions.
249

250 4 SAR2EO PIPELINES

251 In this section, we provide a detailed explanation of our proposed SAR-to-EO pipeline. The SAR-
252 to-EO baseline consists of three main stages: first, a de-noising step to remove the speckle noise in-
253 herent in SAR images, as shown in Figure 4-(b); second, an image-to-image translation module that
254 translates SAR images into EO images, as illustrated in Figure 4-(c); and finally, a post-processing
255 structure that refines the generated images for enhanced quality, as demonstrated in Figure 4-(d).
256

257 4.1 DE-NOISING

258 SAR images inherently contain speckle noise due to the interference of radar signals interacting
259 with multiple scatterers. This noise has a multiplicative nature and is closely linked to the signal
260 itself. Since speckle noise strongly correlates with neighboring pixels, conventional methods that
261 assume noise and signal independence are less effective in removing it. To address this, we adopt a
262 blind-spot method, which predicts the clean value of a pixel based on its surrounding pixels rather
263 than the noisy pixel itself.
264

265 Given the high correlation of speckle noise among neighboring pixels in SAR images, the blind-
266 spot method is particularly effective at distinguishing and removing noise. This de-noising process
267 enhances image quality for SAR-to-EO translation tasks. In our work, we compare two blind-spot-
268 based de-noising methods: Lehtinen et al. (2018) and Zhang et al. (2023).
269

270 4.2 IMAGE TO IMAGE TRANSLATION
271

272 SAR-to-EO image translation poses a complex challenge, requiring the handling of both paired
273 and unpaired settings. Due to changes in ground conditions over time, achieving perfect temporal
274 alignment between SAR and EO images is nearly impossible. For instance, while buildings and
275 fixed structures remain relatively constant, elements like vegetation, moving objects, and lighting
276 conditions vary, complicating precise registration.

277 Considering these factors, SAR-to-EO translation must effectively address both spatial alignment
278 and temporal misalignment. In this paper, we compare paired and unpaired image-to-image trans-
279 lation approaches. Additionally, we propose a partially-paired image-to-image translation method
280 by incorporating objective functions, such as MSE or MAE loss, into the unpaired setting. Given a
281 SAR image I_{sar} and an EO image I_{eo} , the modified loss function is defined as:

$$\mathcal{L}_{\text{total}}(G, D_{eo}, I_{sar}, I_{eo}) = \alpha \mathcal{L}_d(D_{eo}, I_{eo}, G(I_{sar})) + \beta \mathcal{L}_g(G, I_{sar}) + \gamma \mathcal{L}_{mse}(G(I_{sar}), I_{eo}) \quad (2)$$

282 Here, \mathcal{L}_d is the discriminator loss, responsible for distinguishing real EO images I_{eo} from generated
283 EO images $G(I_{sar})$. The discriminator D_{eo} learns this differentiation. \mathcal{L}_g is the generator loss,
284 applied to various unpaired image-to-image translation models such as CycleGAN (Zhu et al., 2017)
285 and CUT (Park et al., 2020). The term \mathcal{L}_{mse} represents the MSE or MAE loss, which aims to
286 minimize the reconstruction error between $G(I_{sar})$ and I_{eo} . By leveraging partially-paired data,
287 this loss encourages the generator to produce EO images that closely resemble the real EO data,
288 thereby reducing the differences between the generated and real images.
289

290 The terms α , β and γ are all hyperparameters, and in all of our experiments, we set α and β to 1,
291 and γ to 0.5.
292

293 5 EXPERIMENTS
294

296 In this section, we validate the SAR2Earth dataset using various image-to-image translation methods
297 and experiment with different preprocessing and postprocessing techniques.
298

299 5.1 IMPLEMENTATION DETAILS
300

301 **Baselines** We selected Pix2Pix (Isola et al., 2017), Pix2PixHD (Wang et al., 2018), and the
302 diffusion-based BBDM (Li et al., 2023) as paired baselines for image-to-image translation. Ad-
303 ditionally, we chose CycleGAN (Zhu et al., 2017), CUT (Park et al., 2020), and StegoGAN (Wu
304 et al., 2024) as unpaired baselines. All hyperparameters strictly followed the default settings of the
305 respective methods ¹²³⁴. We refer to the output of SAR-to-EO models as *SynEO*, and the approach
306 combining paired and unpaired methods is termed the *hybrid* method.
307

308 **Experiments settings** Table 2 presents results without the application of de-noising and post-
309 processing. From Table 4 onward, all experiments include both de-noising and post-processing steps
310 using Hybrid CUT. We used the official codes for Segment Anything Model (SAM) and GeoChat,
311 where the B model was used for SAM, and the 7B model was used for GeoChat.
312

313 **Evaluation metrics** To evaluate the performance of the SAR-to-EO image translation task, we use
314 MAE (Mean Absolute Error), MSE (Mean Squared Error), PSNR (Peak Signal-to-Noise Ratio), and
315 SSIM (Structural Similarity Index Measure) to measure pixel-level accuracy and structural simi-
316 larity. These metrics capture the absolute and squared differences between the generated and real EO
317 images, assess image quality in terms of noise (PSNR), and ensure structural consistency (SSIM),
318 which are crucial for maintaining fidelity in pixel values and structures in SAR-to-EO translation.
319

320 Additionally, we use FID (Fréchet Inception Distance) and LPIPS (Learned Perceptual Image Patch
321 Similarity) to evaluate the perceptual quality and realism of the generated EO images. FID assesses
322

¹<https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>

²<https://github.com/taesungp/contrastive-unpaired-translation>

³<https://github.com/xuekt98/BBDM>

⁴<https://github.com/sian-wusidi/StegoGAN>

324
325
326 Table 2: Results for image-to-image translation baselines on the test set of SAR2Earth. We break
327 down results by training data type: paired training data and unpaired training data. All models are
328 trained on the train set of SAR2Earth.

| Model | Type | MAE ↓ | MSE ↓ | PSNR ↑ | SSIM ↑ | FID ↓ | LPIPS ↓ |
|-------------------------------|--------|--------------|--------------|---------------|--------------|----------------|--------------|
| Pix2Pix (Isola et al., 2017) | pair | 0.172 | 0.051 | 13.818 | 0.085 | 173.751 | 0.569 |
| Pix2PixHD (Wang et al., 2018) | pair | 0.151 | 0.041 | 15.319 | 0.162 | 155.073 | 0.564 |
| BBDM (Li et al., 2023) | pair | 0.161 | 0.047 | 14.772 | 0.163 | 123.051 | 0.477 |
| CycleGAN (Zhu et al., 2017) | unpair | 0.244 | 0.062 | 12.529 | 0.101 | 142.532 | 0.590 |
| CUT (Park et al., 2020) | unpair | 0.236 | 0.086 | 11.172 | 0.094 | 144.312 | 0.592 |
| StegoGAN (Wu et al., 2024) | unpair | 0.214 | 0.073 | 12.041 | 0.152 | 158.930 | 0.595 |
| CycleGAN (Zhu et al., 2017) | hybrid | 0.189 | 0.063 | 13.592 | 0.109 | 142.532 | 0.540 |
| CUT (Park et al., 2020) | hybrid | 0.132 | 0.039 | 16.500 | 0.199 | 140.227 | 0.350 |
| StegoGAN (Wu et al., 2024) | hybrid | 0.197 | 0.059 | 14.213 | 0.161 | 166.325 | 0.593 |

334
335
336 the similarity in feature distributions between the generated and real EO images, while LPIPS fo-
337 cuses on perceptual differences based on deep feature representations, ensuring that the generated
338 images visually resemble real EO data.

340 5.2 COMPARISON OF BASELINE

341
342 Table 2 presents the results of comparing image-to-image translation methods on the SAR2Earth
343 dataset. As observed in the comparison table, methods under the *paired* setting achieved high accu-
344 racy results (MSE, MAE). In contrast, methods under the *unpaired* setting showed lower accuracy
345 (MSE, MAE) but attained higher perceptual scores (FID).

346 The SAR2Earth task aims to accurately *predict* the correct EO image rather than simply *generate*
347 plausible images. Therefore, metrics such as perceptual scores and MSE, MAE are both impor-
348 tant. Accordingly, we combined unpaired baselines that achieved high perceptual scores with paired
349 methods that obtained high MSE and MAE performance. We conducted experiments by applying
350 Eq. 2 on the paired images using existing unpaired methods such as CycleGAN, CUT, and Ste-
351 goGAN.

352 Experimental results showed that the hybrid CUT in Table 2 achieved the highest performance. This
353 is because the SAR2Earth dataset is spatially aligned but temporally unaligned. As a result, objects
354 like buildings are in a paired setting, while moving objects are in an unpaired setting. Therefore, a
355 baseline that considers both settings achieved the best performance.

356 5.3 COMPARISON OF PROCESSING

357 In this section, we compare the performance of SAR-to-EO image-to-image translation according
358 to different de-noising processes.

362 Table 3: Ablation study on de-noising preprocessing methods.

| Model | De-noising | MAE ↓ | MSE ↓ | PSNR ↑ | SSIM ↑ | FID ↓ | LPIPS ↓ |
|-----------------|-------------------------------------|--------------|--------------|---------------|--------------|----------------|--------------|
| CUT (hybrid) | MedianBlur | 0.122 | 0.037 | 16.907 | 0.219 | 140.530 | 0.342 |
| | GaussianBlur | 0.126 | 0.032 | 16.526 | 0.222 | 140.172 | 0.348 |
| | Noise2Noise (Lehtinen et al., 2018) | 0.114 | 0.029 | 16.683 | 0.225 | 144.230 | 0.344 |
| | MM-BSN (Zhang et al., 2023) | 0.107 | 0.022 | 17.431 | 0.236 | 136.684 | 0.332 |

363
364
365
366
367
368
369 **Comparison of de-noising** SAR images contain a large amount of speckle noise. This noise ap-
370 pears as granular interference, obscuring important features and textures in the image. It complicates
371 the feature extraction process in data-driven models by introducing high-frequency artifacts, making
372 it challenging to learn accurate mappings between SAR and EO images. To address this issue, de-
373 noising methods have been applied, but because elements in SAR images that appear as noise can
374 actually be important signals, de-noising methods need to be applied carefully. Table 3 shows the
375 performance variations of SAR-to-EO translation according to different de-noising methods.

376 The results in Table 3 demonstrate that as the de-noising methods become more advanced, per-
377 formance improves. These experimental results indicate that in the SAR-to-EO translation task,
378 employing more advanced de-noising methods positively impacts performance.

378
 379
 380
 Table 4: Results for the regional test set when trained separately on each of the 10 regions or
 combined into urban regions (Charleston-U, Chicago, Paris, Savannah, Sittwe-U) and rural regions
 (Bengbu, Charleston-R, San Francisco, Sittwe-R, Weifang).

| Experiment Setting | Region | MAE ↓ | MSE ↓ | PSNR ↑ | SSIM ↑ | FID ↓ | LPIPS ↓ |
|------------------------------|---------------|--------------|--------------|---------------|--------------|----------------|--------------|
| In-Domain (Single region) | Charleston-U | 0.108 | 0.030 | 17.235 | 0.230 | 130.582 | 0.320 |
| | Chicago | 0.112 | 0.033 | 16.983 | 0.225 | 132.467 | 0.327 |
| | Paris | 0.105 | 0.029 | 17.301 | 0.235 | 128.430 | 0.315 |
| | Savannah | 0.115 | 0.034 | 16.875 | 0.222 | 135.098 | 0.330 |
| | Sittwe-U | 0.109 | 0.031 | 17.102 | 0.229 | 131.744 | 0.322 |
| | Bengbu | 0.098 | 0.025 | 18.512 | 0.240 | 120.320 | 0.300 |
| | Charleston-R | 0.101 | 0.027 | 18.301 | 0.238 | 123.982 | 0.308 |
| | San Francisco | 0.097 | 0.024 | 18.734 | 0.242 | 118.567 | 0.295 |
| | Sittwe-R | 0.099 | 0.026 | 18.589 | 0.239 | 121.765 | 0.305 |
| | Weifang | 0.096 | 0.023 | 18.852 | 0.245 | 117.231 | 0.292 |
| In-Domain | Urban→Urban | 0.106 | 0.028 | 17.478 | 0.240 | 125.345 | 0.310 |
| | Rural→Rural | 0.097 | 0.024 | 18.715 | 0.241 | 115.984 | 0.298 |
| Cross-Domain | Urban→Rural | 0.135 | 0.043 | 16.253 | 0.210 | 145.450 | 0.360 |
| | Rural→Urban | 0.132 | 0.041 | 16.438 | 0.218 | 143.890 | 0.355 |

391
 392
Comparison of refinement We compared the performance of SAR-to-EO translation with respect
 393 to post-processing. For post-processing, we used Zamir et al. (2022), and during training, we aimed
 394 for refinement by adding random deformations (affine transforms, random Gaussian noise) to the
 395 EO images. After that, we applied a refinement model to the images translated from SAR-to-EO.
 396 We observed that the FID score decreased from **136** to **128**, indicating an improvement in perceptual
 397 quality, while the other scores did not change significantly. As observed in the results, we confirmed
 398 that the performance improved slightly. Figure 4 illustrates (a) the original SAR, (b) the denoised
 399 SAR, (c) the synthetic EO, (d) the refined EO, and (e) the ground truth EO. As shown in Figure 4, we
 400 confirmed that the artifacts present in (c) disappeared in (d) through refinement. These experimental
 401 results indicate the cause of the performance improvement due to refinement.

402 403 5.4 MODEL GENERALIZATION EVALUATION

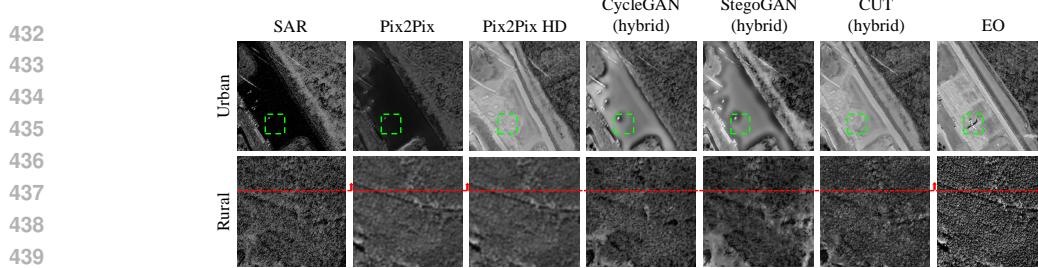
404
 405 The characteristics of SAR images vary significantly by region. This is because SAR relies on radar
 406 backscatter, making it difficult to distinguish between surfaces with similar structures, such as the
 407 ocean and flat areas like runways. Consequently, domain gaps in SAR data are often much larger
 408 than in EO images. To explore this phenomenon, we train and test models within the same region
 409 for in-domain evaluations. Urban areas typically contain many artificial structures with complex
 410 patterns, whereas rural areas have fewer artificial features and simpler, more homogeneous patterns.
 411 Because natural features tend to look similar across different regions, the domain gap in rural regions
 412 is generally smaller.

413 To investigate the domain gaps between urban and rural areas, we conduct a series of experiments,
 414 with the results summarized in Table 4. As shown in the table, rural regions generally outperform
 415 urban regions in in-domain evaluations across all metrics. A notable observation is that when all
 416 urban regions are combined for training and then tested on all urban regions, the performance is
 417 often better than training and testing on a single region. This may be attributed to the increase in
 418 the diversity and amount of training data, which enhances generalization ability from the model.
 419 Conversely, for rural regions, training and testing on individual regions results in better performance
 420 than combining regions. This suggests that for natural features, it is more advantageous for the
 421 model to be tailored to a specific region rather than benefit from a larger, more diverse training set.

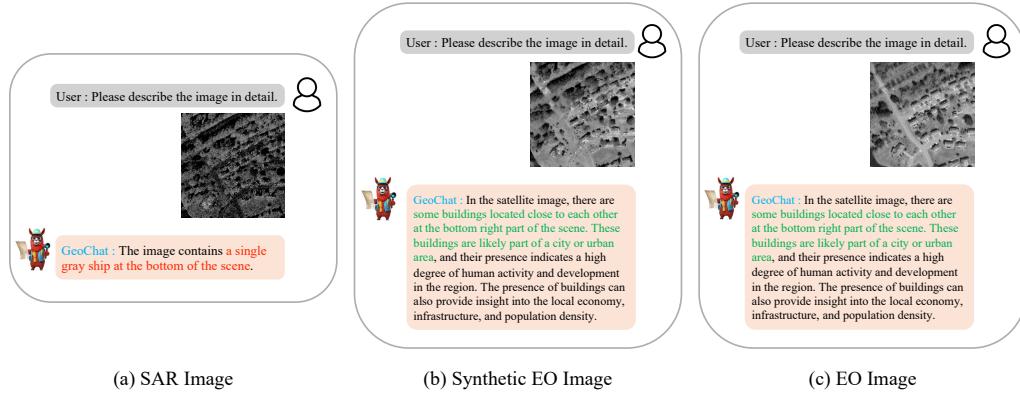
422 In our cross-domain experiments (Urban → Rural and Rural → Urban), we observe a significant
 423 drop in performance. This result highlights the large differences in characteristics between urban
 424 and rural areas. Therefore, for practical applications, it is often more beneficial to collect and train
 425 data suited to the specific regional characteristics rather than simply increasing the quantity of data
 426 from various regions without considering their distinctive features.

427 5.5 QUALITATIVE RESULTS

428
 429 Figure 5 qualitatively compares the results of SAR-to-EO translation across different baselines. As
 430 shown in the figure, CUT (hybrid) produces the most visually plausible results. Specifically, in the
 431 first row, indicated by the green dotted box, the SAR image does not contain an airplane signal
 (due to the time difference, the SAR and EO images do not share the exactly same context), and all



441 Figure 5: Qualitative comparison of various image-to-image translation methods for SAR-to-EO
442 translation in rural and urban cases.



456 Figure 6: Interpretation differences between SAR and EO images. The SAR image (a) leads to an
457 incorrect identification of objects in the scene, such as mislabeling a ship, while the synthetic EO
458 (b) and real EO (c) images accurately capture key urban features, including clusters of buildings.
459

460 baselines successfully avoid generating an airplane in their corresponding SAR-to-EO translation
461 outputs. This experiment demonstrates that, despite the temporally unaligned nature of the SAR-to-
462 EO setting, combining paired and unpaired training approaches effectively mitigates this challenge.
463

464 In the rural example (third row), all baselines produce more plausible images compared to their urban
465 counterparts. However, as highlighted by the red dotted line, fully paired methods like pix2pix and
466 pix2pixHD tend to distort features. This is due to the differing imaging angles between SAR and
467 EO data, where SAR images are often captured from a perspective distinct from that of EO imagery.
468 As a result, the paired models attempt to generate EO-like angles, even for features not present in
469 the original SAR image, creating non-existent structures in the SynEO output. In contrast, baselines
470 that combine paired and unpaired approaches do not exhibit this distortion tendency, maintaining
471 consistency with the original SAR imagery. These results suggest that if the goal is to generate EO-
472 like angles from SAR data, a paired setting is optimal. However, if the aim is to faithfully replicate
473 the appearance of SAR imagery, a combined paired and unpaired training approach is more effective.
474

475 5.6 APPLICATION

476 **GeoChat** Figure 6 illustrates the results of testing SAR images, synthetic EO (SynEO) images
477 obtained through SAR-to-EO translation, and actual EO images using the GeoChat large language
478 model (LLM). As shown in the figure, when a SAR image is input into GeoChat, the responses
479 from the model contain entirely incorrect content. This indicates a failure to interpret the SAR data
480 accurately, primarily because SAR images are excessively noisy and differ significantly from the
481 EO or RGB images on which LLMs are predominantly trained. In contrast, when the SynEO and
482 EO images are provided as input, GeoChat generates the same correct answers, demonstrating its
483 ability to understand and analyze these images effectively.

484 **Segment Anything** We tested the Segment Anything model (Kirillov et al., 2023) on SAR, Sy-
485 nEO, and EO images. As shown in Figure 7, the model struggled to perform effectively on SAR

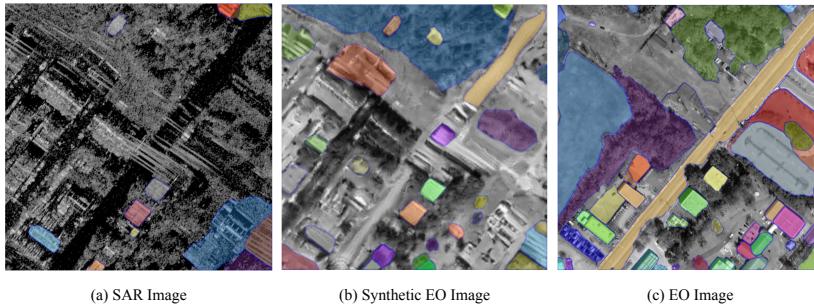


Figure 7: Segmentation results using SAM (Segment Anything Model) for different modalities.

images due to the presence of speckle noise and irregular patterns, which significantly hindered segmentation performance. In contrast, the SynEO images led to successful segmentation by the Segment Anything model. Although the results were still not on par with the EO images, the performance showed substantial qualitative improvement. Improving the SAR-to-EO translation quality could enable generalized models like Segment Anything to be more effectively utilized on SAR data.

6 LIMITATIONS AND BORDER IMPACT

Limitations and future works While SAR images may appear visually similar across regions, their actual representations can differ significantly due to variations in surface roughness. Therefore, regional characteristics play a critical role in SAR-to-EO translation. However, in this work, we do not leverage extra modalities, such as OSM-based label maps or metadata (longitude, latitude, ground sample distance, and date), which can be used to better account for these regional differences. Incorporating metadata, as seen in (Ayush et al., 2021; Khanna et al., 2024), can enhance region-specific accuracy by using geographic and temporal information.

The SAR2Earth dataset does provide metadata information for each image patch level, allowing for future research to utilize this additional data. As part of our future work, we plan to incorporate extra modality-based regional features to enhance the performance of SAR-to-EO translation models by making them more sensitive to regional variations.

Border Impact Our work holds significant implications for remote sensing and related fields. By providing a public benchmark dataset, we aim to accelerate research in SAR-to-EO translation, facilitating advancements in environmental monitoring, disaster response, and urban development. SAR2Earth is particularly valuable in disaster scenarios like floods, where heavy cloud cover renders traditional EO imagery less effective. Since SAR can penetrate clouds and is unaffected by weather conditions or daylight, translating SAR-to-EO images can provide critical information when it is most needed. By translating SAR images into EO-like images, we facilitate the application of advanced AI models developed for EO imagery to SAR data, potentially maximizing the utility of established methodologies.

7 CONCLUSION

In this paper, we present SAR2Earth, a novel public benchmark dataset for Synthetic Aperture Radar to Electro-Optical (SAR-to-EO) translation, aiming to support a wide range of remote sensing applications. We systematically evaluate SAR2Earth by applying various state-of-the-art image-to-image translation models and provide comprehensive benchmark results. Furthermore, we conduct extensive ablation studies—from SAR data pre-processing to model architecture design—to offer valuable insights into the effective utilization of SAR data. Additionally, we validate the versatility of SAR2Earth through experiments with GeoChat and SegmentAnything, demonstrating the potential of SAR-to-EO translation in enhancing data accessibility and utility. Finally, we publicly release our dataset and code to facilitate and encourage future research in this domain. We hope that our research will be widely utilized in tasks such as disaster response and AI for social good.

540 REFERENCES
541

- 542 Zahraa Tarik AlAli and Salah Abdulghani Alabady. A survey of disaster management and sar opera-
543 tions using sensors and supporting techniques. *International Journal of Disaster Risk Reduction*,
544 82:103295, 2022.
- 545 Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and
546 Stefano Ermon. Geography-aware self-supervised learning. *Proceedings of the IEEE/CVF Inter-
547 national Conference on Computer Vision*, pp. 10181–10190, 2021.
- 548 John E Ball, Derek T Anderson, and Chee Seng Chan. Comprehensive survey of deep learning in
549 remote sensing: theories, tools, and challenges for the community. *Journal of applied remote
550 sensing*, 11(4):042609–042609, 2017.
- 551 Marshall Burke, Anne Driscoll, David B Lobell, and Stefano Ermon. Using satellite imagery to
552 understand and promote sustainable development. *Science*, 2021.
- 553 Miriam Cha, Gregory Angelides, Mark Hamilton, Andy Soszynski, Brandon Swenson, Nathaniel
554 Maidel, Phillip Isola, Taylor Perron, and Bill Freeman. Multiearth 2023–multimodal learning for
555 earth and environment workshop and challenge. *arXiv preprint arXiv:2306.04738*, 2023.
- 556 Lei Ding, Kun Zhu, Daifeng Peng, Hao Tang, Kuiwu Yang, and Lorenzo Bruzzone. Adapting
557 segment anything model for change detection in vhr remote sensing images. *IEEE Transactions
558 on Geoscience and Remote Sensing*, 2024.
- 559 Thomas Esch, Wieke Heldens, Andreas Hirner, Manfred Keil, Mattia Marconcini, Achim Roth,
560 Julian Zeidler, Stefan Dech, and Emanuele Strano. Breaking new ground in mapping human
settlements from space—the global urban footprint. *ISPRS Journal of Photogrammetry and Remote
Sensing*, 134:30–42, 2017.
- 561 Mario Fuentes Reyes, Stefan Auer, Nina Merkle, Corentin Henry, and Michael Schmitt. Sar-to-
562 optical image translation based on conditional generative adversarial networks—optimization,
563 opportunities and limits. *Remote Sensing*, 11(17):2067, 2019.
- 564 Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with
565 conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and
566 pattern recognition*, pp. 1125–1134, 2017.
- 567 Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David B
568 Lobell, and Stefano Ermon. Diffusionsat: A generative foundation model for satellite imagery.
569 *The Twelfth International Conference on Learning Representations*, 2024.
- 570 Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete
571 Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *Proceedings
572 of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.
- 573 Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhijit Das, Salman Khan, and
574 Fahad Shahbaz Khan. Geochat: Grounded large vision-language model for remote sensing. *Pro-
575 ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27831–
576 27840, 2024.
- 577 Navneet Kumar, SS Yamaç, and A Velmurugan. Applications of remote sensing and gis in natural
578 resource management. *Journal of the Andaman Science Association*, 20(1):1–6, 2015.
- 579 Jaehyup Lee, Hyebin Cho, Doochun Seo, Hyun-Ho Kim, Jaeheon Jeong, and Munchurl Kim. Cfca-
580 set: Coarse-to-fine context-aware sar-to-eo translation with auxiliary learning of sar-to-nir trans-
581 lation. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- 582 Jaakkko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and
583 Timo Aila. Noise2Noise: Learning image restoration without clean data. *Proceedings of the 35th
584 International Conference on Machine Learning*, 80:2965–2974, 2018.
- 585 Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian
586 bridge diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pat-
587 tern Recognition*, pp. 1952–1961, 2023.

- 594 Yuxuan Li, Xiang Li, Weijie Li, Qibin Hou, Li Liu, Ming-Ming Cheng, and Jian Yang. Sardet-100k:
 595 Towards open-source benchmark and toolkit for large-scale sar object detection. *arXiv preprint*
 596 *arXiv:2403.06534*, 2024.
- 597 Spencer Low, Oliver Nina, Angel D Sappa, Erik Blasch, and Nathan Inkawich. Multi-modal aerial
 598 view image challenge: Translation from synthetic aperture radar to electro-optical domain results-
 599 pbvs 2023. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recog-
 600 nition*, pp. 515–523, 2023a.
- 601 Spencer Low, Oliver Nina, Angel D Sappa, Erik Blasch, and Nathan Inkawich. Multi-modal aerial
 602 view object classification challenge results-pbvs 2023. *Proceedings of the IEEE/CVF Conference*
 603 *on Computer Vision and Pattern Recognition*, pp. 412–421, 2023b.
- 604 Spencer Low, Oliver Nina, Dylan Bowald, Angel D Sappa, Nathan Inkawich, and Peter Bruns.
 605 Multi-modal aerial view image challenge: Sensor domain translation. *Proceedings of the*
 606 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3096–3104, 2024.
- 607 Utkarsh Mall, Cheng Perng Phoo, Meilin Kelsey Liu, Carl Vondrick, Bharath Hariharan, and Kavita
 608 Bala. Remote sensing vision-language foundation models without annotations via ground remote
 609 alignment. *The Twelfth International Conference on Learning Representations*, 2024.
- 610 Youngtack Oh, Minseok Seo, Doyi Ki, and Junghoon Seo. Prototype-oriented unsupervised change
 611 detection for disaster management. *arXiv preprint arXiv:2310.09759*, 2023.
- 612 Lucas Prado Osco, Qiusheng Wu, Eduardo Lopes de Lemos, Wesley Nunes Gonçalves, Ana
 613 Paula Marques Ramos, Jonathan Li, and José Marcato Junior. The segment anything model (sam)
 614 for remote sensing applications: From zero to one shot. *International Journal of Applied Earth*
 615 *Observation and Geoinformation*, 124:103540, 2023.
- 616 Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired
 617 image-to-image translation. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow,*
UK, August 23–28, 2020, Proceedings, Part IX 16, pp. 319–345, 2020.
- 618 Martino Pesaresi, Guo Huadong, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen,
 619 Matina Halkia, Mayeul Kauffmann, Thomas Kemper, Linlin Lu, et al. A global human settlement
 620 layer from optical hr/vhr rs data: Concept and first results. *IEEE Journal of Selected Topics in*
621 Applied Earth Observations and Remote Sensing, 6(5):2102–2131, 2013.
- 622 Clément Rambour, Nicolas Audebert, Elise Koeniguer, Bertrand Le Saux, Michel Crucianu, and
 623 Mihai Datcu. Sen12-flood : a sar and multispectral dataset for flood detection. *IEEE Dataport*,
 624 2020.
- 625 Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-
 626 resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference*
 627 *on computer vision and pattern recognition*, pp. 10684–10695, 2022.
- 628 Michael Schmitt, Lloyd Haydn Hughes, and Xiao Xiang Zhu. The sen1-2 dataset for deep learning
 629 in sar-optical data fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial*
 630 *Information Sciences*, pp. 141–146, 2018.
- 631 Minseok Seo, Youngtack Oh, Doyi Kim, Dongmin Kang, and Yeji Choi. Improved flood insights:
 632 Diffusion-based sar to eo image translation. *arXiv preprint arXiv:2307.07123*, 2023.
- 633 Jacob Shermeyer, Daniel Hogan, Jason Brown, Adam Van Etten, Nicholas Weir, Fabio Pacifici,
 634 Ronny Hansch, Alexei Bastidas, Scott Soenen, Todd Bacastow, et al. Spacenet 6: Multi-sensor
 635 all weather mapping dataset. *Proceedings of the IEEE/CVF conference on computer vision and*
 636 *pattern recognition workshops*, pp. 196–197, 2020.
- 637 Marc Spigai, Céline Tison, and Jean-Claude Souyris. Time-frequency analysis in high-resolution
 638 sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2699–2711, 2011.
- 639 Mélisande Teng, Amna Elmustafa, Benjamin Akera, Yoshua Bengio, Hager Radi, Hugo Larochelle,
 640 and David Rolnick. Satbird: a dataset for bird species distribution modeling using remote sensing
 641 and citizen science data. *Advances in Neural Information Processing Systems*, 36, 2023.

- 648 Haixia Wang, Zhigang Zhang, Zhanyi Hu, and Qiulei Dong. Sar-to-optical image translation with
 649 hierarchical latent features. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12,
 650 2022.
- 651 Junjue Wang, Zhuo Zheng, Ailong Ma, Xiaoyan Lu, and Yanfei Zhong. Loveda: A remote sens-
 652 ing land-cover dataset for domain adaptive semantic segmentation. *Proceedings of the Neural*
 653 *Information Processing Systems Track on Datasets and Benchmarks*, 1, 2021.
- 654 Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-
 655 resolution image synthesis and semantic manipulation with conditional gans. *Proceedings of the*
 656 *IEEE conference on computer vision and pattern recognition*, pp. 8798–8807, 2018.
- 657 Yuanyuan Wang and Xiao Xiang Zhu. The sarptical dataset for joint analysis of sar and optical image
 658 in dense urban area. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing*
 659 *Symposium*, pp. 6840–6843, 2018.
- 660 Sidi Wu, Yizi Chen, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, and
 661 Loic Landrieu. Stegogan: Leveraging steganography for non-bijective image-to-image transla-
 662 tion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.
 663 7922–7931, 2024.
- 664 Hang Xu, Sylvain Barbot, and Teng Wang. Remote sensing through the fog of war: Infrastructure
 665 damage and environmental change during the russian-ukrainian conflict revealed by open-access
 666 data. *Natural Hazards Research*, 4(1):1–7, 2024.
- 667 Zhiyuan Yan, Junxi Li, Xuexue Li, Ruixue Zhou, Wenkai Zhang, Yingchao Feng, Wenhui Diao,
 668 Kun Fu, and Xian Sun. Ringmo-sam: A foundation model for segment anything in multimodal
 669 remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.
- 670 Xi Yang, Jingyi Zhao, Ziyu Wei, Nannan Wang, and Xinbo Gao. Sar-to-optical image translation
 671 based on improved cgan. *Pattern Recognition*, 121:108208, 2022.
- 672 Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-
 673 Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *Proceedings*
 674 *of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.
- 675 Dan Zhang, Fangfang Zhou, Yuwen Jiang, and Zhengming Fu. Mm-bsn: Self-supervised im-
 676 age denoising for real-world with multi-mask based on blind-spot network. *Proceedings of the*
 677 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4189–4198, 2023.
- 678 Yueling Zhang, Chibiao Ding, Xiaolan Qiu, and Fangfang Li. The characteristics of the multi-
 679 path scattering and the application for geometry extraction in high-resolution sar images. *IEEE*
 680 *Transactions on Geoscience and Remote Sensing*, 53(8):4687–4699, 2015.
- 681 Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation
 682 using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference*
 683 *on computer vision*, pp. 2223–2232, 2017.
- 684
- 685
- 686
- 687
- 688
- 689
- 690
- 691
- 692
- 693
- 694
- 695
- 696
- 697
- 698
- 699
- 700
- 701