



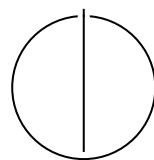
SCHOOL OF COMPUTATION,  
INFORMATION AND TECHNOLOGY —  
INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Data Engineering & Analytics

**Enhancing Photovoltaic Detection Using  
Generative AI for Super-Resolution Satellite  
Imagery**

Ziyad Mourabiti





SCHOOL OF COMPUTATION,  
INFORMATION AND TECHNOLOGY —  
INFORMATICS

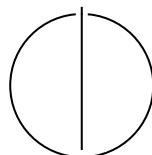
TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Data Engineering & Analytics

**Enhancing Photovoltaic Detection Using Generative  
AI for Super-Resolution Satellite Imagery**

**Verbesserung der Photovoltaik-Erkennung durch  
generative KI für Super-Resolution-Satellitenbilder**

Author: Ziyad Mourabiti  
Examiner: Prof. Dr. Andrea Stocco  
Supervisors: Dr. Jessy Matar, Zhiwei Han, and Dr. Jimmy Tekli  
Submission Date: 04.06.2025



I confirm that this master's thesis is my own work and I have documented all sources and material used.

Munich, 04.06.2025

Ziyad Mourabiti

## Acknowledgments

I extend my heartfelt gratitude to everyone who has supported me throughout my TUM journey and, in particular, during this thesis project. I am especially indebted to my parents, **Hassan Mourabiti and Sabah Mountasir**, whose unwavering love and encouragement have been my foundation; to my sister **Dr. Inasse Mourabiti**; and to my twin nephews **Jad and Iyad**, whose enthusiasm never fails to inspire me.

My deepest appreciation goes to **Dr. Jessy Matar**, whose guidance, patience, and mentorship have shaped every stage of this work. I am likewise grateful to my co-supervisors, **Zhiwei Han** and **Dr. Jimmy Tekli**, whose insightful feedback during our meetings helped refine the technical direction of the project.

I also wish to thank **Dr. Markus Duchon** for championing my collaboration with fortiss and for his constant encouragement; **Stefano Carlo Lambertenghi** for his valuable discussion on evaluation metrics for super-resolution and general image reconstruction evaluation; and **Prof. Dr. Andrea Stocco**, for trusting that I could rise to this challenge and for the constructive dialogue that strengthened this thesis.

Their collective support made this work possible.

# Abstract

The increasing deployment of photovoltaic (PV) installations necessitates reliable and cost-effective approaches for large-scale monitoring. While high-resolution aerial imagery offers precise detection, its limited availability and high acquisition costs make it impractical for broad geographic coverage. In contrast, low-resolution satellite imagery, such as that from Sentinel-2, is freely and consistently available but lacks the spatial detail required for accurate PV segmentation. This thesis investigates whether generative super-resolution techniques can help bridge this resolution gap by enhancing low-resolution imagery to support downstream PV detection.

Three Super Resolution (SR) models (Real-ESRGAN, Satlas, and S2DR3) were evaluated on Sentinel-2 imagery across diverse PV sites in Germany. The super-resolved outputs were assessed using both perceptual quality metrics and segmentation performance with a DeepLabV3+ model built on EfficientNet-B3. SR significantly boosts segmentation accuracy, even with a lightweight model, making it a viable strategy for large-scale PV detection in resource-constrained settings.

Although aerial imagery at sub-meter resolution offers unmatched visual detail, its application remains limited by acquisition costs, file size, and storage demands, which restrict scalability. A single high-resolution tile may occupy as much space as an entire Sentinel-2 scene. In this context, enhanced Sentinel-2 imagery offers a compelling middle ground, balancing improved analytical utility with the efficiency and availability required for operational monitoring.

While high perceptual quality did not always translate to better segmentation, models like Real-ESRGAN achieved consistently strong results across locations. This highlights the need for task-specific evaluation when applying super-resolution methods. The effectiveness of such enhancements using an efficient model backbone also suggests that further gains are possible with larger and more specialized segmentation networks.

Beyond PV segmentation, the methods and insights developed in this work generalize to a wide range of remote sensing applications that benefit from enhanced spatial detail, including land use classification, agricultural monitoring, and infrastructure analysis. The study demonstrates the broader potential of generative super-resolution to increase the usability of publicly available satellite imagery in scalable and cost-sensitive geospatial workflows.

For implementation details, including utility scripts, images, interfaces, and more,

---

*Abstract*

---

this GitHub repository is provided and can be publicly accessed at <https://github.com/mourabitiziyad/pv-sr-detection-thesis>.

# Contents

<b>Acknowledgments</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>1. Introduction</b>	<b>1</b>
1.1. About the Thesis Project . . . . .	1
1.2. Motivation . . . . .	1
1.3. Affiliation with Fortiss . . . . .	2
<b>2. Background</b>	<b>4</b>
2.1. Problem Statement . . . . .	4
2.2. Sentinel-2 . . . . .	5
2.3. Super-Resolution . . . . .	8
2.3.1. Mathematical definition of the SR task . . . . .	8
2.3.2. Models evaluated in this thesis . . . . .	11
2.4. Photovoltaic (PV) Segmentation in Satellite Imagery . . . . .	15
2.4.1. Challenges of Low-Resolution Imagery for Solar-Farm Detection	15
2.4.2. Deep-Learning Approaches for Solar-Farm Segmentation . . . . .	15
2.4.3. Super-Resolution as a Pre-processing Step . . . . .	16
2.4.4. Relevant Metrics . . . . .	17
2.5. Evaluation Metrics . . . . .	17
2.5.1. Super-Resolution Quality Metrics . . . . .	18
2.5.2. PV Segmentation Accuracy Metrics . . . . .	21
<b>3. Methodology &amp; Results</b>	<b>23</b>
3.1. Data Collection and Annotation . . . . .	23
3.2. Super-Resolution Inference and Evaluation . . . . .	27
3.3. Segmentation and Evaluation . . . . .	28
3.4. Interactive User Interface . . . . .	28
3.5. Results . . . . .	36
3.5.1. Super Resolution Results . . . . .	36
3.5.2. PV Segmentation Results . . . . .	44

---

*Contents*

---

<b>4. Discussion &amp; Future Work</b>	<b>50</b>
4.1. Influence of Sun Reflections and Discoloration on Segmentation Performance . . . . .	50
4.2. Perceptual Quality vs. Task-Oriented Utility . . . . .	51
4.3. Effect of Multi-image vs. Single-image Super-Resolution . . . . .	52
4.4. Computational Practicality and Scalability . . . . .	52
4.5. Model Generalization and Geographic Transferability . . . . .	52
4.6. Optimal Resolution Determination for Segmentation . . . . .	53
4.7. Multispectral Utilization and Physical Constraints . . . . .	53
<b>5. Conclusion</b>	<b>56</b>
<b>A. Code Listings</b>	<b>58</b>
A.1. Data Processing Scripts . . . . .	58
A.2. Geospatial Metadata . . . . .	59
A.2.1. Brandenburg Dataset - gdalinfo Output . . . . .	59
<b>B. Super-Resolution Image Comparisons</b>	<b>62</b>
<b>C. Complete Segmentation Results</b>	<b>69</b>
<b>Abbreviations</b>	<b>76</b>
<b>List of Figures</b>	<b>79</b>
<b>List of Tables</b>	<b>81</b>
<b>Listings</b>	<b>82</b>
<b>Bibliography</b>	<b>83</b>

# 1. Introduction

## 1.1. About the Thesis Project

This thesis project brings together a variety of topics, ideas, and research questions. Some are closely related, while others differ significantly in their applications. Despite their differences, all contribute toward a common goal: demonstrating how Super Resolution can enhance remote sensing applications. SR can effectively enhance various remote sensing applications. Specifically, the focus is on improving the segmentation of PV panels, ideally while reducing the reliance on expensive and complex ground truth data. We anticipate that the improvements found in this context will also be applicable, albeit with varying degrees of success, to other remote sensing applications.

To achieve this, we leverage the high availability of Sentinel-2 imagery, which is accessible for a wide range of locations, at various time frequencies, and at a low cost (in most cases, free of charge) [Eur23]. Sentinel-2 imagery is known for its relatively low resolution compared to high-resolution ground-based imagery (as detailed later in Chapter 2). This makes it a prime candidate for the application of SR, given the benefits it can offer for segmentation tasks, particularly in the context of PV segmentation.

## 1.2. Motivation

This work is driven by a shared interest in advancing remote sensing capabilities for PV monitoring, reflecting both personal enthusiasm for sustainable technologies and professional goals in applying AI to real-world challenges. The growing demand for efficient and cost-effective methods for monitoring and evaluating the distribution and performance of PV installations highlights the need for improved remote sensing techniques. High-resolution aerial imagery, while precise, is both expensive and time-consuming to acquire. In contrast, Low Resolution (LR) satellite imagery is abundant but often insufficient for accurate PV detection due to its lower geographic resolution.

This thesis proposes using generative Artificial Intelligence (AI) models to produce synthetic SR images from LR satellite data, aiming to improve PV detection accuracy while reducing the costs associated with collecting high-resolution imagery. Moreover, the application of SR is not limited to PV segmentation; the methodology is designed

to be easily transposable to other remote sensing applications.

From a personal and professional perspective, this work is also motivated by my previous experiences with similar Sentinel-2 remote sensing applications and the challenges encountered with the available data. Additionally, it represents a great learning opportunity and a chance to tackle a real-world problem alongside the same partner I had the pleasure of collaborating with prior to this thesis project, which leads to the next section regarding the affiliation with Fortiss.

### 1.3. Affiliation with Fortiss

This thesis project is in affiliation with fortiss, the research institute for software-intensive systems and services. Particularly, this project is a sub-extension of the NEED project, backed by the Federal Ministry for Economic Affairs and Climate Action (BMWK). As outlined in the NEED project proposal [For23], fortiss brings significant expertise in multiple areas relevant to this research, particularly in the development of AI-based methods for data synthesis and verification. More information about the NEED project can be found on the fortiss website<sup>1</sup>.



Figure 1.1.: Fortiss logo.

Working within fortiss's "Synthesisation and verification, ontology and platform development" sub-project, this thesis directly contributes to Work Package 4 (WP4) of the NEED project. According to the proposal, in WP4 "fortiss will focus on AI-based methods for generating synthetic data from images and with the help of data fusion, as well as developing approaches for verification" [For23, p. 8].

The application of SR to Sentinel-2 imagery for PV panel detection represents a practical implementation of fortiss's research objectives in generating high-quality synthetic data that can replace expensive conventional data collection methods. This aligns with the NEED platform's goal to "fill existing gaps with synthetic data" [For23, p. 9] and reduce the time-consuming and cost-intensive component in the planning of energy technology systems.

By leveraging fortiss's expertise in AI and data synthesis, this thesis aims to contribute practical methodologies to the NEED ecosystem. The work showcases how Super Resolution can be applied to generate synthetic data for PV detection, reducing

---

<sup>1</sup><https://www.fortiss.org/en/research/projects/detail/need>

---

## *1. Introduction*

---

reliance on costly data acquisition and enhancing scalability. This aligns with fortiss's broader mission to support efficient energy planning across various spatial levels. This thesis aims to develop methodologies that can be integrated into the broader NEED ecosystem, enabling more efficient energy planning across multiple scales - from individual buildings to neighborhood and regional levels. This work demonstrates how synthetic data can provide "a robust, low-maintenance and flexible tool for deriving energy measures at different spatial scales" [For23, p. 4].

## 2. Background

PV detection through remote sensing is significant due to its potential for efficient monitoring and assessment of solar installations. High-resolution aerial imagery, such as 40 cm resolution images, provides precise detection results; however, this type of imagery is often limited in availability and expensive, making it impractical for widespread use across various regions. Conversely, satellite imagery, available abundantly and at a lower cost, typically has lower resolution (e.g., 10 m resolution), which poses challenges for accurate PV detection. According to [Guo+23], real-world scenarios introduce complexities such as diverse sensing platforms, sensors, panel categories, and testing regions, which contribute to resolution, size, and foreground-background class imbalances, impeding accurate and generalized PV panel segmentation over large areas.

Recent research on super-resolution techniques applied to satellite imagery, such as Sentinel-2, demonstrates the potential of Generative Adversarial Network (GAN) to enhance resolution. GANs are a type of artificial intelligence that can generate high-resolution images from low-resolution ones. [SRMV20] investigated Sentinel imagery through deep learning techniques applied to allow the spatial enhancement of LR images by recovering high-frequency details to produce High Resolution (HR) images. Thus, super-resolution techniques like GAN, ResNet, and Super Resolution Generative Adversarial Network (SRGAN) could potentially offer promising solutions to bridge the resolution gap.

### 2.1. Problem Statement

The primary challenge addressed in this thesis is the inherent limitation of LR satellite imagery in accurately detecting PV installations. HR aerial imagery, while precise, is not a scalable solution due to its high cost and limited availability across vast regions. Conversely, widely available satellite imagery, such as that provided by Sentinel-2 at a resolution of 10 meters, often lacks sufficient spatial detail for accurate PV detection.

Generative AI techniques, particularly super-resolution methods, offer a viable solution to the limitations of low-resolution satellite imagery for PV detection. For instance, [SRMV20] demonstrated that the Remote Sensing Enhanced Super Resolution Generative Adversarial Network (RS-ESRGAN) model removes upsampling layers to

enhance training with co-registered remote sensing images, showing promising results, particularly in preserving spectral information while enhancing spatial resolution. This thesis explores whether such methods can generate synthetic HR imagery from LR satellite data, achieving comparable accuracy to costly aerial imagery in PV detection tasks. To address this, the applicability of SR techniques in enhancing PV segmentation accuracy will be investigated, alongside an evaluation of their performance under various geographical and environmental conditions.

Although both the application of SR techniques and PV segmentation serve the same overarching objective, they constitute two distinct tasks, each requiring separate methodologies, experimental setups, and design considerations. Treating them separately allows for a clearer evaluation of how super-resolution specifically influences segmentation performance, without conflating the distinct processes involved. This thesis demonstrates that integrating SR techniques leads to measurable improvements in PV segmentation outcomes. As we approach each task separately, it is important to first cover several foundational aspects, including data characteristics, annotation processes, and evaluation methodologies, which will underpin the analyses presented in the subsequent chapters. Such considerations are discussed in the following sections.

## **2.2. Sentinel-2**

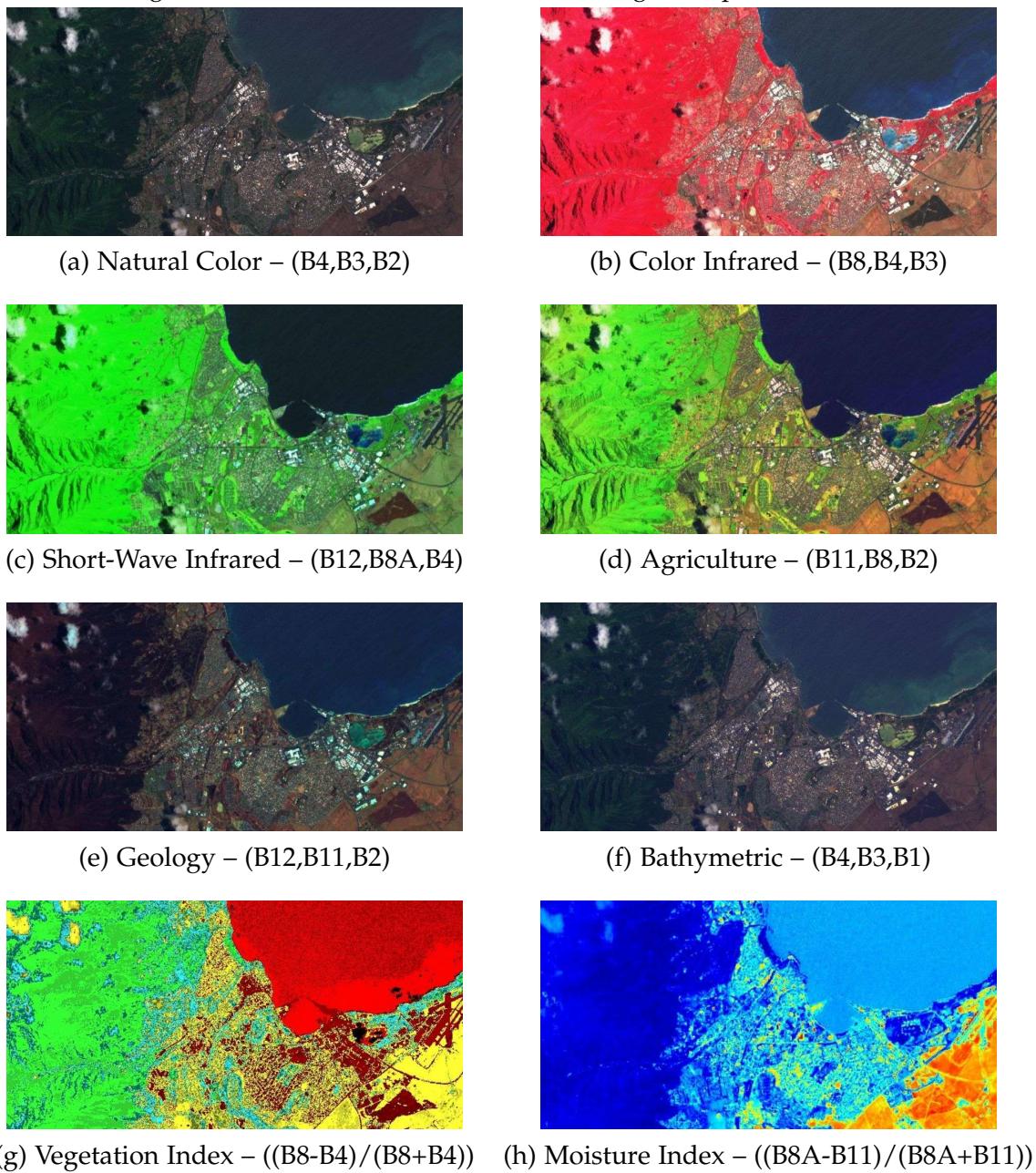
When discussing Sentinel-2 imagery, we're not talking about ordinary photographs. These images capture the Earth in ways far beyond what the human eye can see. While a standard photograph records just three color channels Red, Green, Blue (RGB), Sentinel-2 captures an impressive 13 distinct spectral bands, each revealing different aspects of our environment invisible to the naked eye.

The versatility of Sentinel-2's spectral bands becomes apparent when we combine them in different ways. These combinations, known as band combinations or spectral indices, can highlight specific features of the Earth's surface that might be invisible in individual bands. Figure 2.1 demonstrates several common band combinations and their applications, including the widely used Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) [CG13].

## 2. Background

---

Figure 2.1.: Sentinel-2 band combinations and spectral indices for different applications.  
Each combination highlights specific features of the Earth's surface that might be invisible in individual bands. Images adapted from [GIS19].



## *2. Background*

---

The European Space Agency launched the Sentinel-2 mission as part of the ambitious Copernicus program, a joint initiative with the European Commission (EC) [Phi+20]. The mission consists of two identical satellites (Sentinel-2A and 2B) launched in June 2015 and March 2017, respectively. Working in tandem, these satellites revisit the same spot on Earth every 5 days, providing consistent monitoring of our planet's changing surface [Don+23].

What makes Sentinel-2 particularly remarkable is how it "sees" the Earth across multiple resolutions:

- At 10 meter resolution, it captures the basics: blue, green, red, and Near-Infrared (NIR) light, allowing us to see features roughly the size of a tennis court
- At 20 meter resolution, it records specialized bands including red-edge and Shortwave Infrared (SWIR), which excel at detecting plant health and soil moisture
- At 60 meter resolution, it monitors atmospheric conditions with bands designed to detect aerosols, water vapor, and cirrus clouds

According to [Phi+20], Sentinel-2 represents a substantial improvement over previous satellite systems like SPOT and Landsat.

For applications like detecting solar panels, Sentinel-2's highest resolution (10 meters) presents a significant limitation [Don+23]. At this scale, smaller features like residential PV installations simply disappear into their surroundings. Imagine trying to spot a smartphone on a basketball court from the bleachers.

## 2. Background

---

Table 2.1 offers a closer look at Sentinel-2's "vision" detailing exactly which wavelengths each band captures and how they differ slightly between the twin satellites. These subtle differences might seem trivial, but accounting for them is crucial when processing data from both sources [Eur14; CG13].

Table 2.1.: Characteristics of European Space Agency (ESA) Sentinel-2A and -2B satellite images [Eur14; CG13].

Resolution (m)	Bands	Sentinel-2A		Sentinel-2B	
		Wavelength (nm)	Bandwidth (nm)	Wavelength (nm)	Bandwidth (nm)
10	Band 2—Blue	492.4	66	492.1	66
	Band 3—Green	559.8	36	559	36
	Band 4—Red	664.6	31	664.9	31
	Band 8—NIR	832.8	106	832.9	106
20	Band 5—Red edge	704.1	15	703.8	16
	Band 6—Red edge	740.5	15	739.1	15
	Band 7—Red edge	782.8	20	779.7	20
	Band 8A—Narrow NIR	864.7	21	864	22
	Band 11—SWIR	1613.7	91	1610.4	94
	Band 12—SWIR	2202.4	175	2185.7	185
60	Band 1—Coastal aerosol	442.7	21	442.2	21
	Band 9—Water vapor	945.1	20	943.2	21
	Band 10—SWIR—Cirrus	1373.5	31	1376.9	30

## 2.3. Super-Resolution

SR aims to recover a HR image from one or more LR observations. Early, non-learning pipelines relied on up-sampling, de-blurring and de-noising [AML23], but the field was revolutionized when [Don+15] showed that a deep convolutional network could learn the LR  $\rightarrow$  HR mapping end-to-end. Since then, GAN-based methods such as SRGAN [Led+16] and Enhanced Super Resolution Generative Adversarial Network (ESRGAN) [Wan+18] have further improved perceptual quality, while works in remote sensing (e.g. Deep Sentinel-2 Super-Resolution Network (DSen2) [Lan+18] and Satlas [WBK23]) adapt these ideas to multispectral imagery. In practice, atmospheric effects and mixed-pixel footprints make remote-sensing SR more difficult, and models trained on synthetically degraded data often transfer poorly to real scenes [Arm+21].

### 2.3.1. Mathematical definition of the SR task

Let  $I_{HR} \in \mathbb{R}^{rH \times rW}$  denote the unknown high-resolution image and  $I_{LR} \in \mathbb{R}^{H \times W}$  the observed low-resolution image produced by a degradation operator  $g$ :

---

## 2. Background

---

$$I_{LR} = g(I_{HR})$$

where  $g$  typically applies blur, down-sampling by a factor  $r$ , and noise [Don+15]. Super-resolution seeks a parametric mapping  $f_\theta$  such that

$$I_{SR} = f_\theta(I_{LR})$$

**Reconstruction loss.** A data-fidelity term enforces pixel or spectral accuracy:

$$\mathcal{L}_{\text{rec}} = \|I_{HR} - f_\theta(I_{LR})\|_1$$

with  $p = 1$  preferred for sharper outputs [Su+25].

**Adversarial loss.** If  $f_\theta$  is embedded in a GAN [Goo+14], a discriminator  $D_\phi$  learns to tell real from generated HR images:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}[\log D(I_{HR})] + \mathbb{E}[\log(1 - D(f_\theta(I_{LR})))]$$

ESRGAN replaces the standard GAN objective with a relativistic variant that compares real and fake logits directly, yielding finer textures [Wan+18].

**Perceptual (feature) loss.** To align high-level semantics rather than just pixels, a perceptual loss compares deep features extracted by a fixed network (e.g. VGG-19) [Led+16]:

$$\mathcal{L}_{\text{perc}} = \|\phi(I_{HR}) - \phi(f_\theta(I_{LR}))\|_2^2$$

where  $\phi$  maps an image to its feature representation.

**Full objective.** Most modern SR networks optimize a weighted sum,

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}} \mathcal{L}_{\text{rec}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{perc}} \mathcal{L}_{\text{perc}},$$

striking a balance between distortion and perceptual quality. [BM18] prove that these two goals are fundamentally at odds, formalizing the so-called *perception–distortion trade-off*.

**Universality vs. model-specific choices.** The degradation model  $g$ , the mapping  $f_\theta$  and a reconstruction loss are *universal* to supervised SR. Adversarial and perceptual terms, the choice of the norm  $p$ , and the weights  $\lambda_*$  are *model-dependent*: DSen2 [Lan+18] uses only  $\mathcal{L}_{\text{rec}}$ , whereas Satlas [WBK23] and ESRGAN [Wan+18] adopt all three components to favor sharper, more realistic detail.

Generative adversarial networks (GANs) have recently shown strong potential for satellite SR. Architectures such as the Cascaded SR-WGAN Cascaded Spatial Frequency-Domain Generative Adversarial Network (CSWGAN) and ESRGAN generate sharper reconstructions, especially in high-frequency scenes, which improves PV array detection accuracy [Wan+24].

**CNN baseline: DSen2** A landmark CNN for Earth-observation SR is **DSen2**, introduced by Lanaras *et al.* [Lan+18]. The model is trained on synthetically down-sampled Sentinel-2 tiles and learns two mappings ( $20 \text{ m} \rightarrow 10 \text{ m}$  and  $60 \text{ m} \rightarrow 10 \text{ m}$ ), allowing all thirteen spectral bands to be reconstructed at the sensor’s finest  $10 \text{ m}$  ground-sampling distance. The authors report roughly a 50 % reduction in Root Mean Square Error (RMSE) compared with the state-of-the-art baselines. DSen2 consists of two eight-layer residual CNNs, one up-sampling  $20 \text{ m}$  bands to  $10 \text{ m}$ , the other  $60 \text{ m}$  bands to  $10 \text{ m}$ . Each block uses  $3 \times 3$  convolutions and batch normalization, with a global skip connection from input to output. The network is self-supervised with an  $L_1$  reconstruction loss on synthetically down-sampled Sentinel-2 patches [Lan+18].

**Limitations motivating this thesis** DSen2 is trained with a mean-absolute pixel error ( $L_1$ ) loss only [Lan+18]; this pixel-wise objective delivers accurate but visibly smooth textures. Recent comparative tests on real Sentinel-2 crops describe DSen2 as “blurrier” than other super-resolution methods, especially over urban rooftops [Arm+21]. Furthermore, the network assumes that high-resolution reference bands are perfectly co-registered with the lower-resolution ones [Lan+18], which complicates transfer to sensors that lack such  $10 \text{ m}$  anchors.

**From CNNs to GANs** GAN-based SR methods add an adversarial loss that rewards photo-realistic high-frequency detail. Recent variants, including CSWGAN and ESRGAN, provide higher perceptual quality on remote-sensing benchmarks, yielding sharper panel boundaries and hence better PV segmentation.

**Focus of this work** DSen2 is treated as a conceptual baseline in this thesis. Chapter 3 concentrates on GAN-derived SR models that better address the fine-detail requirements

of PV array detection.

### 2.3.2. Models evaluated in this thesis

#### ESRGAN 4x Super-Resolution

ESRGAN [Wan+18] is adopted as the **generic single-image GAN baseline** for this study.<sup>1</sup> Its generator, known as *RRDBNet*, stacks 23 Residual-in-Residual Dense Blocks to achieve a very deep architecture without Batch-Normalization. Training combines the Visual Geometry Group (VGG)-based perceptual loss, an  $L_1$  reconstruction term, and a relativistic average GAN loss, yielding sharper textures than earlier SRGAN variants while maintaining color fidelity. The VGG network, originally introduced for image classification tasks [SZ15], is commonly used in SR to compute high-level feature similarity between the reconstructed and reference images, enabling perceptual alignment beyond pixel-wise accuracy. Thanks to these qualities, RRDBNet serves as the starting point for a wide family of subsequent super-resolution methods (e.g., Real-ESRGAN, BSRGAN) and remains a widely used baseline in both natural-image and remote-sensing domains. In this thesis, Real-ESRGAN provides the reference against which we compare the multi-image Satlas model and the multispectral 10 $\times$  Sentinel-2 Deep Resolution 3 (S2DR3) approach.

#### Satlas 4x Super-Resolution Model

A special characteristic of Satlas is that it is also a Multi-Image Super-Resolution model, meaning that it can take multiple low-resolution images and super-resolve them to a single high-resolution image, unlike the more commonly used Single-Image Super-Resolution models. Satlas is trained on paired *Sentinel-2* stacks and 1 m airborne photographs from the National Agriculture Imagery Program (NAIP), which provide high-resolution ground-truth supervision for learning the 4 $\times$  mapping [U.S18]. Satlas keeps the RRDBNet generator of ESRGAN but modifies the first convolution so that the network can ingest a stack of *Sentinel-2* images instead of a single frame. The authors state that "our ESRGAN model is an adaptation of the original ESRGAN, with changes that allow the input to be a *time series* of *Sentinel-2* images," and that all released checkpoints upscale by a factor of four [All23, README, "Model"]. Repository weights are provided for 1, 2, 4, 8, or 16 input images, each corresponding to a different sequence length [All23, README, "Model Weights"].

Authors of Satlas and some experimentations yield a few insights that are relevant to this work.

---

<sup>1</sup>In this thesis we use the open-source Real-ESRGAN implementation [Wan+21], a real-image variant of ESRGAN; for brevity we continue to write "ESRGAN" when the distinction is not essential.

## 2. Background

---

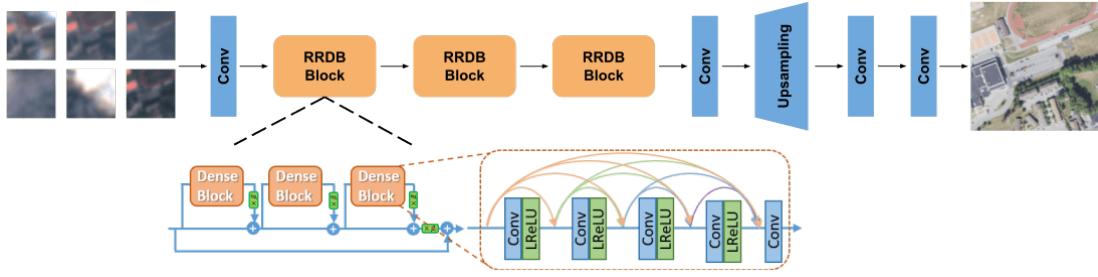


Figure 2.2.: ESRGAN’s RRDBNet generator Adapted for Multi-Image Super-Resolution

### Drawbacks noted by [All23; WBK23]

- **Hallucinated objects.** The model can invent boats on water, cars on highways, or misplaced vegetation–building patterns.
- **Geographic bias.** Training data (the S2–NAIP set) covers only the continental United States from 2019–2020, so performance may drop in landscapes that look "vastly different from the USA," for example Kota, India.
- **Spectral scope.** Public checkpoints are RGB-only; using all twelve Sentinel-2 bands requires retraining with the multispectral configuration.
- **Temporal consistency.** Because the input stack can span several months, super-resolved details may match none of the individual acquisition times, which complicates change detection or time-series analysis.
- **Fixed scale.** Upsampling is fixed at 4×; users who need 6× or 10× (for example to reach 1 m NAIP resolution) must cascade models or choose a different SR method.

Some of these drawbacks are not necessarily limited to Satlas, as it can be the case for a few other models. Moreover, in later chapters, we will be able to delineate more perceptual differences between the models. Additionally, to ensure a more comparable evaluation of the models, a large portion of the methodology in the evaluation between the models will be focused on the RGB bands of the Sentinel-2 imagery.

### S2DR3 10× Super-Resolution

S2DR3 [Akh23] is a **single-image** model that upscales *all twelve* Sentinel-2 bands from their native 10 m/20 m/60 m grids to a uniform 1 m resolution, an effective 10× scale-factor. Like ESRGAN, it employs a fully convolutional RRDB-style backbone,

## 2. Background

---

but its loss is tuned for *spectral fidelity*: training minimizes per-band reflectance errors and a VGG-style perceptual term, rather than pure Peak Signal-to-Noise Ratio (PSNR). Because no true 1 m multispectral ground truth exists, the authors generate synthetic HR labels (pan-sharpened RGB plus simulated narrow-band detail) and report a median  $R^2 = 0.98$  between input and super-resolved spectra on withheld scenes. Qualitative tests suggest negligible hallucination yet occasional geometric distortions in dense urban grids and a tendency to miss objects smaller than 5 m. We use the publicly released inference weights via the authors' Google Colab interface and treat S2DR3 as the high-scale, multispectral endpoint in our comparison matrix.

Below are some examples of the S2DR3 model's performance on Sentinel-2 image coordinates. Thanks to its multispectral capabilities, the model is able to generate a high-resolution image that covers spectral bands beyond the RGB range.

The specific settings ensured minimal cloud coverage and thus the selected date was on 2024-08-28 at coordinates (10.495919, 50.041385), which is in the Haßfurt region of Bavaria, Germany.

For the record, these kinds of specifications may not necessarily be straightforward and there are other considerations to take into account. This is a hint at the amount of work that goes into the selection of the date and location for the acquisition of the imagery, as well as the pre-processing steps that are necessary to ensure the quality and compatibility of the imagery with the models.

Images in Figure 2.3 were generated using the S2DR3 10 $\times$  super-resolution model and visualized with the interactive menu shown above. The menu allows users to select different band combinations and indices for inspection:

- **S2 RGB:** Standard true-color composite using Sentinel-2's red, green, and blue bands.
- **S2 NDVI:** Normalized Difference Vegetation Index, highlighting vegetation health using the NIR and red bands.
- **S2 B11,B8,B5:** A false-color composite using SWIR (B11), NIR (B8), and red-edge (B5) bands

## 2. Background

---



Figure 2.3.: S2DR3 10× Super-Resolution Model outputs: (top) S2 RGB, (middle) S2 NDVI, (bottom) S2 B11,B8,B5

## 2.4. PV Segmentation in Satellite Imagery

### 2.4.1. Challenges of Low-Resolution Imagery for Solar-Farm Detection

High-resolution aerial photographs ( $\leq 0.5$  m Ground-sample distance (GSD)<sup>2</sup>) can delineate individual solar modules, but they are expensive and geographically patchy. Conversely, single pixels in Sentinel-2 imagery represent  $\sim 100$  m<sup>2</sup>, so only **utility-scale farms** emerge as coherent objects, while rooftop arrays vanish in mixed pixels [Kru+21]. Spectral confusion (dark roofs, asphalt, water) further complicates detection, although the sensor's NIR and SWIR bands supply additional discriminative power.

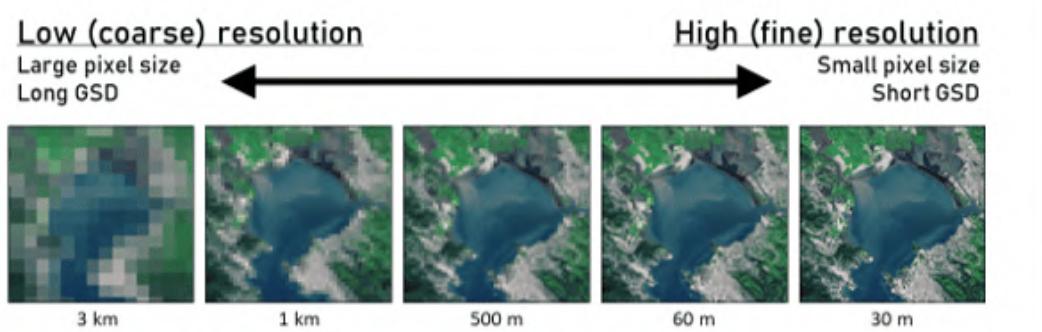


Figure 2.4.: "Examples of low- and high-resolution imagery. Higher resolution imagery allows more detailed spatial patterns to be viewed in the imagery" [U.S24].

### 2.4.2. Deep-Learning Approaches for Solar-Farm Segmentation

**U-Net family.** Encoder-decoder networks with skip connections remain a work-horse for PV mapping; SolarNet segmented 439 Chinese solar farms from high-resolution imagery [Hou+19]; a global U-Net variant achieved 98.6 % precision and 90 % recall on >68 000 installations [Kru+21].

**DeepLab series.** DeepLab networks replace ordinary convolutions with *atrous (dilated) convolutions*, which insert "holes" between kernel weights so that each filter sees a wider area *without* adding extra parameters. Multiple dilation rates are then fused in an *Atrous Spatial Pyramid Pooling (ASPP)* block, and DeepLab V3+ adds a lightweight decoder that sharpens object boundaries with atrous *separable* convolutions. In a recent benchmark, a DeepLab V3+ model ("Solis-seg") reached 96 % mean Intersection-over-Union (IoU) on

<sup>2</sup>GSD is the on-ground distance represented by the spacing of two neighboring pixel centers; a 0.5 m GSD image therefore has pixels that each cover 0.5 m  $\times$  0.5 m on the Earth's surface [U.S24].

European Sentinel-2 scenes, edging out Neural Architecture Search (NAS)-generated variants [Jia+21]. Our thesis adopts DeepLab V3+, pairing it with an EfficientNet backbone [TL19] for a strong accuracy–efficiency trade-off.

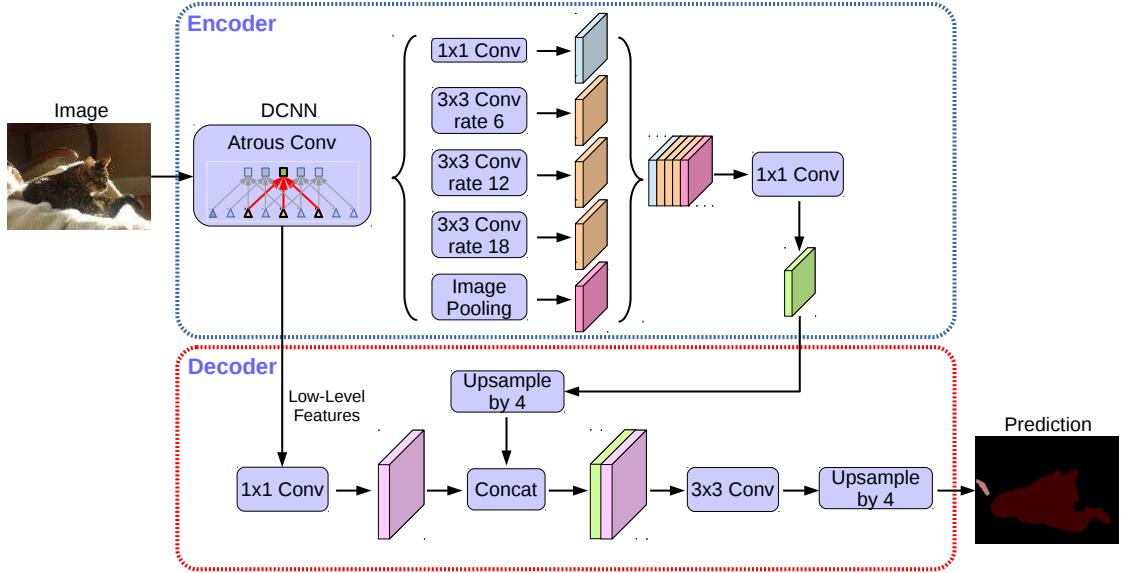


Figure 2.5.: ASPP block used in DeepLabV3+. A  $1 \times 1$  convolution, three  $3 \times 3$  atrous convolutions with dilation rates 6, 12 and 18, and a global-image pooling branch are concatenated to fuse multi-scale context before decoding. *Adapted from [Che+18]*.

**Large-scale databases.** The *DeepSolar* project processed 15 cm imagery to build a U.S. PV inventory, reporting 93 % precision and 88 % recall [Yu+18]. The multi-resolution dataset of [Jia+21] (0.1–0.8 m) quantifies how spatial detail governs segmentation accuracy.

**Foundation-model adaptation.** SolarSAM fine-tunes the Segment-Anything Model (SAM) using lightweight adapters; trained on 40 cm data it attains 76 % IoU, 88 % precision and 83 % recall, and still detects large farms in Sentinel-2 despite coarser input [Mat+24].

#### 2.4.3. Super-Resolution as a Pre-processing Step

Spatial resolution is the principal bottleneck for Sentinel-2 segmentation. CNN-based SR (e.g. DSen2 [Lan+18]) reduces blur but produces smooth textures. GAN-based SR,

such as ESRGAN [Wan+18], adds perceptual and adversarial losses to recover sharper high-frequency details, exactly the patterns that outline solar panel rows. In this thesis we first super-resolve Sentinel-2 images with state-of-the-art SR models (single-image ESRGAN variants and multi-image Satlas), then feed the enhanced imagery to the segmentation network. The SR step enhances the detail of PV panels, by effectively "hallucinating" plausible geometric features such as sharper edges or finer shapes that are not visible in the original low-resolution input. This tends to increase the model's ability to detect panels that were previously too ambiguous or blurry (boosting recall), while maintaining a low false-positive rate (preserving precision). Similar effects have been observed in cross-domain learning studies such as [ZZW20].

#### 2.4.4. Relevant Metrics

Most studies report IoU, precision and recall. SolarSAM's 76 % IoU sits between UNet++ (84 % [Jia+21]) and CrossNet (74 % [ZZW20]) while requiring less task-specific engineering, effectively underlining how foundation-model adaptation can rival bespoke architectures.

**Thesis contribution.** By integrating super-resolution with DeepLabV3 + EfficientNet, our workflow harnesses Sentinel-2's global coverage yet approaches the segmentation fidelity previously confined to aerial imagery, advancing scalable solar-farm monitoring.

### 2.5. Evaluation Metrics

Following up from what we have already covered metrics-wise, and to quantitatively assess the performance of our approach, we adopted standard metrics from each domain. We first outline image fidelity metrics for super-resolution (PSNR, Structural Similarity Index (SSIM), RMSE), then detail region-based segmentation metrics (IoU, Dice, Pixel Accuracy). For completeness, we also mention a potential image similarity metric: Normalized Cross-Correlation (NCC). As well as boundary-focused segmentation metrics, although these are not emphasized in our evaluation. Finally, we look into two no-reference image quality metrics: Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Perceptual Image Quality Evaluator (PIQE), as they prove to be useful in evaluating the quality of Multi-Image Super-Resolution models.

### 2.5.1. Super-Resolution Quality Metrics

#### Peak Signal-to-Noise Ratio (PSNR)

PSNR is a widely-used objective metric that quantifies the reconstruction quality of an image by measuring the error relative to a ground truth reference. It is derived from the Mean Squared Error (MSE) between the high-resolution (HR) reference image  $I$  and the super-resolved image  $\hat{I}$ . If  $I$  and  $\hat{I}$  have dimensions  $H \times W$ , the MSE is defined as

$$\text{MSE}(I, \hat{I}) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (I_{ij} - \hat{I}_{ij})^2,$$

and the PSNR (in decibels) is given by

$$\text{PSNR}(I, \hat{I}) = 10 \log_{10} \left( \frac{L^2}{\text{MSE}(I, \hat{I})} \right),$$

where  $L$  is the maximum possible pixel value (e.g. 255 for 8-bit images). Intuitively, PSNR expresses the ratio between the signal power ( $L^2$ ) and the reconstruction error; higher PSNR (measured in dB) indicates lower distortion. PSNR is simple and interpretable, but it correlates poorly with human perception in some cases (it focuses on pixel-wise error)[Wan+04; HZ10]. Despite its limitations, PSNR remains a core metric in super-resolution research and is reported in virtually all recent SR studies.

#### Structural Similarity Index (SSIM)

To account for perceptual quality, we use the Structural Similarity Index (SSIM) as a complementary metric. SSIM, introduced by Wang *et al.*[Wan+04], evaluates the similarity between two images by comparing local patterns of pixel intensities that have been normalized for luminance and contrast. Rather than treating each pixel independently (as MSE/PSNR do), SSIM computes statistics over local  $8 \times 8$  patches (or windows) of the image: the mean intensity ( $\mu$ ), variance ( $\sigma^2$ ), and cross-covariance ( $\sigma_{xy}$ ) between the reference and test images. The SSIM between two image patches  $x$  and  $y$  is defined as

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

where  $\mu_x, \mu_y$  and  $\sigma_x^2, \sigma_y^2$  are the mean and variance of patch  $x$  and  $y$ , respectively,  $\sigma_{xy}$  is the covariance, and  $C_1, C_2$  are small constants to stabilize the division. SSIM ranges from  $-1$  to  $1$  (with  $1$  indicating identical images). In practice it is often averaged over all patches in the image to give an overall quality score between  $0$  and  $1$ . A higher SSIM

---

## 2. Background

---

means the super-resolved image preserves more structural detail and visual similarity to the ground truth. This metric has been shown to align better with human visual perception than PSNR for many distortions. In our results, we report SSIM alongside PSNR as a key indicator of perceptual image quality.

### Root Mean Square Error (RMSE)

While PSNR is reported in logarithmic decibel units, the Root Mean Square Error provides the same information on a linear scale. RMSE is simply the square root of the MSE:  $\text{RMSE} = \sqrt{\text{MSE}}$ . We include RMSE because it is expressed in the original pixel intensity units, which can be more interpretable. For example, an RMSE of 5 (on an 8-bit scale) means that, on average, the reconstruction error is about 5 intensity levels per pixel. This direct error magnitude can help contextualize the PSNR values. In summary, PSNR and SSIM are our primary SR metrics (high PSNR and SSIM signify good fidelity), with RMSE reported as an intuitive measure of average error.

### Normalized Cross-Correlation (NCC)

We note that some recent works have explored using the Normalized Cross-Correlation (also known as the Pearson correlation coefficient) as an image similarity metric. NCC measures the linear correlation between the predicted image and the ground truth, ranging from  $-1$  (perfectly inversely correlated) to  $1$  (perfectly correlated). The NCC between image arrays  $I$  and  $\hat{I}$  is typically written as:

$$\text{NCC}(I, \hat{I}) = \frac{\sum_i (I_i - \bar{I})(\hat{I}_i - \bar{\hat{I}})}{\sqrt{\sum_i (I_i - \bar{I})^2} \sqrt{\sum_i (\hat{I}_i - \bar{\hat{I}})^2}},$$

which is mathematically equivalent to Pearson's correlation coefficient between the two images.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

A higher NCC indicates that the super-resolved image's pixel values vary in unison with the ground truth. In the context of super-resolution, NCC has been used as a loss function to complement SSIM or MSE losses. However, NCC has important drawbacks as an evaluation metric. In particular, NCC is computed from global image statistics (means and standard deviations), which makes it insensitive to localized errors or structural misalignments. As shown in [Ros+24], optimizing a model solely for NCC can lead to high NCC values but actually *worse* SSIM and PSNR outcomes. This is because a super-resolved image could have the correct overall contrast and intensity

## 2. Background

---

correlation with the ground truth (yielding a high NCC) while still missing fine details or sharp edges that SSIM and PSNR would penalize. Due to these limitations, we do not use NCC as a primary metric in our results. Instead, we focus on PSNR and SSIM, which better capture perceptual and pixel-level fidelity. (For training purposes one might include NCC to constrain global statistics, but for evaluation we prioritize the more interpretable metrics.) This nature of limitations is not specific to NCC alone; it has been shown more broadly that classical image similarity metrics such as PSNR, SSIM, and NCC may fail to reflect task-relevant performance in downstream applications, particularly when fine spatial or semantic details are critical [LS24]. However, much of the literature that we have looked into, continue to use the other metrics, particularly PSNR and SSIM as primary metrics.

### Blind/Reference-less Image Spatial Quality Evaluator (BRISQUE)

For multi-image super-resolution models such as *Satlas*, an exact high-resolution reference is not available, so full-reference metrics (PSNR, SSIM, etc.) cannot be computed confidently. BRISQUE provides a *no-reference* alternative that correlates well with human perception on natural and remote-sensing imagery.

BRISQUE converts the image into a domain of normalized pixel intensities, called MSCN coefficients. These are expected to follow a specific statistical pattern in natural images. The metric fits a statistical distribution (GGD or asymmetric GGD) to these coefficients and extracts features that describe the image's naturalness. A pre-trained support vector regressor then uses these features to predict a quality score between 0 and 100, where **lower values indicate better quality**. For remote sensing imagery, BRISQUE scores below 40 are generally considered acceptable, and scores above 60 suggest significant artifacts [MMB12].

### Perception-based Image Quality Evaluator (PIQE)

Like BRISQUE, PIQE is a no-reference metric that estimates image quality without needing a high-resolution reference. It is more computationally efficient and works well for large images such as Sentinel-2 tiles.

1. The image is divided into non-overlapping  $16 \times 16$  blocks.
2. Blocks that are flat or saturated are ignored.
3. The remaining blocks are scored based on sharpness and distortion patterns, using empirical thresholds.

---

## 2. Background

---

The final PIQE score is the average of all active block scores and also ranges from 0 to 100. **Lower scores indicate better perceptual quality:** 0–20 (excellent), 20–40 (good), 40–60 (fair), and over 60 indicates poor quality [N+15].

**Why use both?** BRISQUE captures unnatural global textures and is sensitive to hallucinated details introduced by some super-resolution models. PIQE, on the other hand, is better at detecting local artifacts like ringing and checkerboard patterns. Together, they provide complementary insights into the perceptual quality of super-resolved images, especially in scenarios where ground truth is not available.

### 2.5.2. PV Segmentation Accuracy Metrics

To evaluate the quality of photovoltaic (PV) segmentation, we report region-based metrics that measure agreement between predicted segmentation masks and manually labeled ground truth. These metrics are well-established in semantic segmentation literature and are particularly informative in class-imbalanced scenarios such as ours, where PV regions typically occupy only a small portion of the image. Our evaluation prioritizes Intersection-over-Union (IoU) and the Dice coefficient, with pixel accuracy included for completeness.

#### Intersection over Union (IoU)

IoU (also known as the Jaccard index) quantifies the overlap between the predicted mask  $P$  and the ground truth mask  $G$ :

$$\text{IoU}(P, G) = \frac{|P \cap G|}{|P \cup G|}.$$

It ranges from 0 to 1, where higher values indicate greater agreement. Unlike accuracy, IoU is robust to class imbalance because true negatives are excluded from the computation. In practice, IoU is the most widely reported metric for semantic segmentation tasks and serves as our principal indicator of PV panel detection quality[TH15].

#### Dice Coefficient

The Dice similarity coefficient (also known as F1 score for binary segmentation) is closely related to IoU and is defined as:

$$\text{Dice}(P, G) = \frac{2|P \cap G|}{|P| + |G|} = \frac{2TP}{2TP + FP + FN}.$$

---

## 2. Background

---

Dice ranges from 0 to 1 and tends to yield slightly higher values than IoU for the same prediction, making it easier to interpret in small-object scenarios. Since it balances precision and recall, it is especially useful for understanding how well PV structures are captured in both presence and extent[Zou+04].

### **Pixel Accuracy**

Pixel accuracy is defined as the proportion of correctly classified pixels:

$$\text{PA} = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives, respectively.

While simple to compute and understand, this metric is sensitive to class imbalance and may overstate performance when PV coverage is sparse. Thus, we treat pixel accuracy as a supplementary indicator.

## 3. Methodology & Results

### 3.1. Data Collection and Annotation

To evaluate the impact of SR on PV segmentation, we collected Sentinel-2 imagery over several PV farms in Germany. Images were selected for the summer period of 2024–2025 to minimize cloud cover and atmospheric distortion. The geographic coordinates of the PV sites were chosen based on publicly available solar plant registries<sup>1</sup> and manual inspection of satellite imagery. We relied on the default image downloading and filtering logic bundled with the SR framework to ensure consistency across models.

Below in Figure 3.1, we show the original Sentinel-2 images that we used for the evaluation. These images represent diverse photovoltaic installations across Germany with the following characteristics:

- **Brandenburg sites:** Three installations showcasing different scales and configurations of solar farms in the region surrounding Berlin, representing both utility-scale and medium-scale PV deployments
- **Mecklenburg-Vorpommern:** Northern German installation demonstrating PV deployment in coastal lowland terrain with distinct agricultural landscape context
- **Saxony-Anhalt:** Central-eastern German site featuring large-scale solar arrays in open terrain, typical of the region's renewable energy infrastructure
- **Lower Saxony:** Northwestern installation showing PV integration within mixed agricultural and industrial landscape patterns

All images were captured during summer 2024-2025 period to ensure optimal lighting conditions and minimal atmospheric interference. Yet, we will see some noteworthy side-effects of this choice, relating to the sun's reflection over the panels.

---

<sup>1</sup><https://list.solar/plants/largest-plants/solar-plants-germany/>

### 3. Methodology & Results

---

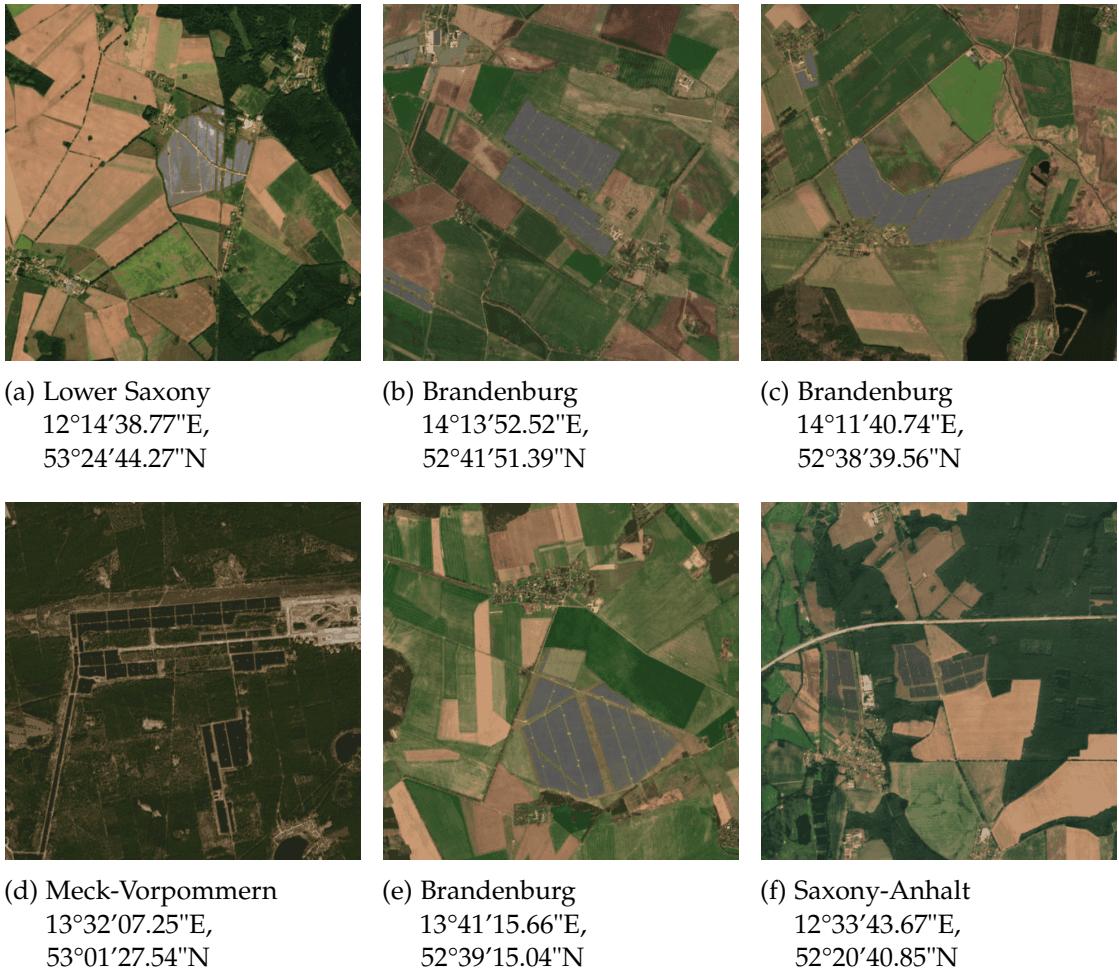


Figure 3.1.: Original Sentinel-2 images used for evaluation across different German states. These images show photovoltaic installations from Brandenburg (3 sites), Mecklenburg-Vorpommern, Saxony-Anhalt, and Lower Saxony, captured during the summer period of 2024–2025 to minimize cloud cover and atmospheric distortion. Coordinates represent the center point of each image tile, demonstrating geographic diversity across northern and eastern Germany. Detailed geospatial metadata for the Brandenburg dataset is provided in Appendix A.2.

### 3. Methodology & Results

Much of the processing pipeline was developed using a Colab notebook I authored<sup>2</sup>. It included image downloading based on the S2DR3 logic (from the Copernicus Hub), patch preparation, model inference (Real-ESRGAN, Satlas, S2DR3), and reconstruction metric computation.

To create ground truth data for segmentation, I manually annotated PV installations using the LabelMe 3.1 graphical interface. The annotation process involved tracing the boundaries of solar arrays using the true-color composites (RGB) as visual references. This ensured accurate labeling of utility-scale solar farms for all the images that we are presenting in this work. The annotations were exported in geo-referenced JSON format.

Figure 3.2 shows the labeling environment used for annotations.

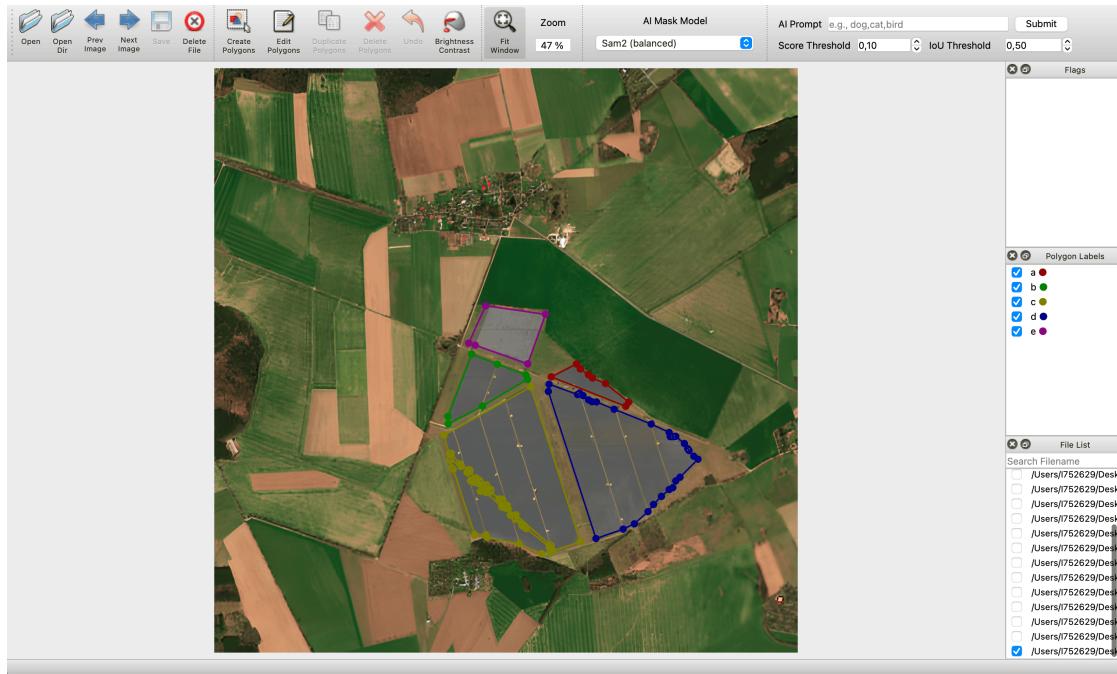


Figure 3.2.: LabelMe GUI used for manual polygon annotation of PV arrays.

<sup>2</sup><https://colab.research.google.com/drive/1oMxrHK2Vsk8p-Ai1yBqhhci1drYTQmmo?usp=sharing>

### 3. Methodology & Results

---

After finishing the manual annotation process, I used this script found in A.1, which I wrote for this thesis, to convert each geo-JSON annotation into a binary mask. The key steps are:

1. Parse JSON into Shapely polygon geometries.
2. Rasterize into a binary array (0/1) matching the original image resolution.
3. Export the result as GeoTIFF to preserve CRS and geotransform. Appendix A.2 provides detailed geospatial metadata for the Brandenburg dataset as an example. For most use cases, however, we simply export to PNG.

Figures 3.3 and 3.4 show the original SR image and its corresponding label mask. Likewise, Figures 3.5 and 3.6 illustrate another annotated pair.



Figure 3.3.: SR Real-ESRGAN Image.

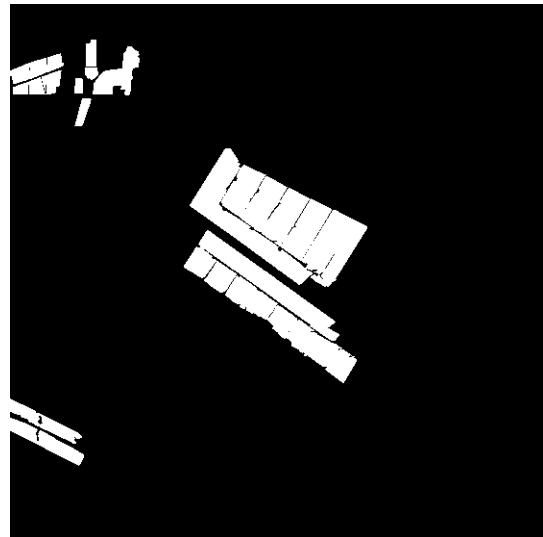


Figure 3.4.: Binary Label Mask.



Figure 3.5.: Super-Resolved Satlas Image.

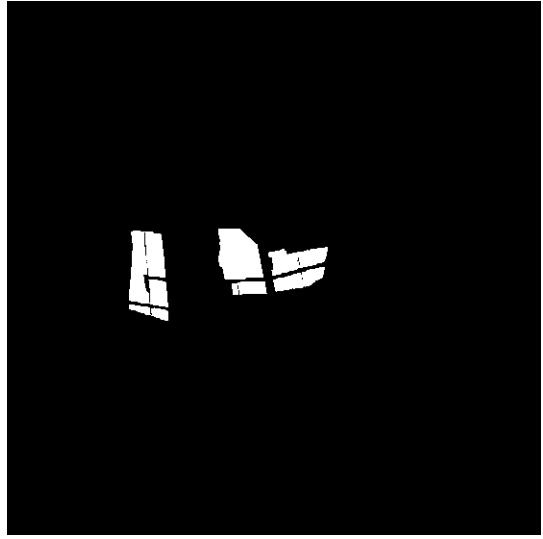


Figure 3.6.: Binary Label Mask.

### 3.2. Super-Resolution Inference and Evaluation

We evaluated three super-resolution models:

- **Real-ESRGAN (4x)**: Single-image GAN-based model using RGB input. The 4x factor corresponds to upscaling from 10 m/pixel to 2.5 m/pixel resolution.
- **Satlas (4x)**: Multi-image temporal model using stacks of Sentinel-2 images.
- **S2DR3 (10x)**: Multispectral upsampler using all 13 Sentinel-2 bands, generating output at 1 m/pixel, from an input of 10 m/pixel.

Each model was run using default parameters and public weights to facilitate reproducibility. For Satlas, we used the official precomputed demo outputs available on their platform<sup>3</sup>. This is due to lack of clarity regarding their best performing model, as they provide multiple models that handle varying levels of Multi-image Super Resolution<sup>4</sup>. This ambiguity caused issues when attempting to replicate their results as their visualisations do not include any information on the model that was used. For S2DR3, inference was done through the authors' Colab interface. All reconstruction quality comparisons used RGB bands only for fairness. This is due to the fact that some models only handles RGB input, such as Real-ESRGAN.

---

<sup>3</sup><https://satlas.allen.ai/superres>

<sup>4</sup><https://github.com/allenai/satlas-super-resolution>

We computed Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Root Mean Square Error (RMSE) per band and reported the mean across bands for each model except for Satlas, where we computed the BRISQUE and PIQE metrics. To triangulate our findings, we also computed the no-reference metrics used in Satlas for the other models as well.

### 3.3. Segmentation and Evaluation

We used DeepLabV3+ with an EfficientNet-B3 encoder for binary PV segmentation. The model was applied without any fine-tuning. Segmentation was performed on both raw Sentinel-2 imagery and super-resolved outputs. Details on the model can be found in Chapter 2.

We used the following evaluation metrics:

- **IoU:** Robust to class imbalance, our main metric.
- **Dice Coefficient:** Balances precision and recall, emphasizes spatial overlap.
- **Pixel Accuracy (PA):** Supplementary indicator; known to overstate accuracy due to class imbalance.

Metrics are also thoroughly defined in Section 2.

### 3.4. Interactive User Interface

We built a Streamlit interface for segmentation visualization and experimentation. The original implementation and basic logic for applying the segmentation model is based on this public repository<sup>5</sup>. We built upon this base to add a lot of functionality, including various user customisations, model parameter tinkering, new visualisations, and many more calculations.

Figure 3.7 shows the interface in action.

---

<sup>5</sup><https://github.com/saizk/Deep-Learning-for-Solar-Panel-Recognition>

### 3. Methodology & Results

---

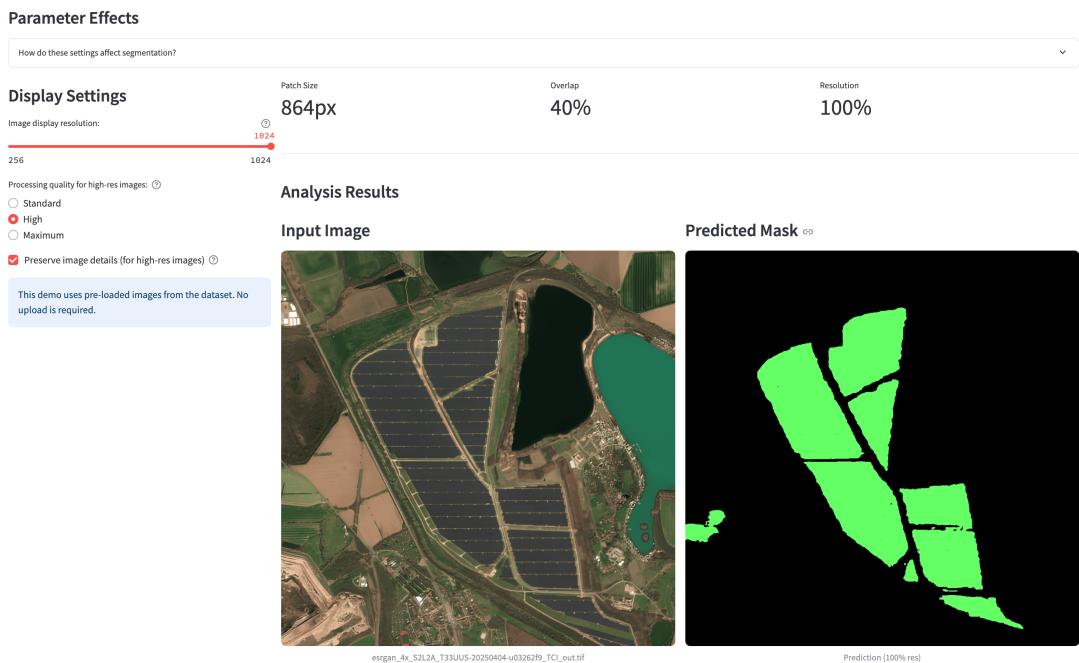


Figure 3.7.: User Interface: SR and segmentation flow.

Overall, the interface enables:

1. Selecting any type of Sentinel-2 imagery, both raw and super-resolved, with support for both GeoTIFF and PNG formats.
2. Applying PV segmentation.
3. Customizing patch size and overlap to handle images of different sizes.
4. Viewing original, super-resolved, segmentation, and labeled overlays side-by-side.

If label masks are uploaded, IoU, Dice, and PA are computed and displayed immediately. The UI also supports resolution scaling and large-tile handling.

Users select an image file and a segmentation model; results, including segmentation masks and evaluation metrics, are rendered interactively.

As shown in 3.8, if the user has provided label images, the system will also display the results of the segmentation and the label images. It will also calculate the relevant segmentation metrics, such as IoU, Dice coefficient, and Pixel Accuracy.

### 3. Methodology & Results

---

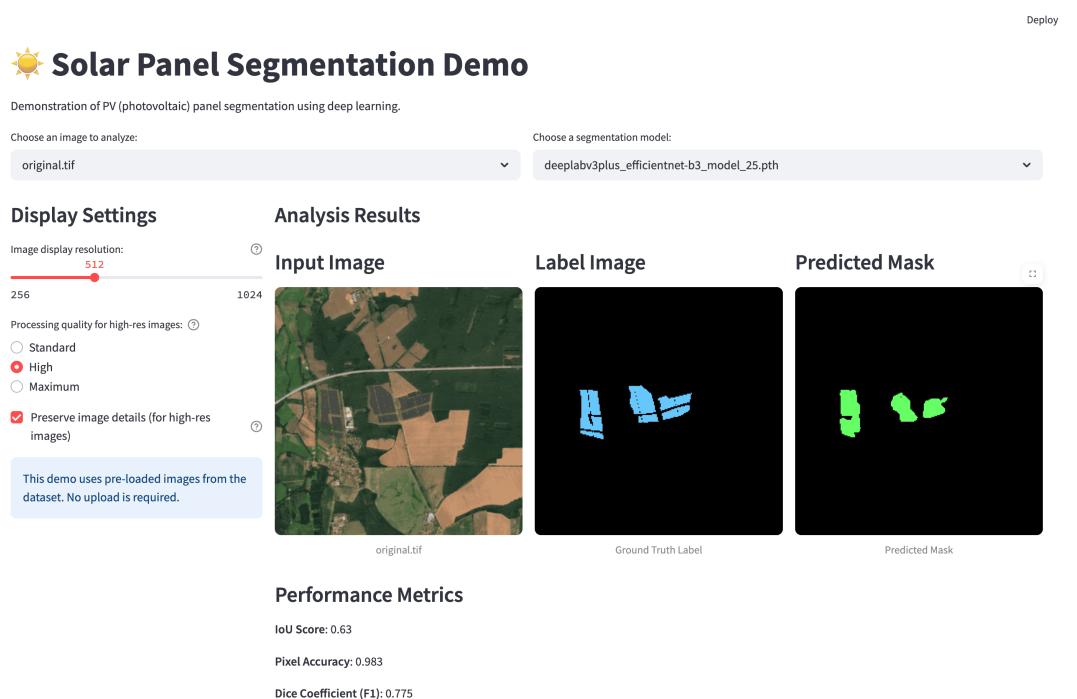


Figure 3.8.: User Interface Demo with Large Image

### 3. Methodology & Results

---

For Super Resolution images, Figure 3.9 shows how the user interface handles these large input images. As it provides a new set of customisations and options to the user. This is crucial as the model will not be able to handle the large input image in one go, as it is not designed to do so.

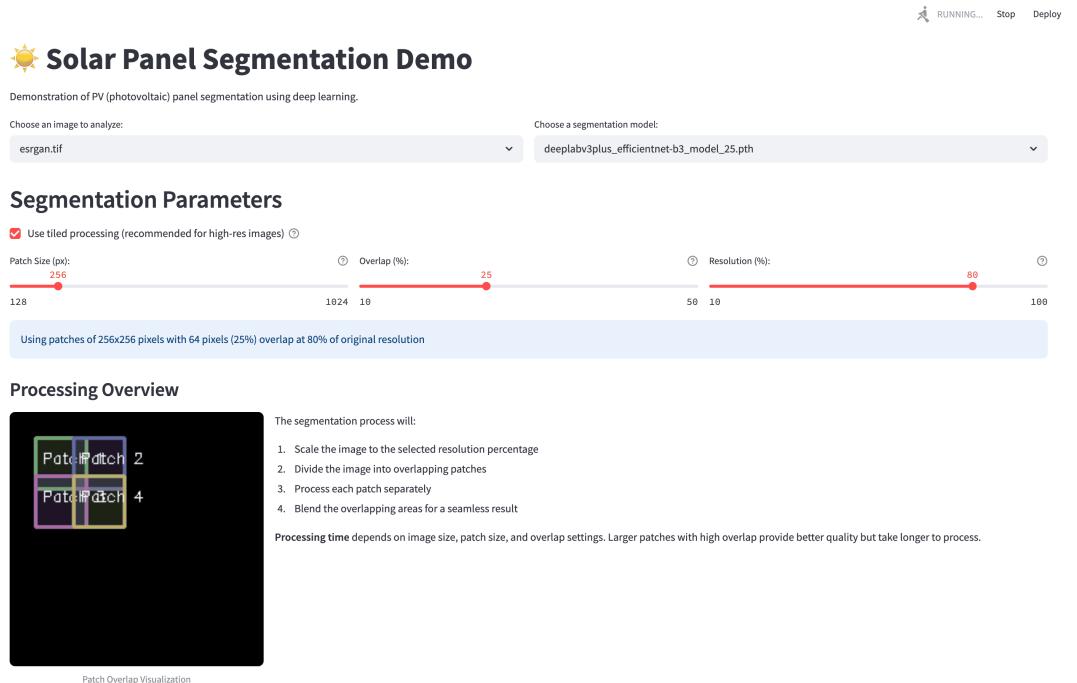


Figure 3.9.: User Interface Parameter Customizations for Large Images

### 3. Methodology & Results

The interface now provides more options on how to run the segmentation. As various models can handle different image sizes and resolutions, we provide sliders that allow for manual customizations of Patch Size and Overlap percentages. We also provide control over how much of the resolution is used for the segmentation. As we will later see that the model can perform better when we lower the resolution of the input image.

The dark mask on the left side in Figure 3.9 shows how big the patches are in relationship to the image, along with how the overlap affects the patches.

Next, in Figure 3.10 we can see the results of the segmentation. Including relevant Segmentation evaluation metrics due to the label image being available for this particular input image.

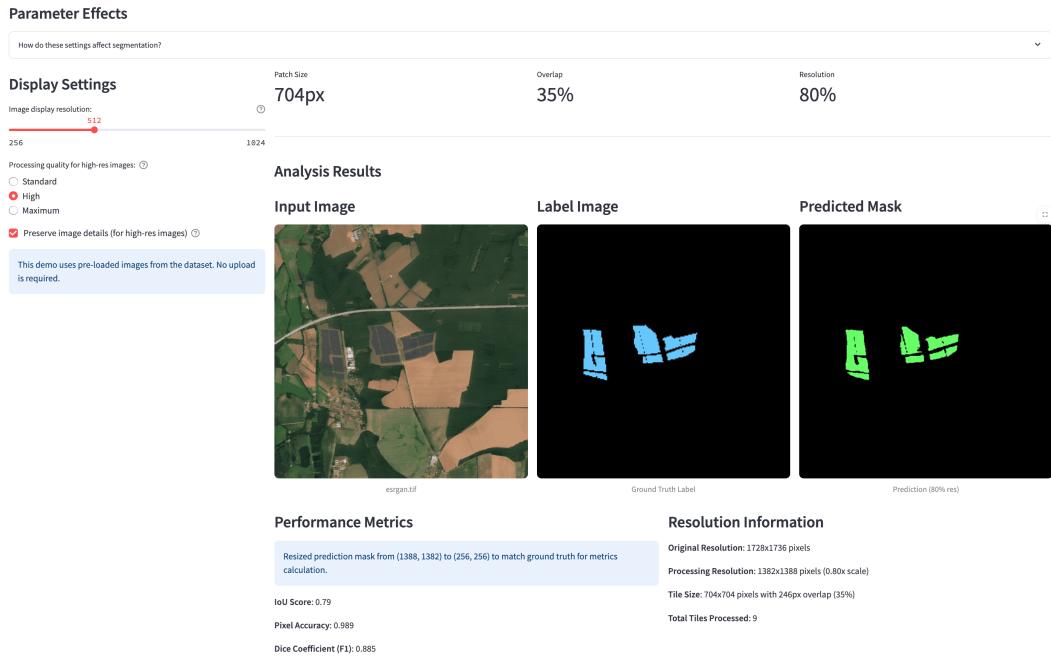


Figure 3.10.: User Interface Demo with Image After Parameter Customization

### 3. Methodology & Results

---

The user interface additionally provides boundary metrics, such as Precision, Recall, and F1-score for the segmentation evaluation, as we will see in the following Figure 3.11.

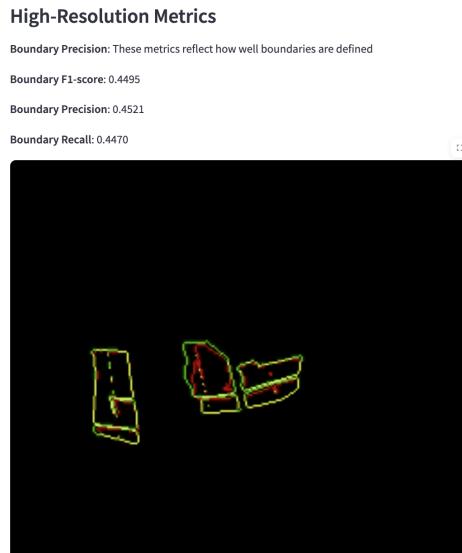


Figure 3.11.: Boundary Metrics Information for the Segmentation

In Figure 3.11, the green boundaries represent the ground truth, and the red boundaries represent the segmentation result.

Additionally, we extend the interface to allow for comparing segmentation results across images of the exact location at different resolutions, as shown in Figure 3.12. This tool helps identify how resolution affects segmentation quality and accuracy across different sets of images simultaneously.

### 3. Methodology & Results

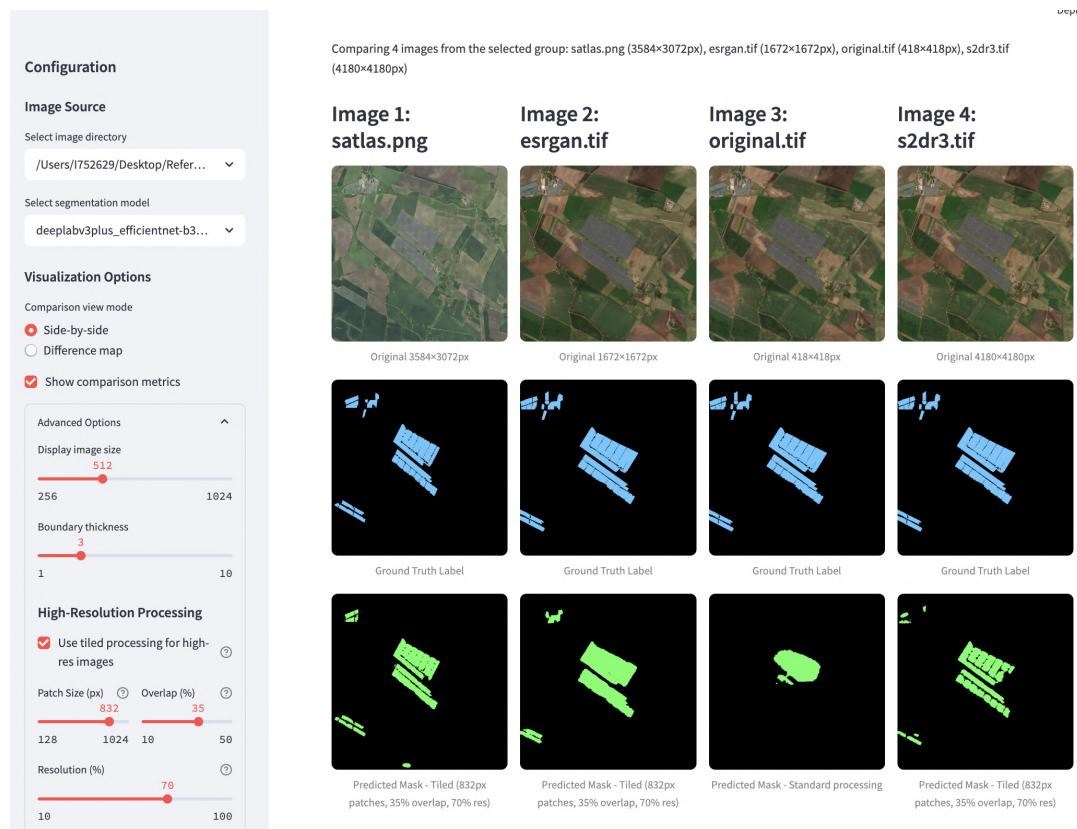


Figure 3.12.: Multi-Resolution Segmentation Comparison

Blue: Ground Truth — Green: Segmentation Result

### 3. Methodology & Results

Finally, Figure 3.13 shows the difference map between the ground truth and the segmentation result. This is a valuable tool for identifying how the segmentation model can see more of the image and visualise this difference in one overall mask. In Figure 3.13, the "Prediction Difference", i.e. the rightmost image above the legend, we can see a difference map that highlights:

- Blue: Segmentation result using the SR image
- Red: Segmentation result of the original Sentinel-2 image
- Green: Where both segmentations overlap.

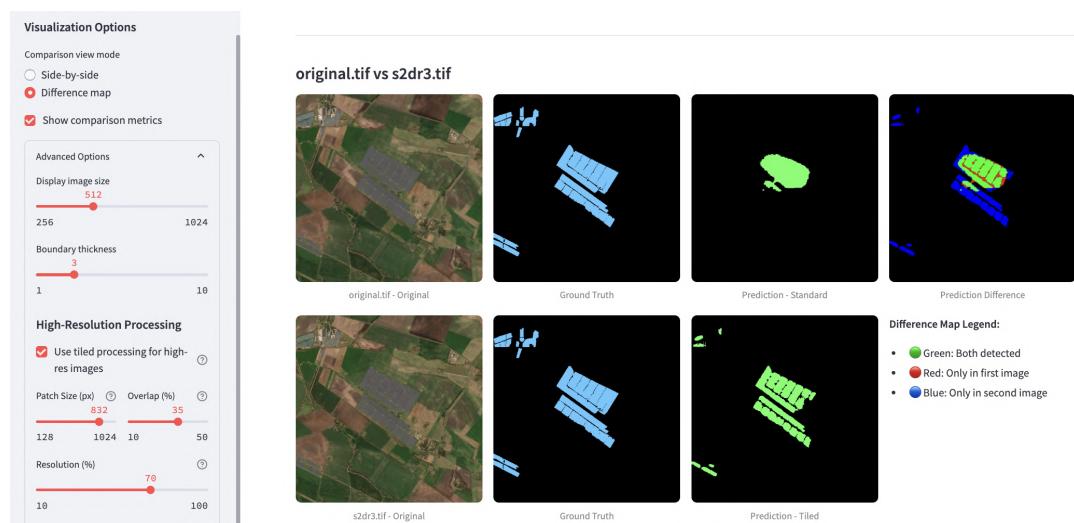


Figure 3.13.: Multi-Resolution Segmentation Difference Map

## 3.5. Results

In this section, we lay out the results in two different areas. First, we cover the results of our experiments with the aforementioned SR models. We evaluate six sets of images across three models and provide the relevant metrics discussed in the methodology. Second, we look at these SR images and how they affect the segmentation accuracy compared to the original images.

### 3.5.1. Super Resolution Results

#### Super-Resolution Visual Comparisons

Given difficulties in visualising the results of the super-resolution models and the fact that performing some compression to make the report accessible has caused some alterations to the colouring of some images, particularly the PVs themselves, we chose to provide the image set as it is. The reader may be able to zoom in to see the significant quality differences across the photos, and we provide public access to all of these image sets here<sup>6</sup>.

Comprehensive visual comparisons of all super-resolution results are provided in Appendix B, which shows the original Sentinel-2 images alongside the outputs from S2DR3 (10× upscaling), Real-ESRGAN (4× upscaling), and Satlas (4× upscaling) for all six image sets.

For the context of this section, we will provide a few side-by-side comparisons of the original Sentinel-2 images and the super-resolved images. Focusing on select zoomed-in Region of Interest (ROI)s of the images.

Figure 3.14 shows the original Sentinel-2 image and the Real-ESRGAN super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N in a zoomed-in area.

---

<sup>6</sup>[https://drive.google.com/file/d/1Y3cMJqyj2Gzg5IOTv\\_Thd-RTwWBHVs2w/view?usp=sharing](https://drive.google.com/file/d/1Y3cMJqyj2Gzg5IOTv_Thd-RTwWBHVs2w/view?usp=sharing)

### 3. Methodology & Results

---



Figure 3.14.: ROI 1 — Original Sentinel-2 image and ESRGAN super-resolved image of Brandenburg,  $13^{\circ}41'15.66''\text{E}$ ,  $52^{\circ}39'15.04''\text{N}$

We can see that the ESRGAN super-resolved image can capture more detail overall, making the photo a lot clearer. However, the ESRGAN image can suffer from artefacts and an overall lack of sharpness and too much smoothing in a way that makes it resemble an oil painting.

Focusing more on the PV panels, we can see that the ESRGAN super-resolved image can capture more detail as well, as seen in Figure 3.15. However, if we focus a bit more, the realism is lacking. The image's lower left region is very smooth and not sharp-edged.

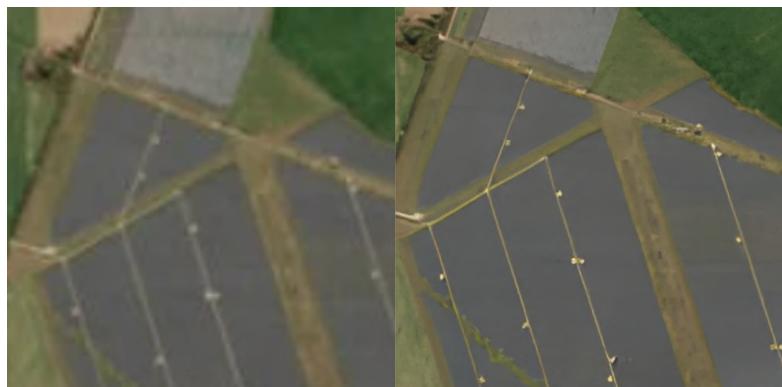


Figure 3.15.: ROI 2 — Original Sentinel-2 image and ESRGAN super-resolved image of Brandenburg,  $13^{\circ}41'15.66''\text{E}$ ,  $52^{\circ}39'15.04''\text{N}$

### *3. Methodology & Results*

---

Moving to Satlas, we can see that the super-resolved image is a lot sharper and more realistic, as seen in Figure 3.16.



Figure 3.16.: ROI 1 — Original Sentinel-2 image and Satlas super-resolved image of Brandenburg,  $13^{\circ}41'15.66"E$ ,  $52^{\circ}39'15.04"N$

As we see above, we can zoom in more on the same ROI as seen in Figure 3.14. The image is much more detailed; buildings are pretty sharp and distinct. However, some particular cut-off points show that the patches are not well aligned, particularly in the lower-middle section of the road. Figure 3.17 shows the PV panels, again on the same ROI as seen in Figure 3.15. We see here the effect of the patching strategy creating a grid-like pattern.



Figure 3.17.: ROI 2 — Original Sentinel-2 image and Satlas super-resolved image of Brandenburg,  $13^{\circ}41'15.66"E$ ,  $52^{\circ}39'15.04"N$

### *3. Methodology & Results*

---

S2DR3 does a really good job at the image reconstruction. It is particularly impressive given that it is a 10x upscaling model. Figures 3.18 and 3.19 show the ROIs of the S2DR3 super-resolved image.



Figure 3.18.: ROI 1 — Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg,  $13^{\circ}41'15.66''\text{E}$ ,  $52^{\circ}39'15.04''\text{N}$



Figure 3.19.: ROI 2 — Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg,  $13^{\circ}41'15.66''\text{E}$ ,  $52^{\circ}39'15.04''\text{N}$

### 3. Methodology & Results

---

It is even more impressive when we look at another ROI of the same model, but in a different location, as seen in Figure 3.20. The image is very sharp and detailed, and the PV panels are very clear.



Figure 3.20.: Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg, 14°13'52.52"E, 52°41'51.39"N

The side-by-side comparison here shows great improvement in the image quality, with no obvious artefacts, oversmoothing, blurred patches, etc.

#### Super-Resolution Quantitative Results

We now present the relevant quantitative metrics introduced in Section 3.2. Due to the lack of precise information regarding the input details used in the Satlas pipeline, we report no-reference image quality metrics instead. These metrics are widely regarded as effective proxies for evaluating perceptual image quality in the absence of a ground truth reference, as properly demonstrated in Section 2.5.1.

Attempting to infer the exact inputs used by Satlas, or manually matching acquisition dates across datasets, would introduce significant uncertainty and potential bias. Therefore, we avoid speculative comparisons and focus on metrics that directly assess the visual fidelity of the super-resolved outputs.

Table 3.1 presents the detailed quantitative results for each image set, showing how the super-resolution models perform across different geographic locations and PV installations. S2DR3 consistently outperforms Real-ESRGAN across all metrics and image sets, achieving higher PSNR values (indicating better signal quality), superior SSIM scores (representing better structural similarity), and lower RMSE values (denoting reduced reconstruction error).

### 3. Methodology & Results

---

Table 3.1.: Individual Set Super-Resolution Results

<b>Set</b>	<b>Location</b>	<b>Model</b>	<b>PSNR</b>	<b>SSIM</b>	<b>RMSE</b>
1	Lower Saxony	Real-ESRGAN	33.16	0.9317	5.37
1		S2DR3	<b>35.77</b>	<b>0.9500</b>	<b>3.97</b>
2	Brandenburg 1	Real-ESRGAN	35.33	0.9384	4.16
2		S2DR3	<b>36.38</b>	<b>0.9466</b>	<b>3.67</b>
3	Brandenburg 2	Real-ESRGAN	33.95	0.9328	4.57
3		S2DR3	<b>35.74</b>	<b>0.9456</b>	<b>3.71</b>
4	Meck-Vorpommern	Real-ESRGAN	29.95	0.9004	7.04
4		S2DR3	<b>34.98</b>	<b>0.9262</b>	<b>3.94</b>
5	Brandenburg 3	Real-ESRGAN	34.71	0.9371	4.53
5		S2DR3	<b>36.33</b>	<b>0.9476</b>	<b>3.76</b>
6	Saxony-Anhalt	Real-ESRGAN	34.61	0.9321	4.31
6		S2DR3	<b>35.47</b>	<b>0.9462</b>	<b>3.88</b>

The aggregated results in Table 3.2 demonstrate S2DR3's superior performance with an average PSNR improvement of 2.16 dB over ESRGAN, while maintaining consistently higher structural similarity (0.9437 vs 0.9288 SSIM) and significantly lower reconstruction error (3.82 vs 4.99 RMSE).

Table 3.2.: Average Super-Resolution Results Across All Sets. "RMSE ( $\downarrow$  = lower is better)"

<b>Model</b>	<b>Avg PSNR</b>	<b>Avg SSIM</b>	<b>Avg RMSE <math>\downarrow</math></b>
Real-ESRGAN	33.62	0.9288	4.99
S2DR3	<b>35.78</b>	<b>0.9437</b>	<b>3.82</b>

Notably, Set 4 shows the largest performance gap between models, with S2DR3 achieving a 5.03 dB PSNR advantage over Real-ESRGAN. This suggests that S2DR3's multispectral approach and 10 $\times$  upscaling capability provide particular advantages for challenging scenarios, potentially including varied terrain or atmospheric conditions.

For Satlas, we report BRISQUE and PIQE as no-reference perceptual quality metrics. These are widely adopted in SR literature and effectively capture visual degradations such as blur, noise, and unnatural texture without requiring ground-truth reference imagery.

**Analysis.** The no-reference image quality evaluation across six super-resolved image sets reveals a clear variation in perceptual fidelity. Set 2 achieves the lowest BRISQUE

### 3. Methodology & Results

---

Table 3.3.: Per-set BRISQUE and PIQE scores for Satlas SR model. Lower scores indicate better perceptual quality.

Set	BRISQUE ↓	PIQE ↓
Set 1	7.72	21.04
Set 2	<b>2.81</b>	<b>14.42</b>
Set 3	4.26	17.18
Set 4	28.54	30.69
Set 5	4.00	17.94
Set 6	14.08	22.70

(2.81) and PIQE (14.42) scores, indicating exceptionally high perceptual quality with minimal distortions or artifacts. Sets 1, 3, and 5 also exhibit strong performance, all with BRISQUE values below 8 and PIQE scores under 22, which is well within the thresholds generally associated with high-quality imagery. In contrast, Set 4 shows a significant degradation, with a BRISQUE score of 28.54 and a PIQE of 30.69, suggesting visible structural or texture inconsistencies that reduce visual realism. Set 6 demonstrates moderate quality with slightly elevated BRISQUE and PIQE values, but remains within acceptable bounds. Overall, the results affirm the effectiveness of the SR model in most cases, while highlighting a potential need for robustness improvements under certain image conditions, as observed in Set 4.

Table 3.4 shows the summary statistics for BRISQUE and PIQE across all sets.

Table 3.4.: Summary statistics for BRISQUE and PIQE across all Satlas image sets. Lower values denote better perceptual quality.

Metric	Mean	Standard Deviation $\sigma$	Range	Interpretation
BRISQUE	10.24	9.86	2.81–28.54	Lower = Better
PIQE	20.66	5.72	14.42–30.69	Lower = Better

For an improved understanding of the results, and to grasp how well the reported non-reference metrics are for satlas, we provide an aggregated table of the results for all the models, as seen in Table 3.5.

### 3. Methodology & Results

---

Table 3.5.: BRISQUE and PIQE scores across super-resolution models (Real-ESRGAN, S2DR3, Satlas) for each evaluation set. Lower scores indicate better perceptual quality. (The best scores per row are highlighted)

Set	Real-ESRGAN		S2DR3		Satlas	
	BRISQUE ↓	PIQE ↓	BRISQUE ↓	PIQE ↓	BRISQUE ↓	PIQE ↓
Set 1	26.41	11.02	18.27	10.95	7.72	21.04
Set 2	24.72	10.37	22.55	10.28	2.81	14.42
Set 3	21.77	10.61	19.43	10.69	4.26	17.18
Set 4	18.20	12.17	12.57	12.40	28.54	30.69
Set 5	16.64	10.90	20.90	10.50	4.00	17.94
Set 6	25.05	10.93	14.00	11.14	14.08	22.70

**Model-wise No-Reference Comparison.** While Satlas is the only model evaluated without reference images, its BRISQUE and PIQE scores frequently outperform those of ESRGAN and S2DR3. In four out of the six evaluated image sets, Satlas achieves the lowest BRISQUE value, which indicates a higher perceptual quality in the absence of pixel-level ground truth. For instance, in Set 2, which features one of the clearest photovoltaic installations, Satlas achieves a BRISQUE score of 2.81, outperforming both ESRGAN (24.72) and S2DR3 (22.55). However, Satlas shows significant perceptual degradation in Set 4, where its BRISQUE score reaches 28.54. This decline is likely caused by temporal inconsistencies in the multi-image stack or scene-specific characteristics such as water or shadow patterns, which may be more challenging for an RGB-only model.

S2DR3 presents more stable perceptual scores across the different sets. Although it does not outperform Satlas in BRISQUE, it generally maintains lower PIQE values compared to ESRGAN and shows reduced variance. This suggests that the multispectral model provides a consistent visual output, even in challenging scenarios. ESRGAN, on the other hand, exhibits the highest PIQE scores in most cases. This may reflect its tendency to introduce smoothing artefacts or unnatural texture patterns, which negatively affect its perceptual quality.

These observations support the interpretation that S2DR3 is the strongest performer in terms of structural and signal fidelity. Satlas tends to produce visually realistic outputs under favorable conditions, whereas ESRGAN acts as a middle ground that suffers more from perceptual degradation. The divergence between no-reference and reference-based metrics also reinforces the perception-distortion trade-off, which was previously discussed in Section 2.

### **3.5.2. PV Segmentation Results**

For the segmentation results, we use the functionality of the User Interface (UI) that was introduced in the previous chapter. We will use the same image sets as the ones used for the super-resolution results. Both the visual and quantitative results will be presented.

As previously mentioned, Pixel Accuracy is reported for completeness only. It is not considered a meaningful indicator in this binary segmentation task due to its sensitivity to class imbalance. Therefore, our analysis focuses on IoU and Dice scores, which better reflect the model's ability to capture photovoltaic regions.

Table 3.6 shows the segmentation results for all the models, across all six locations. It will also show that the Pixel Accuracy (PA) measurement is not a meaningful indicator in this task.

### 3. Methodology & Results

---

Table 3.6.: Segmentation Performance Across All Sets. (The best scores per row are highlighted in bold)

<b>Set</b>	<b>Model</b>	<b>IoU ↑</b>	<b>Pixel Accuracy ↑</b>	<b>Dice Coefficient ↑</b>
1 Lower Saxony	Original	0.00	0.921	0.000
	Real-ESRGAN	0.29	0.944	0.449
	S2DR3	0.45	0.957	0.617
	<b>Satlas</b>	<b>0.66</b>	<b>0.997</b>	<b>0.792</b>
2 Brandenburg 1	Original	0.48	0.917	0.652
	<b>Real-ESRGAN</b>	<b>0.87</b>	0.987	<b>0.932</b>
	S2DR3	0.77	0.963	0.868
	Satlas	0.85	<b>0.987</b>	0.917
3 Brandenburg 2	Original	0.77	0.978	0.870
	Real-ESRGAN	0.84	0.982	0.912
	<b>S2DR3</b>	<b>0.91</b>	<b>0.991</b>	<b>0.956</b>
	Satlas	0.85	0.988	0.919
4 Mecklenburg-Vorpommern	Original	0.65	<b>0.950</b>	0.786
	<b>Real-ESRGAN</b>	<b>0.67</b>	0.944	<b>0.801</b>
	S2DR3	0.53	0.917	0.694
	Satlas	0.46	0.918	0.629
5 Brandenburg 3	Original	0.86	0.992	0.924
	<b>Real-ESRGAN</b>	<b>0.92</b>	<b>0.994</b>	<b>0.957</b>
	S2DR3	0.82	0.970	0.898
	Satlas	0.83	0.985	0.904
6 Saxony-Anhalt	Original	0.63	0.983	0.775
	<b>Real-ESRGAN</b>	<b>0.85</b>	<b>0.995</b>	<b>0.918</b>
	S2DR3	0.82	0.994	0.902
	Satlas	0.66	0.990	0.794

**Analysis.** The segmentation results show a consistent improvement in performance when super-resolution techniques are applied. Across all sets, the original Sentinel-2 images yield the lowest scores in most metrics, confirming that the native 10 m resolution imposes limitations on photovoltaic (PV) detection.

Among the super-resolved outputs, **Real-ESRGAN** achieves the highest average IoU and Dice scores, outperforming all other models in Sets 2, 4, 5, and 6. Its lightweight architecture combined with  $4\times$  upscaling seems to strike an effective balance between detail enhancement and model compatibility.

**S2DR3** delivers top performance in Set 3, with an IoU of 0.91 and Dice of 0.956. Its ability to exploit all 13 Sentinel-2 bands and upscale by a factor of 10 enables it to perform well in structurally complex scenes. However, it underperforms in Set 4, possibly due to artefacts or spectral inconsistencies from multispectral processing.

**Satlas** demonstrates strong performance in Sets 1, 2, and 3, but shows instability in Sets 4 and 6. As a multi-image temporal model, it is more sensitive to alignment errors, especially when dealing with seasonal or illumination variability. Despite that, it occasionally achieves the best perceptual segmentation fidelity, as supported by its high Dice coefficients in multiple sets.

Overall, the results confirm that super-resolution models enhance segmentation quality, with Real-ESRGAN offering the most stable improvements, S2DR3 excelling in specific conditions, and Satlas achieving competitive results but with greater variance.

For a more visual representation of the results, we provide the following figures.

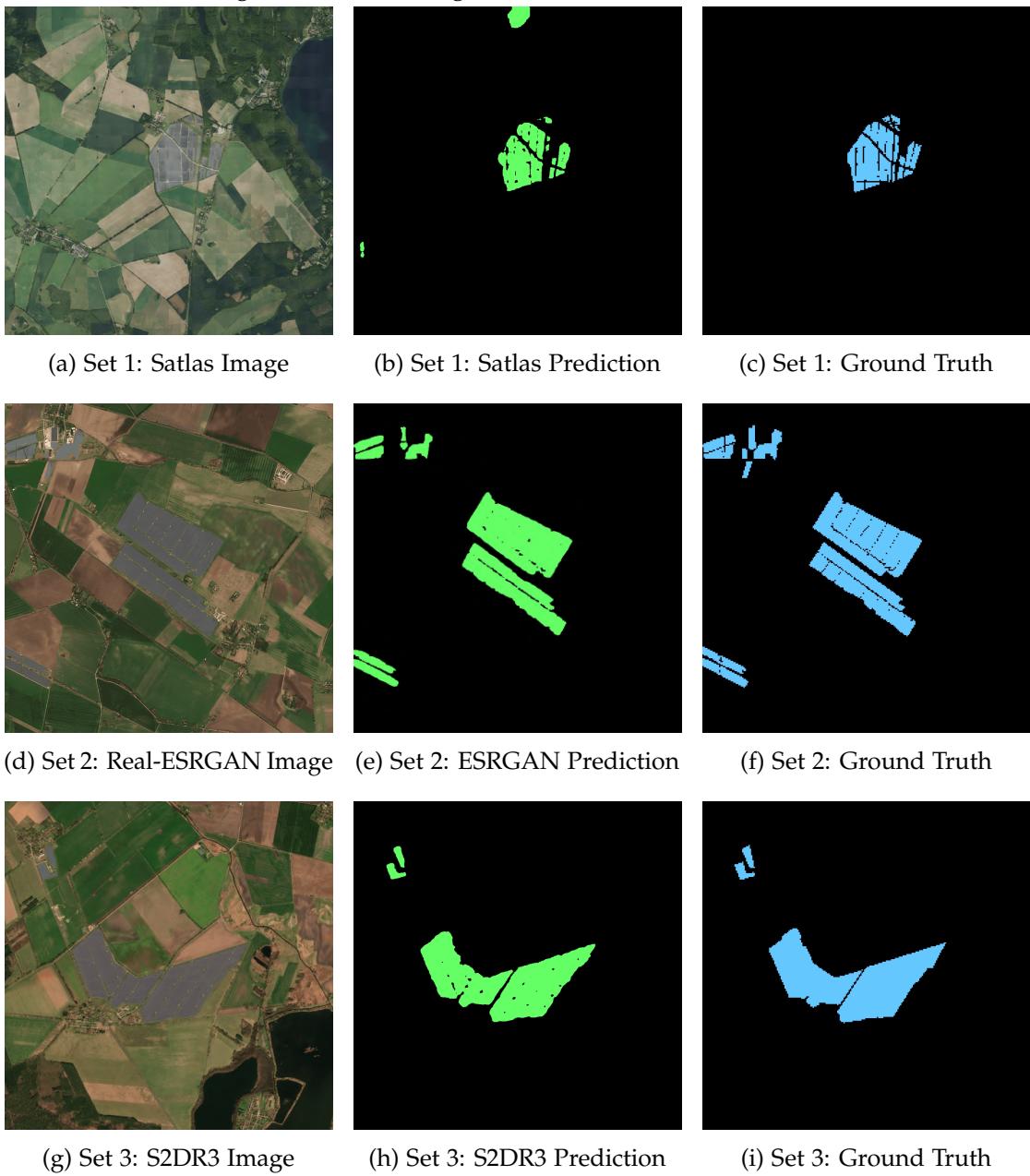
### Visual Segmentation Results

Figures 3.21 and 3.22 present key visual examples of segmentation performance, showing the best performing model for each image set. For each example, we show the input image, the model’s segmentation prediction, and the ground truth mask. Complete visual comparisons for all models across all sets are provided in Appendix C.

### 3. Methodology & Results

---

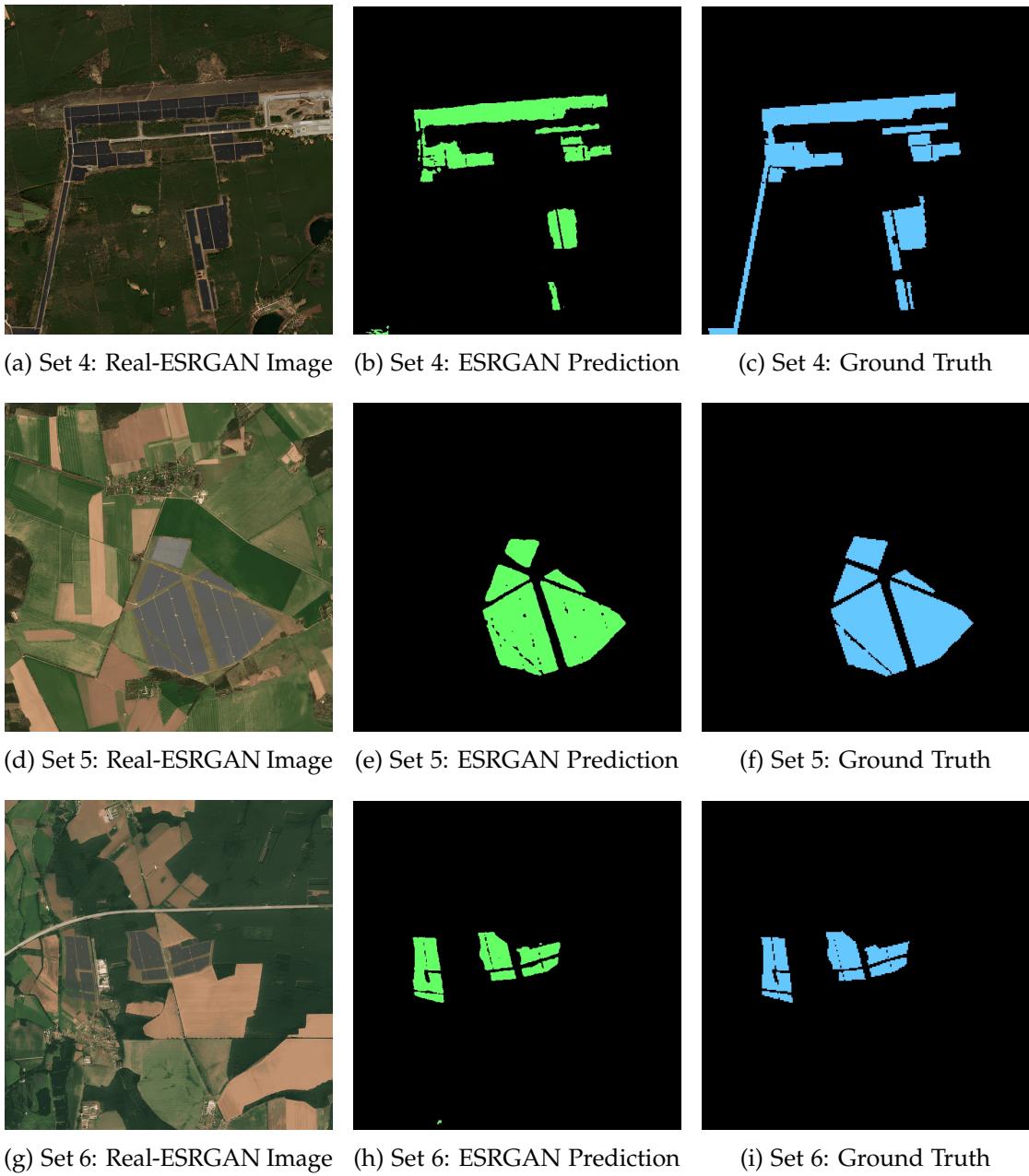
Figure 3.21.: Best Segmentation Results for Sets 1-3



### 3. Methodology & Results

---

Figure 3.22.: Best Segmentation Results for Sets 4-6



---

### *3. Methodology & Results*

---

These visuals support the quantitative trends in Table 3.6: super-resolution consistently enhances segmentation performance, with different models excelling in different scenarios depending on image characteristics and scene complexity. The selected examples represent the best achievable performance for each location, demonstrating the significant potential of super-resolution techniques for improving photovoltaic panel detection and boundary delineation.

## 4. Discussion & Future Work

### 4.1. Influence of Sun Reflections and Discoloration on Segmentation Performance

Sun reflections present a significant preprocessing challenge. Such reflections can manifest as saturated or overly bright regions, potentially complicating accurate panel delineation during segmentation. Results indicated varying degrees of robustness among models. While Real-ESRGAN and S2DR3 experienced relatively lower performance in Set 1, Satlas demonstrated higher resilience, achieving notably stronger segmentation accuracy (IoU: 0.66). Set 3 showed consistently good segmentation performance across all models (IoU exceeding 0.84), suggesting that either the reflections were mild enough to not pose significant disruption or the models managed to handle reflections effectively.

On the other hand, Set 4 (Mecklenburg-Vorpommern) demonstrated a distinct challenge unrelated to sun reflection. The perceptual quality metrics (BRISQUE: 28.54, PIQE: 30.69 for Satlas) and lower segmentation results indicate some degree of image degradation, likely due to temporal misalignment or discoloration effects. Such artifacts could stem from temporal stacking misalignments inherent in multi-image SR methods like Satlas or environmental factors affecting spectral consistency. This highlights that temporal and discoloration issues are distinct, critical factors to consider during preprocessing, significantly influencing segmentation quality independently from sun reflections.

An example of such reflections can be seen in Figure 4.1 where we highlight a bad reflection in the original Sentinel-2 image.

#### 4. Discussion & Future Work

---



Figure 4.1.: Example of poor sun reflection in Set 3 (Brandenburg 2).

Part of preprocessing is to avoid such reflections. Existing tools within Copernicus Hub, are designed to automatically filter out dates where cloud coverage is within a tunable threshold, but there are no tunable parameters for sun reflections.

These observations collectively underscore the complexity of remote sensing image preprocessing, where distinct artifacts (reflective glare, discoloration, and temporal inconsistencies) require targeted mitigation strategies. Future work would explore robust preprocessing methods specifically addressing these issues, potentially incorporating glare-masking techniques, temporal alignment algorithms, and spectral consistency checks to improve overall segmentation performance and reliability in practical deployments.

## 4.2. Perceptual Quality vs. Task-Oriented Utility

While image quality metrics such as PSNR, SSIM, BRISQUE, and PIQE effectively measure perceptual fidelity, our results demonstrate that these metrics alone are not always indicative of downstream task performance, particularly segmentation. For example, Satlas frequently achieved excellent perceptual scores but displayed inconsistent segmentation accuracy, as observed in Set 4. Conversely, Real-ESRGAN, despite moderate perceptual quality, demonstrated robust segmentation performance across diverse

scenarios. This divergence highlights that perceptual realism is not comprehensive for analytical utility. Future research could explore task-specific perceptual loss functions or joint training frameworks explicitly tailored to optimize segmentation outcomes rather than general perceptual quality.

### **4.3. Effect of Multi-image vs. Single-image Super-Resolution**

Satlas's multi-temporal stacking approach showed mixed outcomes. It delivered excellent results in certain datasets yet suffered significant degradation in others (notably Set 4). This suggests inherent instability related to temporal alignment and spectral consistency in multi-image super-resolution techniques. Single-image methods like Real-ESRGAN and S2DR3, while simpler, consistently produced more stable results. Future work should explicitly examine temporal alignment strategies or hybrid approaches combining the strengths of multi-temporal richness with single-image robustness.

### **4.4. Computational Practicality and Scalability**

Although S2DR3 produced superior multispectral resolution enhancements, it required substantial computational resources and complex data management. In contrast, Real-ESRGAN provided competitive segmentation performance at a lower computational cost. This emphasizes the critical trade-off between computational requirements and analytical improvements. Future deployments, particularly in the context of Super Resolution, should prioritize scalability and practical efficiency, potentially favoring simpler models or hybrid approaches that balance computational efficiency with analytical accuracy. This is especially relevant for tasks that involve time-series data, where the computational cost of processing each image can become prohibitive.

### **4.5. Model Generalization and Geographic Transferability**

The variability in segmentation performance across different datasets underscores the challenge of model generalization. Real-ESRGAN consistently demonstrated reliable performance across diverse geographic locations, whereas Satlas and S2DR3 exhibited higher sensitivity to local conditions and artifacts. This highlights the importance of robust training methodologies incorporating geographically diverse datasets, domain adaptation techniques, and fine-tuning processes. Future research should focus on enhancing the generalizability and transferability of segmentation models to ensure reliable performance across varying real-world conditions.

## **4.6. Optimal Resolution Determination for Segmentation**

Interestingly, our experiments revealed scenarios where slightly reducing the resolution of super-resolved imagery improved segmentation accuracy. This suggests an optimal intermediate resolution exists, where too high a resolution introduces unnecessary detail that can overwhelm segmentation models, whereas excessively low resolutions obscure critical features. A systematic resolution sensitivity analysis could help identify ideal scales tailored specifically to segmentation objectives, optimizing the balance between image clarity and processing efficiency.

## **4.7. Multispectral Utilization and Physical Constraints**

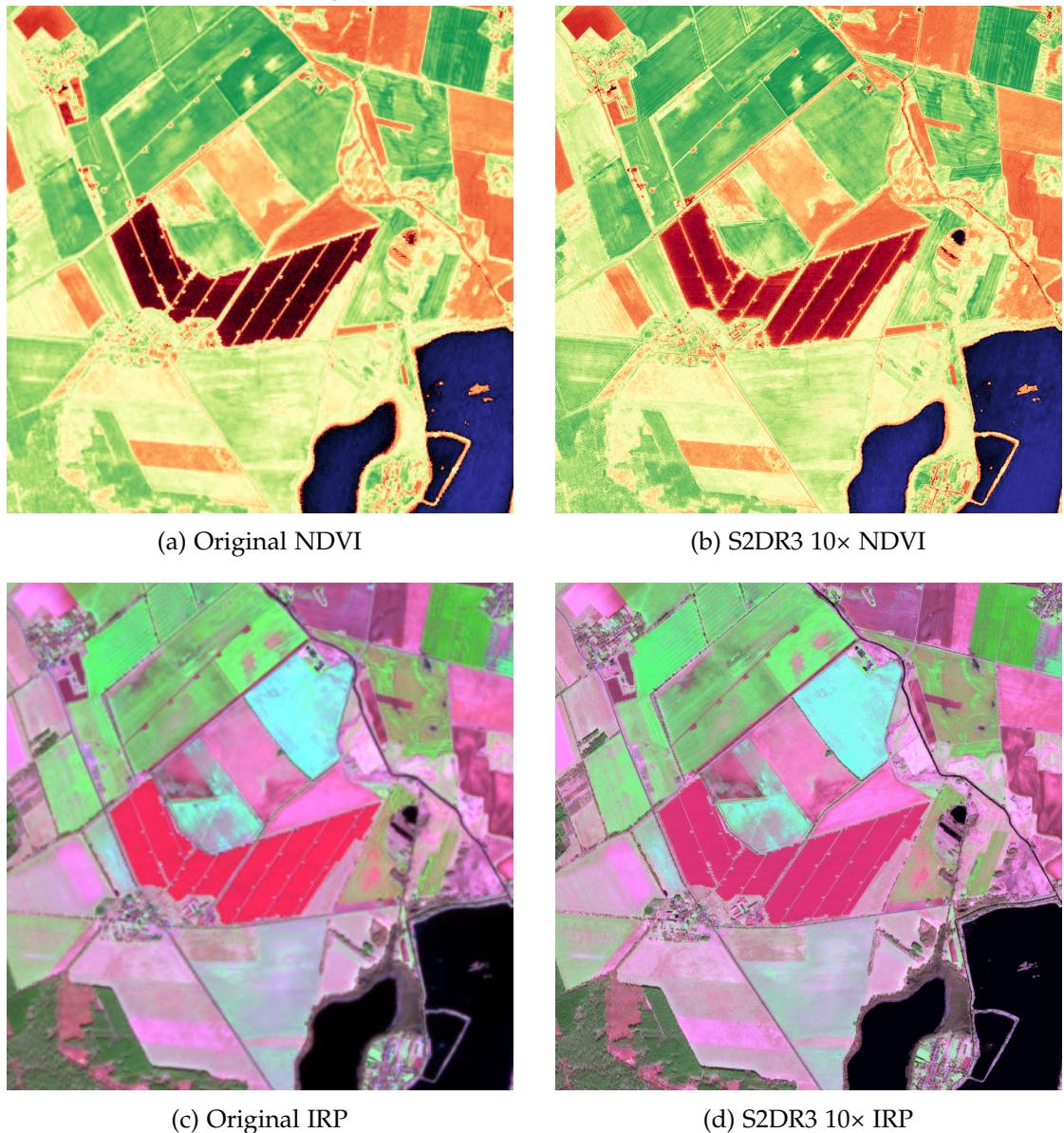
This study primarily focused on RGB spectral bands due to limitations in certain super-resolution models, specifically Real-ESRGAN and Satlas. Nevertheless, as previously discussed in Chapter 2, Sentinel-2 imagery offers additional valuable spectral bands, such as NIR and SWIR, which may enhance segmentation capabilities.

Figures 4.2 shows the original and S2DR3 10x super-resolution results for the NDVI and IRP bands of Set 3 (Brandenburg 2). Results hint at some discoloration, yet the same positive remarks we had about S2DR3 for the RGB bands hold true for the non-RGB bands.

#### 4. Discussion & Future Work

---

Figure 4.2.: Multispectral comparison: Original vs S2DR3 10× Super-Resolution for Set 3 (Brandenburg 2)



---

#### *4. Discussion & Future Work*

---

Preliminary super-resolution results obtained using the multispectral capabilities of the S2DR3 model indicate promising quality improvements across these non-RGB bands. Future studies should further investigate the utility of these multispectral super-resolved outputs, explicitly evaluating whether their inclusion enhances photovoltaic (PV) segmentation performance. Assessing their effectiveness in challenging scenarios, such as differentiating PV panels from similarly reflective backgrounds or reducing the impact of sun reflections and discoloration artifacts, would be particularly beneficial. Additionally, future work could incorporate physical or spectral characteristics specific to PV arrays into super-resolution and segmentation algorithms, potentially improving robustness and minimizing artifacts such as spectral distortions or unrealistic textures. This integration would ultimately support more accurate and reliable segmentation outcomes in practical deployments.

## 5. Conclusion

This thesis explored the application of generative super-resolution techniques to enhance satellite imagery for photovoltaic (PV) detection, with a particular focus on Sentinel-2 imagery. The primary goal was to determine whether enhanced image resolutions, achieved through various generative AI models, could meaningfully improve the accuracy of PV segmentation.

The study demonstrated that super-resolution techniques significantly improved the visual and perceptual quality of satellite images. However, these improvements in image quality metrics did not consistently correspond to enhanced segmentation outcomes. In many cases, segmentation performance depended more on the compatibility of generated textures and structural features with the segmentation model than on perceptual fidelity alone. Real-ESRGAN proved to be a particularly robust and computationally efficient model, delivering stable segmentation results across varied geographic conditions and image scenarios.

Preprocessing challenges such as sun reflections, temporal misalignments, and spectral discoloration were found to affect segmentation quality independently of image resolution. The results also suggested the presence of an optimal intermediate resolution for segmentation tasks. Excessive resolution introduced unnecessary detail or artifacts, while insufficient resolution reduced the visibility of critical features.

Future research should explore the use of multispectral super-resolved imagery, particularly leveraging Sentinel-2's near-infrared and shortwave infrared bands. These bands were super-resolved using S2DR3 in this study but were not used for segmentation due to model limitations. Including these additional bands may enhance segmentation accuracy in more complex or ambiguous environments. Moreover, integrating domain-specific knowledge and physical constraints into both super-resolution and segmentation algorithms may reduce hallucinated details and improve reliability in operational settings.

Although this work focused on PV segmentation, the methodology, findings, and tools developed are applicable to a wide range of remote sensing tasks. Generative super-resolution can support many satellite-based applications, including land use classification, agricultural analysis, urban mapping, and environmental monitoring. By improving the spatial detail of publicly available satellite imagery and examining its interaction with segmentation models, this work contributes to more accurate, scalable,

## *5. Conclusion*

---

and cost-effective Earth observation solutions across diverse domains.

Lastly, a comprehensive GitHub repository is provided and can be publicly accessed at <https://github.com/mourabitiziyad/pv-sr-detection-thesis>.

# A. Code Listings

This appendix contains the complete source code for scripts and algorithms used in this thesis.

## A.1. Data Processing Scripts

```
import json
import numpy as np
import cv2
from PIL import Image
import os

# Path to your LabelMe JSON file
json_path = "path/to/labelme_json.json"

# Load LabelMe JSON
with open(json_path, "r") as f:
    data = json.load(f)

image_shape = (data["imageHeight"], data["imageWidth"])
mask = np.zeros(image_shape, dtype=np.uint8)

for shape in data["shapes"]:
    if shape["shape_type"] != "polygon":
        continue
    points = np.array(shape["points"], dtype=np.int32)
    cv2.fillPoly(mask, [points], 1) # Fill PV area with 1

# Save the binary mask in the same directory as the input JSON file
output_dir = os.path.dirname(json_path)
output_filename = os.path.splitext(os.path.basename(json_path))[0] + ".png"
output_path = os.path.join(output_dir, output_filename)
```

```
mask_img = Image.fromarray(mask * 255) # scale to 0-255 for visibility
mask_img.save(output_path)
print(f"Saved binary mask to {output_path}")
```

Listing A.1: Script to convert LabelMe JSON annotations to PNG masks

## A.2. Geospatial Metadata

This section provides detailed geospatial metadata for the Sentinel-2 imagery used in this thesis, extracted using the GDAL library.

### A.2.1. Brandenburg Dataset - gdalinfo Output

The following output shows the complete geospatial metadata for the Brandenburg photovoltaic site, demonstrating the coordinate system, projection parameters, and geographic bounds of the Sentinel-2 imagery processed through the S2DR3 super-resolution pipeline.

```
Driver: GTiff/GeoTIFF
Files: figures/original/set_2.tif
Size is 418, 418
Coordinate System is:
PROJCRS["WGS_84/UTM_zone_33N",
    BASEGEOGCRS["WGS_84",
        DATUM["World Geodetic System 1984",
            ELLIPSOID["WGS_84",6378137,298.257223563,
                LENGTHUNIT["metre",1]]],
        PRIMEM["Greenwich",0,
            ANGLEUNIT["degree",0.0174532925199433]],
        ID["EPSG",4326]],
    CONVERSION["UTM_zone_33N",
        METHOD["Transverse Mercator",
            ID["EPSG",9807]],
        PARAMETER["Latitude_of_natural_origin",0,
            ANGLEUNIT["degree",0.0174532925199433],
            ID["EPSG",8801]],
        PARAMETER["Longitude_of_natural_origin",15,
            ANGLEUNIT["degree",0.0174532925199433],
```

---

## A. Code Listings

---

```
ID["EPSG",8802]],  
PARAMETER["Scale_factor_at_natural_origin",0.9996,  
SCALEUNIT["unity",1],  
ID["EPSG",8805]],  
PARAMETER["False_easting",500000,  
LENGTHUNIT["metre",1],  
ID["EPSG",8806]],  
PARAMETER["False_northing",0,  
LENGTHUNIT["metre",1],  
ID["EPSG",8807]],  
CS[Cartesian,2],  
AXIS["(E)",east,  
ORDER[1],  
LENGTHUNIT["metre",1]],  
AXIS["(N)",north,  
ORDER[2],  
LENGTHUNIT["metre",1]],  
USAGE[  
SCOPE["Navigation_and_medium_accuracy_spatial_referencing."],  
AREA["Between_12_degrees_E_and_18_degrees_E,_northern_hemisphere_between  
equator_and_84_degrees_N,_onshore_and_offshore._Austria._Bosnia_and  
Herzegovina._Cameroon._Central_African_Republic._Chad._Congo._Croatia.  
Czechia._Democratic_Republic_of_the_Congo_(Zaire)._Gabon._Germany.  
Hungary._Italy._Libya._Malta._Niger._Nigeria._Norway._Poland._San  
Marino._Slovakia._Slovenia._Svalbard._Sweden._Vatican_City_State."],  
BBOX[0,12,84,18]],  
ID["EPSG",32633]]  
Data axis to CRS axis mapping: 1,2  
Origin = (445960.00000000000000,5841000.00000000000000)  
Pixel Size = (10.00000000000000,-10.00000000000000)  
Metadata:  
TIFFTAG_SOFTWARE=S2DR3  
TIFFTAG_COPYRIGHT=GAMMA.EARTH  
S2DR3_AOI_CLOUDS=0  
S2DR3_AOI_NODATA=0  
S2DR3_CLOUDS=0.088984  
S2DR3_CRS=32633  
S2DR3_DATE=20250404  
S2DR3_ID=S2A_33UVU_20250404_1_L2A
```

---

#### A. Code Listings

---

```
S2DR3_MGRS=33UVU
S2DR3_S3PATH=sentinel-s2-l2a-cogs/33/U/VU/2025/4/S2A_33UVU_20250404_1_L2A
AREA_OR_POINT=Area
Image Structure Metadata:
  COMPRESSION=JPEG
  INTERLEAVE=PIXEL
  JPEG_QUALITY=75
  JPEGTABLESMODE=1
Corner Coordinates:
  Upper Left ( 445960.000, 5841000.000) ( 14d11'59.96"E, 52d42'58.29"N)
  Lower Left ( 445960.000, 5836820.000) ( 14d12' 2.43"E, 52d40'43.02"N)
  Upper Right ( 450140.000, 5841000.000) ( 14d15'42.71"E, 52d42'59.73"N)
  Lower Right ( 450140.000, 5836820.000) ( 14d15'44.99"E, 52d40'44.46"N)
  Center ( 448050.000, 5838910.000) ( 14d13'52.52"E, 52d41'51.39"N)
  Band 1 Block=418x16 Type=Byte, ColorInterp=Red
  Band 2 Block=418x16 Type=Byte, ColorInterp=Green
  Band 3 Block=418x16 Type=Byte, ColorInterp=Blue
```

Listing A.2: gdalinfo output for Brandenburg dataset (set 2.tif)

This metadata confirms that the imagery uses the WGS 84 / UTM zone 33N coordinate system (EPSG:32633), has a 10-meter pixel resolution, and covers approximately 4.18 km × 4.18 km. The S2DR3 metadata indicates this is super-resolved Sentinel-2 data processed by GAMMA.EARTH, captured on April 4, 2025, with 8.9% cloud coverage in the original scene.

## B. Super-Resolution Image Comparisons

This section provides visual comparisons of the super-resolution results across all six image sets used in this thesis. Each figure shows the original Sentinel-2 image alongside the outputs from the three super-resolution models: S2DR3 (10 $\times$  upscaling), ESRGAN (4 $\times$  upscaling), and Satlas (4 $\times$  upscaling).

The images demonstrate the varying capabilities of each model in enhancing photovoltaic installations across different German states, including Brandenburg (3 sites), Mecklenburg-Vorpommern, and Saxony-Anhalt. All images were captured during the summer period of 2024–2025 to minimize cloud cover and atmospheric distortion.

*B. Super-Resolution Image Comparisons*

---

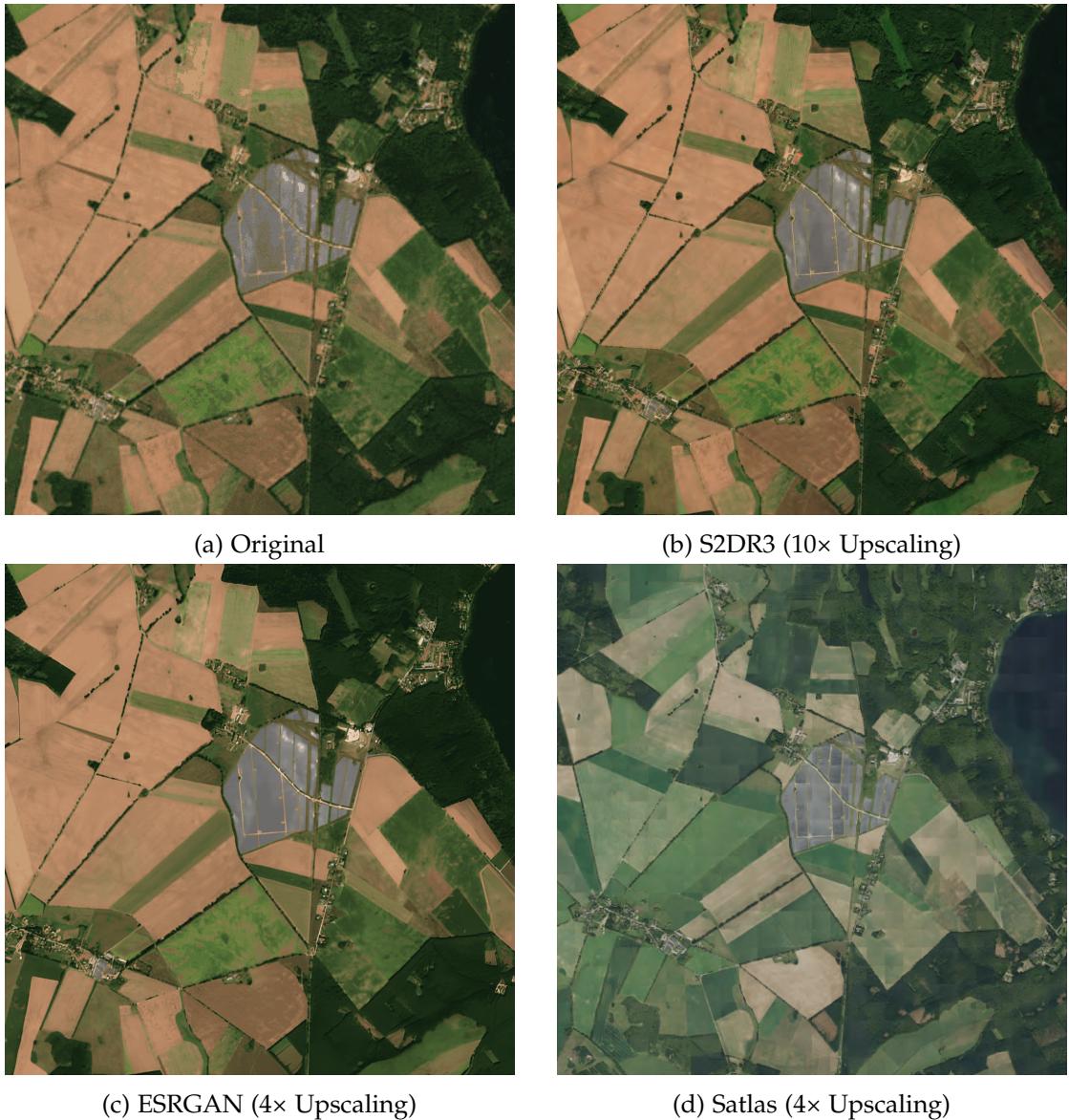


Figure B.1.: Super-resolution comparison for image set 1.

*B. Super-Resolution Image Comparisons*

---



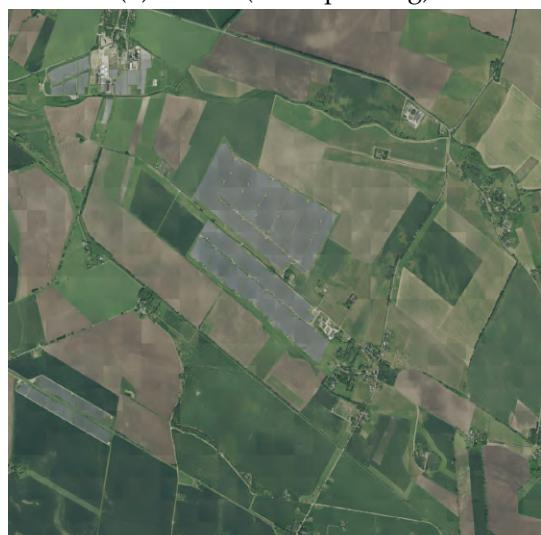
(a) Original



(b) S2DR3 (10 $\times$  Upscaling)



(c) ESRGAN (4 $\times$  Upscaling)

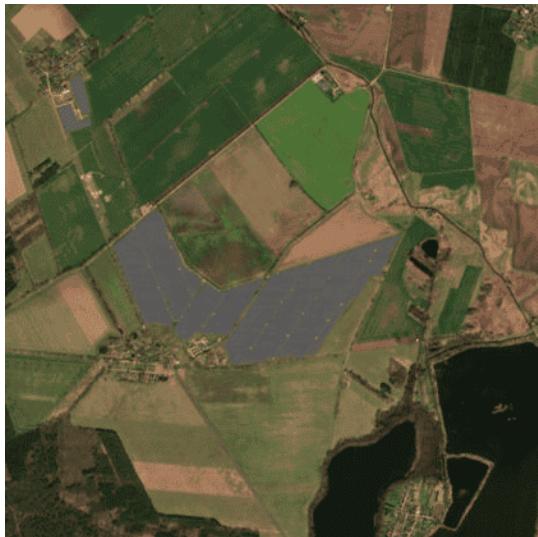


(d) Satlas (4 $\times$  Upscaling)

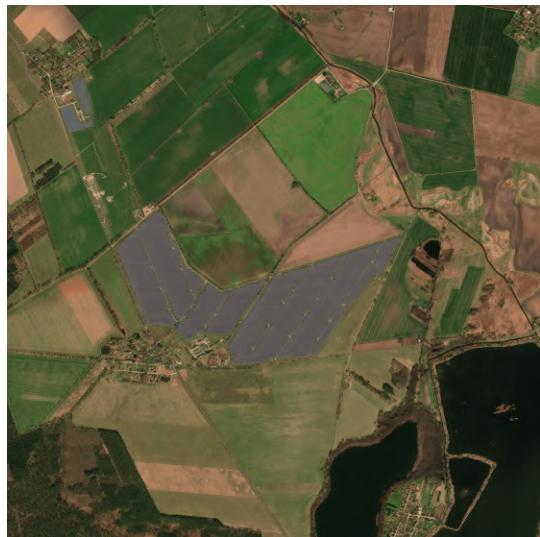
Figure B.2.: Super-resolution comparison for image set 2.

*B. Super-Resolution Image Comparisons*

---



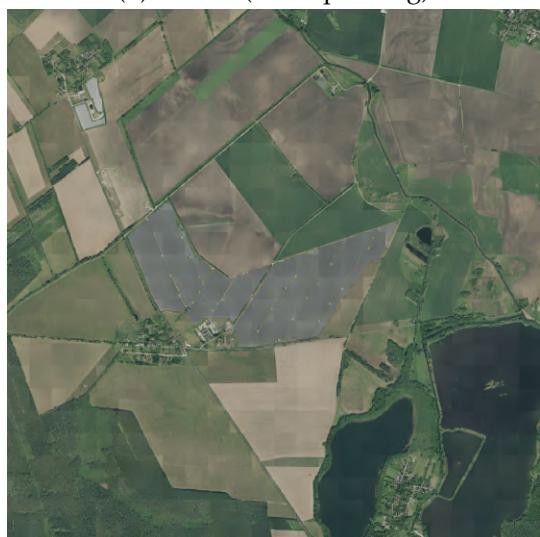
(a) Original



(b) S2DR3 (10 $\times$  Upscaling)



(c) ESRGAN (4 $\times$  Upscaling)



(d) Satlas (4 $\times$  Upscaling)

Figure B.3.: Super-resolution comparison for image set 3.

*B. Super-Resolution Image Comparisons*

---



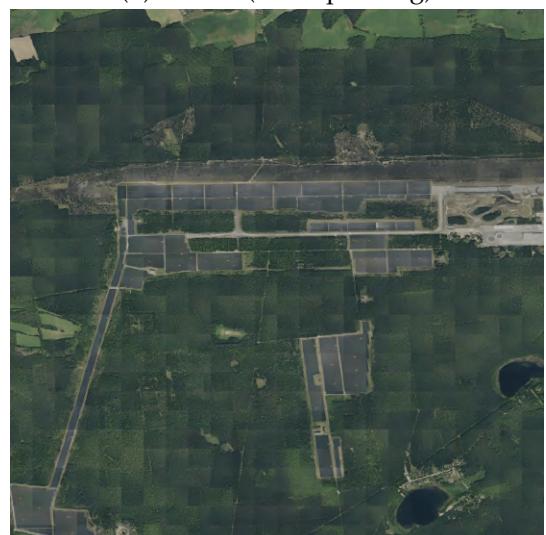
(a) Original



(b) S2DR3 (10 $\times$  Upscaling)



(c) ESRGAN (4 $\times$  Upscaling)



(d) Satlas (4 $\times$  Upscaling)

Figure B.4.: Super-resolution comparison for image set 4.

*B. Super-Resolution Image Comparisons*

---



(a) Original



(b) S2DR3 (10 $\times$  Upscaling)



(c) ESRGAN (4 $\times$  Upscaling)

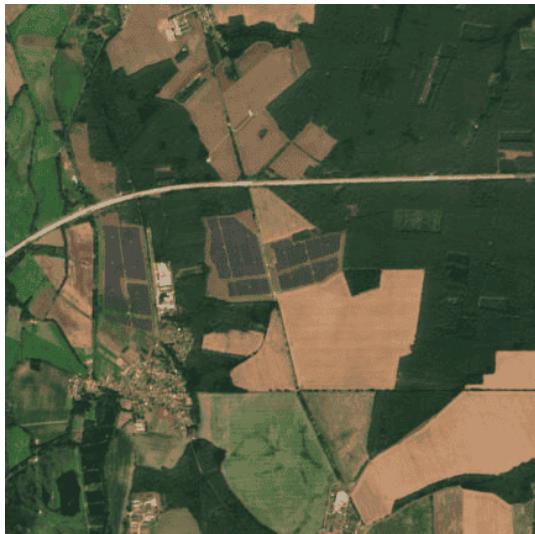


(d) Satlas (4 $\times$  Upscaling)

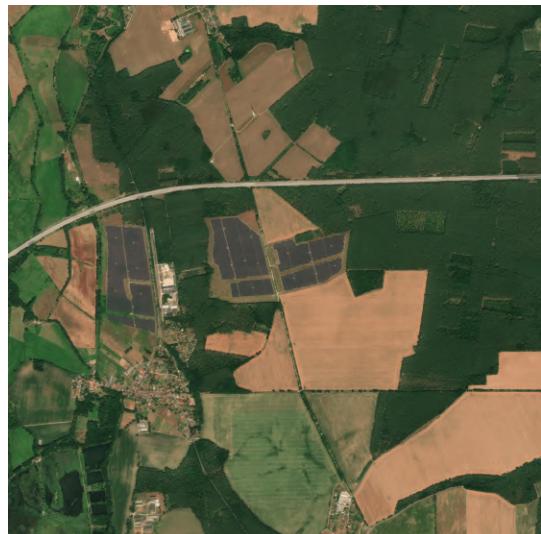
Figure B.5.: Super-resolution comparison for image set 5.

*B. Super-Resolution Image Comparisons*

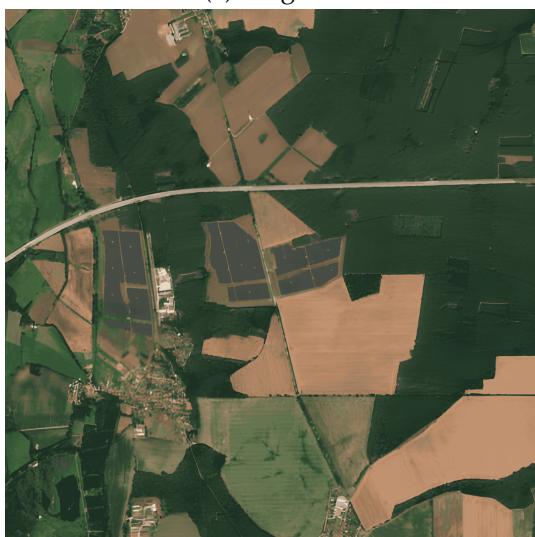
---



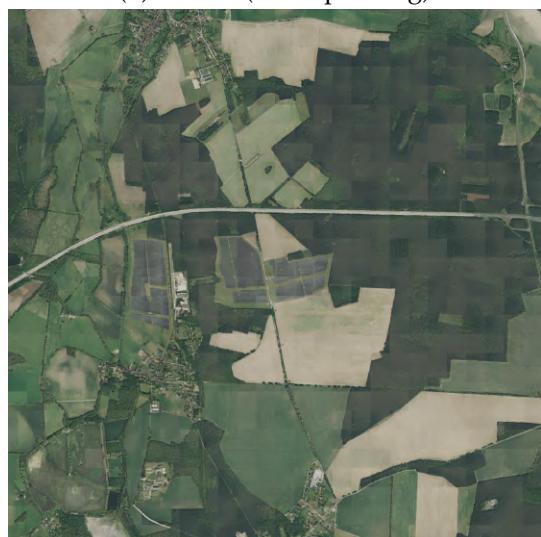
(a) Original



(b) S2DR3 (10 $\times$  Upscaling)



(c) ESRGAN (4 $\times$  Upscaling)



(d) Satlas (4 $\times$  Upscaling)

Figure B.6.: Super-resolution comparison for image set 6.

## C. Complete Segmentation Results

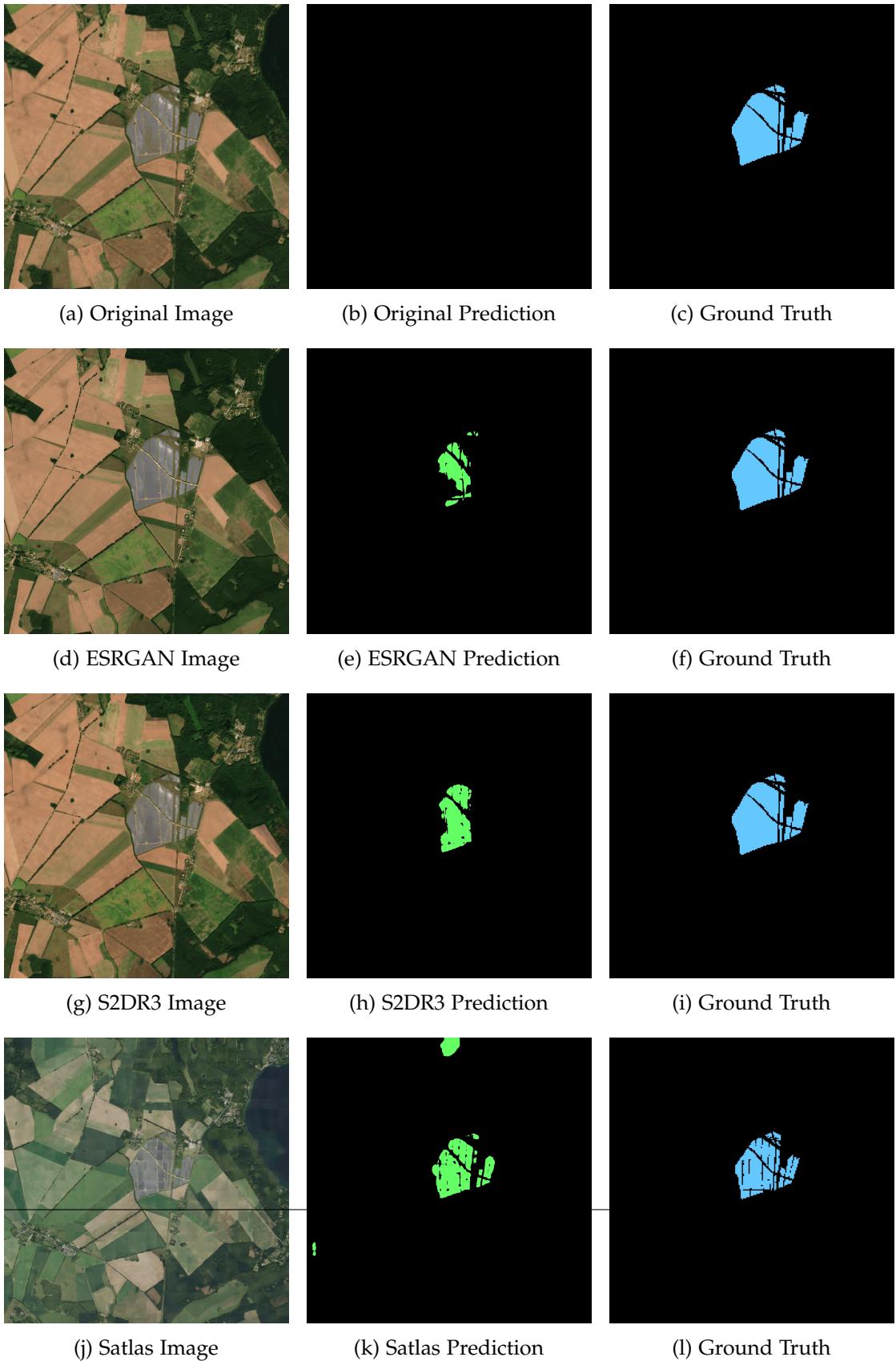
This appendix provides comprehensive visual comparisons of segmentation performance across all image sets and models (ESRGAN, S2DR3, Satlas). For each set, we show the input image, the model's segmentation prediction, and the ground truth mask. The layout is organized with one row per model for easy comparison across different super-resolution approaches.

These comprehensive results demonstrate the varying performance of different super-resolution approaches across all image sets and geographic locations. The complete visual comparison allows for detailed analysis of how each model handles different scene characteristics, lighting conditions, and PV installation types across the German study sites.

### C. Complete Segmentation Results

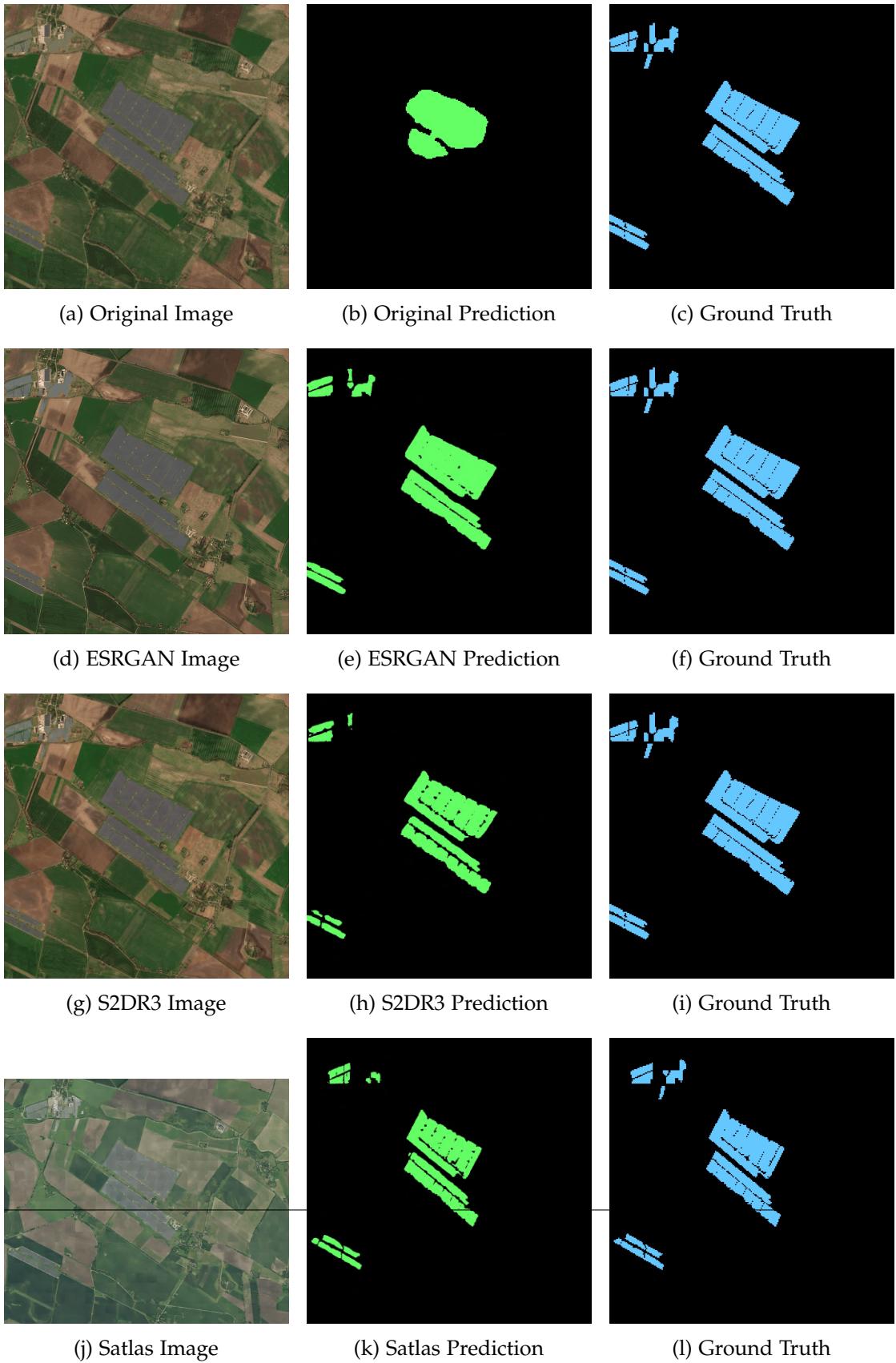
---

Figure C.1.: Set 1 (Lower Saxony) - Complete Segmentation Results



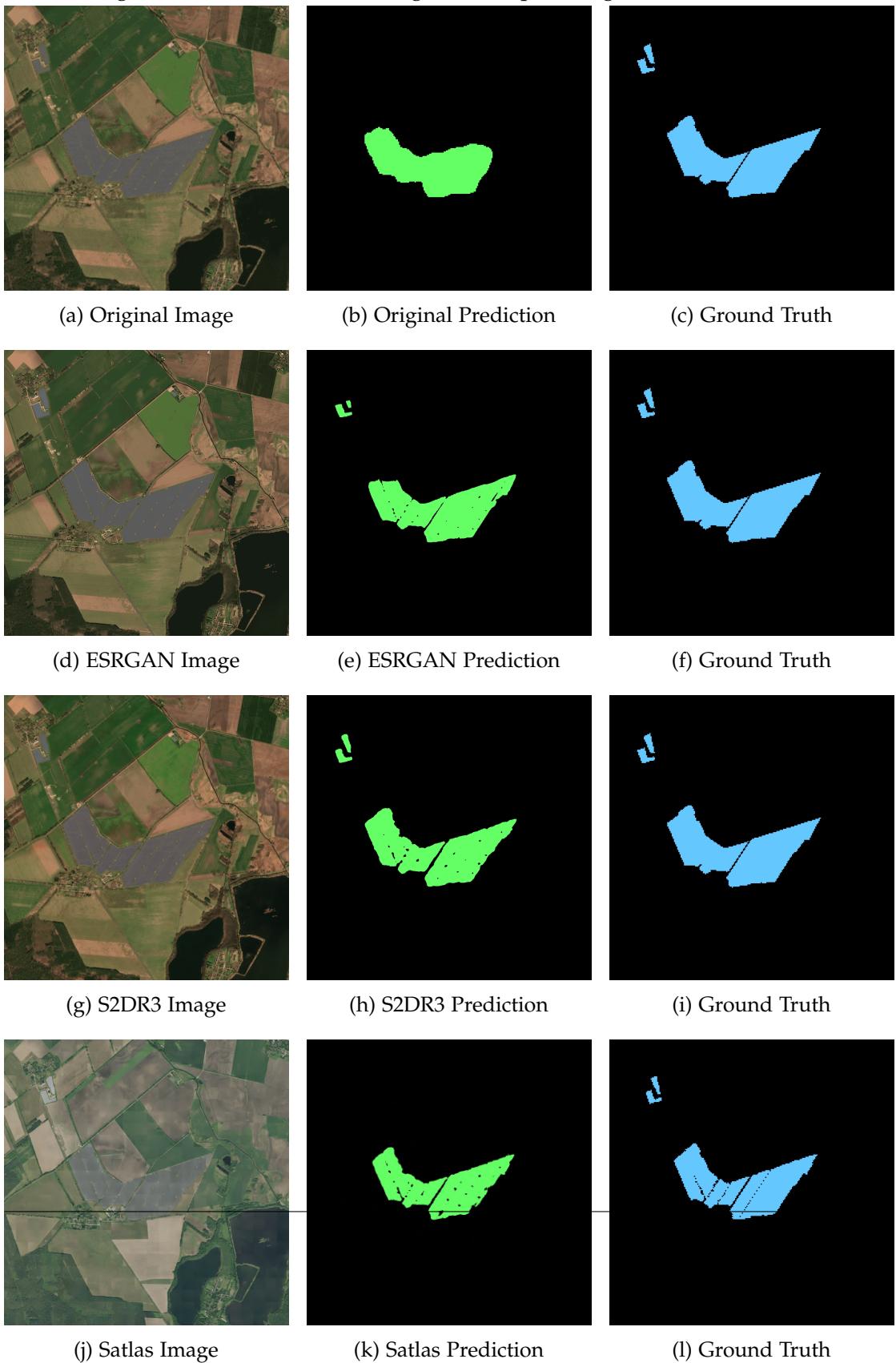
### C. Complete Segmentation Results

Figure C.2.: Set 2 (Brandenburg 1) - Complete Segmentation Results



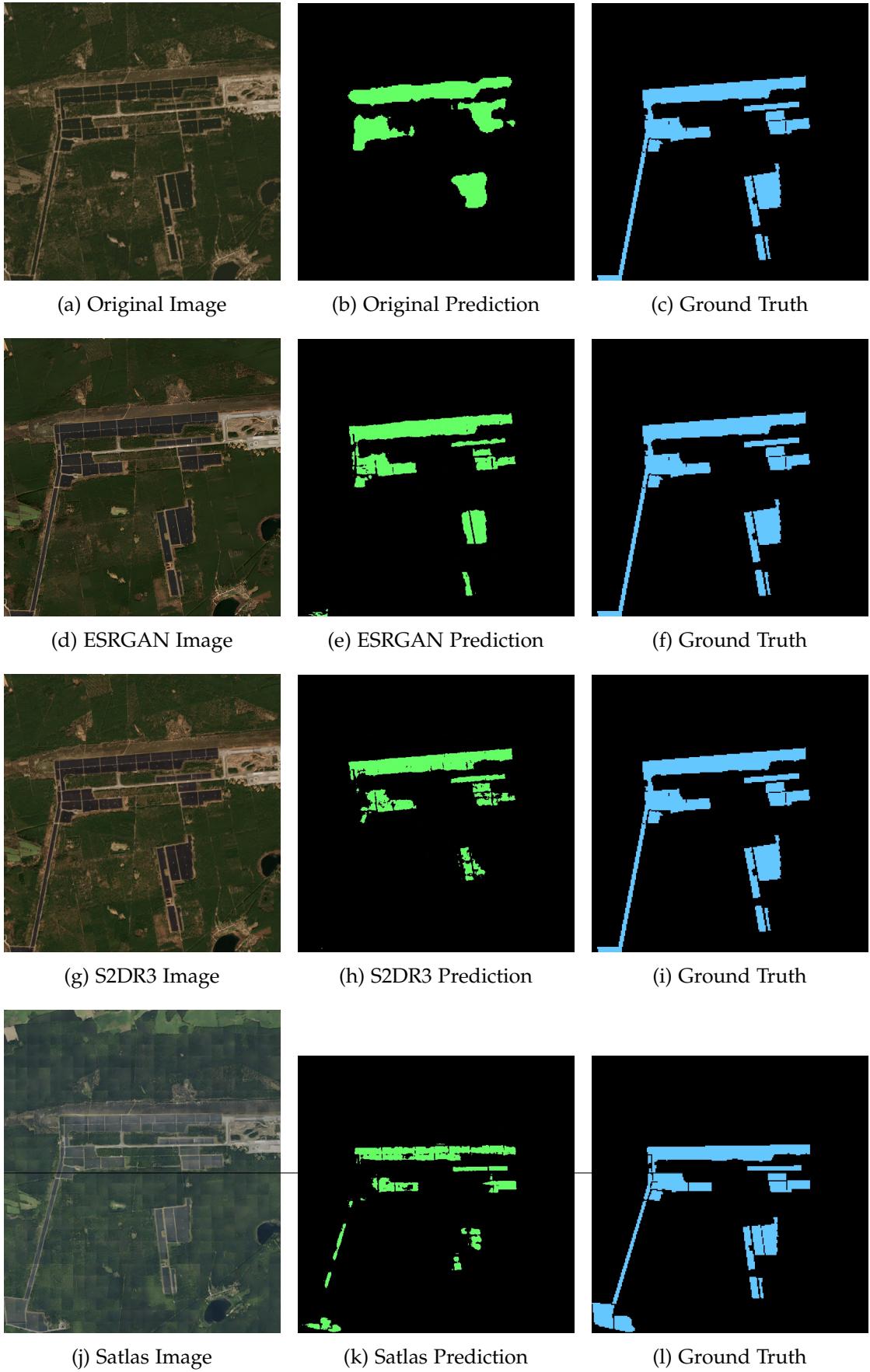
### C. Complete Segmentation Results

Figure C.3.: Set 3 (Brandenburg 2) - Complete Segmentation Results



### C. Complete Segmentation Results

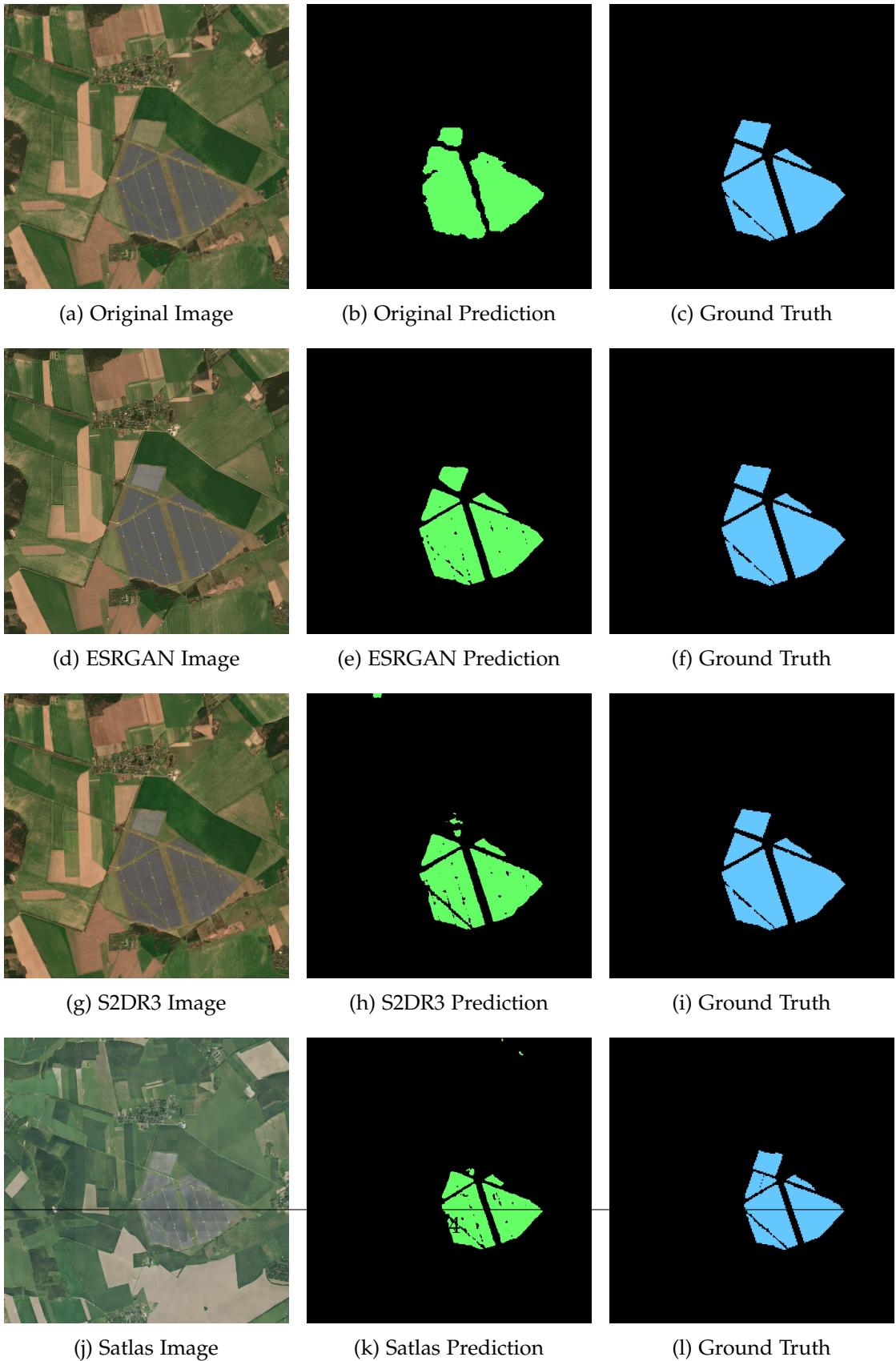
Figure C.4.: Set 4 (Mecklenburg-Vorpommern) - Complete Segmentation Results



### C. Complete Segmentation Results

---

Figure C.5.: Set 5 (Brandenburg 3) - Complete Segmentation Results



### C. Complete Segmentation Results

---

Figure C.6.: Set 6 (Saxony-Anhalt) - Complete Segmentation Results



# Abbreviations

**AI** Artificial Intelligence

**ASPP** Atrous Spatial Pyramid Pooling

**BRISQUE** Blind/Referenceless Image Spatial Quality Evaluator

**CSWGAN** Cascaded Spatial Frequency-Domain Generative Adversarial Network

**DSen2** Deep Sentinel-2 Super-Resolution Network

**EC** European Commission

**ESA** European Space Agency

**ESRGAN** Enhanced Super Resolution Generative Adversarial Network

**GAN** Generative Adversarial Network

**GSD** Ground-sample distance

**HR** High Resolution

**IoU** Intersection-over-Union

**LR** Low Resolution

**NAIP** National Agriculture Imagery Program

**NAS** Neural Architecture Search

---

*Abbreviations*

---

**NCC** Normalized Cross-Correlation

**NDMI** Normalized Difference Moisture Index

**NDVI** Normalized Difference Vegetation Index

**NIR** Near-Infrared

**PA** Pixel Accuracy

**PIQE** Perceptual Image Quality Evaluator

**PSNR** Peak Signal-to-Noise Ratio

**PV** Photovoltaic

**RGB** Red, Green, Blue

**RMSE** Root Mean Square Error

**ROI** Region of Interest

**RS-ESRGAN** Remote Sensing Enhanced Super Resolution Generative Adversarial Network

**S2DR3** Sentinel-2 Deep Resolution 3

**SAM** Segment-Anything Model

**SR** Super Resolution

**SRGAN** Super Resolution Generative Adversarial Network

**SSIM** Structural Similarity Index

**SWIR** Shortwave Infrared

*Abbreviations*

---

**UI** User Interface

**VGG** Visual Geometry Group

**WP4** Work Package 4

# List of Figures

1.1.	Fortiss logo . . . . .	2
2.1.	Sentinel-2 band combinations and spectral indices for different applications. Each combination highlights specific features of the Earth's surface that might be invisible in individual bands. Images adapted from [GIS19].	6
2.2.	ESRGAN's RRDBNet generator Adapted for Multi-Image Super-Resolution	12
2.3.	S2DR3 10× Super-Resolution Model outputs: (top) S2 RGB, (middle) S2 NDVI, (bottom) S2 B11,B8,B5 . . . . .	14
2.4.	"Examples of low- and high-resolution imagery. Higher resolution imagery allows more detailed spatial patterns to be viewed in the imagery"[U.S24]. . . . .	15
2.5.	ASPP block used in DeepLabV3+. A $1 \times 1$ convolution, three $3 \times 3$ atrous convolutions with dilation rates 6, 12 and 18, and a global-image pooling branch are concatenated to fuse multi-scale context before decoding. <i>Adapted from [Che+18].</i> . . . . .	16
3.1.	Original Sentinel-2 images used for evaluation across different German states. These images show photovoltaic installations from Brandenburg (3 sites), Mecklenburg-Vorpommern, Saxony-Anhalt, and Lower Saxony, captured during the summer period of 2024–2025 to minimize cloud cover and atmospheric distortion. Coordinates represent the center point of each image tile, demonstrating geographic diversity across northern and eastern Germany. Detailed geospatial metadata for the Brandenburg dataset is provided in Appendix A.2.	24
3.2.	LabelMe GUI used for manual polygon annotation of PV arrays. . . . .	25
3.3.	SR Real-ESRGAN Image. . . . .	26
3.4.	Binary Label Mask. . . . .	26
3.5.	Super-Resolved Satlas Image. . . . .	27
3.6.	Binary Label Mask. . . . .	27
3.7.	User Interface: SR and segmentation flow. . . . .	29
3.8.	User Interface Demo with Large Image . . . . .	30
3.9.	User Interface Parameter Customizations for Large Images . . . . .	31
3.10.	User Interface Demo with Image After Parameter Customization . . . . .	32

---

*List of Figures*

---

3.11. Boundary Metrics Information for the Segmentation . . . . .	33
3.12. Multi-Resolution Segmentation Comparison . . . . .	34
3.13. Multi-Resolution Segmentation Difference Map . . . . .	35
3.14. ROI 1 — Original Sentinel-2 image and ESRGAN super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	37
3.15. ROI 2 — Original Sentinel-2 image and ESRGAN super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	37
3.16. ROI 1 — Original Sentinel-2 image and Satlas super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	38
3.17. ROI 2 — Original Sentinel-2 image and Satlas super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	38
3.18. ROI 1 — Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	39
3.19. ROI 2 — Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg, 13°41'15.66"E, 52°39'15.04"N . . . . .	39
3.20. Original Sentinel-2 image and S2DR3 super-resolved image of Brandenburg, 14°13'52.52"E, 52°41'51.39"N . . . . .	40
3.21. Best Segmentation Results for Sets 1-3 . . . . .	47
3.22. Best Segmentation Results for Sets 4-6 . . . . .	48
4.1. Example of poor sun reflection in Set 3 (Brandenburg 2). . . . .	51
4.2. Multispectral comparison: Original vs S2DR3 10x Super-Resolution for Set 3 (Brandenburg 2) . . . . .	54
B.1. Super-resolution comparison for image set 1. . . . .	63
B.2. Super-resolution comparison for image set 2. . . . .	64
B.3. Super-resolution comparison for image set 3. . . . .	65
B.4. Super-resolution comparison for image set 4. . . . .	66
B.5. Super-resolution comparison for image set 5. . . . .	67
B.6. Super-resolution comparison for image set 6. . . . .	68
C.1. Set 1 (Lower Saxony) - Complete Segmentation Results . . . . .	70
C.2. Set 2 (Brandenburg 1) - Complete Segmentation Results . . . . .	71
C.3. Set 3 (Brandenburg 2) - Complete Segmentation Results . . . . .	72
C.4. Set 4 (Mecklenburg-Vorpommern) - Complete Segmentation Results . .	73
C.5. Set 5 (Brandenburg 3) - Complete Segmentation Results . . . . .	74
C.6. Set 6 (Saxony-Anhalt) - Complete Segmentation Results . . . . .	75

# List of Tables

2.1. Characteristics of ESA Sentinel-2A and -2B satellite images [Eur14; CG13].	8
3.1. Individual Set Super-Resolution Results . . . . .	41
3.2. Average Super-Resolution Results Across All Sets. "RMSE ( $\downarrow$ = lower is better)" . . . . .	41
3.3. Per-set BRISQUE and PIQE scores for Satlas SR model. Lower scores indicate better perceptual quality. . . . .	42
3.4. Summary statistics for BRISQUE and PIQE across all Satlas image sets. Lower values denote better perceptual quality. . . . .	42
3.5. BRISQUE and PIQE scores across super-resolution models (Real-ESRGAN, S2DR3, Satlas) for each evaluation set. Lower scores indicate better perceptual quality. (The best scores per row are highlighted) . . . . .	43
3.6. Segmentation Performance Across All Sets. (The best scores per row are highlighted in bold) . . . . .	45

# Listings

A.1. Script to convert LabelMe JSON annotations to PNG masks . . . . .	58
A.2. gdalinfo output for Brandenburg dataset (set 2.tif) . . . . .	59

# Bibliography

- [Akh23] Y. Akhtman. *Sentinel-2 Deep Resolution 3.0 (S2DR3)*. [https://medium.com/@ya\\_71389/sentinel-2-deep-resolution-3-0-c71a601a2253](https://medium.com/@ya_71389/sentinel-2-deep-resolution-3-0-c71a601a2253). 2023.
- [All23] Allen Institute for AI. *Satlas Super Resolution: Training and Inference Code*. <https://github.com/allenai/satlas-super-resolution>. Accessed 10 May 2025. 2023.
- [AML23] H. Al-Mekhlafi and S. Liu. “Single image super-resolution: a comprehensive review and recent insight.” In: *Frontiers of Computer Science* 18 (2023), pp. 1–18. doi: 10.1007/s11704-023-2588-9.
- [Arm+21] S. E. Armannsson, M. O. Ulfarsson, J. Sigurdsson, H. V. Nguyen, and J. R. Sveinsson. “A Comparison of Optimized Sentinel-2 Super-Resolution Methods Using Wald’s Protocol and Bayesian Optimization.” In: *Remote Sensing* 13.11 (2021). issn: 2072-4292. doi: 10.3390/rs13112192.
- [BM18] Y. Blau and T. Michaeli. “The Perception-Distortion Tradeoff.” In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, June 2018, 6228–6237. doi: 10.1109/cvpr.2018.00652.
- [CG13] J. G. Clevers and A. A. Gitelson. “Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3.” In: *International Journal of Applied Earth Observation and Geoinformation* 23 (2013), pp. 344–351. doi: 10.1016/j.jag.2012.10.008.
- [Che+18] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” In: *CoRR* abs/1802.02611 (2018). arXiv: 1802.02611.
- [Don+15] C. Dong, C. C. Loy, K. He, and X. Tang. *Image Super-Resolution Using Deep Convolutional Networks*. 2015. arXiv: 1501.00092 [cs.CV].
- [Don+23] R. Dong, L. Zhang, W. Li, S. Yuan, L. Gan, J. Zheng, H. Fu, L. Mou, and X. X. Zhu. “An adaptive image fusion method for Sentinel-2 images and high-resolution images with long-time intervals.” In: *International Journal of Applied Earth Observation and Geoinformation* 121 (2023), p. 103381. issn: 1569-8432. doi: <https://doi.org/10.1016/j.jag.2023.103381>.

## Bibliography

---

- [Eur14] European Space Agency (ESA). *Sentinel-2 Missions-Sentinel Online*. Paris, France, 2014.
- [Eur23] European Space Agency (ESA). *Sentinel-2 Data Collection*. Copernicus Dataspace. 2023. url: <https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2> (visited on 09/30/2023).
- [For23] Fortiss. *NEED: New data for the energy transition*. Tech. rep. Unpublished proposal. Fortiss, 2023, p. 92.
- [GIS19] GISGeography. *Sentinel-2 Bands and Combinations*. GISGeography. 2019. URL: <https://gisgeography.com/sentinel-2-bands-combinations/> (visited on 03/20/2024).
- [Goo+14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].
- [Guo+23] Z. Guo, Z. Zhuang, H. Tan, Z. Liu, L. Peiran, Z. Lin, W. Shang, H. Zhang, and J. Yan. “Accurate and generalizable photovoltaic panel segmentation using deep learning for imbalanced datasets.” In: *Renewable Energy* 219 (2023). doi: 10.1016/j.renene.2023.119471.
- [Hou+19] X. Hou, B. Wang, W. Hu, L. Yin, and H. Wu. “SolarNet: A Deep Learning Framework to Map Solar Power Plants in China From Satellite Imagery.” In: *arXiv e-prints* (2019). eprint: 1912.03685.
- [HZ10] A. Horé and D. Ziou. “Image Quality Metrics: PSNR vs. SSIM.” In: *2010 20th International Conference on Pattern Recognition*. 2010, pp. 2366–2369. doi: 10.1109/ICPR.2010.579.
- [Jia+21] H. Jiang, L. Yao, N. Lu, J. Qin, T. Liu, Y. Liu, and C. Zhou. “Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery.” In: *Earth System Science Data* 13.11 (2021), pp. 5389–5401. doi: 10.5194/essd-13-5389-2021.
- [Kru+21] L. Kruitwagen, K. Story, J. Friedrich, L. Byers, S. Skillman, and C. Hepburn. “A global inventory of photovoltaic solar energy generating units.” In: *Nature* 598 (2021), pp. 604–610. doi: 10.1038/s41586-021-03957-7.
- [Lan+18] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler. “Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (Dec. 2018), 305–319. issn: 0924-2716. doi: 10.1016/j.isprsjprs.2018.09.018.

---

*Bibliography*

---

- [Led+16] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." In: *CoRR* abs/1609.04802 (2016). arXiv: 1609.04802.
- [LS24] S. C. Lambertenghi and A. Stocco. "Assessing Quality Metrics for Neural Reality Gap Input Mitigation in Autonomous Driving Testing." In: *2024 IEEE Conference on Software Testing, Verification and Validation (ICST)*. 2024, pp. 173–184. doi: 10.1109/ICST60714.2024.00024.
- [Mat+24] J. Matar, M. Albrecht, A. Kandekari, F. Kotthoff, and M. Duchon. "Mapping Solar Potential: Comprehensive PV System Detection with Aerial and Satellite Images Using SolarSAM." In: *2024 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 2024, pp. 33–39. doi: 10.1109/SmartGridComm60555.2024.10738071.
- [MMB12] A. Mittal, A. K. Moorthy, and A. C. Bovik. "No-Reference Image Quality Assessment in the Spatial Domain." In: *IEEE Transactions on Image Processing* 21.12 (2012), pp. 4695–4708. doi: 10.1109/TIP.2012.2214050.
- [N+15] V. N. P. D. M. C. Bh, S. S. Channappayya, and S. S. Medasani. "Blind image quality evaluation using perception based features." In: *2015 Twenty First National Conference on Communications (NCC)*. 2015, pp. 1–6. doi: 10.1109/NCC.2015.7084843.
- [Phi+20] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage. "Sentinel-2 Data for Land Cover/Use Mapping: A Review." In: *Remote Sensing* 12.14 (2020), p. 2291. doi: 10.3390/rs12142291.
- [Ros+24] L. Rossi, V. Bernuzzi, T. Fontanini, M. Bertozzi, and A. Prati. *Swin2-MoSE: A New Single Image Super-Resolution Model for Remote Sensing*. 2024. arXiv: 2404.18924 [cs.CV].
- [SRMV20] L. Salgueiro Romero, J. Marcello, and V. Vilaplana. "Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks." In: *Remote Sensing* 12.15 (2020), p. 2424. doi: 10.3390/rs12152424.
- [Su+25] H. Su, Y. Li, Y. Xu, X. Fu, and S. Liu. "A review of deep-learning-based super-resolution: From methods to applications." In: *Pattern Recognition* 157 (2025), p. 110935. issn: 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2024.110935>.
- [SZ15] K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].

## Bibliography

---

- [TH15] A. A. Taha and A. Hanbury. “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.” In: *BMC medical imaging* 15 (2015), pp. 1–28.
- [TL19] M. Tan and Q. V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” In: *ICML*. 2019, pp. 6105–6114.
- [U.S18] U.S. Geological Survey (USGS). *USGS EROS Archive – National Agriculture Imagery Program (NAIP)*. <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-national-agriculture-imagery-program-naip>. Accessed 1 April 2025. 2018.
- [U.S24] U.S. Geological Survey. *Introduction to Aquatic Remote Sensing—Resolution Types*. <https://www.usgs.gov/centers/western-geographic-science-center/introduction-aquatic-remote-sensing>. Defines spatial resolution and ground-sample distance (GSD); accessed 11 May 2025. 2024.
- [Wan+04] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. “Image quality assessment: from error visibility to structural similarity.” In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. ISSN: 1941-0042. DOI: 10.1109/TIP.2003.819861.
- [Wan+18] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang. “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks.” In: *CoRR* abs/1809.00219 (2018). arXiv: 1809.00219.
- [Wan+21] X. Wang, L. Xie, C. Dong, and Y. Shan. *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*. 2021. arXiv: 2107.10833 [eess.IV].
- [Wan+24] X. Wang, Z. Ao, R. Li, Y. Fu, Y. Xue, and Y. Ge. “Super-Resolution Image Reconstruction Method between Sentinel-2 and Gaofen-2 Based on Cascaded Generative Adversarial Networks.” In: *Applied Sciences* 14.12 (2024), p. 5013. DOI: 10.3390/app14125013.
- [WBK23] P. Wolters, F. Bastani, and A. Kembhavi. “Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing.” In: *arXiv preprint arXiv:2311.18082* (2023).
- [Yu+18] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal. “DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States.” In: *Joule* 2.12 (2018), pp. 2605–2617. DOI: 10.1016/j.joule.2018.11.021.

## Bibliography

---

- [Zou+04] K. Zou, S. Warfield, A. Bharatha, C. Tempany, M. Kaus, S. Haker, W. Wells, F. Jolesz, and R. Kikinis. “Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index.” In: *Academic radiology* 11 (Feb. 2004), pp. 178–89. doi: 10.1016/S1076-6332(03)00671-8.
- [ZZW20] L. Zhuang, Z. Zhang, and L. Wang. “Automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method.” In: *Applied Soft Computing* 92 (2020), p. 106283. doi: 10.1016/j.asoc.2020.106283.