SENTINELKILNDB: A Large-Scale Dataset and Benchmark for OBB Brick Kiln Detection in South Asia Using Satellite Imagery

Rishabh Mondal[1], Jeet Parab[2], Heer Kubadia[1], Shataxi Dubey[1], Shardul Junagade[1], Zeel B Patel[1], Nipun Batra[1]

[1]IIT Gandhinagar, India, [2]IIIT Surat, India

Abstract

Air pollution was responsible for 2.6 million deaths across South Asia in 2021 alone, with brick manufacturing contributing significantly to this burden. In particular, the Indo-Gangetic Plain; a densely populated and highly polluted region spanning northern India, Pakistan, Bangladesh, and parts of Afghanistan sees brick kilns contributing 8–14% of ambient air pollution. Traditional monitoring approaches, such as field surveys and manual annotation using tools like Google Earth Pro, are time and labor-intensive. Prior ML-based efforts for automated detection have relied on costly high-resolution commercial imagery and non-public datasets, limiting reproducibility and scalability. In this work, we introduce SENTINELKILNDB, a publicly available, hand-validated benchmark of **62,671 brick kilns** spanning **three kiln types** Fixed Chimney Bull's Trench Kiln (FCBK), Circular FCBK (CFCBK), and Zigzag kilns - annotated with **oriented bounding boxes (OBBs)** across **2.8 million km^2** using free and globally accessible **Sentinel-2 imagery**. We benchmark state-of-the-art oriented object detection models and evaluate generalization across *in-region*, *out-of-region*, and *super-resolution* settings. SENTINELKILNDB enables rigorous evaluation of geospatial generalization and robustness for low-resolution object detection, and provides a new testbed for ML models addressing real-world environmental and remote sensing challenges at a continental scale. Datasets and code are available in SentinelKilnDB Dataset and SentinelKilnDB Benchmark, under the *Creative Commons Attribution–NonCommercial 4.0 International License*.

1 Introduction

Air pollution is responsible for seven million deaths annually. In 2021, 15% of all global deaths among children under five were linked to air pollution [60], with India alone accounting for 22% of these casualties [59]. Nearly 2.3 billion people across South-East Asia are exposed to hazardous levels of air quality. To understand and mitigate this impact, air quality researchers rely on physics-based chemical transport models, such as CAMx [1], which require accurate emission inventories [18]. Brick kilns are among the major contributors to ambient air pollution in the Indo-Gangetic Plain, accounting for 8-14% of total emissions in South Asia [67]. Beyond their environmental footprint, brick kilns are also socio-economically significant, employing an estimated 15 million workers, including children, thus intersecting directly with UN Sustainable Development Goal 8.7 on ending forced labor and child exploitation [42, 6]. In India, approximately 45% of brick production is concentrated in the Indo-Gangetic Plain (IGP), commonly referred to as the "brick belt" of the country. The region is home to three primary kiln types: Circular Fixed Chimney Bull's Trench Kilns (CFCBKs), Fixed Chimney Bull's Trench Kilns (FCBKs), and Zigzag kilns. FCBKs are the most prevalent, accounting for 70-75% of kilns, while Zigzag kilns make up 20-25%. Zigzag kilns are estimated to be up to

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Track on Datasets and Benchmarks.

40% more fuel-efficient than FCBKs, making them a target for cleaner technology transition. In Bangladesh, significant shifts have occurred with the number of Zigzag kilns rising from 150 to 4,247 and FCBKs declining from 4,500 to 2,373 between 2009 and 2017 [49]. Pakistan, the third-largest brick producer in Asia, had around 20,000 kilns as of 2021 [13], while Afghanistan primarily relies on traditional kiln methods [13]. Monitoring small, unregulated brick kilns through traditional field surveys is labor-intensive, resource-consuming, and inherently unscalable. To maintain emission inventories, air quality experts manually annotate satellite imagery. According to domain experts, scanning and labeling an area of 3,600 km^2 can take 6–8 hours, implying over 7,000 hours of manual effort to cover India. The dynamic nature of kiln locations further complicates repeat surveys. Recent studies [29] have explored transfer learning using pretrained CNNs such as VGG16 [22], ResNet [21], and EfficientNet [53] for brick kiln classification from satellite imagery. However, classification approaches require an additional post-processing step to localize kilns spatially. In contrast, object detection models provide an end-to-end solution for both identification and geo-localization. Prior work [35, 6] has used axis-aligned object detectors for this task, but has not leveraged recent advances in oriented bounding box (OBB) detection. OBBs allow for more accurate estimation of kiln area and spatial extent—critical for downstream tasks such as estimating production capacity and emission factors. In our previous research work, *Space to Policy* [38], we detected 30,638 brick kilns using Planet imagery [41] across five Indo-Gangetic states, covering an area of 520K km^2. We did not release the dataset due to Planet Labs' policy, which is open to educational and research purposes but limited to non-commercial and non-shareable [58, 40]. Our paper focuses on building a large-scale dataset of brick kilns in the Indian subcontinent with oriented bounding box (OBB) labels in low-resolution Sentinel-2 imagery. We use satellite images from Sentinel-2 (10 m/pixel), covering a total area of approximately **2.8 M km^2**. We begin with the dataset from our previous work [38], which identified 30,638 brick kilns across five Indian states using Planet Imagery [41]. We trained the YOLOv11L-OBB model [27] and predicted in our area of interest. We manually validated the model's predictions using Esri Wayback imagery [2]. Our brick kiln detection pipeline follows an iterative process consisting of: (i) training, (ii) prediction, (iii) manual validation of the predicted labels, and (iv) appending the validated results to the existing dataset. Through four rounds of this process, we expanded the dataset to **62,671 brick kilns** across India, Bangladesh, Pakistan, and Afghanistan. For accurate label creation, we utilized moderate-resolution Planet imagery [39] and mapped the identified brick kilns onto corresponding Sentinel-2 imagery to compile the final dataset. To the best of our knowledge, no prior work has provided oriented bounding box (OBB) annotations on low-resolution satellite imagery with hand validation. We release the **Sentinel-2 imagery with OBB labels** in YOLO OBB format, axis-aligned (AA) labels in YOLO and DOTA [70] formats.

To assess the utility of SENTINELKILNDB for real-world deployment and model development, we benchmark several state-of-the-art object detection models. Our evaluation spans multiple tasks: (i) *in-region detection*, where models are tested on regions included in the training distribution; (ii) *out-of-region generalization*, to assess robustness on geographically unseen areas; (iii) *temporal generalization*, where we evaluate robustness to seasonal dynamics using data from all four seasons in the Lucknow airshed: Winter (W), Pre-Monsoon (PM), Monsoon (M), and Post-Monsoon (PoM); and (iv) *super-resolution adaptation*, where we evaluate model performance on enhanced-resolution imagery to quantify the gains achievable through super-resolution techniques. This multi-task evaluation framework enables a rigorous analysis of both detection performance and geospatial generalization, critical for scalable and transferable applications in low-resolution remote sensing. Our dataset has broad utility across both environmental science and machine learning research. From a domain perspective, the oriented bounding box (OBB) annotations allow for precise estimation of kiln area, which can be used to **estimate emissions at the individual kiln level** and enhance regional emission inventories. This supports more accurate air quality modeling and policy interventions. The main contributions of this paper are:

1. We introduce SENTINELKILNDB, a Sentinel-2 benchmark containing bounding boxes of 62,671 brick kilns across South Asia to facilitate training and evaluation of object detection models on satellite imagery (see Section 4.4).

2. We provide a benchmark of 28 state-of-the-art object detection models developed over the past decade, along with standard evaluation (see Table 3).

3. We show that pre-training remote sensing foundation models on satellite imagery yields significant gains over prior models on benchmarks and downstream tasks (see Table 3).

4. We present insights into out-of-region generalization, seasonal dynamics, and super-resolution performance (see Section 4.4).

Our work is fully reproducible. All code and experimental details are available in our GitHub repository[1]. Our dataset and metadata are available on Kaggle[2] under the *Creative Commons Attribution–NonCommercial 4.0 International License* [10].

2 Background & Related work

In this section, we provide a background on brick kilns and the pertinent relevant work.

2.1 Brick Kilns

Approximately 45% of India's brick production is concentrated in the Indo-Gangetic Plain (IGP), one of South Asia's most densely populated and polluted regions [56]. Kilns are mainly CFCBK (oldest, fuel-intensive, high emissions), FCBK (newer, more efficient), and Zigzag (natural/forced draft), with Zigzag cutting energy use and emissions by $\approx 40\%$ compared to FCBK but adopted by only 25–30% of kilns [56]. In India, FCBKs produce 60–70% of bricks, with more than 30,000 kilns across the IGP [4]. Bangladesh has a high share of Zigzag kilns due to policy push, while Afghanistan and Pakistan (especially Punjab) remain FCBK-dominant but are transitioning to Zigzag.

2.2 Related Work

Object Detection for Aerial Imagery: Object detection in aerial imagery has recently captured attention in the object detection community. It is hard for two main reasons: i) targets are often sparse and non-uniformly distributed, making detection inefficient, and ii) focal objects like pedestrians are very small and blend into the surrounding backgrounds unlike natural images in MS COCO dataset [32] where objects are easily identifiable. Our dataset has similar challenges and thus falls within the area of 'object detection from aerial imagery'.

Remote Sensing OBB Datasets: Popular remote sensing datasets such as VEDAI, HRSC2016, DOTA, FGSD, and DIOR-R [43, 33, 70, 7, 8] primarily focus on object categories like vehicles, ships, and urban infrastructure. However, they give limited attention to industrial emissions—an increasingly critical issue in the context of global warming and climate policy. Our dataset fills this gap by uniquely representing large-scale, unorganized sources of air pollution such as brick kilns.

Oriented Object Detection: Detecting objects with arbitrary orientations (Oriented Bounding Boxes (OBB) remains a challenging task in remote sensing. In recent years, significant advancements have been made in object detection, including notable progress in oriented object detection [65, 76, 14, 19, 85, 83]. However, most existing methods benchmark their performance on high-resolution datasets such as HRSC-2016 [33] and DOTA [70]. While these datasets are valuable, their high resolution makes it comparatively easier to identify object orientations. In contrast, detecting orientation in low-resolution imagery is more difficult due to reduced spatial detail. Our dataset provides a valuable benchmark for advancing research in *Oriented Object Box (OBB) Detection in Low-Resolution Remote Sensing Imagery*.

Brick Kiln Detection from Satellite Imagery: Machine learning techniques applied to satellite imagery have been utilized to identify brick kilns in South Asian regions. A recent work [35] employs a gated neural network that decouples classification and object detection tasks to identify brick kilns. The study utilizes a deep learning architecture inspired by Inception-ResNet [52] and the You Only Look Once (YOLO) [44] object detector to locate brick kilns and evaluate it over a 3300 km^2 area. In contrast, we evaluate our models on more than 2.8M km^2 region. A study by Lee et al. (2021) for brick kiln classification in Bangladesh, identifying 3,345 non-geotagged brick kilns [29]. A recent work [6] uses YOLOv3 axis-aligned bounding box model to detect the FCBK kilns in the Indo-Gangetic plain. Another notable work, SustainBench [81], is a benchmark for kiln detection using Sentinel-2 imagery. It framed the task as tile-level classification over 64×64×13 patches (6,329 positives, 67,284 negatives; Bangladesh, 2018–19). In contrast, our dataset provides oriented bounding boxes, capturing the direction and shape of each kiln accurately with positive samples.

[1] https://github.com/rishabh-mondal/SENTINELKILNDB_NeurIPS_2025

[2] https://www.kaggle.com/datasets/rishabhsnip/sentinelkiln-dataset

ESA WorldCover v200: For negative examples, we curated 41,068 Sentinel-2 tiles using ESA WorldCover v200 [84], selecting land-cover–pure regions with at least 99.90% class confidence across eight relevant classes. All candidate tiles were manually verified against Esri basemaps, and any image containing visible kilns was discarded.

3 SentinelKilnDB Dataset

This section outlines the dataset construction process and key statistics.

3.1 Imagery Sources

Sentinel Imagery: Sentinel-2 [66] provides optical imagery at 10–60 m resolution across 13 spectral bands. Each tile spans 110×110 km [12]. Using Google Earth Engine (GEE), we accessed the Sentinel-2 Surface Reflectance product (`COPERNICUS/S2_SR_HARMONIZED`), retrieving bands B4 (Red), B3 (Green), and B2 (Blue) at 10 m resolution with <1% cloud cover and masking clouds via QA60. We focused on November 2023–February 2024, when brick kilns are most active and contrast with surrounding farmland is highest. We empirically set a patch size of 128×128 pixels with a 0.0027° ($\approx$30-pixel) sliding window (overlaps) to capture edge kilns and later deduplicated by geographic coordinates. The pipeline first downloads $11{,}000 \times 11{,}000$-pixel Sentinel-2 tiles, then splits them into overlapping 128×128 patches, yielding the image samples shown in Figure 1.

Planet Imagery: We utilize satellite imagery from Planet Labs [28], which offers high-cadence imagery products 4.77-meter monthly and quarterly mosaics[3]. The imagery is accessed through Planet's Research and Education Program and downloaded via the Mosaics API [3]. To preserve object integrity, large 4096×4096 images are split into overlapping 640×640 crops with 64-pixel overlap. We use imagery from Q1 2024 to ensure cloud-free conditions during peak kiln activity.

Esri Wayback Imagery: Esri provides high-resolution (up to 30 cm) timestamped basemaps via a Web Map Service (WMS) [16]. We used imagery same time window to match the Planet Q1 mosaics. Due to WMS limitations, this imagery was not used for training and does not support bulk downloads.

ESA WorldCover v200: For negative examples, we curated 41,068 Sentinel-2 tiles using ESA WorldCover v200 [84], selecting land-cover–pure regions with at least 99.90% class confidence across eight relevant classes. All candidate tiles were manually verified against Esri basemaps, and any image containing visible kilns was discarded. Representative negative sample images are shown in Figure 15 in the Appendix.

3.2 Label Creation Pipeline

Inter-Annotator Agreement to Quantify Labeling Uncertainty: We did not compute formal inter-annotator agreement, but followed a quality-control workflow inspired by AI-TOD [74, 38]. *S1)* Experts created a seed set and a standardized manual; *S2)* trained volunteers annotated accordingly; *S3)* one annotator screened boxes and a second verified tightness and orientation; *S4)* doubtful cases were flagged; *S5)* an air-quality expert adjudicated flags and assigned final labels. While not every image had multiple annotators, all uncertain cases underwent expert review to ensure consistency and reduce labeling uncertainty.

Scalable Brick Kiln Mapping through Data Collection and Validation across South Asia In our previous research, *Space to Policy* [38], we identified and annotated 30,638 brick kilns across five Indo-Gangetic states (Punjab, Bihar, Haryana, Uttar Pradesh, and West Bengal) using Planet imagery. We trained a model on four manually labeled small airsheds and applied it to predict kilns across all five states. The bounding boxes and kiln types were then manually validated using Esri Wayback Imagery [2]. For the dataset creation process in this paper, we began with the 30,638 labeled brick kilns and trained our model on that dataset. Using Planet imagery, we subsequently predicted kiln locations in Rajasthan, Gujarat, Jharkhand, Assam, and Delhi, as well as in three neighboring South Asian countries (Bangladesh, Pakistan, and Afghanistan), covering 54.76% of South Asia. To ensure the accuracy of the predicted oriented bounding boxes (OBBs) and kiln type classifications, each prediction was manually validated using a custom-developed user interface built with Leafmap [69] for Esri Imagery visualization (see Figure 8). We repeated the brick kiln detection pipeline, namely (i) training, (ii) prediction, (iii) manual validation, and (iv) appending to the existing data, three times.

[3] `https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-a-mosaic.htm`

As a result, we expanded our dataset to **62,671** brick kilns, categorized into three types: CFCBK (**1,944**), FCBK (**33,963**), and Zigzag (**26,764**). The complete dataset is shown in Figure 2, with high-resolution examples included in Appendix Figure 3 [38].

Iteration Limits and Manual Validation Effort We limited the brick-kiln detection pipeline to three iterations based on diminishing returns, the cost of manual validation, and external validation indicating dataset completeness. Table 1 uses the following notation: M_0^K is the baseline kiln count from our prior study [38]; M_i^{NPP} denotes the newly predicted positives in pass i; ΔM_i^K is the subset confirmed by manual validation. The cumulative confirmed kilns after pass i are $M_i^K = M_{i-1}^K + \Delta M_i^K$. Finally, M_i^T denotes the manual-validation time in pass i, which scales with M_i^{NPP}. As shown in Table 1, each successive iteration of the detection pipeline produced fewer

Table 1: Progress across three iterations of the brick kiln detection pipeline, showing predicted positives, confirmed kilns, and manual hand-validation effort.

Country	M_0^K	M_1^{NPP}	ΔM_1^K	M_2^{NPP}	ΔM_2^K	M_3^{NPP}	ΔM_3^K	M_1^T	M_2^T	M_3^T	M_3^K	Total T
India	30,638	8,799	7,105	6,542	2,985	8,554	2,254	36h 08m	27h 17m	35h 38m	42,982	99h 03m
Bangladesh	0	6,399	6,200	599	296	692	407	26h 40m	2h 30m	2h 10m	6,903	31h 20m
Pakistan	0	8,786	8,051	3,334	1,417	6,534	2,709	36h 37m	13h 57m	27h 13m	12,177	77h 47m
Overall	30,638	23,984	21,356	10,475	4,698	15,780	5,370	99h 25m	43h 44m	65h 01m	62,062	208h 10m

newly confirmed kilns per predicted positive, reflecting diminishing returns. For example, in India the number of newly confirmed kilns decreased from $\Delta M_1^K = 7{,}105$ in the first pass to $\Delta M_3^K = 2{,}254$ in the third, while in Pakistan the decrease was from $\Delta M_1^K = 8{,}051$ to $\Delta M_3^K = 2{,}709$. Overall yield fell from approximately 0.81 kilns per predicted positive in round 1 to 0.34 in round 3, indicating saturation. Each predicted box also underwent a two-stage manual review with expert adjudication for edge cases. This rigorous process ensured near-zero false positives (as discussed under sampling bias) but required substantial human effort, with 62,062 instances manually verified over 208 hours, underscoring our emphasis on precision and auditability. Finally, strong external validation further supports the completeness of our dataset: prior work [38] showed that our district-level kiln counts in Uttar Pradesh align closely with official data from the 2023 Uttar Pradesh Pollution Control Board survey (19,671 kilns), with a Pearson correlation of $r = 0.94$, suggesting low false negatives and high coverage.

3.3 OBB Label Transfer from Planet to Sentinel-2

After completing the label creation, we overlaid the Planet imagery labels onto the cropped, geo-referenced Sentinel-2 images. Planet imagery is in the Web Mercator projection (EPSG:3857), while Sentinel-2 uses the Universal Transverse Mercator (UTM) projection (EPSG:326XX, depending on the zone). To align the datasets, we reprojected the oriented bounding box (OBB) coordinates from Web Mercator to the corresponding UTM zone of the Sentinel-2 tiles using spatial reference transformation based on geolocation metadata. To ensure accurate alignment, we performed both visual inspections using Geemap [68] and quantitative validation using Intersection over Union (IoU) metrics. Note that one can extract geo-referenced bounding boxes of brick kilns from our dataset and overlay them on any geo-referenced satellite imagery, regardless of its resolution.

Figure 1: A few samples of Sentinel-2 imagery with Oriented Bounding Boxes (OBBs) around brick kilns. CFCBK (Circular Fixed Chimney Bull's Trench Kiln) is marked in red, FCBK (Fixed Chimney Bull's Trench Kiln) in blue, and Zigzag kilns in green.

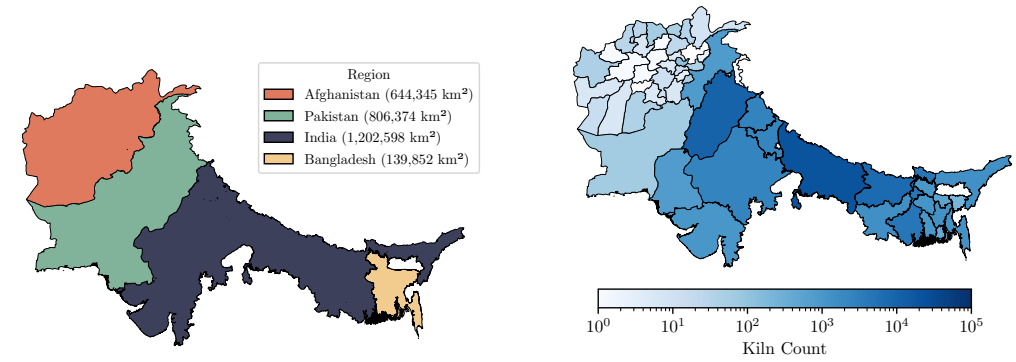
Table 2: Physically present, unique class-wise kiln counts by country/region (left) and dataset split summary (right). Image files are `.png`; label files are `.txt`. `BBoxes` denotes the total number of bounding boxes (the lines across label files).

Country	State/Region	CFCBK	FCBK	Zigzag	Total	Neg. Samples
India	9 states	1,939	21,451	19,592	42,982	30,650
Bangladesh	8 divisions	2	1,461	5,440	6,903	5,548
Pakistan	4 provinces	3	10,443	1,731	12,177	3,970
Afghanistan	34 provinces	0	608	1	609	900
Total	–	**1,944**	**33,963**	**26,764**	**62,671**	**41,068**

Split	Images (.png)	Label files (.txt)	No. of BBoxes
Train	71,856	47,214	63,787
Val	23,952	15,738	21,042
Test	18,492	10,278	12,819
Total	**114,300**	**73,239**	**97,648**

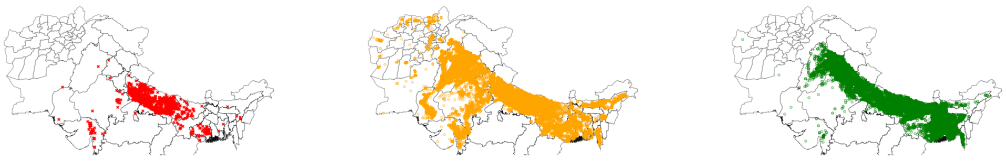
3.4 Overall Dataset Statistics

Table 2 summarizes the distribution of brick kilns physically present across South Asia—India, Bangladesh, Pakistan, and Afghanistan(see Appendix Figs. 4, 5, 6, 7)—totaling 62,671 annotated kilns. India dominates the dataset, contributing over 42,000 kilns, primarily from Uttar Pradesh, Bihar, and West Bengal. The distribution highlighted a regional preference for different kiln types: India shows a relatively balanced spread among FCBK and Zigzag kilns, while CFCBKs remain sparse. In contrast, Bangladesh exhibits a strong dominance of Zigzag kilns, reflecting the country's regulatory push toward cleaner technologies. Pakistan, particularly Punjab (Pak), reports a high number of FCBKs, while Afghanistan's entries are minimal and largely limited to FCBKs. This diverse geographical and typological spread underlines the need for tailored policy and technological interventions across different regions to effectively monitor and regulate kiln emissions.



(a) Coverage of our dataset over various countries

(b) Brick kiln count per state



(c) CFCBK

(d) FCBK

(e) Zigzag

Figure 2: Spatial distribution of brick kilns across South Asia. (a) Dataset extent covering Afghanistan, Pakistan, India, and Bangladesh. (b) State-wise kiln counts (log scale). (c–e) Distribution by type: CFCBKs concentrated in Uttar Pradesh, Bihar, and West Bengal; FCBKs widespread; Zigzags absent in Afghanistan.

3.5 Dataset Format

We release the Sentinel-2 patches of size 128×128 with 10 meter-per-pixel native resolution and corresponding bounding boxes as the dataset. The images are provided in .png format and are saved with `lat,lon.png` naming convention where `lat` and `lon` are the center coordinates of the images. The labels are provided in both OBB and AABB formats, following the YOLO [45], DOTA [71]

format. The labels are saved in `lat,lon.txt` format to match the corresponding `lat,lon.png` image. We will release our dataset under CC BY NC 4.0 license.

4 SentinelKilnDB Benchmark

In this section, we discuss the benchmarking of models on our dataset across different tasks.

4.1 Benchmarks

`T1`**: In-region detection** We benchmarked detectors for localizing and segregation of kilns in 10 m Sentinel-2 imagery, where a 300 m kiln spans 30 pixels. We fine-tuned one-stage, two-stage, transformer-based, and remote-sensing foundation models with Faster R-CNN heads for OBB and AA variants. One-/two-stage detectors and Faster R-CNN–based foundation models require NMS to prune duplicates; DETR-style transformers output a fixed set without NMS. All models were evaluated using their standard training inference pipelines.

`T2`**: Out-of-the-region detection** Geographic domain adaptation is an important experiment in remote sensing data, as the RGB distribution in satellite imagery can change across different regions. After we identified the best-performing models on our dataset, we tested how well these models work in different geographic regions. We first trained and tested the models on the same region to measure in-region performance. Then, we trained the models on one region and tested them on a different region to measure out-of-region performance.

`T3`**: Temporal Generalization** Seasonal variation plays a crucial role in remote sensing analysis, as land-cover appearance and atmospheric conditions change throughout the year. To assess model robustness under such temporal dynamics, we designed two complementary setups: (i) a leave-one-season-out (LOSO) experiment using data from Winter (W), Pre-Monsoon (PM), Monsoon (M), and Post-Monsoon (PoM), and (ii) a Winter-trained model (the peak brick kiln activity period) evaluated across the remaining seasons. Together, these setups capture the challenge of achieving reliable, year-round monitoring of brick kilns under diverse seasonal conditions.

`T4`**: Super resolution** Sentinel-2 imagery is low resolution. We applied super-resolution methods using only pretrained models at inference on our Sentinel-2 data (see Appendix Figure 16). This addresses resolution limits without requiring costly high-resolution commercial imagery. We then evaluated the best-performing detector on the enhanced images to test whether super resolution improved detection performance.

4.2 Models

In this section, we discuss about various models to perform benchmark on our dataset. We chose the set of models as per the following criteria: i) OBB v/s AA bounding box support; ii) two-stage methods; iii) one-stage methods; iv) DETR-based methods; and v) remote sensing foundation models. Below, we provide a brief summary of each model used for benchmarking.

`One-Stage Methods:` One-stage oriented object detection methods provide efficient frameworks for localizing objects while addressing the challenges of angle discontinuity and annotation costs. Phase-Shifting Coder (PSC) [82] reformulates angle prediction as a periodic problem, using discrete circular representations to predict accurate oriented bounding box (OBB) detection. H2RBox [79] leverages only horizontal box supervision, bridging the gap between easy annotation and precise localization through hybrid representations. RoI Transformer [15] further improves alignment by learning transformations that match rotated regions of interest with objects, thereby resolving misalignment between horizontal proposals and rotated targets.

`Two-Stage Methods:` Two-stage detectors generate region proposals in the first stage and refine them in the second for classification and bounding box regression. Rotated FCOS [55] is an anchor-free detector that predicts object angles together with center points and sizes, while DCFL [73] adopts a two-stage learning strategy that first identifies coarse locations and then refines them. CSL [75] reformulates angle regression as classification with circular smooth labels to mitigate periodicity issues. Rotated RetinaNet [31] builds on an anchor-based framework by incorporating orientation prediction for rotated object detection. GWD [77] introduces a loss based on Gaussian Wasserstein Distance to capture geometric differences between oriented boxes. R^3Det [76] progressively refines region proposals with rotation-aware losses, while S^2A-Net [20] integrates spatial feature alignment and angle-sensitive mechanisms. More recent architectures include ConvNeXt [34], which adapts con-

volutional networks with transformer design elements for dense prediction tasks; YOLO-World [9], which enables open-vocabulary detection via vision-language pretraining; YOLOv12L [54], optimized for real-time high-resolution inference; and YOLO-OBB [25], which extends YOLO to oriented bounding box detection by predicting object rotations. Recent loss-based enhancements such as KFIoU [80] and KLD [78], further improve orientation-aware regression by incorporating geometric and probabilistic modeling techniques.

`DETR-Based Methods:` DETR-based detectors formulate object detection as a direct set prediction problem, where a fixed number of object queries attend to image features through transformers to produce bounding boxes and classifications without the need for anchors or non-maximum suppression. Several extensions refine this framework. DETA [89] modifies object-query matching to stabilize training and improve accuracy without complex post-processing. RT-DETR [88] combines convolutional backbones with lightweight transformers to enable real-time detection. RF-DETR [47] adopts a region-focused strategy that enhances query assignment, improving performance on dense and small-object detection tasks.

`Remote Sensing Foundation Models:` Foundation models for remote sensing aim to learn transferable representations from large-scale, multi-modal Earth observation data. SatMAE [11] extends masked autoencoders to temporal and multi-spectral imagery. TerraMind [24] introduces dual-scale generative modeling across nine modalities for zero- and few-shot transfer. Galileo [57] uses masked modeling with contrastive losses for multi-scale feature learning. DOFA [72] employs a dynamic hypernetwork to integrate diverse sensors, including unseen modalities. CROMA [17] fuses optical and SAR data through contrastive and reconstruction objectives with attention-biasing strategies. CopernicusFM [64] leverages an 18.7M-image dataset from Sentinel missions with dynamic hypernetworks and a systematic benchmark for EO, weather, and climate tasks.

4.3 Evaluation

For task `T1`, we conducted experiments using the data setup shown in Table 2. For fair evaluation, we created a clean test set that does not share any `.png`, `.txt`, or bounding boxes `BBoX` with the train and validation sets. For task `T2`, we use two setups: (i) *Region-wise*: Uttar Pradesh data split 60:20:20 (stratified) for train/val/test, evaluating in-region (UP) and out-of-region (Dhaka, Bangladesh; Punjab, Pakistan); (ii) *Leave-One-Country-Out (LOCO)*: among four countries, hold out one for testing and train on the remaining three. For T3, we conduct a LOSO experiment using Winter (W), Pre-Monsoon (PM), Monsoon (M), and Post-Monsoon (PoM) data, with a Winter-trained model tested on other seasons. This setup reflects the practical challenge of year-round kiln monitoring under seasonal variability in the Lucknow airshed (3,962 km^2 [38]). For task `T4`, we apply various super-resolution methods to Sentinel-2 images across the Delhi-NCR dataset (60:20:20 stratified train/val/test) and evaluate our best-performing detector on the enhanced imagery. We report class-agnostic mAP_{50} and class-aware AP_{50} across three classes. Model hyperparameters are provided in the appendix. All experiments were run on an NVIDIA A100 GPU.

4.4 Results

T1: Benchmarking of Different Models: Table 3 (see Appendix Fig. 9) summarizes results across detector families. Two-stage methods (e.g., RoI Transformer) achieve moderate performance (CA mAP_{50} = 70.74), one-stage YOLO variants improve further (YOLOv8L-WORLDv2: 80.07), and DETR-based models raise the bar (RTDETR: 86.17). Remote-sensing foundation models outperform generic detectors: Galileo and TerraMind lead, with TerraMind achieving the highest CA mAP_{50} (86.91) and strong class-wise AP_{50}, especially for Zigzag (75.55). Overall, AA-based models consistently surpass OBB-specific designs; Zigzag is the easiest class, while CFCBK remains the hardest due to limited ground-truth examples. DETR-style heads and ViT backbones drive the largest gains, and remote-sensing pretraining provides the strongest lift, particularly for minority classes. Notably, existing models still struggle with certain classes in low-resolution imagery, highlighting opportunities for unified AA/OBB formulations, label-efficient learning.

T2: Spatial Shift Performance: Table 4 represents the region-wise experiments showing that AA-based detectors with transformer backbones generalize more effectively across different regions than OBB-specific and two-stage designs, with YOLO models providing stable performance as competitive baselines. LOCO validation further confirms that models leveraging ViT architectures transfer more robustly across countries, whereas older convolutional approaches exhibit clear limitations. These results highlight that transformer-based AA detectors are better suited for spatial generalization across

Table 3: Comparison of different object detection methods based on AA (Axis-Aligned) and OBB (Oriented Bounding Box). The CA mAP_{50} refers to the class-agnostic mAP value, while CFCBK, FCBK, and Zigzag indicate class-wise AP_{50} scores. In the table, we mark the highest score in **bold** and underline the second-highest score in each column.

Category	Method	Publication	Backbone	BBoX	CA mAP_{50} ↑	Class-wise AP_{50} ↑		
						CFCBK	FCBK	Zigzag
Two stage	PSC [82]	CVPR-23	Res50	OBB	31.95	15.76	16.57	21.41
	H2RBox [79]	ICLR-23	Res50	OBB	45.91	23.81	28.85	30.81
	RoI Transformer [15]	CVPR-19	Swin-T	OBB	70.74	40.45	51.84	55.23
One stage	CSL [75]	ECCV-20	Res50	OBB	18.19	8.80	9.48	10.35
	DCFL [73]	CVPR-23	Res50	OBB	22.36	0.01	10.01	13.98
	Rotated FCOS [55]	ICCV-19	Res50	OBB	30.28	16.61	17.82	20.67
	KFIoU + Rotated-Retinanet [80]	ArXiv-22	Res50	OBB	30.63	0.09	10.89	22.59
	Rotated-Retinanet [31]	ICCV-17	Res50	OBB	35.62	12.89	18.91	24.26
	R^3Det [76]	AAAI-21	Res50	OBB	43.37	0.02	24.68	29.08
	GWD [77]	ICML-21	Res50	OBB	44.18	0.12	22.85	28.41
	KLD + Rotated-Retinanet [78]	NeurIPS-22	Res50	OBB	51.42	37.31	29.51	36.80
	S^2A-Net [20]	TGRS-21	Res50	OBB	56.55	31.12	32.99	43.71
	ConvNeXt [34]	CVPR-22	Res50	OBB	66.32	28.11	41.10	46.12
	YOLOv11L-OBB [26]	arXiv-24	CSPDr53	OBB	75.56	59.19	56.87	50.91
	YOLOv12L [54]	arXiv-25	CSPDr53	AA	78.17	52.05	56.52	47.62
	YOLOv8L-WORLDv2 [9]	CVPR-24	CSPDr53	AA	80.07	55.28	60.81	54.26
DETR Based	DETA [86]	ICCV-23	Res50	AA	66.31	44.76	49.56	61.21
	RFDETR [47]	arXiv-25	Dinov2	AA	81.24	63.07	59.19	65.78
	RTDETR [88]	CVPR-24	Res101	AA	86.17	58.38	62.46	61.34
Remote Sensing Foundation Model	SatMAE++ [36]	CVPR-24	ViT-L	AA	36.63	27.43	21.01	29.74
	CROMA [17]	NeurIPS-23	ViT-B	AA	63.89	17.69	44.66	56.40
	Prithvi [23]	arXiv-23	ViT-L	AA	59.26	16.86	40.63	52.77
	Panopticon [61]	CVPR-25	ViT-B	AA	74.87	43.13	50.61	55.11
	SatMAE [11]	NeurIPS-22	ViT-L	AA	76.36	40.65	50.72	56.03
	CopernicusFM [64]	ICCV-25	ViT-B	AA	77.22	61.09	59.73	67.78
	Scale-MAE [46]	ICCV-23	ViT-L	AA	78.43	52.10	60.77	65.18
	Galileo [57]	ICML-25	ViT-B	AA	<u>86.66</u>	**72.02**	<u>69.81</u>	<u>72.19</u>
	TerraMind [24]	ICCV-25	ViT-B	AA	**86.91**	<u>69.04</u>	**70.54**	**75.55**

regions and national boundaries. Looking forward, future research could explore domain adaptation strategies to reduce cross-region performance gaps, integrate multi-sensor inputs for robustness, and develop unified AA/OBB frameworks to strengthen transferability under varying geographic contexts.

Table 4: Summary of results: (a) region-wise performance, (b) LOCO validation where I = India, P = Pakistan, B = Bangladesh, and A = Afghanistan, and (c) seasonal generalization. In seasonal generalization (**Winter-only**), models are trained on winter (Nov–Feb) and tested across other seasons. Reported metrics are class-agnostic mAP50 (**CA, mAP**50 ↑). The best results are shown in **bold**, and the second-best are underlined.

Spatial Shift								Temporal Shift	
(a) Region-wise				(b) LOCO Validation				(c) Seasonal Generalization	
Model	UP	Dhaka	Punjab	Train→Test	YOLO-OBB11L	RT-DETR	YOLOv8L-WORLDv2	Train→Test	CA (mAP_{50} ↑)
	CA (mAP_{50} ↑)				CA (mAP_{50} ↑)				
YOLOv8L-WORLDv2	<u>83.11</u>	<u>59.64</u>	<u>68.84</u>	I+B+P→A	46.34	53.11	53.70	W(2024)→PM(2024)	55.56
YOLOv11L-OBB	79.49	51.49	67.81	I+B+A→P	75.02	84.40	78.22	W(2024)→M(2024)	60.21
RTDETR	**84.71**	**61.50**	**75.58**	I+A+P→B	72.20	86.66	78.92	W(2024)→PoM(2024)	60.02
RoI Transformer	63.83	13.68	27.24	B+A+P→I	53.39	57.63	58.94	W(2024)→W(2023)	79.21

T3: Temporal Shift Performance: Table 4 shows that models trained on a single season generalize poorly to unseen periods, underscoring strong seasonal variability in satellite imagery. Multi-season training enhances robustness (see Appendix Table 8 and Fig. 14), although performance gaps remain when compared to in-season evaluations. Maintaining reliable kiln detection across atmospheric and land-surface changes remains difficult. Future work can explore domain adaptation, temporal ensembling, and physics-informed augmentations to capture seasonal dynamics. Integrating cross-year data and temporal self-supervised learning may further strengthen year-round generalization.

T4: Super-Resolution: Table 5(see Appendix Fig. 11) highlights the effectiveness of scaling and super-resolution techniques in improving detection performance on low-resolution satellite imagery. Both interpolation-based and learned super-resolution methods significantly enhance model accuracy, with advanced approaches like HiT achieving substantial gains across all kiln categories. These findings suggest that incorporating super-resolution as a pre-processing step is a promising direction for scaling object detection in low-resolution, resource-constrained remote sensing scenarios.

Methods	Resolution	mAP_{50} ↑				PSNR ↑	SSIM ↑
		CA mAP	CFCBK	FCBK	Zigzag		
Original resolution	128×128	65.02	0.00	0.00	63.18	–	–
Bilinear interpolation	512×512	87.64	22.77	34.00	86.11	–	–
Stable Diffusion [48]	512×512	83.46	38.50	27.04	79.60	26.71	0.6785
SwinIR [30]	512×512	89.52	37.01	48.11	86.54	27.14	0.7780
ESRGAN [63]	512×512	90.12	47.43	42.89	87.35	27.16	0.5678
HiT [87]	512×512	**90.74**	**53.79**	**53.88**	**88.28**	34.44	0.9168

Table 5: Super-resolution methods on our dataset: CA mAP_{50} and Class-wise AP_{50} CFCBK, FCBK, Zigzag for the YOLOv11L-OBB model and perceptual quality (PSNR, SSIM).

5 Applications

Our dataset opens new opportunities for the ML and domain applications.

A Benchmark Dataset for Oriented Object Detection: We propose that our dataset can serve as a benchmark for evaluating various oriented object detection (OBB) models. As shown in Table 6, the released dataset is comparable in scale to widely used benchmarks such as HRSC2016 and DOTA. To the best of our knowledge, this is the first publicly available dataset designed specifically to support the development and evaluation of OBB methods in low-resolution satellite imagery.

Dataset	Imagery	Classes	Quantity	Instances	GSD
VEDAI [43]	Aerial Imagery	9	1,210	3,640	0.125 m
HRSC2016 [33]	Google Earth	25	1,070	2,976	$0.4 \sim 2$ m
DOTA-V1.0 [70]	Google Earth	15	2,806	188,282	$0.1 \sim 4.5$ m
FGSD [7]	Google Earth	43	5,634	2,612	$0.12 \sim 1.93$ m
DIOR-R [8]	Google Earth	20	23,463	192,518	$0.5 \sim 1$ m
SENTINELKILNDB (Ours)	**Sentinel-2**	**3**	**1,14,300**	**97,648**	**10m**

Table 6: Comparison of our dataset with other well-known object detection datasets. We have the largest number of images compared to the specified datasets.

Geographical Domain Adaptation and Active Learning: Models trained in one region often degrade in distant geographies due to distribution shifts. Our dataset enables rigorous study of geographic robustness and development of geography-aware SSL [5] for cross-region consistency. Given extensive hand-validation and severe class imbalance (many negatives), it also serves as a strong benchmark for active learning in object detection to cut labeling cost while preserving accuracy.

Air Quality Modeling: Air quality modeling often requires well-maintained emission inventories. Accuracy of the conventional Chemical Transport Models (CTMs) [1] heavily depends on the quality of the emission inventory. Due to the dynamic nature of the brick kilns, it is hard to maintain inventory by manual annotations. Air quality experts can take the data and methods proposed in this paper to maintain a dynamic inventory; they can further use the areas of the identified brick kiln and their type to accurately estimate the emission factors.

6 Limitations and Future Work

SentinelKilnDB dataset is constructed using freely available Sentinel-2 imagery to enable scalable and reproducible research. While 10 m resolution limits fine-grained detail, it reflects practical constraints in large-scale environmental monitoring. Future work could explore combining Sentinel-2 with high-resolution sources for hybrid detection or multi-scale fusion. Temporally, our dataset focuses on peak kiln activity (Nov 2023–Feb 2024); extending to multiple seasons can support time-series analysis and kiln lifecycle modeling. From a benchmarking standpoint, while we evaluate in-region, out-of-region, and super-resolution settings, further work may incorporate uncertainty quantification, geography-aware learning, and human-in-the-loop active learning.

7 Conclusion and Broader Impact

In this paper, we release a dataset containing 62,671 brick kilns along with their subtypes across South Asia. To the best of our knowledge, this is the largest publicly available hand-validated dataset related to brick kilns. We benchmark the dataset using various object detection methods, including both oriented and axis-aligned bounding box models, on low-resolution imagery. The results show strong baseline performance but indicate room for improvement, particularly in class-wise detection under low-resolution settings. The dataset supports future work in machine learning and geospatial research across environmental monitoring, policy design, and cross-border impact analysis, serving as a resource for both academic and applied studies.

References

[1] CAMX - The Composites and Advanced Materials Expo. https://www.camx.com/. Accessed: YYYY-MM-DD.

[2] Esri wayback imagery — esri.com. https://www.esri.com/arcgis-blog/products/arcgis-living-atlas/imagery/wayback-server-connection-in-pro/. [Accessed 10-11-2024].

[3] Planet api reference. https://developers.planet.com/docs/basemaps/reference/. (Accessed on 11/08/2024).

[4] Brick kilns performance assessment, 2015. Accessed: 2024-06-02.

[5] Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10181–10190, 2021.

[6] Doreen S Boyd, Bethany Jackson, Jessica Wardlaw, Giles M Foody, Stuart Marsh, and Kevin Bales. Slavery from space: Demonstrating the role for satellite remote sensing to inform evidence-based action related to un sdg number 8. *ISPRS journal of photogrammetry and remote sensing*, 142:380–388, 2018.

[7] Zhiming Chen, Kean Chen, Weiyao Lin, John See, Hui Yu, Yan Ke, and Cong Yang. Piou loss: Towards accurate oriented object detection in complex environments. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 195–211. Springer, 2020.

[8] Gong Cheng, Jiabao Wang, Ke Li, Xingxing Xie, Chunbo Lang, Yanqing Yao, and Junwei Han. Anchor-free oriented proposal generator for object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.

[9] Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, and Ying Shan. Yolo-world: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16901–16911, 2024.

[10] Creative Commons. Creative commons attribution-noncommercial 4.0 international public license. https://creativecommons.org/licenses/by-nc/4.0/, 2013. Accessed: YYYY-MM-DD.

[11] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[12] Copernicus SentiWiki. Sentinel-2 products. https://sentiwiki.copernicus.eu/web/s2-products, 2023. Accessed: 2024-06-02.

[13] Jessica D DeWitt, Peter G Chirico, Marissa A Alessi, and Kathleen M Boston. Remote sensing inventory and geospatial analysis of brick kilns and clay quarrying in kabul, afghanistan. *Minerals*, 11(3):296, 2021.

[14] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2849–2858, 2019.

[15] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[16] Esri. Arcgis geographic information system, 2024. Accessed: 2024-11-27.

[17] Anthony Fuller, Koreen Millard, and James Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. *Advances in Neural Information Processing Systems*, 36:5506–5538, 2023.

[18] Sarath K Guttikunda, KA Nishadh, Sudhir Gota, Pratima Singh, Arijit Chanda, Puja Jawahar, and Jai Asundi. Air quality, emissions, and source contributions analysis for the greater bengaluru region of india. *Atmospheric Pollution Research*, 10(3):941–953, 2019.

[19] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia. Align deep features for oriented object detection. *IEEE transactions on geoscience and remote sensing*, 60:1–11, 2021.

[20] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[22] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018.

[23] Johannes Jakubik, Sujit Roy, CE Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, et al. Foundation models for generalist geospatial artificial intelligence. *arXiv preprint arXiv:2310.18660*, 2023.

[24] Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. Terramind: Large-scale generative multimodality for earth observation. *arXiv preprint arXiv:2504.11171*, 2025.

[25] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[26] Hussain Muhammad Khanam, Rahima. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.

[27] R Khanam and M Hussain. Yolov11: An overview of the key architectural enhancements. arxiv 2024. *arXiv preprint arXiv:2410.17725*.

[28] Planet Labs. Education and research program. `https://www.planet.com/industries/education-and-research/`. [Accessed 08-11-2024].

[29] Jihyeon Lee, Nina R Brooks, Fahim Tajwar, Marshall Burke, Stefano Ermon, David B Lobell, Debashish Biswas, and Stephen P Luby. Scalable deep learning to identify brick kilns and aid regulatory capacity. *Proceedings of the National Academy of Sciences*, 118(17):e2018863118, 2021.

[30] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

[31] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017.

[32] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[33] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[34] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976, Los Alamitos, CA, USA, June 2022. IEEE Computer Society.

[35] Usman Nazir, Usman Khalid Mian, Muhammad Usman Sohail, Murtaza Taj, and Momin Uppal. Kiln-net: A gated neural network for detection of brick kilns in south asia. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3251–3262, 2020.

[36] Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024.

[37] Jeffrey Ouyang-Zhang, Jang Hyun Cho, Xingyi Zhou, and Philipp Krähenbühl. Nms strikes back. *arXiv preprint arXiv:2212.06137*, 2022.

[38] Zeel B Patel, Rishabh Mondal, Shataxi Dubey, Suraj Jaiswal, Sarath Guttikunda, and Nipun Batra. Space to policy: Scalable brick kiln detection and automatic compliance monitoring with geospatial data. *arXiv preprint arXiv:2412.04065*, 2024.

[39] Planet Labs PBC. Planet science programs: Satellite imagery access for researchers.

[40] Planet Labs PBC. Planet explorer terms of service. `https://www.planet.com/explorer-terms-of-service/`, 2022. Accessed: 2025-05-14.

[41] Planet Labs PBC. Planet basemaps: Visual. `https://docs.planet.com/data/imagery/basemaps/visual/`, 2024. Accessed: 2025-05-14.

[42] Uma Rajarathnam, Vasudev Athalye, Santhosh Ragavan, Sameer Maithel, Dheeraj Lalchandani, Sonal Kumar, Ellen Baum, Cheryl Weyant, and Tami Bond. Assessment of air pollutant emissions from brick kilns. *Atmospheric Environment*, 98:549–553, 2014.

[43] Sebastien Razakarivony and Frederic Jurie. Vehicle detection in aerial imagery: A small target detection benchmark. *Journal of Visual Communication and Image Representation*, 34:187–203, 2016.

[44] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[45] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

[46] Colorado J Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Kurt Keutzer, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4088–4099, 2023.

[47] Isaac Robinson, Peter Robicheaux, and Matvei Popov. Rf-detr. `https://github.com/roboflow/rf-detr`, 2025. SOTA Real-Time Object Detection Model.

[48] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

[49] Brannon Seay, Anna Adetona, Natasha Sadoff, Marcus C Sarofim, and Michael Kolian. Impact of south asian brick kiln emission mitigation strategies on select pollutants and near-term arctic temperature responses. *Environmental research communications*, 3(6):061004, 2021.

[50] Shuai Shao, Zeming Li, Tianyuan Zhang, Chao Peng, Gang Yu, Xiangyu Zhang, Jing Li, and Jian Sun. Objects365: A large-scale, high-quality dataset for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8430–8439, 2019.

[51] Piyush Sharma, Nan Ding, Sebastian Goodman, and Radu Soricut. Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2556–2565, 2018.

[52] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[53] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

[54] Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*, 2025.

[55] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. FCOS: Fully Convolutional One-Stage Object Detection . In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9626–9635, Los Alamitos, CA, USA, November 2019. IEEE Computer Society.

[56] Kushal Tibrewal, Chandra Venkataraman, Harish Phuleria, Veena Joshi, Sameer Maithel, Anand Damle, Anurag Gupta, Pradnya Lokhande, Shahadev Rabha, Binoy K Saikia, et al. Reconciliation of energy use disparities in brick production in india. *Nature Sustainability*, 6(10):1248–1257, 2023.

[57] Gabriel Tseng, Anthony Fuller, Marlena Reil, Henry Herzog, Patrick Beukema, Favyen Bastani, James R Green, Evan Shelhamer, Hannah Kerner, and David Rolnick. Galileo: Learning global & local features of many remote sensing modalities. In *TerraBytes - ICML 2025 workshop*, 2025.

[58] UBC Research Commons. Planet imagery. `https://researchcommons.library.ubc.ca/planet-imagery/`, 2022. Accessed: 2025-05-14.

[59] UNEP. Emissions gap report 2019, 11 2019.

[60] UNICEF – Centre of Excellence on Humanitarian Action. Spotlight on risk: Air pollution. `https://ceh.unicef.org/spotlight-risk/air-pollution`, 2024. Accessed: 2025-09-19.

[61] Leonard Waldmann, Ando Shah, Yi Wang, Nils Lehmann, Adam Stewart, Zhitong Xiong, Xiao Xiang Zhu, Stefan Bauer, and John Chuang. Panopticon: Advancing any-sensor foundation models for earth observation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2204–2214, 2025.

[62] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7464–7475, 2023.

[63] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[64] Yi Wang, Zhitong Xiong, Chenying Liu, Adam J Stewart, Thomas Dujardin, Nikolaos Ioannis Bountos, Angelos Zavras, Franziska Gerken, Ioannis Papoutsis, Laura Leal-Taixé, et al. Towards a unified copernicus foundation model for earth vision. *arXiv preprint arXiv:2503.11849*, 2025.

[65] Haoran Wei, Yue Zhang, Zhonghan Chang, Hao Li, Hongqi Wang, and Xian Sun. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:268–279, 2020.

[66] Wikipedia contributors. Sentinel-2. `https://en.wikipedia.org/wiki/Sentinel-2#:~:text=Sentinel%2D2%20is%20an%20Earth,a%20third%2C%20Sentinel%2D2C.`, 2023. Accessed: 2024-06-02.

[67] WorldBank. Dirty stacks, high stakes: An overview of brick sector in south asia, 2020.

[68] Qiusheng Wu. geemap: A python package for interactive mapping with google earth engine. *Journal of Open Source Software*, 5(51):2305, 2020.

[69] Qiusheng Wu. Leafmap: A python package for interactive mapping and geospatial analysis with minimal coding in a jupyter environment. *Journal of Open Source Software*, 6(63):3414, 2021.

[70] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018.

[71] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[72] Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired multimodal foundation model for earth observation. *arXiv preprint arXiv:2403.15356*, 2024.

[73] Chang Xu, Jian Ding, Jinwang Wang, Wen Yang, Huai Yu, Lei Yu, and Gui-Song Xia. Dynamic coarse-to-fine learning for oriented tiny object detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7318–7328, 2023.

[74] Chang Xu, Jinwang Wang, Wen Yang, Huai Yu, Lei Yu, and Gui-Song Xia. Detecting tiny objects in aerial images: A normalized wasserstein distance and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190:79–93, 2022.

[75] Xue Yang and Junchi Yan. Arbitrary-oriented object detection with circular smooth label. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 677–694. Springer, 2020.

[76] Xue Yang, Junchi Yan, Ziming Feng, and Tao He. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 3163–3171, 2021.

[77] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian. Rethinking rotated object detection with gaussian wasserstein distance loss. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11830–11841. PMLR, 18–24 Jul 2021.

[78] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[79] Xue Yang, Gefan Zhang, Wentong Li, Xuehui Wang, Yue Zhou, and Junchi Yan. H2rbox: Horizontal box annotation is all you need for oriented object detection. *arXiv preprint arXiv:2210.06742*, 2022.

[80] Xue Yang, Yue Zhou, Gefan Zhang, Jirui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian. The kfiou loss for rotated object detection. *arXiv preprint arXiv:2201.12558*, 2022.

[81] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B Lobell, and Stefano Ermon. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*, 2021.

[82] Yi Yu and Feipeng Da. Phase-shifting coder: Predicting accurate orientation in oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13354–13363, 2023.

[83] Yi Yu and Feipeng Da. On boundary discontinuity in angle regression based arbitrary oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[84] Daniele Zanaga, Ruben Van De Kerchove, Dirk Daems, Wanda De Keersmaecker, Carsten Brockmann, Grit Kirches, Jan Wevers, Oliver Cartus, Maurizio Santoro, Steffen Fritz, Myroslava Lesiv, Martin Herold, Nandin-Erdene Tsendbazar, Panpan Xu, Fabrizio Ramoino, and Olivier Arino. Esa worldcover 10 m 2021 v200, October 2022.

[85] Ying Zeng, Yushi Chen, Xue Yang, Qingyun Li, and Junchi Yan. Ars-detr: Aspect ratio-sensitive detection transformer for aerial oriented object detection. *IEEE transactions on geoscience and remote sensing*, 62:1–15, 2024.

[86] Ji Zhang, Lianli Gao, Xu Luo, Hengtao Shen, and Jingkuan Song. Deta: Denoised task adaptation for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11541–11551, October 2023.

[87] Xiang Zhang, Yulun Zhang, and Fisher Yu. Hit-sr: Hierarchical transformer for efficient image super-resolution. In *European Conference on Computer Vision*, pages 483–500. Springer, 2024.

[88] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16965–16974, 2024.

[89] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] They do.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   [Yes] Please see the limitations section

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   [NA]

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   [Yes] We included a public URL containing code, data, and instructions needed to reproduce the main experimental results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes] We included a public URL containing code, data, and instructions needed to reproduce the main experimental results.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes] yes we specify them

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes] We mentioned in evaluation section

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA] In air pollution prospective

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer:[NA] No

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes] Yes CC-BY NC 4.0

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes] Yes its a new assets

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer:[NA]

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

A Benchmarking Models and Hyperparameters

1) PSC: Phase-Shifting Coder (PSC) [82] proposed a novel angle coding method that accurately predicted object orientation. PSC encoded angles in a periodic manner to resolve the boundary discontinuity problem. In this work, we used the Faster PSC with a ResNet-50 backbone pre-trained on the ImageNet dataset. The model was trained using the SGD optimizer with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.0001.

2) H2RBox: H2RBox [79] was a weakly supervised oriented object detection framework that learned rotated bounding boxes (RBox) from horizontal box (HBox) annotations. It utilized a dual-branch architecture combining weak and self-supervised learning to predict object orientations by enforcing consistency between different views. We employed the H2RBox model with a ResNet-50 backbone. The model was fine-tuned on our dataset using the AdamW optimizer with a learning rate of 0.0001, $\beta = (0.9, 0.999)$, and a weight decay of 0.05.

3) RoI Transformer: RoI Transformer [15] was an OBB detection framework that addressed the misalignment between horizontal region proposals and objects with arbitrary orientations. It introduced a Rotated RoI (RRoI) learner that transformed Horizontal RoIs (HRoIs) into Rotated RoIs, enabling more accurate localization of oriented objects. We utilized the RoI Transformer with a Swin Transformer (Swin-T) backbone. The model was trained using the AdamW optimizer with a learning rate of 0.0001 and $\beta = (0.9, 0.999)$.

4) Rotated FCOS: Rotated FCOS [55] extends the anchor-free FCOS detector to predict oriented bounding boxes by directly regressing angles, enabling efficient detection of rotated objects. We implement Rotated FCOS with a ResNet-50 backbone, and the SGD optimizer with a learning rate of 0.0025, momentum of 0.5, and weight decay of 0.0001.

5) DCFL: Dynamic Coarse-to-Fine Learning (DCFL) [73] enhanced oriented tiny object detection by addressing label assignment bias. It was implemented through a coarse matching stage followed by fine refinement. We used the model with a ResNet-50 backbone, trained using the SGD optimizer with a learning rate of 0.0025, momentum of 0.9, and weight decay of 0.0001.

6) CSL: Circular Smooth Label (CSL) [75] addressed the boundary discontinuity problem inherent in angle regression methods by transforming the angle prediction task from regression to classification. It handled the periodic nature of angles by using a circular smooth label encoding. We implemented CSL with a ResNet-50 backbone pre-trained on the ImageNet dataset. The model was trained using the SGD optimizer with a learning rate of 0.0025, momentum of 0.9, and weight decay of 0.0001.

7) Rotated RetinaNet: Rotated RetinaNet [31]was an extension of the original RetinaNet framework, designed for detecting oriented bounding boxes (OBBs). It overcame the limitations of axis-aligned boxes by predicting rotated bounding boxes. The model employed a ResNet-50 backbone with a Feature Pyramid Network (FPN). Training was conducted using the SGD optimizer with a learning rate of 0.0025, momentum of 0.9, and weight decay of 0.0001.

8) GWD: Gaussian Wasserstein Distance (GWD) [77] was a regression loss function designed to counter the problems of boundary discontinuity and square-like issues. GWD represented rotated bounding boxes as 2D Gaussian distributions, enabling the computation of a differentiable distance metric that approximated the rotational IoU loss. In our experiments, we integrated GWD into a detection framework with a ResNet-50. The model was trained using the SGD optimizer with a learning rate of 0.0025, momentum of 0.9, and weight decay of 0.0001.

9) R^3Det: Refined Single-Stage Detector R^3Det [76] introduced a progressive regression strategy that transitioned from coarse to fine granularity and an approximate SkewIoU loss function to facilitate effective rotation estimation, amplifying localization accuracy for objects with varying orientations. We implemented R^3Det with a ResNet-50 backbone. The model was trained using the SGD optimizer with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.0005.

10) S^2A-Net: Single-shot Alignment Network S^2A-Net [20] introduced two key modules: the Feature Alignment Module (FAM), which adaptively aligned convolutional features with anchor boxes using an Alignment Convolution, and the Oriented Detection Module (ODM), which employed active rotating filters to encode orientation information and produced both orientation-sensitive and orientation-invariant features. In our experiments, we implemented S^2A-Net with a ResNet-50 backbone. The model was trained using the SGD optimizer with a learning rate of 0.0025, momentum of 0.9, and weight decay of 0.0005.

11) ConvNeXt: ConvNeXt [34] is a modern convolutional network that adapts principles from vision transformers (ViT) to standard ConvNet architectures, achieving state-of-the-art performance on image classification and object detection tasks. In our experiments, we implemented ConvNeXt with a ResNet-50 backbone pre-trained on the ImageNet dataset. The model was trained using the SGD optimizer with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.0005.

12) YOLO-World: YOLO-World [9] is a recent general-purpose expansion of the YOLO model from CVPR 2024, built for zero-shot object detection. The YOLO-World model introduces an advanced, real-time YOLOv8-based approach for Open-Vocabulary Detection tasks. Upon random trials with our images, it was not able to detect brick kilns in zero-shot. However, to leverage the capabilities of this model, we used the model pre-trained on Object365 [50] and CC3M-Lite [51] and then fine-tuned it on our dataset with the same hyperparameters as YOLO.

13) YOLO: YOLO (You Only Look Once) models are known for their speed of inference. YOLO allows detecting objects in a single forward pass. We experimented with the most updated 'v12L' [62] versions of YOLO in this work.

14) YOLO-OBB: YOLO-OBB [25] is a YOLO variant capable of detecting oriented bounding boxes. The model predicts an additional variable, 'angle', along with the center coordinates of the box, width, and height. We use YOLO-OBB from Ultralytics [25] pre-trained on the DOTA v1.0 [70] dataset. We experiment with four variants of YOLO-OBB in this work: 'v11L'.

15) DETA: DEtection **T**ransformers with **A**ssignment [37] follows the two-stage DETR framework [89] and replaces the one-to-one Hungarian matching loss with a one-to-many IoU-based assignment loss in both stages. DETA has precisely the same architecture as its end-to-end counterparts and uses NMS (Non-Maximum Suppression) for removing the overlapping bounding boxes. We have used DETA pretrained on the Objects 365 [86] dataset with a ResNet-50 backbone. We fine-tuned the model on our dataset for 100 epochs with the AdamW optimizer, using a learning rate of 1e-5 for the backbone and 1e-4 for the rest of the parameters.

16) RT-DETR: Real-Time DEtection Transformer (RT-DETR) [88] was optimized extension of the original DETR framework that enhances its speed and performance, particularly for real-time object detection tasks. RT-DETR introduces an efficient hybrid encoder and a modified decoder that reduces the computational cost of the attention mechanism. In our experiments, we use RT-DETR with a ResNet-101 backbone pre-trained on the ImageNet dataset. The model is fine-tuned on our custom dataset with the AdamW optimizer, a learning rate of 1e-5, and a weight decay of 0.01. This setup allows for efficient training and inference while ensuring robust detection performance on complex object detection tasks.

17) Remote Sensing Foundation Model Hyperparameters: All models are trained with Faster R-CNN detectors using default input resolutions per backbone and 4-level pyramids. Anchor sizes are $32, 64, 128, 256$ with aspect ratios $(0.5, 1.0, 2.0)$ and ROIAlign of size 7 (sampling 2). Optimization follows AdamW with backbone and head learning rates in the range of $[1 \times 10^{-5}, 1 \times 10^{-4}]$, weight decay 0.05, cosine learning rate schedule ($T_{\max} = 50$), batch sizes between 4–32, and total 50 training epochs.

18) KFIoU: KFIoU [80] is a regression loss function designed for rotated object detection, which improves localization accuracy by considering the geometric properties of rotated bounding boxes. It enhances the learning of object orientation by incorporating angle-aware IoU calculations. We implemented KFIoU using the Rotated RetinaNet framework with a ResNet-50 + FPN backbone. The model was trained using the SGD optimizer with a learning rate of **0.0025**, momentum of **0.9**, and weight decay of **0.0001**.

19) KLD: KLD [78] introduces a Kullback-Leibler Divergence-based loss for rotated bounding box regression, enabling better modeling of uncertainty in object orientation. It improves detection performance by aligning predicted distributions with ground truth. We used the Rotated RetinaNet architecture with a ResNet-50 + FPN backbone for KLD-based training. The model was optimized using SGD with a learning rate of **0.0025**, momentum of **0.9**, and weight decay of **0.0001**.

B Additional figures/tables

The additional figure from our previous research[38] shows an example of a bounding-box plot over Esri imagery and Planet Labs imagery.

(a) CFCBK - Esri (b) CFCBK - Planet (c) FCBK - Esri (d) FCBK - Planet (e) Zigzag - Esri (f) Zigzag - Planet

Figure 3: Satellite view of brick kilns with bounding boxes. CFCBK is Circular Fixed Chimney Bull's Trench Kiln, and FCBK is Fixed Chimney Bull's Trench Kiln. We use Esri's high-resolution basemap (zoom level 17-19, 1.19-0.3 m per pixel) to create geo-referenced bounding boxes and hand-validate the predicted bounding boxes. We use Planet's moderate-resolution imagery (zoom level 15, 4.77 m per pixel) to train object detection models. Planet Imagery © 2024 Planet Labs Inc. Esri imagery © Esri, TomTom, Garmin, Foursquare, METI/NASA, USGS.

Table 7: Class-wise brick kiln counts for all states and provinces covered in our dataset.

Country	State/Region	CFCBK	FCBK	Zigzag	Total
India	Gujarat	57	961	45	1063
	Assam	5	1239	109	1353
	Jharkhand	17	1229	191	1437
	Haryana	1	159	2231	2391
	Rajasthan	34	2137	232	2403
	Punjab	1	444	2074	2519
	West Bengal	67	1575	2776	4418
	Bihar	58	2038	5392	7488
	Uttar Pradesh	1699	11669	6542	19910
Bangladesh	Sylhet	0	9	216	225
	Barisal	0	109	256	365
	Mymensingh	0	108	411	519
	Rangpur	0	112	894	1006
	Khulna	0	388	645	1033
	Rajshahi	0	251	880	1131
	Dhaka	2	92	1205	1299
	Chittagong	0	392	933	1325
Pakistan	Balochistan	1	63	1	65
	Sindh	1	788	2	791
	Khyber Pakhtunkhwa	0	846	2	848
	Punjab (PK)	1	8746	1726	10473
Afghanistan	34 Provinces	0	608	1	609
Total		**1944**	**33963**	**26764**	**62671**

Table 8: Intra-Season Generalization (**YOLO11L-OBB**). Each model is trained on multi-season data and tested on the held-out season to assess whether seasonal diversity improves robustness and year-round performance. Results are reported in class-agnostic mAP_{50} (**CA**) and per-class AP_{50} for **CFCBK**, **FCBK**, and **Zigzag**.

Train Data (2024)	Test Data (2024)	CA mAP_{50}	Per-Class AP_{50}		
			CFCBK	FCBK	Zigzag
W + PM + M	PoM	60.62	35.07	29.67	51.83
W + PM + PoM	M	61.31	57.98	34.97	39.92
W + M + PoM	PM	66.46	68.39	46.32	56.22
PM + M + PoM	W	52.42	53.28	25.06	46.35

B.1 Different Satellite Imagery and Enhanced Image Comparisons

Figure 16 presents a comparative visualization of super-resolution techniques applied to a specific patch as captured by Sentinel-2 imagery. The top row includes the original Sentinel-2 image (a), followed by higher-resolution imagery from ESRI Wayback (b) and Planet Labs (c), serving as external visual references. The second row features imagery from Google Earth (d), standard bilinear interpolation (e), and the SwinIR deep learning-based super-resolution model (f). The final row shows

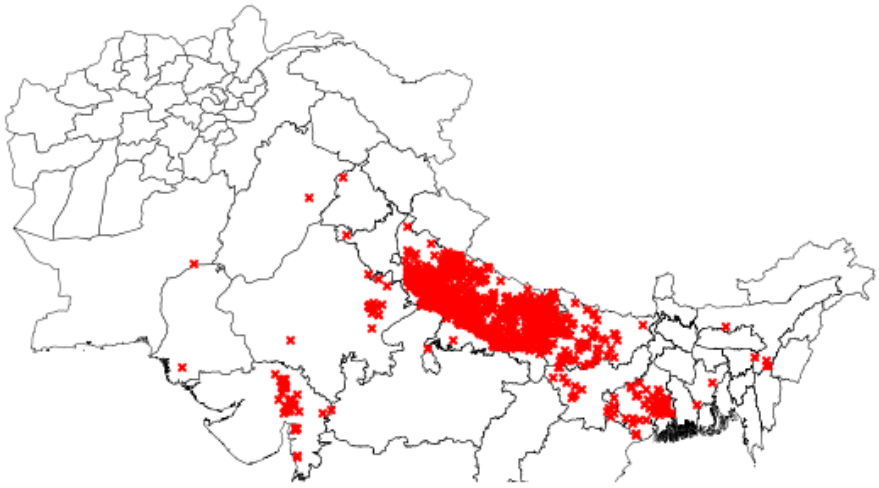


Figure 4: Spatial Distribution of CFCBK Brick Kilns in our dataset.

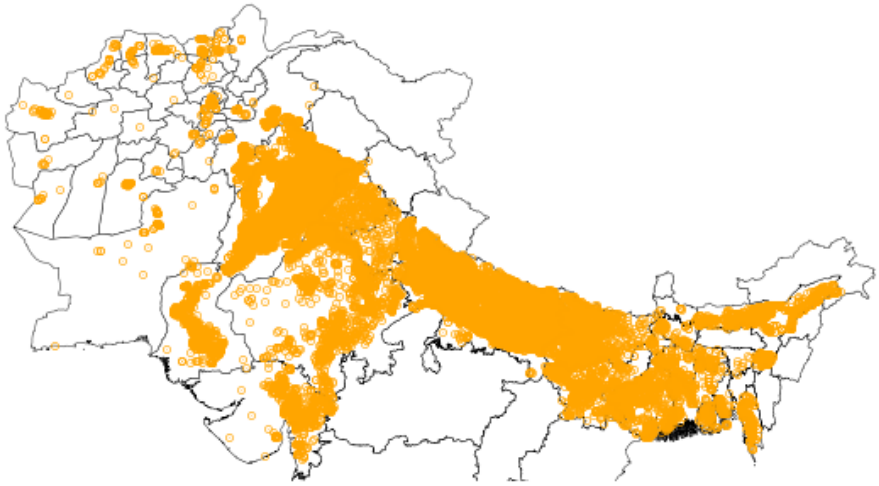


Figure 5: Spatial Distribution of FCFCBK Brick Kilns in our dataset.

outputs from ESRGAN (g), HIT (h), and Stable Diffusion (i), representing various state-of-the-art generative and enhancement models.

This 3×3 grid offers a side-by-side comparison, highlighting visual differences in clarity, structure, and detail enhancement across traditional and modern super-resolution methods. It underscores the potential of advanced models to bridge the resolution gap between publicly available low-res satellite imagery and high-res commercial sources, which is critical for accurate remote sensing applications such as object detection and environmental monitoring.

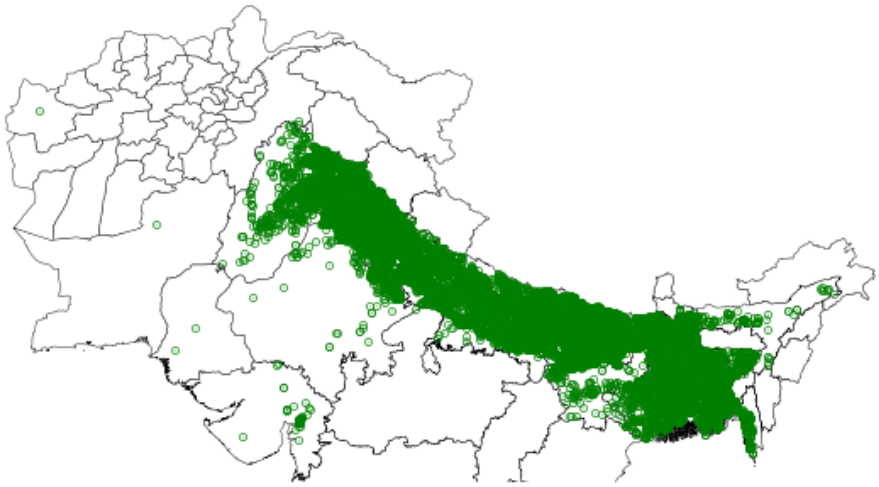


Figure 6: Spatial Distribution of Zigzag Brick Kilns in our dataset.

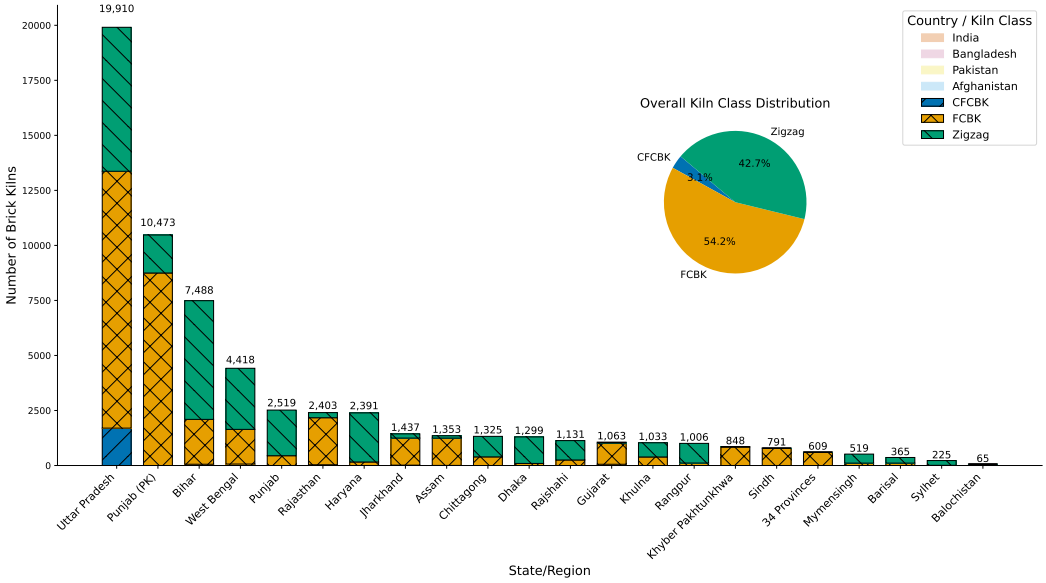


Figure 7: Visualization of the regional brick kiln counts across different states/regions, categorized by kiln classes (CFCBK, FCBK, and Zigzag). The stacked bar chart illustrates the distribution of each kiln class within each region, while the accompanying pie chart represents the overall proportion of each kiln class aggregated across all regions, providing a clear comparative overview of the kiln class distribution at both regional and aggregate levels.

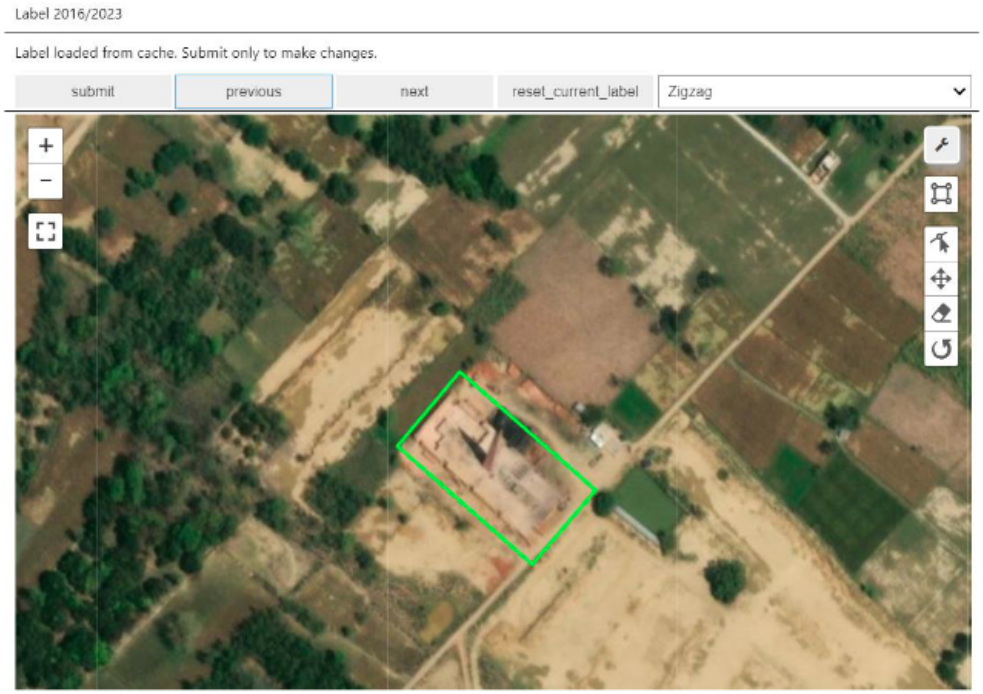


Figure 8: Snippet of the custom hand-validation interface developed in-house. The interface overlays bounding box annotations on ESRI Wayback Imagery, displaying a Zigzag brick kiln.

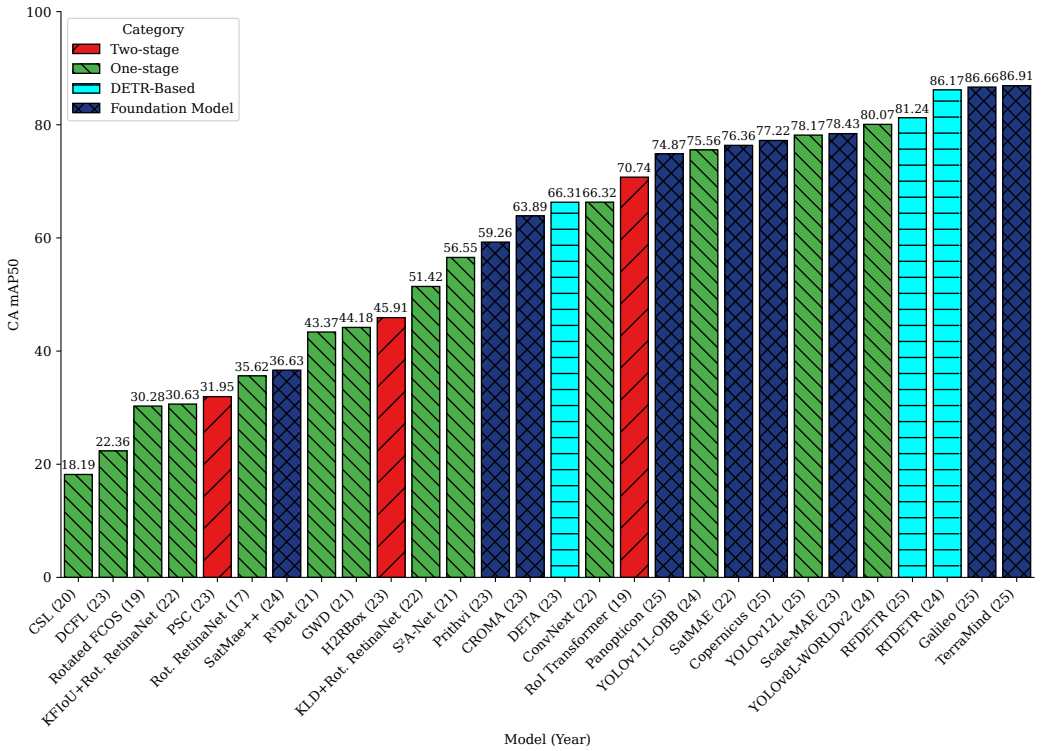


Figure 9: Comparison of class-agnostic mAP 50 for one-stage, two-stage, and transformer-based (DETR) object detection methods.

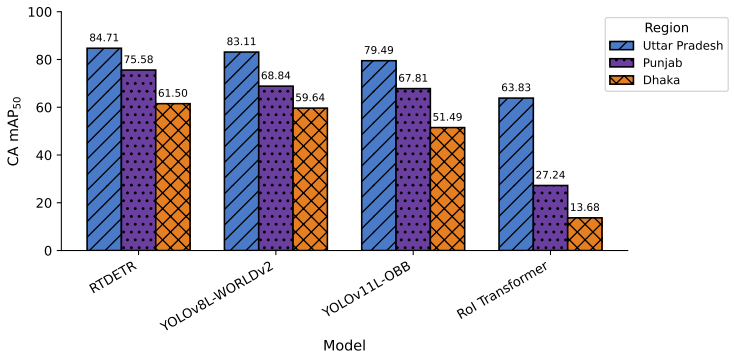


Figure 10: Comparison of class-agnostic mAP 50 for out-of-region performance experiment using the Uttar Pradesh dataset. Results are shown on test sets of Uttar Pradesh, Punjab (Pakistan), and Dhaka using the best-performing methods from Table 3.

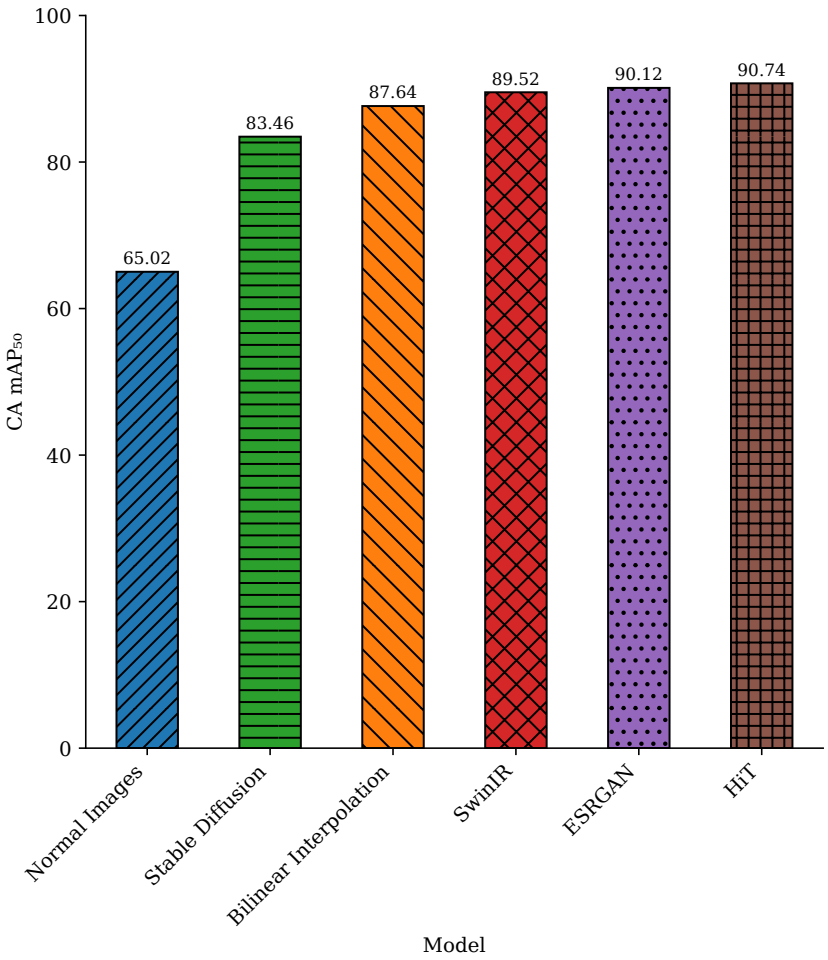


Figure 11: Class-agnostic mAP 50 comparison of different super-resolution methods applied to the YOLOv11L-OBB detection pipeline.

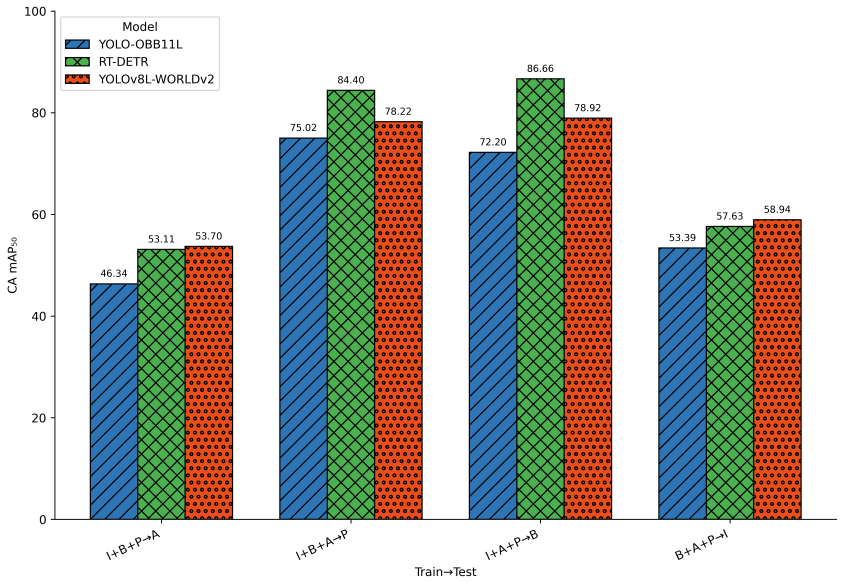


Figure 12: Cross-country spatial generalization (LOCO) test results for brick kiln detection: class-agnostic mAP_{50} performance of three models trained on three South Asian countries and tested on the fourth (I = India, B = Bangladesh, P = Pakistan, A = Afghanistan).

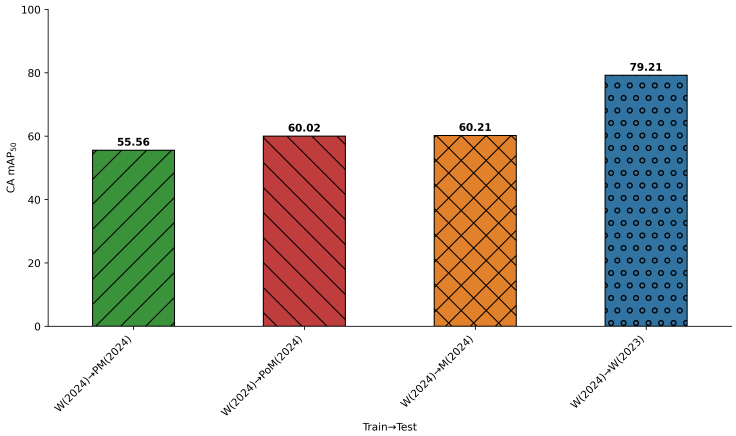


Figure 13: Class-agnostic mAP_{50} evaluation of brick kiln detection models across single-season train/test splits. Seasons with their years are indicated in parentheses (Season (Year)). Seasonal codes represent Winter (W), Monsoon (M), Pre-monsoon (PM), and Post-monsoon (PoM).

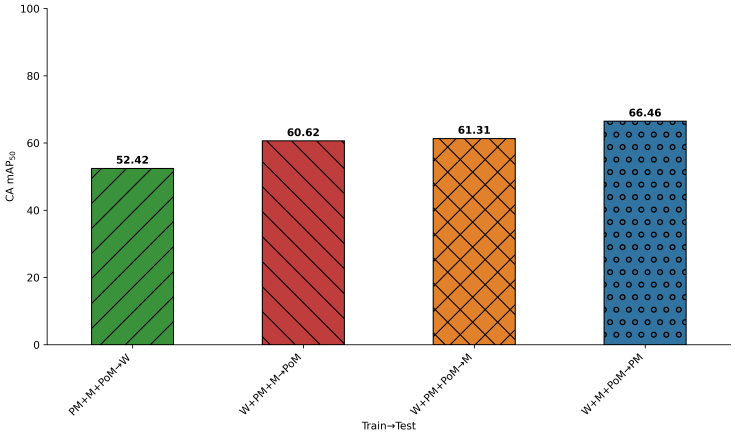


Figure 14: Class-agnostic mAP_{50} evaluation of brick kiln detection models across multi-seasonal train/test splits. Seasonal codes represent Winter (W), Monsoon (M), Pre-monsoon (PM), and Post-monsoon (PoM).

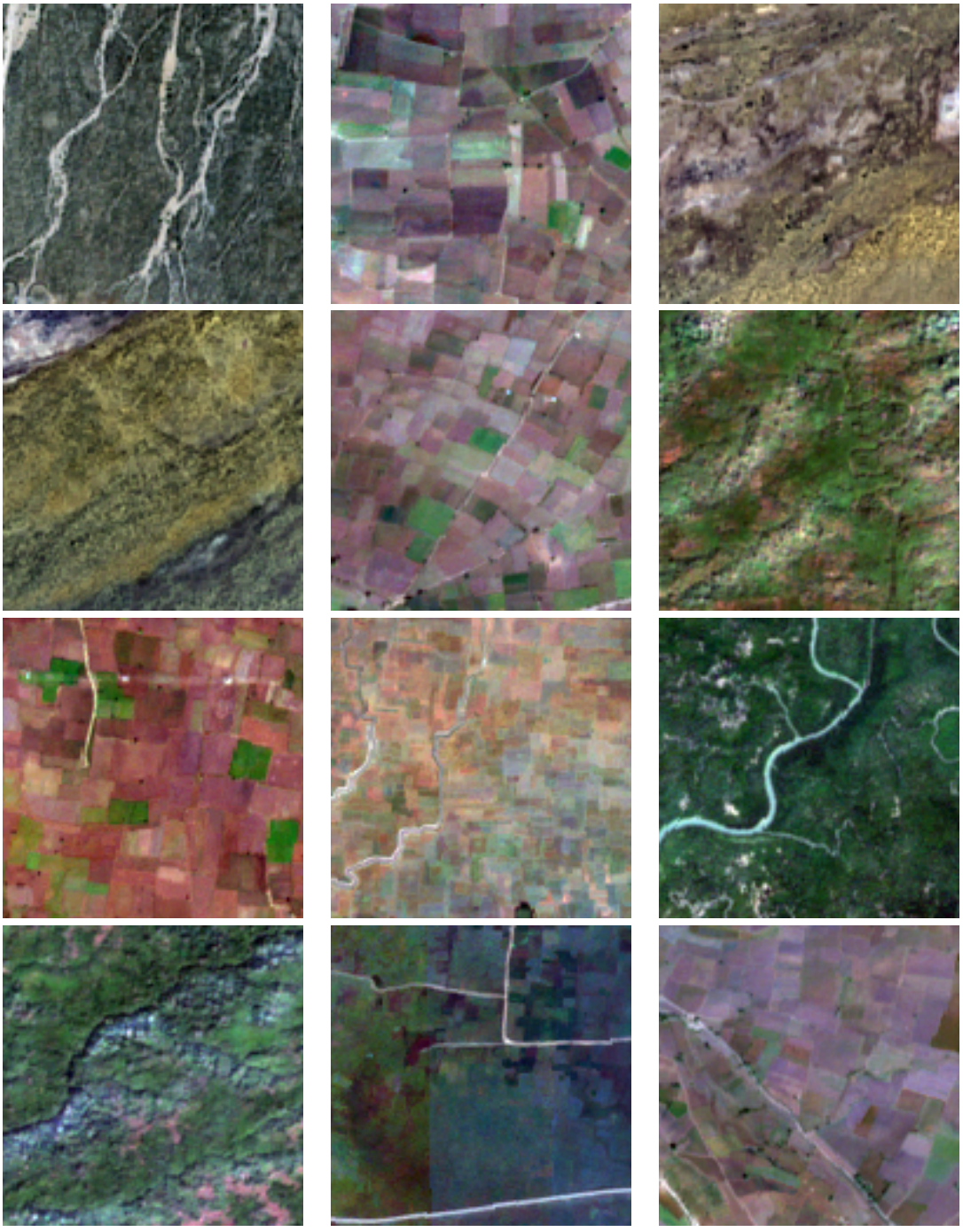
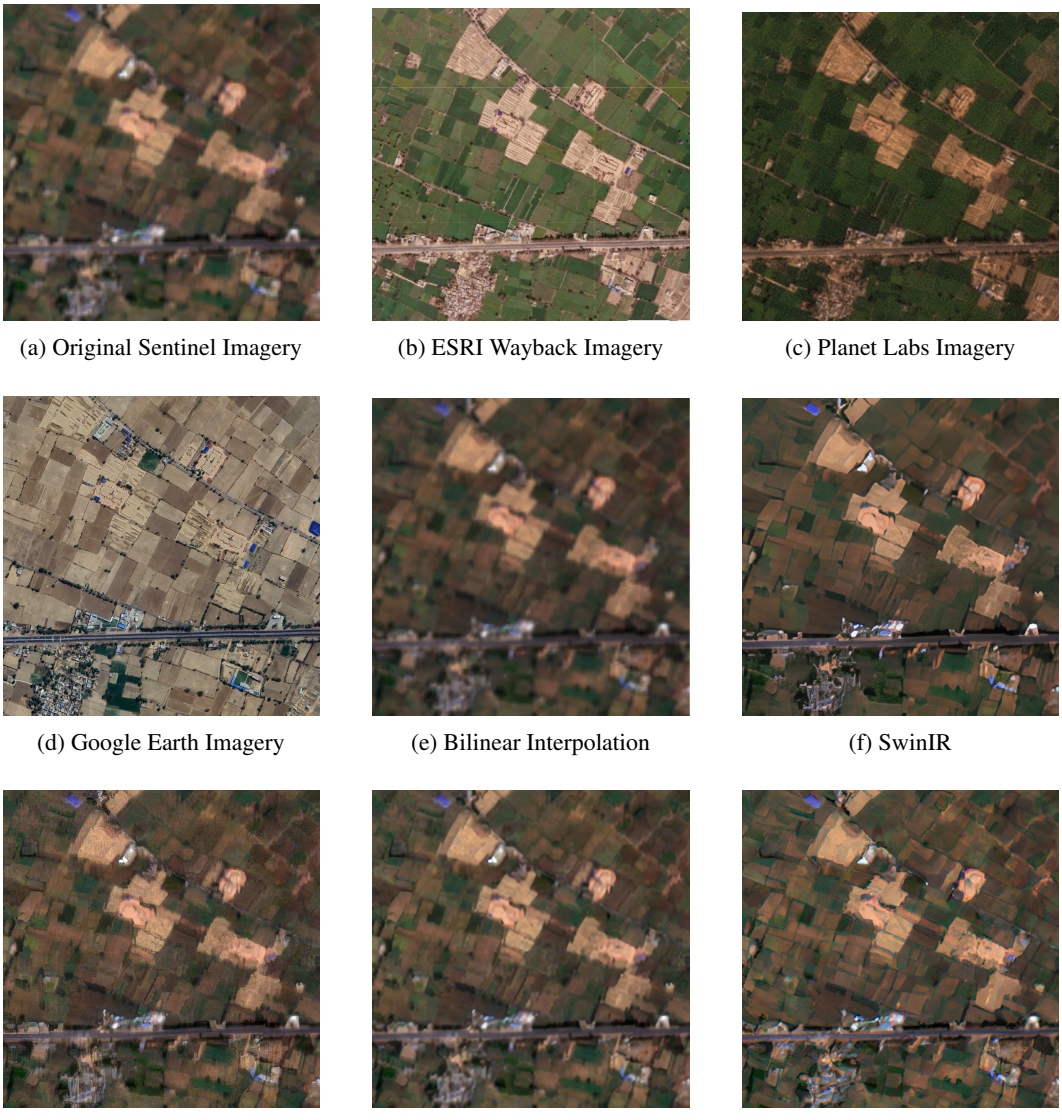


Figure 15: Visualization of negative examples using ESA WorldCover v200 [84], selecting land-cover–pure regions with at least 99.90% class confidence across eight relevant classes. All candidate tiles were manually verified against Esri basemaps, and any image containing visible kilns was discarded.



(a) Original Sentinel Imagery (b) ESRI Wayback Imagery (c) Planet Labs Imagery

(d) Google Earth Imagery (e) Bilinear Interpolation (f) SwinIR

(g) ESRGAN (h) HIT (i) Stable Diffusion

Figure 16: Comparison of nine images over the same target Sentinel-2 tile: (i) original resolution, (ii) external sources (ESRI, Planet Labs, Google Earth), and (iii) super-resolution outputs from interpolation and deep learning methods (Bilinear, SwinIR, ESRGAN, HIT, Stable Diffusion).