

Space4iCity: Land Use/Land Cover Segmentation and Land-Surface Temperature Super-Resolution

Hamed Hemati, Michael Mommert

Stuttgart

Zusammenfassung

Analysis of multi-sensor satellite data, such as imagery from Sentinel-2 and Sentinel-3, is essential for monitoring Earth's surface dynamics, including Land Use/Land Cover (LULC) changes and Land Surface Temperature (LST) patterns. However, practical applications face significant challenges: LULC segmentation accuracy can be compromised by temporal mismatches between satellite image acquisition and label validity, while LST data often suffers from low spatial resolution and performance degradation under domain shifts (e.g., varying atmospheric or seasonal conditions). This work addresses these challenges through two distinct methodological contributions.

First, for LULC segmentation we use Sentinel-2 data and train U-Net-based architectures with the help of the Ben-Ge dataset labels. Additionally, we introduce a computationally inexpensive, data-driven label refinement technique. By leveraging the Normalized Difference Vegetation Index (NDVI) calculated from the input image, we automatically correct potentially outdated labels for dynamic classes like cropland and grassland, improving the consistency between image features and training labels. Experiments show this refinement enhances segmentation performance of spectrally ambiguous classes like "bare/sparse vegetation".

Second, for LST super-resolution of coarse Sentinel-3 data guided by auxiliary Sentinel-2 imagery, we propose a lightweight and robust Mixture of Linear Experts (MoLE) model. This approach utilizes an ensemble of efficiently trainable Ridge Regression models, each specializing in specific conditions represented by different training dates. At inference time, a dynamic expert selection or averaging strategy, based on the Wasserstein distance between LST histograms of the input and training data, allows the model to adapt to the current input conditions, thereby enhancing robustness against domain shifts while maintaining high computational efficiency suitable for CPU execution.

Together, these contributions demonstrate effective strategies for improving the reliability and accuracy of critical information derived from satellite remote sensing data, addressing common issues related to label quality and domain adaptation in segmentation and super-resolution tasks, respectively.

Inhaltsverzeichnis

1	Vegetation Segmentation	3
1.1	Problem Formulation	3
1.2	Data	3
1.3	Methodology	3
1.4	Results	6
1.5	Conclusion	7
2	Land Surface Temperature Super-Resolution	8
2.1	Introduction and Motivation	8
2.2	Problem Formulation	8
2.3	Data and Preprocessing	9
2.4	Methodology: Mixture of Linear Experts for LST Super-Resolution	10
2.5	Results	11
2.6	Conclusion	12
	Literatur	14

1 Vegetation Segmentation

Accurate and automated Land Use/Land Cover (LULC) segmentation is crucial for monitoring environmental changes across the globe. Deep learning models, particularly Convolutional Neural Networks (CNNs) applied to multispectral satellite imagery like Sentinel-2, have demonstrated state-of-the-art performance in this domain [1, 2]. This section details our approach to LULC segmentation, focusing on the U-Net architecture and addressing the common challenge of imperfect or outdated training labels through a novel data-driven refinement technique.

1.1 Problem Formulation

The objective is semantic segmentation of multispectral satellite images. Given an input image patch $X \in \mathbf{R}^{H \times W \times C}$, where H and W are the height and width, and C is the number of spectral bands, the goal is to predict a corresponding LULC map $Y \in \{0, 1, \dots, K - 1\}^{H \times W}$. Each pixel (i, j) in Y is assigned a class label k from a set of K predefined LULC categories. We aim to learn a mapping function $f_\theta : X \mapsto Y$, parameterized by θ , typically implemented as a deep neural network, that minimizes a suitable loss function between the predicted segmentation map $\hat{Y} = f_\theta(X)$ and the ground truth map Y .

1.2 Data

We utilize the **ben-ge dataset** [3] for training and evaluation. This dataset pairs Sentinel-2 multispectral imagery, primarily acquired in 2017 and 2018 from the BigEarthNet dataset [4], with LULC labels derived from the ESA WorldCover product generated for the year 2021. This temporal mismatch between imagery and labels introduces potential inaccuracies, particularly in dynamic areas like agricultural land, motivating our label refinement strategy.

The input satellite imagery consists of four Sentinel-2 bands:

- Band 2 (Blue, 490 nm)
- Band 3 (Green, 560 nm)
- Band 4 (Red, 665 nm)
- Band 8 (Near Infrared - NIR, 842 nm)

Thus, the input tensor X has $C = 4$ channels.

The ben-ge dataset utilizes a subset of the ESA WorldCover schema, containing 8 distinct LULC classes relevant to the covered regions. Notably, three classes (“shrubland”, “bare/sparse vegetation”, and “herbaceous wetland”) initially constitute less than 1% of the total labeled pixels in the original dataset, posing a challenge for standard training procedures due to class imbalance.

1.3 Methodology

Our methodology employs the U-Net architecture [5] for segmentation and introduces a label refinement technique to enhance training data quality. We conduct comparative experiments using both the original and refined datasets.

1.3.1 Model Architecture: U-Net

Our adopted U-Net consists of:

- An **encoder path** (contracting path) that captures context through successive convolutional and max-pooling layers, progressively reducing spatial resolution while increasing feature channels.
- A **decoder path** (expansive path) that enables precise localization through transposed convolutions (upsampling) combined with feature maps from the corresponding encoder layer via skip connections. These skip connections allow the network to reuse high-resolution features from the encoder, which is crucial for detailed segmentation.
- A final 1×1 convolution layer maps the feature vector at each pixel to the desired number of output classes ($K = 11$).

The U-Net is well-suited for image segmentation tasks and has proven effective for satellite image analysis due to its ability to integrate multi-scale contextual information with fine-grained spatial detail. In Figure 1 we show an example of such segmentation.

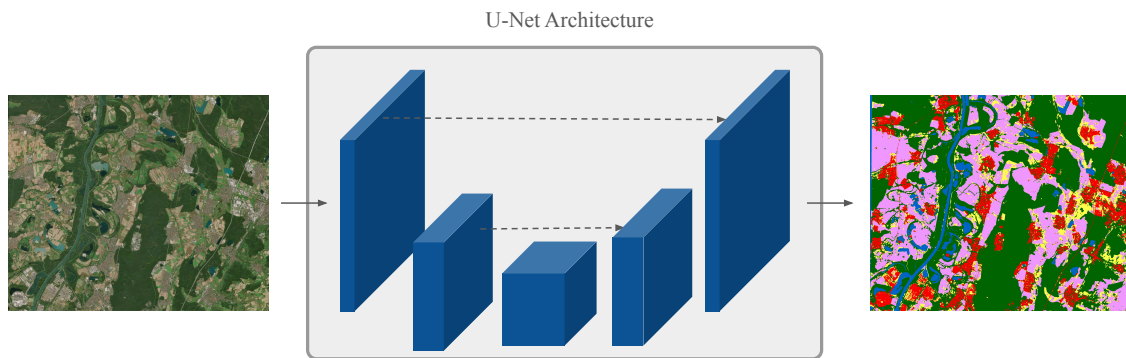


Abb. 1: Illustration of the U-Net architecture used for LULC segmentation.

1.3.2 Label Refinement Technique

Datasets combining satellite imagery with LULC labels often suffer from temporal discrepancies. Labels valid for one point in time may not accurately reflect the land cover state at the exact time the image was captured, especially for dynamic classes affected by agriculture (plowing), seasonal changes (drying grass), or other natural events. While a plowed field remains semantically "cropland", its spectral signature resembles bare soil, potentially confusing a model trained to distinguish between actual bare soil areas and temporarily bare agricultural land based solely on outdated labels.

To address this, we introduce a data-driven label refinement technique based on the **Normalized Difference Vegetation Index (NDVI)**, a common indicator of live green vegetation:

$$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{Band\ 8 - Band\ 4}{Band\ 8 + Band\ 4}$$

The refinement rule is applied pixel-wise to the original ben-ge labels:

Algorithm 1 NDVI-based Label Refinement

Require: Training set labels Y , Sentinel-2 Bands 4 (Red) and 8 (NIR), NDVI threshold τ

Ensure: Refined labels Y_{refined}

```

1: for each pixel  $(i, j)$  in the training set do
2:   Calculate  $NDVI(i, j) \leftarrow \frac{NIR(i, j) - Red(i, j)}{NIR(i, j) + Red(i, j)}$ 
3:   if  $(Y(i, j) = \text{"Cropland"} \vee Y(i, j) = \text{"Grassland"}) \wedge (NDVI(i, j) < \tau)$  then
4:      $Y_{\text{refined}}(i, j) \leftarrow \text{"Bare/sparse vegetation"}$ 
5:   else
6:      $Y_{\text{refined}}(i, j) \leftarrow Y(i, j)$ 
7:   end if
8: end for
9: return  $Y_{\text{refined}}$ 

```

Here, τ is a predefined threshold indicating low vegetation presence (a value like 0.2 or 0.3 is often used, but should be specified if determined). This process generates a *refined* version of the ben-ge training labels, aiming for better alignment between the spectral information in the image and the corresponding class label at the time of image acquisition. This refinement guides the learning process by providing more consistent examples, particularly improving the model’s understanding of spectrally similar but semantically distinct areas under different temporal conditions.

1.3.3 Experimental Setup

We evaluate the impact of our label refinement technique through two primary experiments:

1. **Baseline:** A U-Net model trained and evaluated using the *original* ben-ge dataset labels.
2. **Refined:** A U-Net model trained and evaluated using the *refined* ben-ge dataset labels generated by our NDVI-based method.

To ensure robustness of the results, we perform three independent training runs for each experiment, initializing the model with different random seeds. We report the mean and standard deviation of performance metrics across these three runs.

Evaluation is performed using the F1-score, considering:

- **Macro F1:** Unweighted mean of the F1-scores across all K classes. Sensitive to performance on rare classes.
- **Micro F1:** Calculated globally by counting the total true positives, false negatives, and false positives across all classes. Equivalent to pixel accuracy in multi-class settings.
- **Weighted F1:** Mean of the F1-scores weighted by the support (number of true instances) for each class. Reflects performance on more frequent classes.
- **Per-class F1:** F1-score for each individual LULC class.

1.4 Results

The experimental results demonstrate the effectiveness of the proposed label refinement technique in improving segmentation performance, particularly for classes affected by temporal changes.

1.4.1 Quantitative Analysis

Table 1 summarizes the key performance metrics on the test set for the baseline (original labels) and refined label experiments.

Tab. 1: Comparison of segmentation performance (Mean F1-score $\pm$ Std Dev across 3 runs) on the ben-ge test set using original vs. refined labels.

Metric	Baseline (Original Labels)	Refined Labels
Macro F1	59% $\pm$ 1%	65% $\pm$ 1%
Micro F1	91.5% $\pm$ 0.4%	92.1% $\pm$ 0.3%
Weighted F1	91.4% $\pm$ 0.4%	91.8% $\pm$ 0.3%
<i>F1: Bare/sparse veg.</i>	34% $\pm$ 6%	94.3% $\pm$ 0.1%
<i>F1: Grassland</i>	93.5% $\pm$ 0.3%	94.2% $\pm$ 0.3%
<i>F1: Cropland</i>	76% $\pm$ 1%	61% $\pm$ 2%

Key observations include:

- The **Macro F1 score increases significantly** from 59% to 65% with refined labels, indicating substantially improved performance on average across all classes, especially benefiting rarer ones.
- Micro F1 and Weighted F1 scores show marginal increases, suggesting that the overall pixel accuracy and performance on dominant classes are only slightly affected.
- The most dramatic improvement is seen for the **"Bare/sparse vegetation"** class, whose F1 score jumps from 34% to 94.3%. This is strongly linked to the label refinement process, which increased the representation (support) of this class in the training data from 0.017% to 4.5% by correctly re-labeling areas that appeared bare in the imagery.
- The "Grassland" class shows a slight improvement (93.5% to 94.2%).
- Conversely, the F1 score for the **"Cropland"** class decreases from 76% to 61%. This is attributed to the refinement process reducing the support for this class (from 7.3% to 4.2% of pixels) as many cropland areas exhibiting low NDVI were re-labeled. While potentially reducing the score for the original semantic class, this reflects a better alignment with the image's spectral reality at the time of capture.
- Analysis of confusion matrices (not shown) reveals reduced confusion between classes for the model trained on refined labels.

1.4.2 Qualitative Analysis

Visual inspection of the segmentation outputs on the test set supports the quantitative findings. The model trained with refined labels produces segmentation maps that show much better agreement with the visual appearance of the corresponding Sentinel-2 images, especially in agricultural regions. Areas that are visibly bare or sparsely vegetated in the image are more consistently and confidently classified as "Bare/sparse vegetation" compared to the baseline model, which often misclassifies them based on the outdated original labels. Figures 2 and 3 illustrate corrected examples of grassland and cropland respectively.

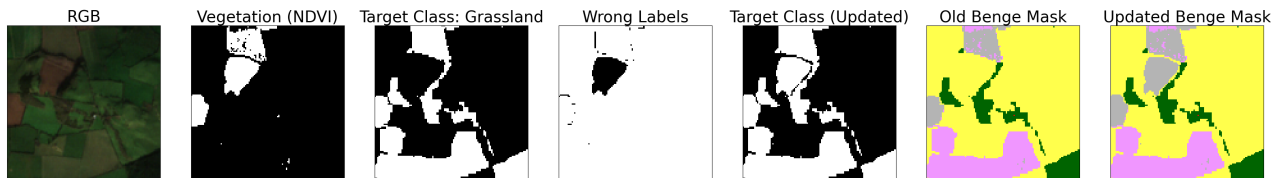


Abb. 2: Updated ben-ge LULC labels after correcting wrong grassland parts.

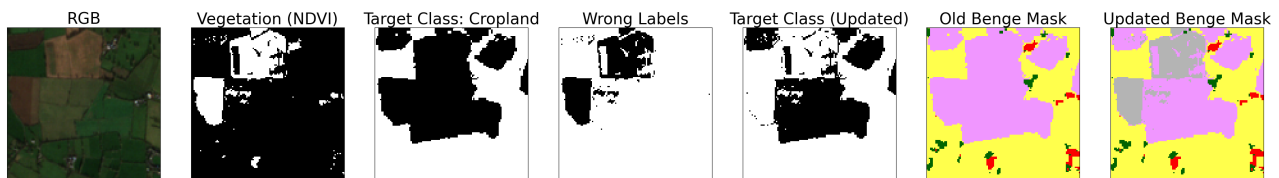


Abb. 3: Updated ben-ge LULC labels after correcting wrong cropland parts.

The increased confidence in identifying “Bare/sparse vegetation” is a key qualitative improvement, enabling reliable mapping of a class that was poorly recognized by the baseline model due to its initial rarity and inconsistent labeling in the original dataset.

1.5 Conclusion

The results demonstrate that our simple, data-driven label refinement technique enhances LULC segmentation performance when dealing with datasets exhibiting temporal mismatches between imagery and labels. By leveraging NDVI calculated directly from the input Sentinel-2 data, we can automatically correct labels for dynamic classes like “Cropland” and “Grassland” to better reflect their appearance at the image acquisition time. This leads to substantial gains in Macro F1 score and drastically improves the identification of underrepresented but spectrally distinct classes like “Bare/sparse vegetation”.

While the refinement causes a drop in the reported F1 score for the original “Cropland” class definition, this reflects the model learning a representation more consistent with the spectral data it observes. The method requires minimal computational resources and no external data beyond the input imagery itself. Future work could explore using other spectral indices to refine labels related to different phenomena (e.g., water indices for temporary water bodies) or incorporating temporal information more explicitly into the model architecture.

2 Land Surface Temperature Super-Resolution

2.1 Introduction and Motivation

Predicting Land Surface Temperature (LST) at high spatial resolution provides valuable insights for diverse applications, including urban planning (e.g., urban heat island effect), precision agriculture, environmental monitoring, and construction management. While satellite remote sensing, particularly from platforms like Sentinel-3, offers global LST measurements, these are often available at coarse spatial resolutions (typically 1 kilometer for Sentinel-3 SLSTR). This limits their direct use in applications requiring finer detail (e.g., 10-100 meters). Consequently, considerable research effort focuses on developing super-resolution (SR) techniques to enhance the spatial resolution of satellite-derived LST maps, often aiming for an increase of one or two orders of magnitude [6–8].

Common approaches leverage deep learning, particularly Convolutional Neural Network (CNN) based architectures like U-Nets, adapted for regression tasks. While most capable models are proprietary, the open source options can be prone to inaccuracies, generate visual artifacts, and suffer from overfitting. A critical challenge arises from domain shifts: variations in atmospheric conditions, seasonality, or instrument calibration between training data and new input data can lead to performance degradation, especially when training data is scarce or does not fully represent the target domain’s variability. Complex models often require substantial computational resources for training and retraining, making adaptation to new data streams costly.

Addressing these limitations, this section proposes a **lightweight, ensemble-based super-resolution model** specifically designed for Sentinel-3 LST data. Our approach prioritizes robustness against domain shifts and computational efficiency. Key contributions include:

- An ensemble model composed of simple, efficiently trainable experts.
- A distance-based model selection mechanism at inference time to adapt predictions to the input data characteristics.
- Demonstrated improvement in prediction accuracy, particularly under conditions of domain shift, compared to naive baselines.
- High computational efficiency, allowing for training and adaptation on standard CPUs without costly retraining of the entire system.

2.2 Problem Formulation

The LST super-resolution task aims to estimate a high-resolution LST map $Y \in \mathbf{R}^{H \times W}$ from a corresponding low-resolution LST input $x_{S3} \in \mathbf{R}^{h \times w}$, where $H > h$ and $W > w$. In our approach, we leverage auxiliary high-resolution data, specifically multispectral imagery from Sentinel-2, denoted as $x_{S2} \in \mathbf{R}^{H \times W \times C'}$, where C' is the number of Sentinel-2 bands used (Blue, Green, Red, NIR; $C' = 4$).

Unlike end-to-end deep learning models that might directly learn the mapping $f : x_{S3} \rightarrow Y$, our strategy involves two stages:

1. **Upsampling:** The low-resolution Sentinel-3 LST map x_{S3} is first upsampled to match the target high resolution $H \times W$, resulting in $x'_{S3} \in \mathbf{R}^{H \times W}$. This is achieved using a

simple interpolation method (Nearest Neighbor, as indicated by `order=0` in the code's `resize` function).

2. **Pixel-wise Value Refinement:** A model g is trained to predict the high-resolution LST value $Y(i, j)$ for each pixel (i, j) based on features derived from the upsampled Sentinel-3 LST $x'_{S3}(i, j)$ and the corresponding high-resolution auxiliary Sentinel-2 data $x_{S2}(i, j, :)$.

$$Y(i, j) = g(x'_{S3}(i, j), x_{S2}(i, j, 1), \dots, x_{S2}(i, j, C'))$$

The core of our method lies in the specific implementation of the refinement model g as an adaptive ensemble.

2.3 Data and Preprocessing

Our experiments utilize the following datasets:

- **Low-Resolution Input:** Sentinel-3 SLSTR LST data (e.g., `sent3_*.tif`), originally at approx. 1 km resolution.
- **High-Resolution Auxiliary Input:** Sentinel-2 MSI data (e.g., `sent2_*.tif`) at 10m resolution, specifically Bands 2 (Blue), 3 (Green), 4 (Red), and 8 (NIR). A single Sentinel-2 image (`sent2_2024-08-28.tif`) is used as a static high-resolution reference (`ref_rgb`) across different dates in this implementation.
- **High-Resolution Target:** Ground Truth (GT) LST data (e.g., `gt_*.tif`), assumed to be available at a higher resolution corresponding to the target Sentinel-2 resolution (e.g., derived from airborne campaigns, higher-resolution sensors like Landsat, or model outputs).

Data from specific dates are used for training and testing, e.g.:

- Training: `date_list_train = ["2024-01-27", ..., "2024-12-26"]`
- Testing: `date_list_test = ["2024-03-20", ..., "2024-09-19"]`

allowing evaluation under potential domain shifts represented by the test dates.

Preprocessing steps are critical:

- **Resampling/Alignment:** All datasets are resampled or cropped to a common high-resolution grid ($H \times W$) defined by the reference Sentinel-2 image. Sentinel-3 LST is upsampled using nearest neighbor interpolation `resize(..., order=0)`. GT LST is resized using potentially higher-order interpolation `resize(..., antialiasing=True)`.
- **Unit Conversion:** Sentinel-3 LST is converted from Kelvin to Celsius by subtracting 273.15.
- **Normalization:** LST data (upsampled S3 and GT) are normalized to the range $[0, 1]$ using predefined global minimum and maximum values (`s3_min`, `s3_max`, `gt_min`, `gt_max`). Sentinel-2 bands are also normalized (e.g., divided by expected maximum reflectance values and clipped).

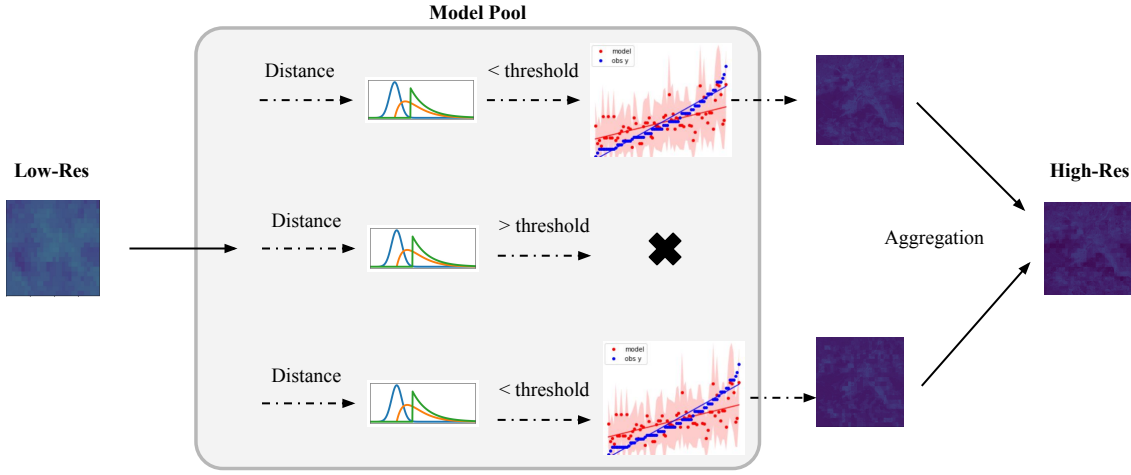


Abb. 4: A schematic of the Mixture of Linear Experts model for pixel-wise mapping of values for super-resolution.

- **Feature Vector Creation:** For each pixel (i, j) , a 5-dimensional feature vector $\mathbf{x}(i, j)$ is constructed by concatenating the normalized, upsampled S3 LST value $x'_{S3}(i, j)$ and the four normalized Sentinel-2 band values $x_{S2}(i, j, :)$. This vector forms the input to the refinement model g .

$$\mathbf{x}(i, j) = [x'_{S3}(i, j), x_{S2}(i, j, B), x_{S2}(i, j, G), x_{S2}(i, j, R), x_{S2}(i, j, N)]^T$$

2.4 Methodology: Mixture of Linear Experts for LST Super-Resolution

We propose a Mixture of Experts (MoE) approach where the “experts” are simple linear models, trained efficiently and selected dynamically at inference time based on data characteristics. In Figure 4 we show a schematic of the proposed model.

2.4.1 Expert Model Architecture: Ridge Regression

Each expert in our ensemble is a **Ridge Regression** model. This is a linear regression model that incorporates L2 regularization to prevent overfitting and improve generalization, especially important when dealing with potentially collinear features (like multispectral bands). For a given pixel (i, j) with feature vector $\mathbf{x}(i, j)$, the expert k predicts the normalized high-resolution LST value $y_{pred}^{(k)}(i, j)$ as:

$$y_{pred}^{(k)}(i, j) = \mathbf{w}_k^T \mathbf{x}(i, j) + b_k$$

where $\mathbf{w}_k$ is the weight vector and b_k is the intercept (bias) learned by expert k . The Ridge model minimizes the squared error loss with an added penalty term proportional to the squared magnitude of the weights $\mathbf{w}_k$:

$$\min_{\mathbf{w}_k, b_k} \sum_{(i, j)} \|(\mathbf{w}_k^T \mathbf{x}(i, j) + b_k) - Y(i, j)\|^2 + \alpha \|\mathbf{w}_k\|^2$$

where α is the regularization strength (e.g., `alpha=50.5` in the code). Using simple linear experts makes the model extremely lightweight and computationally inexpensive to train and run.

2.4.2 Ensemble Training

Instead of training a single complex model on all data, we train an ensemble of experts. Specifically, **one independent Ridge Regression expert is trained for each date** (or scene) available in the training set (`date_list_train`). Each expert k learns the relationship between the 5D feature vector and the target GT LST specifically for the conditions prevalent on its corresponding training date d_k . This training process is highly parallelizable and efficient, suitable for execution on standard CPUs. The result is a list of trained models (`model_list`), where `model_list[k]` corresponds to the expert trained on data from `date_list_train[k]`.

2.4.3 Distance-Based Inference and Domain Adaptation

To predict the high-resolution LST for a new, unseen low-resolution Sentinel-3 input image x_{S3}^{new} , we employ a dynamic model selection or averaging strategy based on data similarity. This mechanism provides robustness to domain shifts:

1. **Calculate Input Histogram:** Compute the histogram of the normalized, upsampled LST values $x'_{S3,new}$ from the new input image.
2. **Compute Distances:** Calculate the distance between the input histogram and the pre-computed histograms of the upsampled S3 LST data from each training date d_k . The code uses the **Wasserstein distance** (also known as Earth Mover's Distance or EMD) which measures the minimum "cost" to transform one distribution into another, providing a robust measure of similarity between the LST distributions. Let D_k be the distance between the new input's LST histogram and the histogram corresponding to expert k .
3. **Prediction Strategy:** Two strategies are possible:
 - a) **Single Best Expert (`ensemble_prediction=False`):** Identify the expert k^* with the minimum distance: $k^* = \arg \min_k D_k$. The final prediction Y_{pred} is generated solely by this single, best-matching expert: $Y_{pred}(i, j) = \mathbf{w}_{k^*}^T \mathbf{x}_{new}(i, j) + b_{k^*}$.
 - b) **Ensemble Averaging (`ensemble_prediction=True`):** Select all experts k whose corresponding distance D_k is below a predefined **threshold**. Generate predictions using each selected expert, and compute the final prediction Y_{pred} by averaging the outputs of these selected experts pixel-wise.

$$Y_{pred}(i, j) = \frac{1}{|S|} \sum_{k \in S} (\mathbf{w}_k^T \mathbf{x}_{new}(i, j) + b_k) \quad \text{where } S = \{k \mid D_k < \text{threshold}\}$$

This distance-based selection/averaging allows the model to adapt its prediction based on which historical training conditions are most similar to the current input data, effectively handling domain shifts captured by variations in the overall LST distribution.

2.5 Results

This section presents the evaluation of the proposed Mixture of Linear Experts (MoLE) model for LST super-resolution. We compare its performance against baseline methods and analyze its effectiveness, particularly on test dates representing domain shifts.

2.5.1 Quantitative Analysis

Table 2 below summarizes the performance of nine distinct models in the ensemble. Each model was trained and evaluated on its corresponding temporal data group, and the results reflect the model’s effectiveness on that specific dataset.

Overall, the models demonstrate strong performance, achieving consistently low Mean Squared Error (MSE) values, generally on the order of 10^{-3} to 10^{-4} . This indicates a high degree of precision and low average error in the super-resolved outputs.

While the coefficient of determination (R^2) shows some variability, the model for Group 7 performed well, explaining nearly 50% of the variance ($R^2 \approx 0.493$) with a minimal MSE of 0.00065. The less effective performance in groups 3 and 8 likely points to unique data characteristics on those dates rather than a general failure of the modeling approach, which is validated by the low MSE across the board.

Tab. 2: Performance Metrics of the Trained Models per Group

Group ID	Date	MSE	R^2
0	2024-01-27	0.00086	0.456
1	2024-02-29	0.00089	0.272
2	2024-04-29	0.00117	0.394
3	2024-05-10	0.00245	-0.037
4	2024-07-29	0.00125	0.447
5	2024-08-28	0.00104	0.485
6	2024-10-29	0.00213	0.192
7	2024-11-29	0.00065	0.493
8	2024-12-26	0.00066	-0.063

2.5.2 Qualitative Analysis

To visually assess the performance of our super-resolution model, we present a side-by-side comparison of the model’s output against the original low-resolution input and the high-resolution ground truth. Figures 5 and 6 displays two randomly selected cases from the test set.

In both cases, it is evident that the model successfully enhances the spatial resolution of the original Sentinel-3 LST maps. The predicted images are visibly sharper and contain significantly more detail than the coarse input. Crucially, the model effectively preserves the structural integrity and spatial patterns present in the ground truth data. Fine details, such as the layout of agricultural fields and the distinct thermal signatures of urban areas, are reconstructed with high fidelity. This visual evidence confirms that our approach not only increases resolution but also maintains the essential contextual and structural information of the landscape.

2.6 Conclusion

In this section, we introduced a lightweight Mixture of Linear Experts (MoLE) approach for the super-resolution of Land Surface Temperature maps. By combining an ensemble of simple Ridge Regression models with a dynamic, distance-based selection mechanism at inference, our

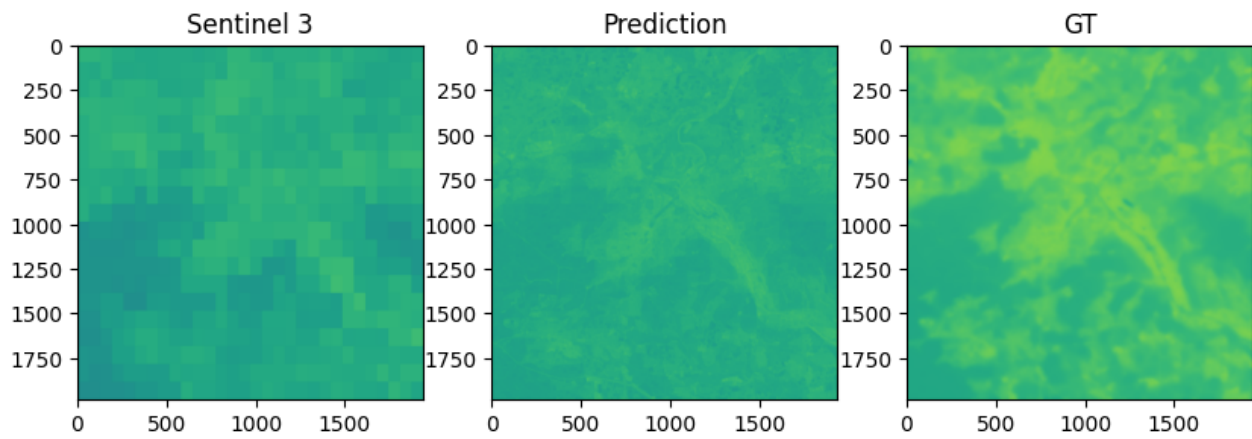


Abb. 5: Super-resolution: Example 1

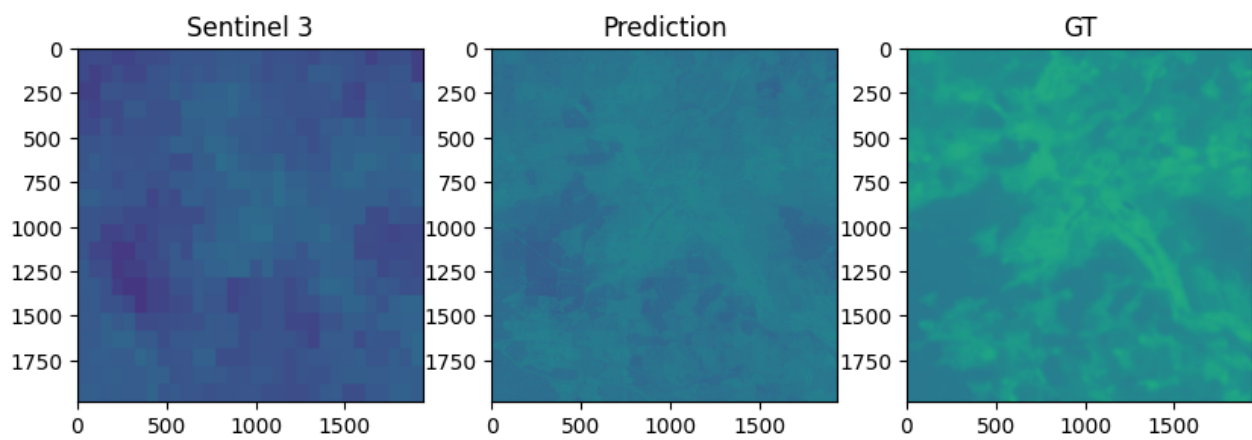


Abb. 6: Super-resolution: Example 2

method demonstrated a promising ability to enhance spatial resolution while adapting to the distributional characteristics of new input data. The quantitative results showed consistently low reconstruction errors, and the qualitative analysis confirmed that essential structural details were well-preserved. While the approach proved computationally efficient and robust, the variability in performance across different test dates suggests that its effectiveness is still sensitive to the diversity of the training ensemble. This work presents a viable, adaptable alternative to monolithic deep learning models, though future investigation into more advanced expert models and dynamic updating strategies may yield further improvements.

Literatur

1. ZHU, X. X. u. a.: Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*. 2017, Jg. 5, Nr. 4, S. 8–36. Abger. unter DOI: 10.1109/MGRS.2017.2762307.
2. HELBER, P. u. a.: EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. In: 2019, Bd. 12, S. 2217–2226. Nr. 7. Abger. unter DOI: 10.1109/JSTARS.2019.2918242.
3. MOMMERT, M. u. a.: Ben-ge: Extending BigEarthNet with geographical and environmental data. In: *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, S. 1016–1019.
4. SUMBUL, G. u. a.: BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*. 2021, Jg. 9, Nr. 3, S. 174–180.
5. RONNEBERGER, O. u. a.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, Cham, 2015, Bd. 9351, S. 234–241. Lecture Notes in Computer Science. Abger. unter DOI: 10.1007/978-3-319-24574-4_28.
6. LI, Z.-L. u. a.: Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*. 2013, Jg. 131, S. 14–37. Abger. unter DOI: 10.1016/j.rse.2012.12.008.
7. GHENT, D. u. a.: Global Land Surface Temperature from the Along-Track Scanning Radiometers. *Journal of Geophysical Research: Atmospheres*. 2017, Jg. 122, Nr. 23, S. 12, 167–12, 193. Abger. unter DOI: 10.1002/2017JD027161.
8. WU, P. u. a.: A New Land Surface Temperature Downscaling Strategy Based on Deep Learning and Its Application in the Heihe River Basin. *Journal of Geophysical Research: Atmospheres*. 2019, Jg. 124, Nr. 21, S. 11385–11403. Abger. unter DOI: 10.1029/2019JD031086.