PATCH-WISE GROUND-LEVEL NO_2 POLLUTION ESTIMATION

Gonzalez Avilés Ruben Octavio

University of St. Gallen HSG

ABSTRACT

This paper addresses the critical environmental challenge of estimating ground-level Nitrogen Dioxide (NO_2) concentrations, a key issue in public health and environmental policy. Motivated by the limitations of existing methods, particularly the computational intensity of point-wise estimates in satellite-based monitoring, this study introduces an innovative patch-wise technique. Our approach seeks to balance the accuracy of high-resolution estimates with the practicality of computational constraints, thereby enabling efficient and scalable global environmental monitoring. Utilizing deep learning algorithms and remote sensing data from satellites such as Sentinel-2 and Sentinel-5P, the patch-wise method generates a grid of estimates in a single inference, significantly reducing the computational resources required to provide estimates for larger areas. Notably, our approach surpasses the results of existing point-wise methods by a significant margin, achieving a Mean Absolute Error (MAE) of 4.98 $\mu g/m^3$. This demonstrates both high accuracy and computational efficiency, highlighting the applicability of our method for global environmental monitoring. Furthermore, the results indicate the method's adaptability and robustness, particularly in its successful application to diverse geographic regions. By offering a viable solution to the computational challenges of large-scale environmental monitoring, this approach represents a substantial contribution to the field of remote sensing and environmental science.

For further insights and practical implementation, the source code related to this research can be found on Github.

***Index Terms*—** Nitrogen Dioxide Estimation, NO2, Deep Learning, Remote Sensing, Air Pollution, Environmental Monitoring

1. INTRODUCTION

Air quality degradation, particularly due to NO_2, poses a significant risk to environmental sustainability and public health. This toxic and highly reactive gas, primarily emitted from the combustion of fossil fuels, is a major contributor to air pollution and climate change [1]. Moreover, NO_2 exposure is associated with a range of health issues, as highlighted by various studies demonstrating the correlation between NO_2 levels and the prevalence of respiratory and cardiovascular diseases [2]. Therefore, the accurate monitoring of NO_2 levels is crucial to safeguard public health and to support the development of effective policies aimed at mitigating its detrimental effects.

For these reasons, NO_2 concentration monitoring has been conducted through localized ground-based air quality stations. While these stations provide accurate and reliable data, they offer limited spatial coverage and entail high operational costs. These factors result in various notable data gaps, particularly in remote or less developed areas [3].

To address these limitations, researchers have increasingly turned to satellite-based remote sensing as a viable alternative for comprehensive environmental monitoring [4, 5]. Satellites such as Sentinel-5P provide extensive global coverage, enabling the monitoring of atmospheric constituents, including NO_2, on a much broader scale [6]. However, a significant challenge persists in effectively converting satellite observations into precise ground-level pollutant concentrations. This difficulty is particularly pronounced when estimating over extensive areas, due to the inherent spatial sparsity of available ground truth data, which are often confined to localized measurements.

This paper presents a novel method for ground-level NO_2 estimation, utilizing a patch-wise technique that leverages the capabilities of advanced deep learning models and the comprehensive data acquired from multispectral satellite imagery. This approach employs a unique sampling technique to address the spatial sparsity of ground truth data, ensuring data continuity in areas without direct measurements. The patch-wise approach efficiently computes estimates for larger areas, significantly reducing the computational demands compared to individual point-wise estimations and enabling broader, more frequent monitoring. This method marks a notable improvement in environmental monitoring, providing a scalable, efficient solution for global NO_2 estimation, which can in turn support enhanced environmental policy-making.

2. RELATED WORK

Local monitoring of NO_2 concentrations, primarily through ground-based stations, has been the traditional method for assessing air quality. Despite their precision, these localized stations face significant limitations, including their inability to provide comprehensive spatial coverage. This results in notable data blind spots, especially in remote and underdevel-

oped regions where setting up and maintaining such stations is logistically challenging and costly. Complementing this, the utilization of satellite data for environmental monitoring introduces its own set of complexities. While satellite imagery offers expansive geographical coverage, transforming these high-altitude observations into accurate, ground-level pollutant concentration estimates poses a significant challenge.

In the face of these limitations, researchers have increasingly turned to deep learning as a potential solution. Convolutional neural networks (CNNs), in particular, have demonstrated their efficacy in remote sensing applications, from land cover classification to air quality estimation [7]. These advancements in deep learning have opened up new possibilities for interpreting and utilizing satellite data more effectively for environmental monitoring purposes.

One notable example of such innovation is presented in the work of Scheibenreif et al. (2022) [4]. This study introduces a deep learning-based approach for estimating ground-level NO_2 concentrations using remote sensing data, which we will refer to as the point-wise approach. The point-wise approach excels in producing highly accurate NO_2 concentration estimates at specific global locations. Despite its precision and improvement over traditional methods, this approach is constrained by its focus on single spatial locations. When applied to expansive areas, the point-wise method necessitates repeated processing for each location, resulting in considerable computational demands for global-scale application.

Addressing the limitations inherent in the point-wise approach, our research introduces a more efficient alternative: the patch-wise approach. This novel method aims to maintain the accuracy of point-wise estimations while significantly reducing computational requirements. Instead of producing individual estimates, the patch-wise approach simultaneously generates a grid of estimates for a designated area, effectively decreasing the total number of iterations required for larger-scale NO_2 concentration assessments. The patch-wise approach thus adeptly strikes a balance between the need for accurate estimates with the practical limitations of computational resources.

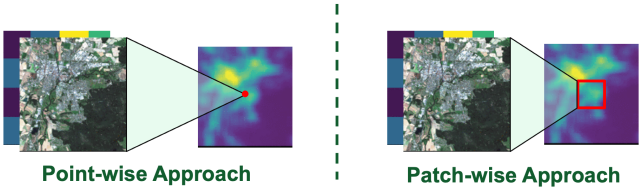

Fig. 1: Comparison of point-wise and patch-wise approaches for NO_2 concentration estimation from satellite imagery. The point-wise approach (left) generates predictions at a single location, denoted by the red dot. The patch-wise approach (right) estimates NO_2 concentrations over a designated area, indicated by the red square, enhancing computational efficiency for regional-scale estimations.

3. APPROACH

Figure 2 provides an overarching view of our approach to estimating ground-level NO_2 concentrations, combining advanced deep learning techniques with multispectral satellite imagery. In the following subsections, we will explore each component of this methodology in detail, explaining how they collectively contribute to an efficient and scalable environmental monitoring solution.

3.1. Handling Sparse Annotations with Uniformly Random Distributed Offset Sampling

3.1.1. The Challenge of Sparse Annotations

One of the primary challenges in estimating ground-level NO_2 concentrations from satellite imagery is the inherent sparsity of ground truth annotations. Typically, ground truth measurements are only available for a specific geographical location, which corresponds to the coordinates of the measurement station. Therefore, the ground truth only covers a single point within the expansive area captured by the satellite images. This sparsity presents a significant challenge in training models to accurately infer NO_2 concentrations across a broader area.

Another challenge within our dataset is the centralized positioning of ground truth annotations – each being at the center of the corresponding image. This unique setup poses a risk of the model developing a bias towards predicting accurate NO_2 levels only for the central pixel, while potentially yielding less reliable estimates for the remaining pixels in the image. Such a scenario is problematic for our objective of estimating NO_2 concentrations across an entire patch, rather than just at a single point. If unaddressed, this could result in a model that, while precise at the center, disperses random or less accurate estimations across the rest of the patch, undermining the goal of achieving comprehensive and uniform coverage in NO_2 concentration estimation.

3.1.2. Uniformly Random Offset Sampling Technique

To address this issue, we employed a novel sampling technique that introduces variability into the spatial relationship between the satellite imagery and the ground truth measurements. This method involves sampling each image with a fixed-size window, ensuring that every sample maintains the same spatial dimensions (128 x 128 pixels). However, instead of centering this window on the measurement location, we implemented a uniformly random offset. This offset shifts the sampling window from the origin by an amount sampled from a uniformly random distribution. Crucially, the range of this offset is constrained to ensure that the ground truth measurement locations are uniformly distributed over a designated area, henceforth referred to as the prediction area. This

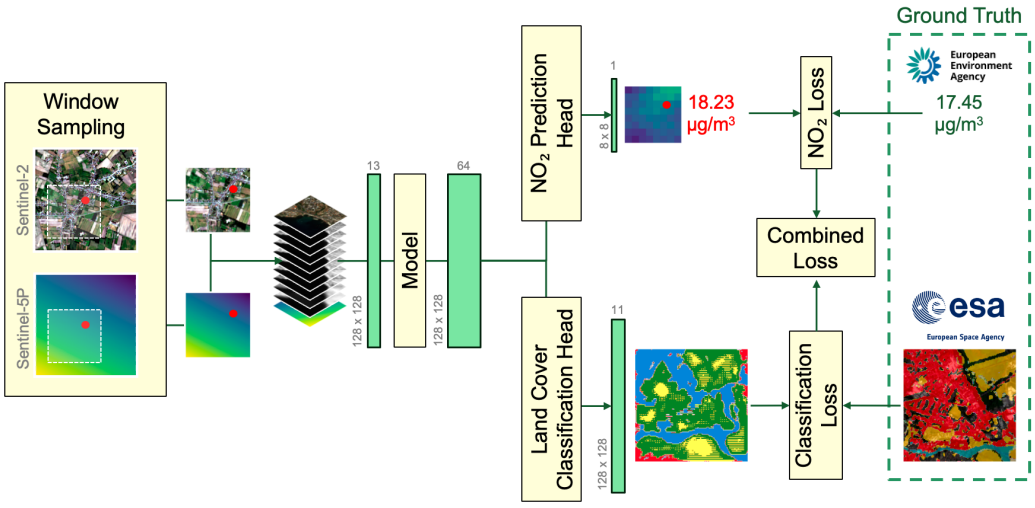

Fig. 2: High-level overview showcasing the process of estimating NO_2 concentrations using a deep learning model. The model inputs are randomly sampled windows consisting of satellite data from Sentinel-2 and Sentinel-5P, which then pass through a convolutional neural network to extract relevant features. The NO_2 Regression Head produces a concentration estimate, which is subsequently aligned with ground truth data for validation. Concurrently, the Land Cover Classification Head processes the same features to classify land cover, aiding in the interpretation of NO_2 distribution. Losses from both heads are combined to optimize the model, with the ultimate goal of providing accurate, high-resolution estimations of ground-level NO_2.

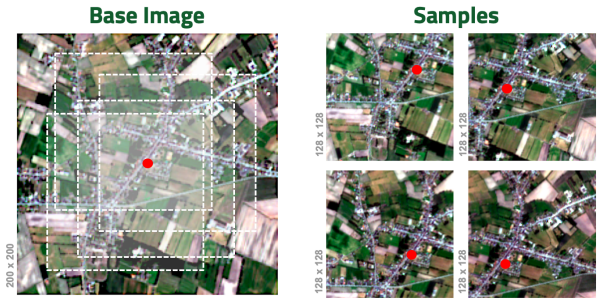

Fig. 3: Illustration of the uniformly random offset sampling technique. The base image (left) shows the full 200 x 200 pixels area from which samples are taken. Individual samples (right) are 128 x 128 pixels, each randomly offset within the larger image. This sampling strategy diversifies the pixel location of the measurement point while retaining the original geospatial station coordinates.

prediction area forms the scope of our model's prediction, ensuring that the model learns to estimate NO_2 concentrations across a broader and more representative area rather than being limited to a single central point.

Despite the random offset, each sample retains the information about the new pixel coordinates of the measurement location. This ensures that the ground truth remains accurately associated with the original geospatial location within the corresponding image data.

3.2. Model Architecture and Dual-Task Implementation

To experiment with various outcomes in estimating ground-level NO_2 concentrations, we developed two distinct neural network models, each exhibiting unique architectural capabilities. These models are designed to process and interpret satellite data, outputting results that are further utilized by two specialized heads for dedicated tasks.

3.2.1. Chosen Models

3.2.1.1. Model I - UNet

The first model is based on the UNet architecture, widely acclaimed for its effectiveness in image segmentation tasks. The UNet is characterized by its distinctive U-shaped design, incorporating a contracting path for capturing and downsampling features, complemented by an expansive path for reconstruction. In our adaptation of this architecture, padding has been purposefully incorporated, particularly within the contracting path. This modification ensures that the spatial dimensions of the input can be accurately reconstructed.

3.2.1.2. Model II - Autoencoder

The second model is an Autoencoder architecture, derived from the UNet but with a significant modification: the exclusion of skip connections. This simplification aims to test the model's capability in learning and reconstructing features without the direct transfer of information between corresponding layers of the encoder and decoder. Notably, the absence of skip connections in the Autoencoder avoids the

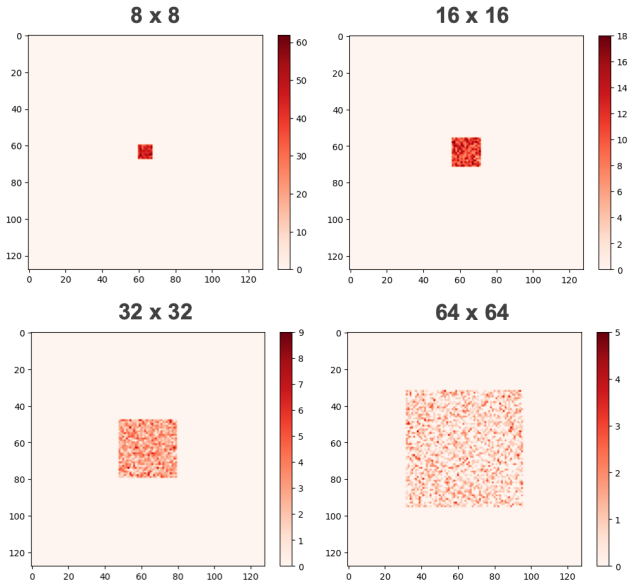

Fig. 4: Distribution of ground-truth pixel coordinates of 2871 samples within the sampling window for varying prediction space sizes.

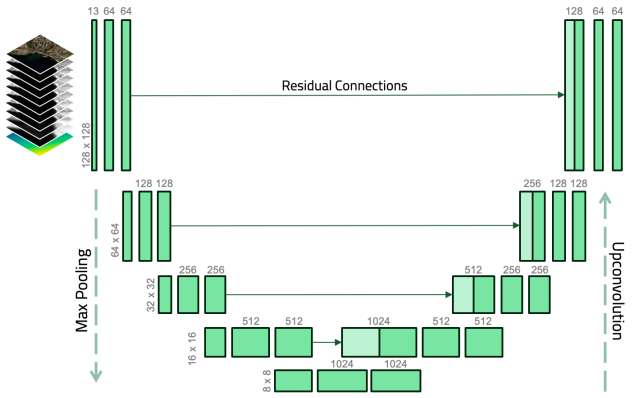

Fig. 5: Schematic of the UNet architecture showcasing the flow from multispectral input through contracting and expansive paths with residual connections and upconvolutions for feature map generation.

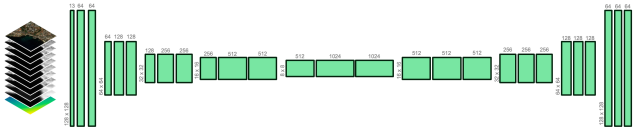

Fig. 6: Depiction of the Autoencoder architecture, adapted from the UNet, illustrating the flow of multispectral data through encoding and decoding sequences for feature map generation.

introduction of fine-grained features that the UNet might otherwise capture. This characteristic is particularly relevant for estimating NO_2 concentrations, where it is expected that NO_2 levels do not exhibit strong variations from pixel to pixel. By eliminating these connections, the Autoencoder can potentially provide a more generalized and smoother representation of the NO_2 distribution, which could be more accurate for this specific task.

3.2.2. Input and Output Specifications

Both the UNet and Autoencoder models process inputs comprising the 12 multispectral bands from Sentinel-2 data and an additional plane representing Sentinel-5P data. These inputs encompass the geographical area determined by the random sampling technique previously described in Section 3.1.2.

Each model, the UNet and the Autoencoder, generates a feature map with dimensions of 128 x 128 x 64 in its output. Both models benefit from the expanded depth of 64 channels in their outputs, which allows for a richer representation of processed satellite imagery. However, despite having identical dimensional outputs, each model offers a distinct contextual representation of the data.

3.2.3. Dual-Task Heads

Upon processing the input data, the outputs of these models are then fed into two distinct task heads, each serving a specific purpose:

3.2.3.1. NO_2 Regression Head

This head utilizes the model output to estimate the NO_2 concentration at the ground level. It focuses on producing a 2-dimensional regression plane, reflecting the spatial distribution of NO_2 concentrations across the sampled area. The output from this head is then further center-cropped to match the dimensions of the prediction area. This prediction area is determined by the distribution of measurement coordinates, a result of the random sampling technique as detailed in Section 3.1.2. Such cropping ensures that the model's output comprehensively encompasses the entire area where the ground truth data coordinates are distributed in, thereby aligning the predictions accurately with the spatial scope of the ground truth.

After the center-cropping process, the output of the head consists of a plane with the spatial dimensions of the prediction area and a depth of 1. Thus, the final output effectively covers the entire area of interest, providing a detailed map of ground-level NO_2 concentration predictions.

3.2.3.2. Land Cover Segmentation Head

The second head is responsible for classifying the land cover within the geographical area captured by the input sample. While the land cover segmentation results are utilized exclusively during the training phase and not in the inference process, they potentially enhance the model's learning. This approach is rooted in the hypothesis that NO_2 concentration may be closely related to the type of land cover. By training the model to recognize these land cover categories,

it is hypothesized that the model might develop a more nuanced understanding of potential NO_2 distribution patterns. This understanding, based on the underlying environmental context, could potentially improve the overall accuracy and effectiveness of the NO_2 concentration estimation. Utilizing the feature map generated by the models, this head outputs a 128 x 128 x 11-dimensional representation, where each pixel is categorized into one of 11 distinct land cover classes, thus providing a detailed prediction of the land cover diversity within the sampled area.

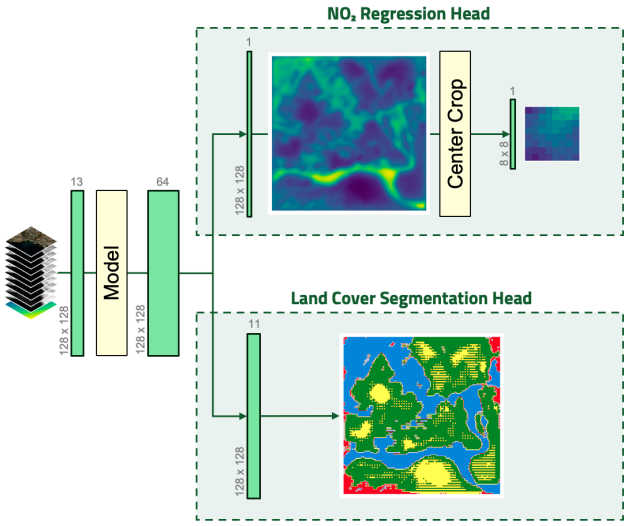

Fig. 7: Diagram showcasing the dual-task approach where one head is dedicated to the NO_2 estimation, and the other head focuses on land cover classification.

In summary, the dual-task approach of these models leverages two specialized heads – the NO_2 Regression Head and the Land Cover Segmentation Head – each contributing uniquely to the overall objective. The NO_2 Regression Head is crucial for providing a precise and detailed estimation of ground-level NO_2 concentrations, aligning its output with the prediction area defined by the ground truth data distribution. On the other hand, the Land Cover Segmentation Head, although used only during training, may play a vital role in enhancing the model's understanding of the relationship between land cover types and NO_2 concentrations. This dual-task framework enables the models to not only estimate NO_2 levels but also incorporate crucial environmental context, potentially leading to a more comprehensive and nuanced understanding of NO_2 distribution patterns. The synergy between these two heads thus aims to improve the precision and contextual relevance of our predictions.

3.3. Loss Function Formulation

A critical component of the model's training process is the choice of the loss function, which guides the optimization of the models. For this project, a combined loss function was implemented, tailored to accommodate the dual objectives of the NO_2 concentration estimation and land cover segmentation.

3.3.1. Loss Function Components

The combined loss function comprises two primary components: the Squared Error Loss for the NO_2 regression task, and the Cross Entropy Loss for the land cover segmentation task. Each component is designed to optimize the model's performance in its respective task.

3.3.1.1. Squared Error Loss (NO_2 Regression)

For the NO_2 concentration estimation, the Squared Error Loss (SEL) is used, mathematically represented as:

$$\text{SEL}(y, \hat{y}_{i,j}) = (y - \hat{y}_{i,j})^2 \tag{1}$$

where y is the ground truth NO_2 concentration and $\hat{y}_{i,j}$ is the model's predicted concentration at the pixel location (i, j). This loss function penalizes the discrepancies between the predicted and actual NO_2 levels, with a larger penalty for more significant errors.

3.3.1.2. Cross Entropy Loss (Land Cover Segmentation)

The Cross Entropy Loss (CEL) is employed for the land cover segmentation task, defined as:

$$\text{CEL}(\hat{c}, c) = -\sum_{i=1}^{C} w_i \cdot 1(c = i) \cdot \log\left(\frac{\exp(\hat{c}_i)}{\sum_{j=1}^{C} \exp(\hat{c}_j)}\right) \tag{2}$$

In this formula, $\hat{c}$ represents the predicted class probabilities, c is the true class, C is the number of classes, w_i are the class weights, and $1(c = i)$ is an indicator function that is 1 if $c = i$ and 0 otherwise. . This loss function is a weighted variant of the standard cross entropy loss, designed to account for the class imbalance present in the ground truth data. The weighting w_i for each class compensates for the prevalence of certain land cover types, ensuring a more balanced learning process and improving the model's ability to classify each pixel accurately into the correct land cover category.

3.3.2. Combined Loss Function

The final loss function used in the model training is a weighted sum of these two components:

$$\text{L}(y, \hat{y}_{i,j}, c, \hat{c}, \lambda) = \text{SEL}(y, \hat{y}_{i,j}) + \lambda \times \text{CEL}(\hat{c}, c) \tag{3}$$

where λ is a regularization parameter that balances the contribution of each loss component. This combined loss

function allows simultaneous optimization for both NO_2 concentration estimation and land cover segmentation, ensuring that the model is effectively trained to perform both tasks.

The design of this combined loss function is instrumental in guiding the model to achieve the dual objectives of this approach. The Squared Error Loss ensures that the model is sensitive to the variations in NO_2 concentrations, while the Cross Entropy Loss enhances the model's capability in understanding the environmental context through accurate land cover classification. The use of the regularization parameter λ provides flexibility in balancing these two aspects, enabling a tailored approach to model training based on the specific characteristics of the dataset and model.

4. EXPERIMENTAL RESULTS

4.1. Data Used for Training

A pivotal aspect of the model's performance is the nature and quality of the training data. This project utilizes a comprehensive dataset, encompassing ground-level NO_2 measurements and satellite imagery, to ensure a robust and effective training process.

4.1.1. Ground-Level NO_2 Measurements

The core of the training data consists of surface-level NO_2 concentration measurements, expressed in micrograms per cubic meter ($\mu g/m^3$). These measurements were sourced from 3087 unique locations, providing a diverse and geographically varied dataset. The data consists of measurements taken at an hourly frequency, which have been averaged over a time frame spanning from 2018 to 2020 [8].

4.1.2. Satellite Imagery

The ground truth measurements were complemented with high-resolution satellite imagery from two distinct satellite missions:

4.1.2.1. Sentinel-2 Data

This includes 12 multispectral bands with a resolution of 10 meters, offering detailed visual and spectral information about the Earth's surface. Each image segment corresponds to a 200 x 200 pixel area (approximately 2 x 2 km), cropped around the measurement location. To ensure data quality, the data was pre-selected to only include images with minimal cloud coverage and artifacts [8].

4.1.2.2. Sentinel-5P Data

Sentinel-5P data provides insights into trace gases and aerosols in the atmosphere, offering a spatial resolution of 5 x 3.5 km^2. To facilitate compatibility with Sentinel-2 imagery, this data has been linearly interpolated to a 10-meter resolution [8]. A significant challenge when working with Sentinel-5P data is the complexity in deducing ground-level NO_2 concentrations from tropospheric column density measurements. This difficulty arises because the NO_2 distribution within the atmospheric column is not necessarily uniform. There is a possibility that the concentration of NO_2 at ground level significantly differs from that in higher atmospheric layers, making it challenging to accurately infer ground-level concentrations based solely on column density data. For this reason, the Sentinel-5P data is utilized in conjunction with Sentinel-2 data, combining the detailed spectral information from the latter with the atmospheric data from the former to provide a more comprehensive input for the model.

4.1.3. WorldCover Data

Additionally, the WorldCover dataset [9] was utilized for land cover classification information. This dataset provides a classification of each pixel into one of 11 classes and was aligned to cover the same area as the Sentinel-2 data. The integration of WorldCover data supports the model's understanding of land cover variations and their potential impact on NO_2 distribution.

4.1.4. Data Preprocessing and Augmentation

Prior to training, all data underwent preprocessing to ensure consistency and compatibility. This included normalization of the satellite imagery and ground truth measurements. Furthermore, the uniformly random distributed offset sampling technique, as detailed in section 3.1.2, was applied to augment the dataset and mitigate the issue of sparse annotations.

4.2. Dataset Specifications and Training Methodology

4.2.1. Dataset Size and Split

The full dataset comprises a total of 3087 samples, each representing a unique combination of ground-level NO_2 measurements and corresponding satellite data provided by Scheibenreif et al. (2022) [8]. However, since the land cover ground truth data was not available for every measurement location, the effective dataset resulted in 2871 complete samples. The division of this dataset was strategically executed as follows:

- Training set: 70% (approximately 2009 samples)
- Validation set: 15% (approximately 431 samples)
- Test set: 15% (approximately 431 samples)

This split ensures a substantial training dataset for model learning, while providing adequate and equal-sized data for validation and testing.

4.2.2. Hyperparameter Tuning and Evaluation Metrics

Hyperparameter tuning was conducted using the validation dataset, focusing on optimizing parameters such as learning rate, batch size, prediction space and the regularization parameter λ in the loss function. The primary evaluation metric was the Mean Absolute Error (MAE) for the NO_2 concentration estimation on the validation set.

4.2.3. Baseline and Training Procedure

The baseline for this study was established based on previous research conducted by Scheibenreif et al. (2022) [4]. It is important to note that the aim was not to surpass this baseline but to explore the feasibility and effectiveness of the proposed approach, considering that the prediction area in this study is larger, thus posing a more challenging task.

The training procedure involved assessing both models using two distinct loss functions — either a combined loss approach (3) or solely the squared error loss (1) computed on the resulting NO_2 predictions. The models were trained using a GPU-accelerated environment to ensure efficient learning. Regular checkpoints were saved for model performance evaluation and to facilitate the selection of the best-performing model based on validation data.

4.3. Evaluation

In this subsection, we conduct a comprehensive comparative analysis of the point-wise and patch-wise approaches, incorporating both qualitative and quantitative evaluations.

4.3.1. Qualitative Analysis

The qualitative analysis focuses on the characteristics and visual aspects of the results generated by each approach, as illustrated in Figure 8 which showcases the different outcomes based on the same input area. The input area for this analysis comprises data with spatial dimensions of 1200 x 1200 pixels, corresponding to an area of 1.2km x 1.2km. For the patch-wise approach, predictions were achieved by iteratively creating prediction patches of 8x8 pixels and concatenating these to form the final output. In contrast, the point-wise results were generated by iteratively estimating several points within the input area and employing linear interpolation to achieve a cohesive prediction across the entire area [4].

4.3.1.1. UNet Model

The UNet model, employed in the patch-wise approach, exhibits a tendency to generate a more fine-grained distribution of NO_2 concentrations. While this high level of detail can be advantageous in certain applications, it may not be ideally suited for NO_2 prediction. This finer granularity could potentially introduce localized variations, which might not be representative of the broader environmental context of NO_2 distribution.

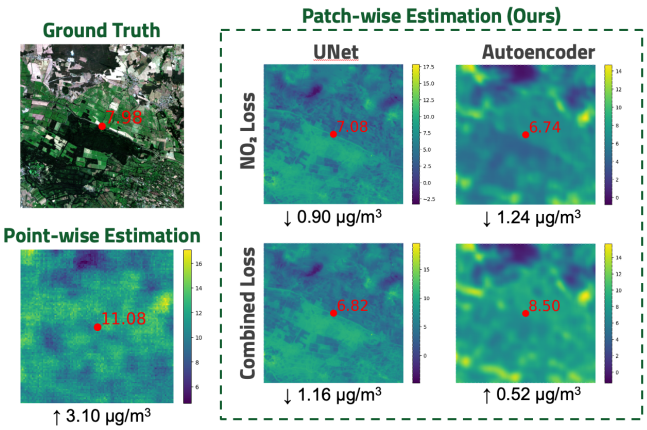

Fig. 8: Comparative visualization of NO_2 estimation methods: the ground truth with a marked measurement point, point-wise estimation showing per-pixel predictions, and patch-wise estimation by UNet and Autoencoder models with different loss objectives. Deviations from the ground truth are indicated below each estimate.

4.3.1.2. Autoencoder Architecture

In contrast, the Autoencoder architecture demonstrates a distribution pattern that aligns more closely with the results from the point-wise approach. This similarity suggests that the Autoencoder is capable of capturing the essential features relevant to NO_2 concentration without the excessive detail that might complicate the interpretation or application of the data. Its performance indicates a balance between detail and contextual accuracy, potentially making it more suitable for reliable NO_2 predictions.

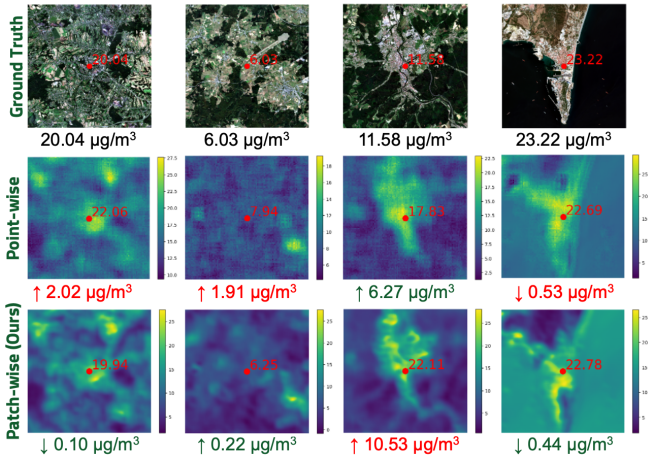

Fig. 9: Visualization of ground truth NO_2 levels and corresponding point-wise and patch-wise estimations using the Autoencoder model with NO_2 loss, showing close approximation to actual concentrations and similar distributions as the point-wise estimates.

4.3.2. *Quantitative Analysis*

In the quantitative analysis, we examine the performance metrics of both approaches, including MAE, R2-Score, and Mean Squared Error (MSE). These evaluation results are depicted in Table 1 and enable a comprehensive understanding of each model's accuracy and predictive capabilities.

Model	**Loss**	**R2 (↑)**	**MSE (↓)**	**MAE (↓)**
UNet (Ours)	Combined	0.49	47.25	5.06
UNet (Ours)	NO_2	0.49	**46.83**	5.03
AE (Ours)	Combined	0.47	50.94	**4.98**
AE (Ours)	NO_2	0.46	49.55	4.99
Point-wise [4]		0.55	61.25	5.65
Point-wise (Pre-trained) [4]		**0.57**	58.47	5.50

Table 1: Quantitative Results: Point-wise vs Patch-wise.

The quantitative analysis reveals that our patch-wise approach not only matches but surpasses the accuracy of the point-wise approach, achieving significantly lower values in both MAE and MSE metrics. This is particularly noteworthy considering our model's operational conditions, which included a reduced training set and an expanded test set (10% for point-wise [4] and 15% for our patch-wise approach), further underscoring the effectiveness and robustness of our method.

4.3.3. *Summary*

In summary, the qualitative and quantitative analyses highlight the strengths and advantages of the patch-wise approach over the point-wise method. While the point-wise approach already excelled in accuracy, our patch-wise method, particularly with the Autoencoder architecture, now not only equals but surpasses it in both visual and quantitative metrics. This advancement is achieved with significantly lower computational costs. Our patch-wise method achieves this by processing patches of 64 pixels (8x8) collectively, in stark contrast to the point-wise approach which requires individual computation for the generation of each pixel. Moreover, our approach exhibits a notable improvement in MAE and MSE, outperforming the point-wise method in a context of a reduced training set and an expanded test set. These enhancements in accuracy and efficiency position the patch-wise approach as a highly effective and scalable alternative for large-scale environmental monitoring.

4.3.4. *Application to a New Geographical Region - US West Coast*

An integral part of evaluating the robustness of our approach involved testing the model on a dataset consisting of locations along the US West Coast. This subset of data comprises 91 samples and was not included in the initial training or validation phases, ensuring an unbiased evaluation of the model's robustness. It features unique characteristics not encountered in the European data used for model training, such as dense, rectangular street patterns and expansive desert areas, which are distinctive to the US West Coast [4].

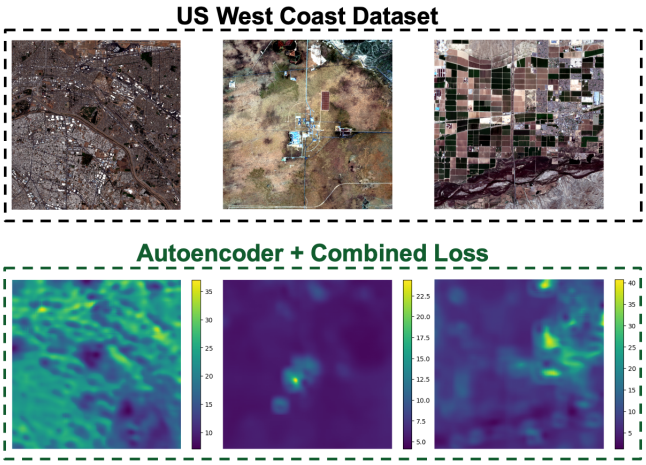

Fig. 10: Application of the Autoencoder model with a combined loss on the US West Coast dataset, showcasing the model's capability to adapt and estimate NO_2 levels in diverse geographic settings.

Model	**Loss**	**R2 (↑)**	**MSE (↓)**	**MAE (↓)**
UNet (Ours)	Combined	-1.41	140.69	10.17
UNet (Ours)	NO_2	-1.42	140.74	10.09
AE (Ours)	Combined	0.15	**62.18**	**6.19**
AE (Ours)	NO_2	-0.02	71.43	6.69
Point-wise [4]		**0.28**	89.92	7.86

Table 2: Evaluation on US West Coast Dataset.

When evaluated based on the MAE, the patch-wise approach, using the Autoencoder architecture, not only adapted to this new environment but also surpassed the quantitative results of the point-wise approach.

The performance of the patch-wise approach in this new region is indicative of its potential for broader application and underscores the benefits of an adaptable and efficient model for global-scale environmental monitoring.

4.3.5. *Impact of Increased Prediction Space*

In a continued effort to optimize our model's performance, we explored the effects of increasing the prediction space on both the quality and efficiency of our NO_2 concentration estimations.

Given the qualitative superiority observed in the Autoencoder architecture's output compared to that of the UNet, the former was selected for this phase of evaluation. Similarly, the Combined Loss was employed over the Squared Error Loss due to its slightly superior quantitative performance, particularly in terms of MAE.

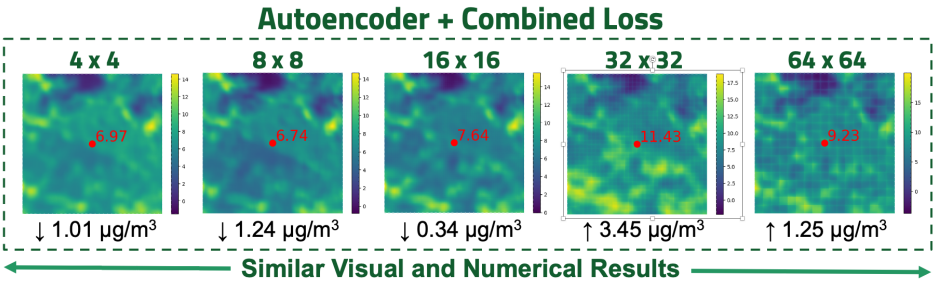

Fig. 11: Performance comparison of the Autoencoder model utilizing the combined loss across varying prediction spaces, demonstrating consistent visual and numerical accuracy as the estimation grid size increases.

Prediction Space	R2 (↑)	MSE (↓)	MAE (↓)
2 x 2	0.49	50.21	**4.95**
4 x 4	0.49	49.11	5.04
8 x 8	0.47	50.94	4.98
16 x 16	0.46	50.15	5.00
32 x 32	**0.50**	**47.12**	**4.95**
64 x 64	0.49	47.13	4.99

Table 3: Evaluation of varying prediction space dimensions.

Our investigation revealed that enlarging the prediction space did not result in a noticeable compromise in the model's performance. Both quantitative metrics and qualitative assessments remained consistently similar to those obtained with smaller prediction spaces. This finding suggests that the model preserves its predictive integrity and spatial resolution of NO_2 distribution patterns, even as the scope of prediction broadens.

The computational benefits are increasingly apparent with the expansion of the prediction space. As the area of estimation grows, the model's efficiency improves markedly, affirming the scalability of our approach in processing larger geographical extents with reduced computational demands.

However, it was noted that upon extending the prediction space to 32 x 32 pixels and beyond, visual artifacts began to emerge within the qualitative results. While the quantitative measures remained stable, these visual irregularities indicate a threshold beyond which the increase in prediction space may begin to adversely affect the model's ability to render the finer details of NO_2 distribution.

In summary, our findings underscore the Autoencoder architecture's capacity to maintain accurate NO_2 concentration estimations across varying scales of prediction space, up to a certain limit where visual fidelity is challenged by expansive computational efficiency.

5. CONCLUSION

This study introduced an innovative patch-wise approach for estimating ground-level NO_2 concentrations using deep learning and remote sensing data. This method represents a significant advancement in the domain of environmental monitoring, offering a scalable and efficient alternative for global NO_2 pollution estimation.

The patch-wise approach was developed in response to the limitations of existing point-wise methods, which, while accurate, are computationally intensive and not feasible for global-scale application. By generating a grid of estimates for a designated area simultaneously, the patch-wise approach significantly reduces the computational demands associated with large-scale environmental monitoring.

Our findings demonstrate that this approach not only provides a substantially improved precision compared to point-wise estimations but also enhances computational efficiency significantly. This is particularly evident in scenarios requiring large-scale analysis, where the patch-wise method processes extensive geographical areas with reduced computational requirements. The approach has shown remarkable results in both qualitative and quantitative assessments, enhancing accuracy and maintaining the spatial resolution of NO_2 distribution patterns across various scales of prediction space.

Moreover, the applicability of this method was successfully tested on a new geographical region - the US West Coast. The dataset, featuring unique characteristics not encountered in the European data used for model training, provided an unbiased evaluation of the model's robustness. The patch-wise approach adapted to this new environment and also surpassed the quantitative results of the point-wise approach in terms of MAE. This underscores the potential of the patch-wise method for broader application and its adaptability to different environmental contexts.

In summary, the patch-wise approach presented in this study offers a viable solution for efficient and scalable global environmental monitoring. It balances the need for detailed, high-resolution data with the practical constraints of computational resources, making it a significant contribution to the field of remote sensing and environmental monitoring. This approach opens new avenues for global estimations and improved environmental policy-making, showcasing the potential of deep learning in addressing critical environmental challenges.

6. REFERENCES

[1] Tze-Ming Chen, Ware G. Kuschner, Janaki Gokhale, and Scott Shofer, "Outdoor air pollution: nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects," *The American Journal of the Medical Sciences*, vol. 333, no. 4, pp. 249–256, 4 2007.

[2] Ute Latza, Silke Gerdes, and Xaver Baur, "Effects of nitrogen dioxide on human health: Systematic review of experimental and epidemiological studies conducted between 2002 and 2006," *International Journal of Hygiene and Environmental Health*, vol. 212, no. 3, pp. 271–287, 5 2009.

[3] L. N. Lamsal, Randall V. Martin, Aaron Van Donkelaar, Martin Steinbacher, E. A. Celarier, E. J. Bucsela, E. J. Dunlea, and Joseph P. Pinto, "Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument," *Journal of Geophysical Research*, vol. 113, no. D16, 8 2008.

[4] Linus Scheibenreif, Michael Mommert, and Damian Borth, "Toward global estimation of Ground-Level NO2 pollution with deep learning and remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 1 2022.

[5] M. J. Cooper, Randall V. Martin, Chris A. McLinden, and Jeffrey R. Brook, "Inferring ground-level nitrogen dioxide concentrations at fine spatial resolution applied to the TROPOMI satellite instrument," *Environmental Research Letters*, vol. 15, no. 10, pp. 104013, 9 2020.

[6] Marina Vîrghileanu, Ionuţ Săvulescu, Bogdan Mihai, Constantin Nistor, and Robert Dobre, "Nitrogen Dioxide (NO2) Pollution Monitoring with Sentinel-5P Satellite Imagery over Europe during the Coronavirus Pandemic Outbreak," *Remote Sensing*, vol. 12, no. 21, pp. 3575, 10 2020.

[7] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer, "Deep Learning in Remote Sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 12 2017.

[8] Linus Scheibenreif, Michael Mommert, and Damian Borth, "Estimation of Air Pollution with Remote Sensing Data: Revealing Greenhouse Gas Emissions from Space," Jan. 2022.

[9] 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," Oct. 2022.