

UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science master thesis

**Urban Air Quality Assessment: Developing a Pipeline for NO₂
Prediction Using Street View Imagery and Deep Learning Models**

First examiner:

Jules Kerckhoffs

Candidate:

Floris Videler

Second examiner:

Tim Ophelders

In cooperation with:

Faculty of Medicine, University of
Augsburg

June 27, 2024

Abstract

Urban air pollution, particularly nitrogen dioxide (NO₂), poses significant health risks and remains a widespread concern. Traditional stationary monitoring stations often fail to capture the spatial variability of pollutants in urban environments, leading to incomplete exposure assessments. Mobile measurement campaigns provide a more dynamic and detailed picture of air pollution distribution.

This study leverages street view imagery and deep learning models to develop a pipeline for predicting NO₂ concentrations at a hyperlocal level. By utilizing complex pre-trained models without the need for retraining, this approach significantly lowers the technical barrier to entry. Using data from a mobile measuring campaign in Augsburg, Germany, pre-trained deep learning models were employed to extract features from 360-degree street view images at both object detection and semantic segmentation levels. These features served as inputs for secondary models, including Linear Regression, Support Vector Regression (SVR), and XGBoost, to predict NO₂ levels.

Despite the innovative approach, significant data quality issues, such as low revisit frequency and spatial inconsistencies, led to poor model performance. Models trained on a subset with higher revisit frequencies demonstrated improved results but still fell short of expectations. These findings underline the critical importance of data quality and revisit frequency in mobile air quality monitoring campaigns.

The results suggest that while the proposed methodology has potential, data quality significantly impacts model accuracy. Future research should focus on improved data integration, better planning of measurement campaigns, and applying this pipeline to other urban datasets. This study contributes to the ongoing effort to enhance urban air quality monitoring using advanced image analysis techniques, offering a scalable solution with the potential to provide more detailed pollution assessments in various urban settings.

Contents

1	Introduction	4
2	Data and Methods	6
2.1	Data Collection and Sources	6
2.2	Data Preprocessing and Cleaning	6
2.3	Modelling Methods and Techniques	11
2.4	Implementation and Execution	16
2.5	Summary	17
3	Results	18
3.1	Data Summary	18
3.2	Model Performance	19
3.3	Issues and Limitations	22
3.4	Summary	23
4	Discussion	24
4.1	Interpretation of Results	24
4.2	Implications of Findings	25
4.3	Data and Model Limitations	26
4.4	Recommendations	30
4.5	Conclusion	31
Appendix		
A	Full spatial Overview of NO₂	34
B	Data Integration examples	35
C	Code	36
D	Good Examples of Spatial NO₂ Distribution	37
E	Examples of Anomalies in the NO₂ Distribution	38
F	Full Spatial Revisit Count Distribution	41

G Final Data Sample Spatial Overview of NO₂	42
H Final Data Sample Split	43
I Pre-trained Features	44
J UFP Results	46
Bibliography	51

1. Introduction

Air pollution is a significant global issue, affecting millions of people and causing various health problems. Among the many pollutants, nitrogen dioxide (NO₂) stands out as particularly harmful. NO₂ is a critical component of traffic-related air pollution (TRAP), which is predominantly emitted by vehicles. This pollutant not only contributes to respiratory and cardiovascular diseases but also exacerbates existing health conditions, making it a major concern for public health worldwide. Studies have shown that exposure to NO₂ is associated with increased mortality rates and adverse health effects, especially among vulnerable populations [1–4].

The spatial mapping of NO₂ and other air pollutants is crucial for accurately assessing exposure levels and understanding their health impacts. Traditional stationary monitoring stations provide limited spatial coverage and may not capture the variability of pollutant concentrations in different environments. This limitation is particularly significant in urban areas, where NO₂ levels can vary dramatically over short distances due to traffic density, road configurations, and building structures. Mobile measurement campaigns, which involve portable sensors on vehicles or individuals, offer a more detailed and dynamic picture of air pollution distribution, overcoming these shortcomings [5–8]. These campaigns can identify pollution hotspots, track temporal changes, and provide critical data for developing targeted mitigation strategies. High-resolution spatial data from mobile measurements can better correlate with health outcomes and improve the accuracy of exposure assessments compared to stationary monitors alone [9].

The use of mobile monitoring for creating hyperlocal pollution mappings is becoming increasingly common [5–8]. Studies confirm that nearly continuous measurements effectively reveal both intra- and inter-city pollution dynamics. They consistently identify higher pollutant concentrations near major roads, busy areas, and industrial zones, aligning with longstanding expectations about these locations [10–12]. Additionally, some studies highlight a direct correlation between traffic density and pollutant levels, noting substantial variations in pollution based on traffic proximity to the measurement points [10–14].

Land-use Regression (LUR) is one of the most common methods for modeling air pollution [15, 16]. LUR models, especially when combined with hyperlocal data, have proven effective in predicting the concentration of multiple air pollutants, including NO₂ [16–22]. Typically, LUR models incorporate features at the street or road level, such as road type, traffic density, and proximity to industrial sites. However, these features mostly act as proxies for the sources of pollutants. This research proposes using street view imagery to predict and map hyperlocal NO₂ values, generating a more detailed local feature space and potentially uncovering direct connections between NO₂ levels and objects seen in the images.

Leveraging street view imagery to predict NO₂ and other spatially dependent variables is not a new concept and has shown promise in previous research, achieving high model performance [23–29]. For example, a 2022 study using a ResNet model achieved R² values ranging from 0.51 to 0.9 across five cities when training and testing within the same city [24]. Similarly, research in Toronto used street view imagery to extract architectural features for an air pollution prediction model, achieving R² values between 0.59 and 0.64 [30]. Another study combined multiple pre-trained models to predict high-level features, which were then used in a Gradient Boosted Regression model, yielding an R² value of 0.48 for noise pollution [23].

This research proposes utilizing pre-trained models to extract high-level features from street view imagery as input for a simpler secondary model to predict NO₂ concentrations. This method significantly reduces the need for large amounts of data typically required for training extensive models, although a comprehensive and diverse dataset remains essential [23]. Evaluating whether this approach, successful in noise pollution studies, can achieve comparable accuracy for NO₂ prediction is particularly interesting. Using data from a mobile measurement campaign in Augsburg, this study aims to provide a scalable and efficient method for NO₂ prediction that lowers the technical barrier to entry compared to training large models from scratch, leveraging advanced image analysis techniques to enhance environmental monitoring efforts. The final pipeline and model aims to answer the following research question: *"How can features extracted from street view images using state-of-the-art deep learning models predict air quality in urban environments?"*

2. Data and Methods

As discussed in the Introduction, this research utilizes data from a mobile NO₂ measuring campaign together with street view images. This chapter will go into detail about this data and about the methodological steps taken to prepare the data for further analysis and how the models were utilized to extract insights from this data.

2.1 Data Collection and Sources

The data utilized in this research is a set of geographically tagged nitrogen dioxide (NO₂) measurements in micrograms per cubic meter ($\mu\text{g}/\text{m}^3$) and 360-degree street view imagery, which are not geographically identifiable. This data was gathered in a mobile measuring campaign utilizing a modified car that was equipped with a GoPro camera and air pollution monitoring devices that drove randomly through Augsburg. The collection period spanned 15 days during July and August 2022, resulting in around 330,000 NO₂ measurements and 230,000 photographic captures. Figure 2.1 displays a zoomed in map of the study area, showing the NO₂ measurements aggregated into 50-meter grids, a full overview of the study area is provided in appendix A.

2.2 Data Preprocessing and Cleaning

Before the data can be analyzed it needs to be prepared. This section describes the steps taken to preprocess the data.

2.2.1 Data Cleaning

The set of NO₂ values received for this study was pre-cleaned of outliers by winsorizing the data. During the winsorizing process, the maximum NO₂ values were capped at $88.1 \mu\text{g}/\text{m}^3$, with all values higher than this threshold rounded down to $88.1 \mu\text{g}/\text{m}^3$. The distribution of the data received is shown in Figure 2.2

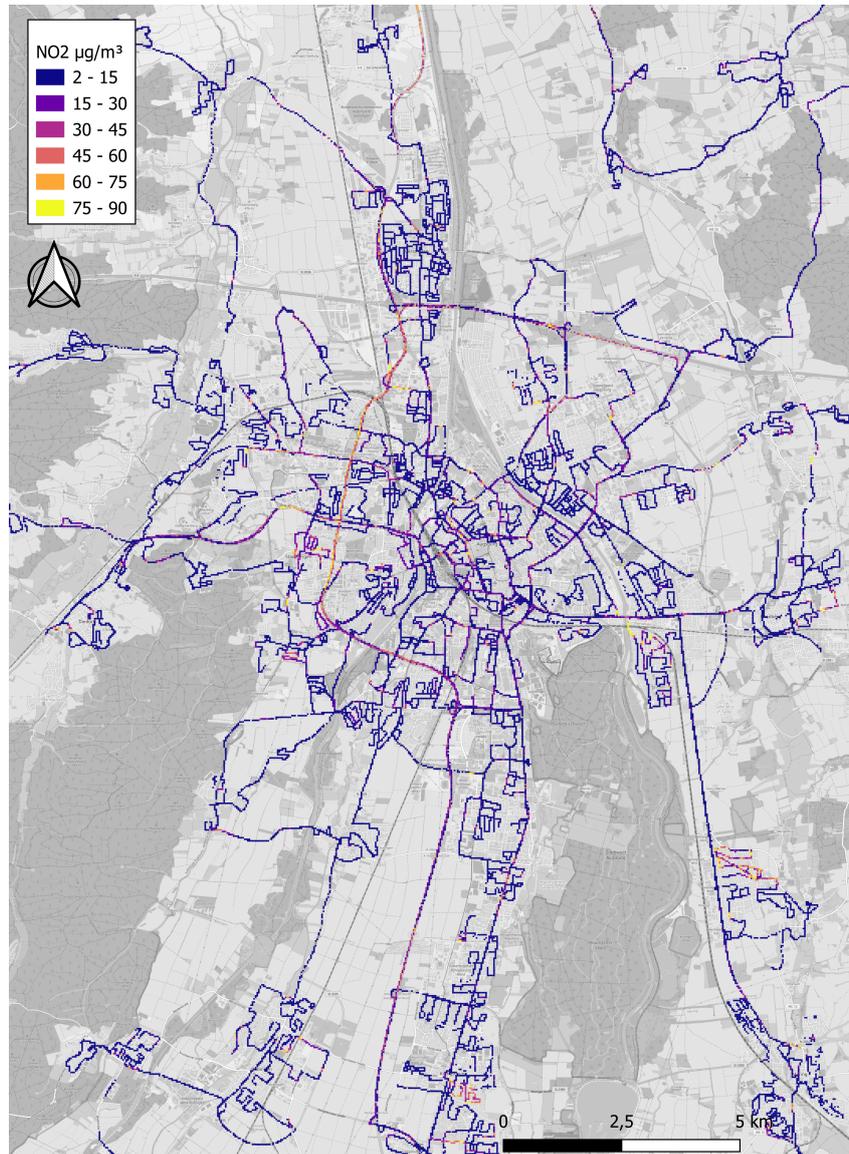


Figure 2.1: Zoomed in map of Augsburg and direct surroundings showing the mean NO₂ levels in 50 meter grid cells.

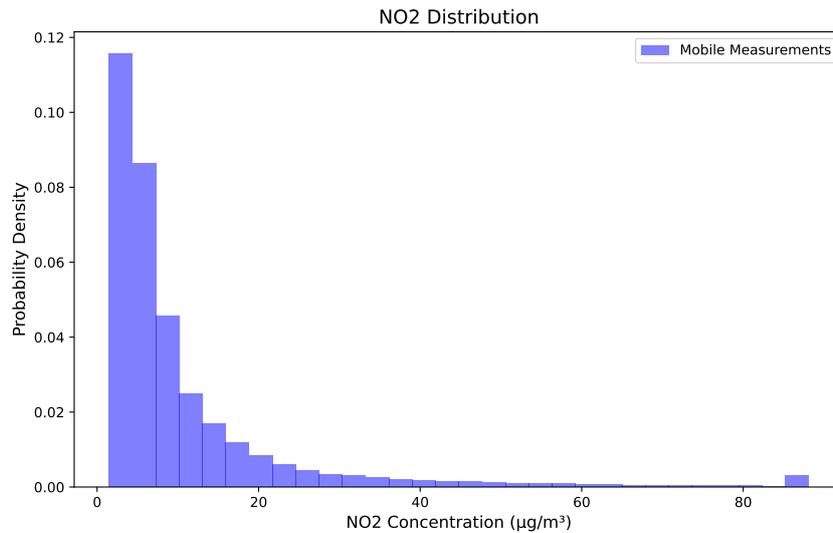


Figure 2.2: Distribution of received NO₂ values.

2.2.2 Data Transformation

The NO₂ values were aggregated to grids of 50 meter to average the measurements. This grid size was chosen to maintain the hyperlocality of the data while smoothing out variations by averaging. There is not a clear standard for aggregating this type of data, earlier research describe segments from 30 meters [29] all the way up to 540 meters [8]. A research using data that was gathered with the same equipment used 50 meters as the smallest segment [5], making it an appropriate reference for this research. Despite receiving a clean dataset, averaging NO₂ measurements per grid was crucial for describing spatial patterns. Ultimately the aggregating resulted in a total of 26,567 grid cells.

The image data required specific preparation due to its initial format as flattened 360-degree panoramas, which introduced two main biases. The first was a disproportionate emphasis on the center of the flattened image and the second was the obscuration of nearly half the image by the roof of the car.

To correct these biases, the images were reprojected into a 90-degree field of view (FOV) with an upward tilt of the camera. This process involved three steps:

1. Spherical Projection: Assuming the image represents a sphere, the correlation of longitude and latitude from the panorama with positions on the image was established.
2. Directional Reprojection: Adjustments were made for the specified FOV, cam-

era tilt, and desired direction to compute new viewing angles. A 90-degree FOV was selected, allowing for the computation of four directions of 90 degrees each, summing up to a complete 360-degree image. The tilt was set at 12 degrees, which testing indicated was optimal to minimize roof visibility while maintaining as full a view as possible

3. Mapping to Original Panorama: All angles were mapped back to the original panorama to extract specific pixels and create a new perspective image.

Each directionally adjusted image was then resized to 716 by 1333 pixels, accommodating model requirements and reducing the storage size. An overview of this process can be seen in Figure 2.3

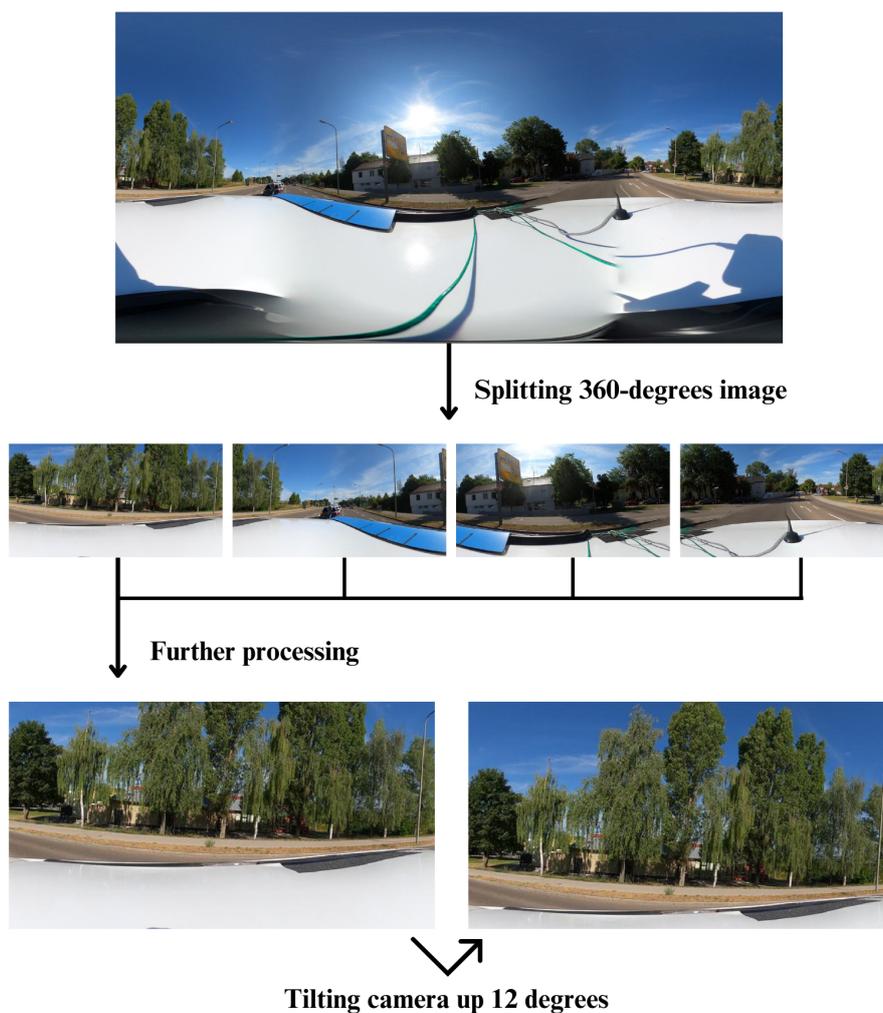


Figure 2.3: Image preprocessing pipeline.

2.2.3 Data Integration

The data was supplied in two parts: the NO₂ measurements and the street view images. Given that both data sources were captured using separate devices, it had to be integrated to create one dataset where each image could be linked to the closest NO₂ measurement. To facilitate this, the time displayed on a smartphone, held by individuals in the images taken at the beginning of each day, was used as a reference. However, this method presented challenges. On some days, the time was either not displayed or was only shown at the minute level, which is not accurate enough, examples of this can be seen in Appendix B. The offset vastly differs between days, ranging from 7 to 94 seconds.

Ultimately, successful synchronization of 7 days worth of images with their corresponding NO₂ measurements was achieved, utilizing 51.85% (120,462) of the available images. The failure to match the remaining data points not only resulted in a loss of potential data but also impacted the temporal resolution of the dataset.

2.2.4 Data Selection

Due to the vast number of images in the dataset and the computational and storage limitations, it was necessary to make a selection of the data to be used. Two sampling methods were utilized to compare the performance of both:

1. Time-based: For the time-based data selection, 221 images were sampled from each hour to ensure a balanced representation across different times. This approach yielded 40,000 transformed images and their corresponding NO₂ levels.
2. Revisit-based: The revisit-based data selection uses the amount of unique days a certain grid was visited by the measurement car. All grid cells had to be visited more than once on different days to be selected. This threshold was chosen to be able to create a sizeable dataset, as can be seen in Table 2.1, the average days a grid has been revisited is around three. If a higher threshold would have been chosen too much data would be lost. This data selection yielded a sample of about 8,500 transformed images and their corresponding NO₂ level.

Apart from being smaller in size than the time-based dataset, the revisit-based dataset also has a much smaller study area. Since a substantial amount of cells

do not have a revisit frequency higher than one, these have to be discarded for this selection, Appendix G shows the grids selected, note how most of the cells are clustered near busy places in and near the city center.

Table 2.1: Summary of the frequency of NO₂ measurement grid visits.

Statistic	Value
5th percentile	1.0
25th percentile	1.0
Median	2.0
Mean	3.19
75th percentile	3.0
95th percentile	12.0

2.2.5 Train, Validation and Test Split

To ensure robust model development and evaluation, the dataset was divided into training, validation, and test sets, comprising 60%, 20%, and 20% of the data, respectively. To realize this split each 50 meter grid was randomly assigned to one of the 3 sets using a weighted random function to approximately get the 60-20-20 split. This approach helps prevent images and NO₂ measurements within the same grid from being placed in different sets, thereby avoiding data leakage which could potentially bias the models. A map detailing the final dataset split is provided in appendix H.

2.3 Modelling Methods and Techniques

This study focuses on a multi-step modeling approach. While each individual step is simpler, the combination of these steps can become complex. This section provides detailed explanations of the methodologies used for modeling.

2.3.1 Model Selection

The model selection process was conducted in two phases. Initially, suitable pre-trained models were identified to generate the necessary features for further analysis and predictions. Then, an evaluation was carried out to determine which secondary models would be most compatible with these features.

2.3.1.1 Pre-trained Models

Before selecting the pre-trained models, two feature levels were established: object-level features and semantic-level features. This decision was influenced by earlier research in another field, which identified these levels as the most important [23]. Limiting to two levels also helped maintain low complexity. Object-level features provide information about what and how many objects are visible in a given image. Semantic-level features classify each pixel into a category, followed by computing the relative amount of pixels belonging to each class. This results in a percentage value for each category, indicating how much of the image is covered by that classified semantic.

For the object detection the pre-trained DEtection TRansformer (DETR) (End-to-End Object Detection) model with ResNet-101 backbone [31] trained by Meta (Facebook) was chosen. The DETR model was trained on the COCO dataset, an important resource in the field of computer vision designed to provide a comprehensive coverage of 91 object categories with diverse image representation [32]. By training on the COCO dataset the DETR model is able to robustly predict object boundaries and categories across a wide range of contexts, as demonstrated in its foundational study [31]. Apart from its robust performance on the COCO dataset, it is also easily accessible through Hugging Face [33], a collaborative platform for machine learning models and datasets, which further reduces the technical complexity. The DETR model uses a ResNet-101 as backbone, which provides a strong feature extractor with deep layers, enhancing the model's ability to handle a wide range of images and object scales [34]. What makes this model unique is that it uses end-to-end object detection, this streamlines the training and inference pipeline by removing the need for manually tuned components. It also leverages transformer layers, which are known for their effectiveness in handling relational reasoning and parallel processing. All this makes the performance accuracy and run time on par with the well-established and highly-optimized Faster RCNN baseline on the COCO object detection dataset [31]. All the aforementioned reasons make the DETR model a reliable model to use for object detection in this study.

For semantic segmentation, the Mask2Former model trained on the Cityscapes dataset [35, 36] was selected. This model is fine-tuned on the Cityscapes dataset, which contains diverse urban street scenes. The dataset consists of around 20,000 images across 30 classes from 50 different cities [37]. This dataset's focus on ur-

ban street scenes, aligns perfectly with this study's requirements. Apart from this, the model incorporates an advanced Mask2Former architecture with a Swin Transformer backbone, enhancing its ability to handle complex visual scenes and achieve state-of-the-art results in semantic segmentation [35, 36]. This model is also easily available on Hugging Face [38]. Given that both models can be found in the same place, again lowers the barrier to entry to applying such large and complex models.

2.3.1.2 Secondary Models

The secondary models were not as complex as the pre-trained models, but they had to be trained from scratch. To choose an adequate model, there were three main requirements that a secondary model needed to fulfill:

1. **Simplicity:** The secondary model should be straightforward, avoiding unnecessary complexity since the pre-trained models have already handled the intricate feature extraction.
2. **Regression Capability:** As the target variable (NO₂ levels) is continuous, the model must be able to perform regression tasks. This aligns with the methodological approaches of most relevant research, thus making it easier to compare the results with similar research.
3. **Inclusion of Linear and Non-linear Models:** It is crucial to assess both linear and non-linear models to determine which best represents the underlying patterns in the data.

The simplest model that will be applied is linear regression. Linear regression models the relationship between a dependent variable and one or more independent variables using a linear equation. This model's transparency and ease of implementation make it a good choice for interpreting straightforward relationships in the data, suitable for continuous outcomes like NO₂ levels [39].

Support Vector Regression (SVR) uses the same principles as the Support Vector Machine (SVM) for classification but applies them to regression tasks. It creates a set of hyperplanes in a high dimensional space, which can be linear or non-linear, depending on the chosen kernel [40]. SVR offers a compromise between model simplicity and the ability to model non-linear relationship through kernels making it robust in handling non-linear data and predicting continuous outcomes. It was chosen to determine whether a moderately complex model can better capture non-

linear relationships within the data.

The last secondary model that will be implemented is the XGBoost regressor. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. It sequentially constructs trees, where each new tree attempts to correct the errors made by the previous trees. With this it also incorporates regularization to avoid overfitting [41]. The model may be more complex, but remains efficient and scalable, making it manageable with larger and complex datasets. It was included to evaluate whether a more advanced non-linear model can better make out the patterns missed by simpler models.

2.3.2 Methodological Approach

This subsection gives an overview of how each selected model was methodologically utilized.

2.3.2.1 Pre-trained Models

The methodological approach for the pre-trained models focuses on their output, which serves as input for the secondary models. After predicting both the object detection and the semantic segmentation of the four angles of all the images, the outputs were parsed in the following ways:

- Object detection: All detected objects across the angles were aggregated for each base image and summed to compile a complete count per image.
- Segmentation: The percentages of pixels classified into specific categories were calculated for each angle. These percentages were then summed and averaged across the four angles to normalize the semantic data for each image.

A detailed compilation of all extracted features is available in appendix I.

2.3.2.2 Secondary Models

For the secondary models, the methodological approach focuses on the training of the models. Since the nature of each model is different, slightly different approaches were taken for each one.

- Linear Regression: The features used were chosen using backward elimina-

tion.

- SVR: The input data was normalized to have a mean of zero and an unit variance. The model was configured with the RBF (Radial Basis Function) kernel to find non-linear patterns. Hyperparameters were tuned using a grid search with cross-validation.
- XGBoost: Hyperparameters were tuned using a grid search with cross-validation.

2.3.3 Validation and Testing

For evaluating model performance, two metrics were selected: the coefficient of determination (R^2) and the root mean squared error (RMSE). These metrics are widely used in similar research [24, 25], making it easier to compare the results, they also work well for regression problems like ours. Equation 2.1 shows how the R^2 score is calculated. The R^2 is a statistical measure indicating the proportion of variance in the dependent variable that can be predicted from the independent variables. Understanding this variance is crucial for assessing the impact of various predictors on NO_2 levels. Since the R^2 only indicates the proportion of variance explained, it does not provide information about the absolute size of the errors, this is where the RMSE comes in. As shown in Equation 2.2, RMSE quantifies the square root of the average squared differences between the predicted and actual values, providing a measure of how accurately the model predicts the target variable. This metric is especially important when modeling NO_2 or any air pollutant, as minimizing large prediction errors is critical due to their potential health and regulatory implications. Furthermore, since RMSE is expressed in the same units as the target variable, it allows for straightforward interpretation of the magnitude of prediction errors.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.2)$$

2.3.3.1 Baseline

As a benchmark for model evaluation, this study employs the mean of the target variable (NO₂ levels) from the training set. This simple baseline model serves as a foundational performance metric. Utilizing this baseline allows for a clear comparison of model efficacy. Any advanced model implemented should demonstrate a substantial performance improvement over this baseline to justify the increased complexity.

2.4 Implementation and Execution

All of the programming has been done in Python 3.11.5 and all the code can be found in a Git repository linked to in Appendix C. The pre-trained models employed in this study are accessible via the Hugging Face Pipeline API in Python [42] to easily process a large volume of images with minimal coding. The three secondary models used in this study were implemented using Python: Linear Regression and SVR with the scikit-learn library [43], and XGBoost using the xgboost library [44]. The function employed to realize the grid search, scaling and the calculation of the metrics was all done with function and classes from the scikit-learn library, all other data wrangling and computational needs were handled using Pandas [45] and Numpy [46].

For the linear regression model, the features used were chosen using backward elimination. Initially, all variables were included in the model. For each iteration, the variable with the least significant impact on the model's performance, assessed via the p-value, was removed. This process was repeated iteratively until only statistically significant variables remained, optimizing the model by retaining only the most impactful features.

The SVR model was configured with the RBF (Radial Basis Function) kernel to find non-linear patterns. Key hyperparameters such as C (regularization parameter), gamma (kernel coefficient), and epsilon were optimized using a grid search with cross-validation. The features in the dataset were standardized to enhance the model's sensitivity to kernel variations. Table 2.2 displays the unique values tested in the grid search.

For the XGBoost model, no additional data preparations were required, as the dataset contained no categorical variables that needed encoding. The hyperpa-

Table 2.2: Unique Values for Each Hyperparameter in SVR Tuning

C	Gamma	Epsilon
0.1	0.001	0.0
1	0.01	0.1
10	0.1	0.2
100	1	0.5
1000	scale [47]	1.0
	auto [47]	

parameters were again tuned using grid search, focusing on parameters critical for controlling the model’s complexity and fitting behavior. These parameters include max depth (maximum depth of a tree), min child weight (minimum sum of instance weight needed in a child node), gamma (minimum loss reduction required to make a further partition on a leaf node), learning rate (step size shrinkage used to prevent overfitting), and n estimators (number of trees in the ensemble). Table 2.3 provides an overview of the unique values explored in the grid search.

Table 2.3: Unique Values for Each Hyperparameter in XGBoost Tuning

Max Depth	Min Child Weight	Gamma	Learning Rate	N Estimators
3	1	0	0.01	100
5	3	0.1	0.1	300
7	5		0.2	500

2.5 Summary

To summarize, this research utilizes data from mobile NO₂ sensors along with street view images. Using this data, it was proposed to employ two pre-trained models: a DETR (End-to-End Object Detection) model with a ResNet-101 backbone [31] and a Mask2Former model trained on Cityscapes semantic segmentation [35, 36]. These models extract high-level features from street view images, such as objects and semantic content. These features will then serve as input for a simpler secondary model, in this case, a linear regression [39], Support Vector Regression (SVR) [40] and XGBoost regressor [41], to predict the NO₂ values per grid cell. This method and similar ones have shown promising results in earlier research [20, 23, 25], while lowering the overall complexity of the models that need to be trained. Model performance are evaluated using R² and RMSE metrics.

3. Results

This chapter focusses on all the findings done after designing the methodology. This includes data related insights, issues, model performance and an analysis of this performance. A more detailed interpretation of the results will be given in the Discussion chapter (4)

3.1 Data Summary

Since this research relies on two datasets: the original dataset, which consists of NO₂ measurements and street view images and a dataset build from extracted features from these street view images using pre-trained models, both of these datasets will be described.

3.1.1 Original Dataset

The distribution of the original NO₂ measurements was already shown in 2.2. To look at this in more detail and to compare it to the distribution of the new sample, Table 3.1 highlights the summary statistics of both the original NO₂ measurements and those from the two data selection methods. Note how the time-based sample and the full data share very similar summary statistics. In contrast, the revisit-based sample does not. This indicates that the revisit sample might not be as representative of the full dataset as the time-based sample is.

Table 3.1: Summary of NO₂ values for different sampling methods: time-based sample, revisit-based sample, and full data from Augsburg mobile measurements.

	Time-based sample	Revisit-based sample	Augsburg mobile (full data)
5th percentile	2.1	4.8	1.6
25th percentile	4.1	7.6	3.6
Median	7.0	11.8	6.1
Mean	11.0	14.7	10.6
75th percentile	12.2	17.9	11.4
95th percentile	34.6	33.9	36.4

Further analysis showed that, almost all elevated NO₂ measurements seem out of

place. Such as unusually high NO₂ values in tranquil neighborhoods or abrupt spikes following periods of low measurements. In appendix E specific examples can be found. Overall, the spatial pattern of the dataset appears irregular and out of place compared to established trends. It was also concluded that these anomalies were mostly removed when examining the revisit-based sample. This might indicate issues with measurements at locations that have not been revisited multiple times, potentially biasing the spatial pattern.

3.1.2 Feature Extraction Dataset

As was discussed in the Data and Methods chapter (2), features extracted using pre-trained models will be used to train, validate and test the secondary models. The combined number of extracted features is 21. From those 21 features, eight come from the object detection model and 13 come from the semantic segmentation model. The extracted features did not exhibit any note worthy anomalies and were thus transformed and combined into one dataset. A comprehensive overview of all the features extracted and their frequency can be seen in Appendix I.

3.1.3 Combined Dataset

To finalize the dataset the extracted feature set had to be merged with the correct NO₂ measurement. This was done by merging both datasets based on the image name. Each feature was extracted from an image and thus connected to that image and as described in the Data Integration (2.2.3), every NO₂ measurement was also connected to an image. This way the features were correctly matched up with their corresponding target variable.

3.2 Model Performance

To compare the performance of all the models as fair as possible over both the data selection methods, all the secondary models have been tuned and trained separately on each set. All six models (three models for each of the two sets), have been benchmarked against the baseline that corresponds to their dataset over the same two metrics.

3.2.1 Model Descriptions

Equation 3.1 and 3.2 show the equations for the linear regression models fitted on the time and revisit-based data samples respectively, after the features were selected using backward elimination. This shows what variables impacted the linear regression the most. It is note worthy that the 'SuitcaseObject' in the time-based equation (3.1) shows a relatively large value compared to the other variables, this presumably has to do with the model not being able to find a clear pattern and thus being more likely to assign larger values to variables that may not be considered important. It is good to mention that Augsburg is not littered with suitcases, but that the detection model tends to classify large garbage bins as suitcases.

$$\begin{aligned}y = & 6.049 - 0.066 \times \text{CarObject} + 0.25 \times \text{TrafficLightObject} \\ & + 1.258 \times \text{BusObject} - 6.727 \times \text{SuitcaseObject} \\ & + 0.338 \times \text{TrainSegment} + 0.031 \times \text{VegetationSegment} \\ & + 0.101 \times \text{CarSegment} + 0.042 \times \text{BuildingsSegment} \\ & + 0.498 \times \text{RoadSegment} + \epsilon\end{aligned}\tag{3.1}$$

$$\begin{aligned}y = & 34.378 - 0.061 \times \text{CarObject} + 0.187 \times \text{TrafficLightObject} \\ & - 2.026 \times \text{MotorcycleObject} - 0.170 \times \text{VegetationSegment} \\ & - 0.610 \times \text{PersonSegment} - 0.214 \times \text{SkySegment} \\ & - 0.779 \times \text{SidewalkSegment} - 0.122 \times \text{TerrainSegment} \\ & - 0.176 \times \text{VehicleSegment} - 0.232 \times \text{BuildingsSegment} + \epsilon\end{aligned}\tag{3.2}$$

Table 3.2 and 3.3 show the final tuned settings for the SVR and XGBoost model respectively. These settings have been used to build the final models.

Table 3.2: Optimal Settings for SVR for Different Data Samples

Parameter	Time sample	Revisit sample
C	10	1000
Gamma	1	0.5
Epsilon	1	0.1

Table 3.3: Optimal Settings for XGBoost for Different Data Samples

Parameter	Time sample	Revisit sample
Max Depth	5	3
Min Child Weight	5	1
Gamma	0	0
Learning Rate	0.01	0.01
N Estimators	100	100

3.2.2 Performance Metrics

When all model were trained, all metrics could be computed. Table 3.4 displays the R^2 and RMSE scores of all the models in both of the data samples.

Table 3.4: Comparison of Model Performance on Validation and Test Sets for Different Data Samples

Model	Time sample				Revisit sample			
	Validation Set		Test Set		Validation Set		Test Set	
	R^2 Score	RMSE	R^2 Score	RMSE	R^2 Score	RMSE	R^2 Score	RMSE
Baseline	0.00	11.89	0.00	12.98	0.00	9.65	-0.06	8.65
Linear Regression	0.05	11.96	0.02	12.33	0.06	9.35	0.05	8.20
SVR	0.02	12.12	0.00	12.42	0.06	9.36	0.14	7.81
XGBoost	0.03	12.05	0.02	12.30	0.08	9.23	0.11	7.96

3.2.3 Analysis of the Performance

Examining the results achieved by the models in Table 3.4, there are three main aspects that stand out:

1. Revisit Sample Performance Improvement Over Time Sample: All the models that were trained and tested on the revisit dataset outperformed the time sample dataset. This can be seen by the fact that the R^2 scores are higher and the RMSE values are closer to zero, indicating that the revisit sample might make it easier for models to pick up on certain patterns.

2. Revisit Sample Performance improvement Over The Baseline: All the models, especially those trained on the revisit sample, outperformed the baseline, indicating that they predict better than merely using the average value from the training set. However, the marginal improvements suggest that while the models learn more than averaging, they do not capture the underlying data patterns.
3. General Underperformance: Despite outperforming the baseline, all models exhibited generally poor performance. This can be seen by the R^2 scores being very close to zero and the RMSE values being not far from the median of the respective samples, as seen in Table 3.1.

The overall poor performance of all the models could indicate underfitting, meaning that the complexity and patterns in the data were not able to be captured by the models.

3.3 Issues and Limitations

There is some concern about issues in the data that could be limiting the full potential of the models. Those concerns were analyzed in this section.

3.3.1 Data Quality Issues

During this research four main issues have been noted with the quality of the data provided:

1. Overall Low Values: The summary statistics in Table 3.1 and the distribution in Figure 2.2 confirm that the overall NO_2 levels are fairly low. An increase in these levels is observed in the revisit sample, suggesting that this might be part of the issue.
2. Lack of spatial pattern: The zoomed in map of the study area as seen in Figure 2.1 and the full map of the study area found in Appendix A show that there is no clear spatial pattern at play. Most of the grids are the same color and there are not a lot of clear hotspots and especially not in places where these would be expected, traffic dense areas for example. This issue also becomes less pronounced with the revisit sample.
3. Spatial Outliers: During analysis of the spatial pattern, some spatial outliers

were also found. These are grid cells that exhibit high NO₂ levels but are located in places where the opposite is expected. Specific examples can be found in Appendix E. This also is partly aided by the revisit sample, this can be confirmed by looking at the summary statistics shown in Table 3.1, here the revisit sample has higher values for all statistics, except for the 95th percentile. This lower value for this statistic indicates that some higher values have been removed from this set.

4. Sparse Revisit Time: The amount of times a certain grid cell has been revisited on different days is very sparse. Most grids have only been visited two or three times, as can be seen in Table 2.1. As discussed in the other points the low frequency of revisits on most grids seems to have role in the data related issues, this can be seen in that using a sample based on the amount of times a grid has been visited aids the other data issues.

3.3.2 Impact on Performance

It is difficult to determine the exact impact of each of the mentioned issues on the performance of the models. However, the improved performance of models trained on the revisit sample suggests that these issues affect the models in some way, as this data sample eliminated part of these issues. Other than that, all the performance metrics of the models indicate underfitting which shows that the models fail to capture the complexity and patterns of the data.

3.4 Summary

As shown in Table 3.4, the models are very much underperforming and thus limiting the research is general. The issues with the quality of the data seem to play a major part in the underperformance of the models. Further examination of these issues will be conducted in the Discussion chapter (4).

4. Discussion

This study aimed to construct a pipeline that was able to efficiently predict NO₂ values by utilizing street view imagery. The proposed pipeline consisted of pre-trained state-of-the-art models to extract features and simpler secondary models to predict the NO₂ levels, this way the overall complexity remained low while still utilizing advanced models. The final pipeline and model created would be used to answer the following research question: *"How can features extracted from street view images using state-of-the-art deep learning models predict air quality in urban environments?"*.

In the Results chapter (3), it was observed that the models achieved poor results that did not align with outcomes reported by similar research using comparable methods and data. Alongside these poor results, multiple data quality issues were identified that potentially limited the models. In this chapter, further exploration will be conducted to understand why this might have happened and how this study differs from others that achieved better results.

4.1 Interpretation of Results

The models achieved poor performance over all the metrics and for both data samples, as is evident in Table 3.4. Apart from showcasing this poor performance of the models, Table 3.4 shows a more important finding in the difference between the metrics for the time and revisit samples. The revisit sample outperforms the time-based sample in every model on every metric, the reason this is interesting is that the revisit sample is, according to summary statistics (Table 4.2), less representative of the entire dataset.

There are nuances to be made about the revisit time achieving better model performance, such as that this set has a smaller study area and thus is easier to learn since the grid cells are closer together than those in the time sample and the overall RMSE being lower. However, when objectively examining the issues with the data and their impact on model performance, as discussed in the Issues and Limitations

section (3.3), it becomes evident that all the observed issues appear to be reduced or disappear entirely in the revisit sample. The combination of better model performance and the overall lesser effect of issues on the revisit set suggests that the frequency of grid revisits is a determining factor for the stability and predictability of the data.

4.1.1 Comparison with Other Studies

When comparing the results from this study to the results of other studies, it becomes more evident that the performance of the models used in this study is not expected. Similar studies that employed comparable methods reported R^2 scores as high as 0.95 and RMSE values as low as 2.71, depending on the city [24, 25]. This big difference in model performance raises concerns about the suitability of the feature set and the model's capacity to generalize from the training data. Apart from the data being gathered in a different city and the modeling methodology being slightly different, there are not many striking differences in the way these studies were conducted.

This stark difference in expected performance is evident in all the studies focusing on NO_2 or other pollutant predictions, which consistently achieve better metric scores compared to this study. Given that good results are consistently being achieved with studies from all around the world, utilizing vastly differing techniques and data, underlines the unexpectedness of the achieved performance.

4.2 Implications of Findings

The findings suggest that there is little to no predictability that can be extracted from street view images using this methodology. However, since the results achieved in this study were not in line with expected outcomes and the poor results seem to be correlated with the data sample used, a definitive conclusion cannot be drawn at this time. The difference in performance underscores the importance of data quality and illustrates how various segments of the data can vary significantly.

Because this study had a more practical focus, the main contribution to the field is the pipeline and methodology created for future research to utilize and refine.

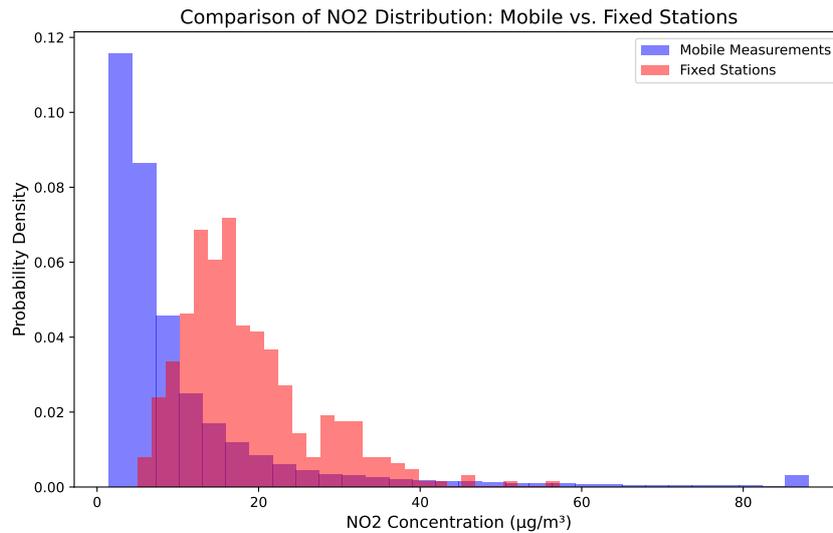


Figure 4.1: Comparative histogram of NO₂ concentration distributions from mobile measurements and fixed stations in Augsburg.

4.3 Data and Model Limitations

As was noted in the Results chapter (3) there are some data quality issues that influence the model performance. These subsections will provide more detail for each of the points mentioned and explain why they are considered issues.

4.3.1 Values

At first glance the values in the full NO₂ dataset seem to be lower than would be expected. This suspicion is confirmed when comparing the NO₂ levels from the mobile measuring campaign with the averaged levels of NO₂ from all fixed measurement stations in Augsburg [48] over the same time period. Figure 4.1 shows a histogram of these values. The histograms, normalized to show distribution shapes, highlight the differences in distribution between the two measurement sources in Augsburg. In this histogram, it is evident that the mobile measurements are substantially lower than the fixed measurements.

This difference is further highlighted when comparing the levels from the mobile campaign in Augsburg with mobile measurements from Amsterdam and Copenhagen, using the same measurement tools [5]. These measurements were not taken in the same period, but do give an idea of the difference in values. Table 4.1 shows the summary values from both of the measurement sources in Augsburg and the mobile campaigns from Amsterdam and Copenhagen. The fixed station values in

Augsburg fall between the mobile measurements of Amsterdam and Copenhagen, suggesting an ordinary distribution of the data. However, the Augsburg mobile measurements show significantly lower values across all summary statistics compared to the other values.

Table 4.1: NO₂ levels in Augsburg (mobile and fixed), Amsterdam and Copenhagen

	Augsburg mobile	Augsburg fixed	Amsterdam [5]	Copenhagen [5]
5th percentile	1.6	8.5	14.0	7.0
25th percentile	3.6	12.5	19.0	10.0
Median	6.1	14.9	24.0	13.0
Mean	10.6	15.8	28.0	17.0
75th percentile	11.4	18.4	33.0	20.0
95th percentile	36.4	25.3	55.0	39.0

Research indicated that mobile and stationary measurements can differ by as much as 48%, depending on the proximity [49]. These findings align with this, as the mean mobile measurement (10.6) was approximately 48% lower than that of fixed stations (15.8). Contrary to expectations, the stationary measurements were higher than the mobile measurements, which is the opposite of typical findings. This atypical difference and the overall lower values could be part of the reason why the models did not perform as expected.

4.3.2 Spatial pattern

Compared to patterns observed in previous studies [5, 8, 10, 11, 17, 18, 20, 22, 24, 29], which consistently reveal distinct spatial correlations around heavy traffic areas, industrial zones, and residential areas, the data in this study presents a less coherent pattern [10, 11, 13, 50]. The NO₂ measurements in this study are predominantly lower and do not align with the expected patterns identified in other research.

However, two roads, detailed in appendix D do display a pattern of elevated NO₂ levels consistent with these studies [5, 8, 20, 29]. Apart from these, almost all elevated measurements seem out of place. Such as unusually high NO₂ values in tranquil neighborhoods or abrupt spikes following periods of low measurements. In appendix E specific examples can be found. Overall, the spatial pattern of the dataset seems irregular and out of place compared to established trends.

The lack of a spatial pattern and the data not exhibiting known trends in NO₂ mea-

surements can be concluded to be a significant reason why the models were unable to perform better. Machine learning models are designed to extract patterns from data, when these patterns are absent or irregular, models tend to struggle.

4.3.3 Revisits

In the Results chapter (3) the stark difference between the data sample based on when the photos were taken and how often a grid was revisited on different days, already highlighted that the revisit frequency is an important aspect of the data. The increased performance from the revisit sample in combination with most of the data issues dissipating is an indication that this should be further investigated.

The revisit dataset showed fewer anomalies, this is supported by the updated summary statistics. Table 4.2 illustrates that the refined sample more closely aligns with the expected values from fixed stations, with improvements across most statistical measures except for the 95th percentile, which indicates a reduction in high outliers.

Table 4.2: Summary of NO₂ values from different data sets in Augsburg: revisit sample and full data from mobile measurements, and data from fixed measurement stations.

	Augsburg mobile (revisit sample)	Augsburg mobile (full data)	Augsburg fixed
5th percentile	4.8	1.6	8.5
25th percentile	7.6	3.6	12.5
Median	11.8	6.1	14.9
Mean	14.7	10.6	15.8
75th percentile	17.9	11.4	18.4
95th percentile	33.9	36.4	25.3

Figure 4.2 displays a comparative histogram that visually confirms this shift towards the expected distribution, bringing the sampled data closer in line with fixed station measurements.

As illustrated in Table 2.1 the average number of unique days a grid cell was revisited is approximately three, the majority of these, relatively high, average revisit locations centered in Augsburg, as depicted in the spatial plot of revisit days in Appendix F. Previous studies have indicated that for a long term air quality model, approximately nine drive days per grid are necessary to achieve stable NO₂ measurement [5]. Although the focus was on a short-term model, the frequency of grid

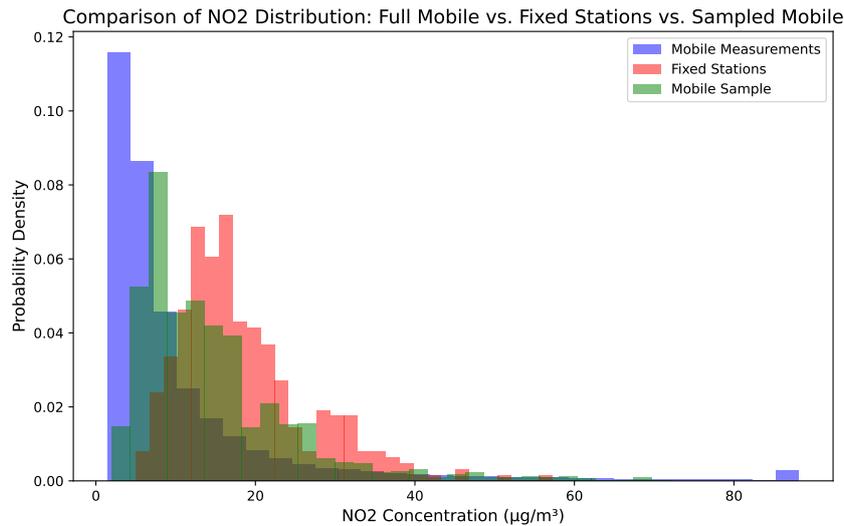


Figure 4.2: Comparative histogram of NO₂ concentration distributions from full mobile measurements, samples mobile measurements and fixed stations in Augsburg.

visits seems to have a high impact on the data patterns observed.

Observing that a sample with a minimum of just two revisit days already mitigates data quality issues and increases model performance, it can be concluded that a stable spatial pattern is essential for building an adequate model. In this case, the frequency at which a grid has been revisited is key to achieving this stability.

4.3.4 Location

The location of the measurement campaign might also significantly influence the data patterns observed. Augsburg, with a population of approximately 300,000 [51], is considerably smaller than cities like Amsterdam and Copenhagen, which have populations of about 900,000 [52] and 650,000 [53] respectively. Studies conducted in these larger cities have consistently demonstrated expected NO₂ patterns and values [5], utilizing the same measurement tools as in Augsburg.

However, a contrasting example can be found in a mobile measuring campaign conducted in 2017 in parts of Oakland, CA [29], which also showed very clear patterns, despite the overall city population being roughly similar to Augsburg at about 420,000 [54]. It is important to note that the specific areas studied in Oakland were not representative of the entire city, potentially putting the population of the study area closer to that of Augsburg. This suggests that mere city size or population numbers may not fully explain the differences observed in data patterns.

The absence of distinct patterns and the anomalously low values in the Augsburg data likely result from a complex interplay of factors including location, population density, land use, and the distribution of highways, residential areas, and green spaces. These elements collectively influence urban air quality patterns, indicating that a different approach is necessary to understand and predict NO₂ levels effectively. Understanding the dynamics at play is crucial to adapt air quality monitoring strategies to the specific characteristics of each urban area.

4.3.5 Ultrafine Particles Comparison

To ensure the validity of the results and to rule out the possibility of a faulty measurement device, parallel experiments were conducted with ultrafine particles (UFP) using a different device as part of the same mobile measurement campaign. The models followed nearly the same methodology, the only difference being that UFP was only trained, validated, and tested on the revisit sample due to time constraints. NO₂ and UFP levels are often correlated [5], and in this case, they exhibited a Pearson correlation coefficient of approximately 0.38. According to the findings detailed in Appendix J, the predictive performance for UFP was marginally better than for NO₂, but still suboptimal. This similarity in performance patterns between UFP and NO₂ suggests that the anomalies observed in the study are likely unrelated to equipment malfunction.

4.3.6 Summary

To summarize, the models utilized seem to be mostly limited by quality issues related to the data. The section above explored where these issues might come from and why they can influence the performance of the models. Other possibilities were also discussed, such as the location and characteristics of the study area compared to other studies, and the possibility of the NO₂ sensor malfunctioning. Taking all of the mentioned points into account, it can be concluded that data quality issues are the biggest obstacle for the models, and the root cause of this issue is likely the frequency at which a grid has been revisited.

4.4 Recommendations

To combat the data and model related limitation four recommendations can be made for future research:

1. **Plan Out Driving Routes:** Since the drive days are important, planning ahead to determine how often each location will be visited is essential for future mobile measurement campaign efforts. It is recommended to have a minimum of nine drive days per segment to be included in the analysis.
2. **Improved Data Integration:** This study faced challenges with integrating NO₂ measurements and images, resulting in the removal of about half the images. Future research could improve alignment by setting the correct time on both the measurement and image-taking devices or by consistently using a timestamp with a minimum precision of seconds.
3. **Temporal Correction:** Future research could benefit from implementing temporal corrections to the NO₂ measurements. Temporal adjustments aim to minimize the influence of background concentrations at specific times, thereby isolating local NO₂ levels more effectively. This adjustment involves using data from a fixed monitoring station to calibrate the mobile measurements, effectively normalizing for time specific variations. The methodology for such corrections has been extensively explained in previous studies [5, 55, 56]. Applying these corrections could enhance the accuracy of mobile air quality monitoring, providing a clearer picture of local pollution dynamics.
4. **Apply Pipeline on an Existing Dataset:** To verify that the results obtained from the experiments in this study are not due to errors in the experimental setup, it would be useful to conduct the same set of experiments on a dataset from a different city. Ideally, this dataset should come from a study that has successfully utilized a similar methodological approach. If the models also fail to deliver accurate predictions with this new dataset, it would suggest that either the methodological approach may not be optimal for predicting NO₂ levels or that there may be fundamental flaws in the experimental setup itself.

4.5 Conclusion

Reflecting on the original research question as stated in the Introduction (1), "*How can features extracted from street view images using state-of-the-art deep learning models predict air quality in urban environments?*", it becomes evident that the study did not fully achieve its intended objectives. Despite extensive experimentation, none of

the models tested managed to deliver considerable scores on the chosen metrics. The primary challenges, as extensively discussed in the Data and Methods (2) and Discussion chapters (4), come from data related issues that can be traced back to the revisit frequency of a grid cell.

Nevertheless, this research successfully developed a pipeline capable of formatting street view images, extracting features using state-of-the-art models, and utilizing these features as inputs for secondary models. This pipeline significantly lowers the technical barriers typically associated with developing and training complex deep neural networks. This research also underlines the importance of data quality, especially in spatial patterns, for models to perform as expected.

In conclusion, while this study did not succeed in accurately predicting air quality in urban environments as initially hoped, it established a foundational pipeline that facilitates further exploration and refinement. Future research, as outlined in the Discussion chapter (4), can expand upon this groundwork to enhance the pipeline's effectiveness and potentially achieve more accurate predictions of urban air quality.

Appendices

A. Full spatial Overview of NO₂

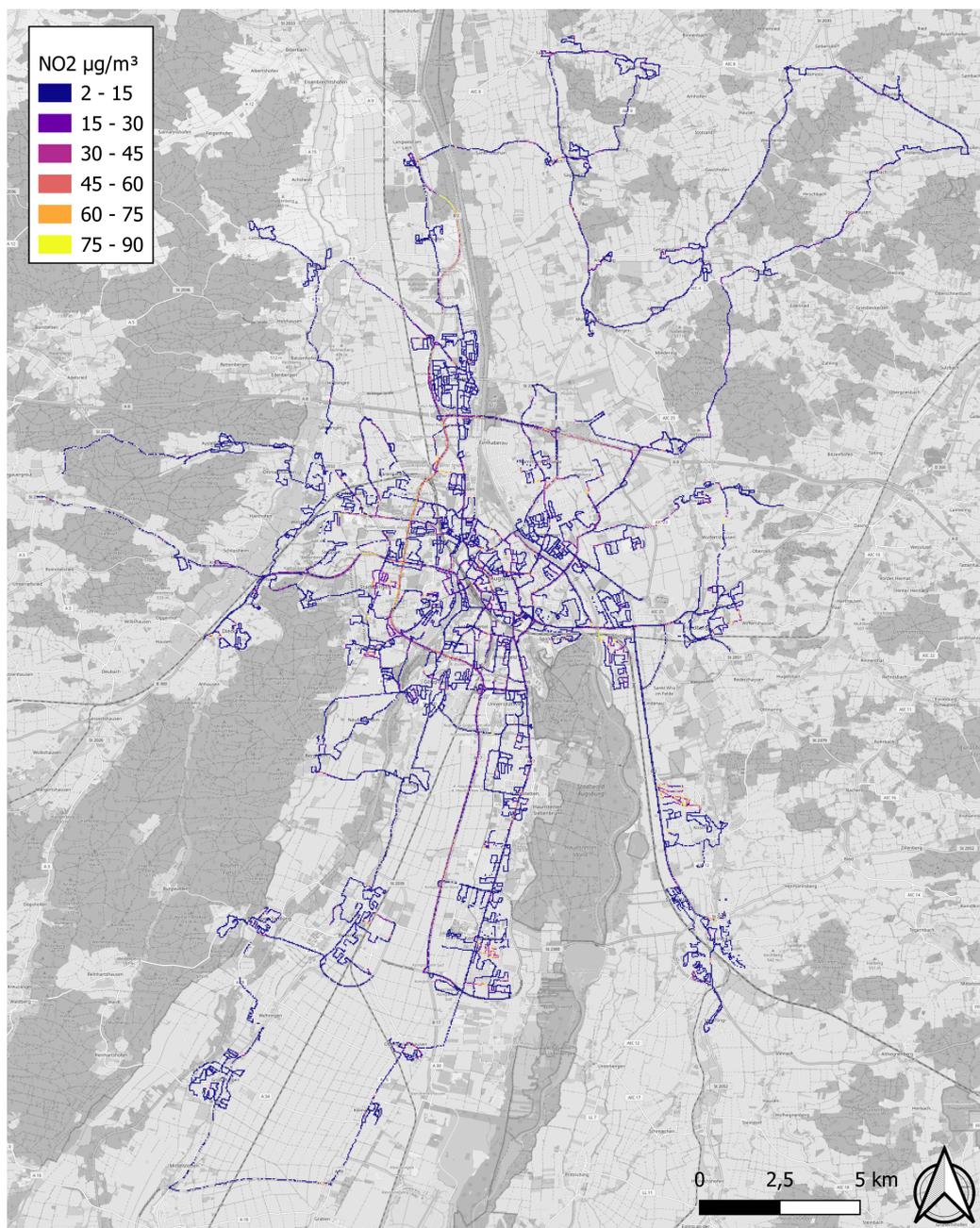


Figure A.1: Map of Augsburg and direct surroundings showing the mean NO₂ in 50 meter grid cells.

B. Data Integration examples



Figure B.1: Example of a good image where time is shown with second precision



Figure B.2: Example of a bad image where time is shown with minute precision

C. Code

https://mbees.med.uni-augsburg.de/gitlab/mbees/airview_no2_augsburg

D. Good Examples of Spatial NO₂ Distribution

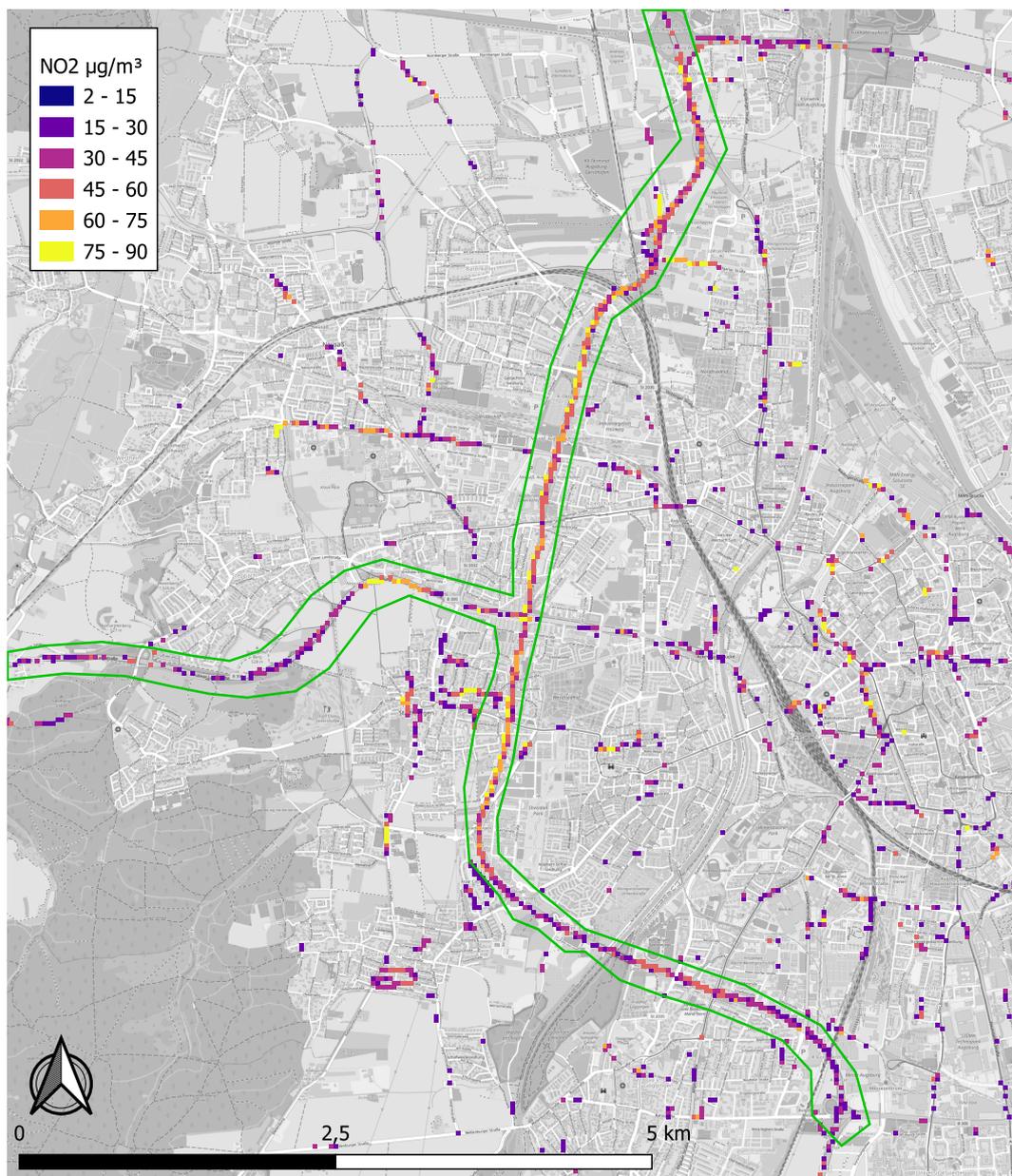


Figure D.1: Example of expected spatial pattern, highlighted by the green area.

E. Examples of Anomalies in the NO₂ Distribution

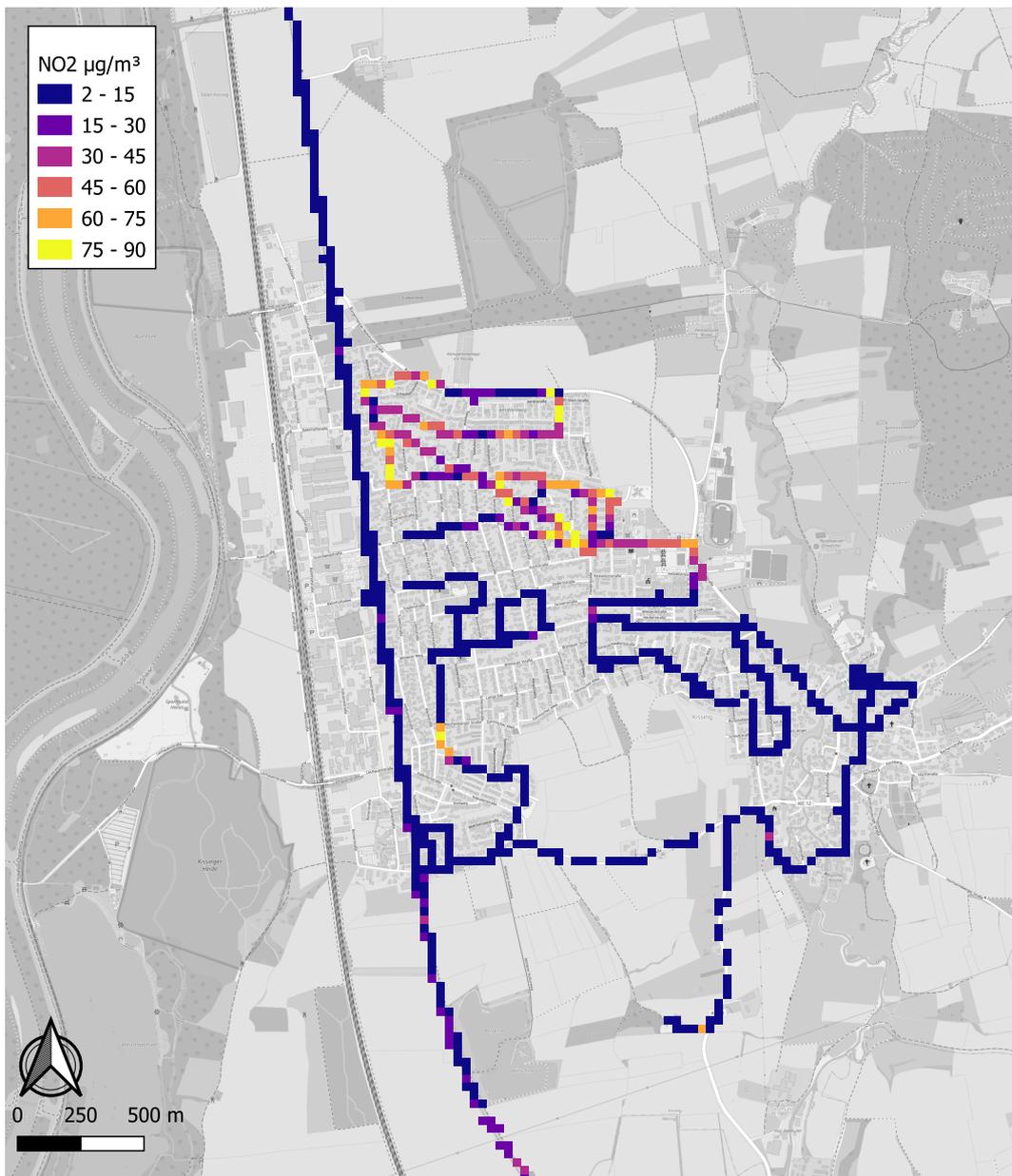


Figure E.1: Example of anomaly pattern.

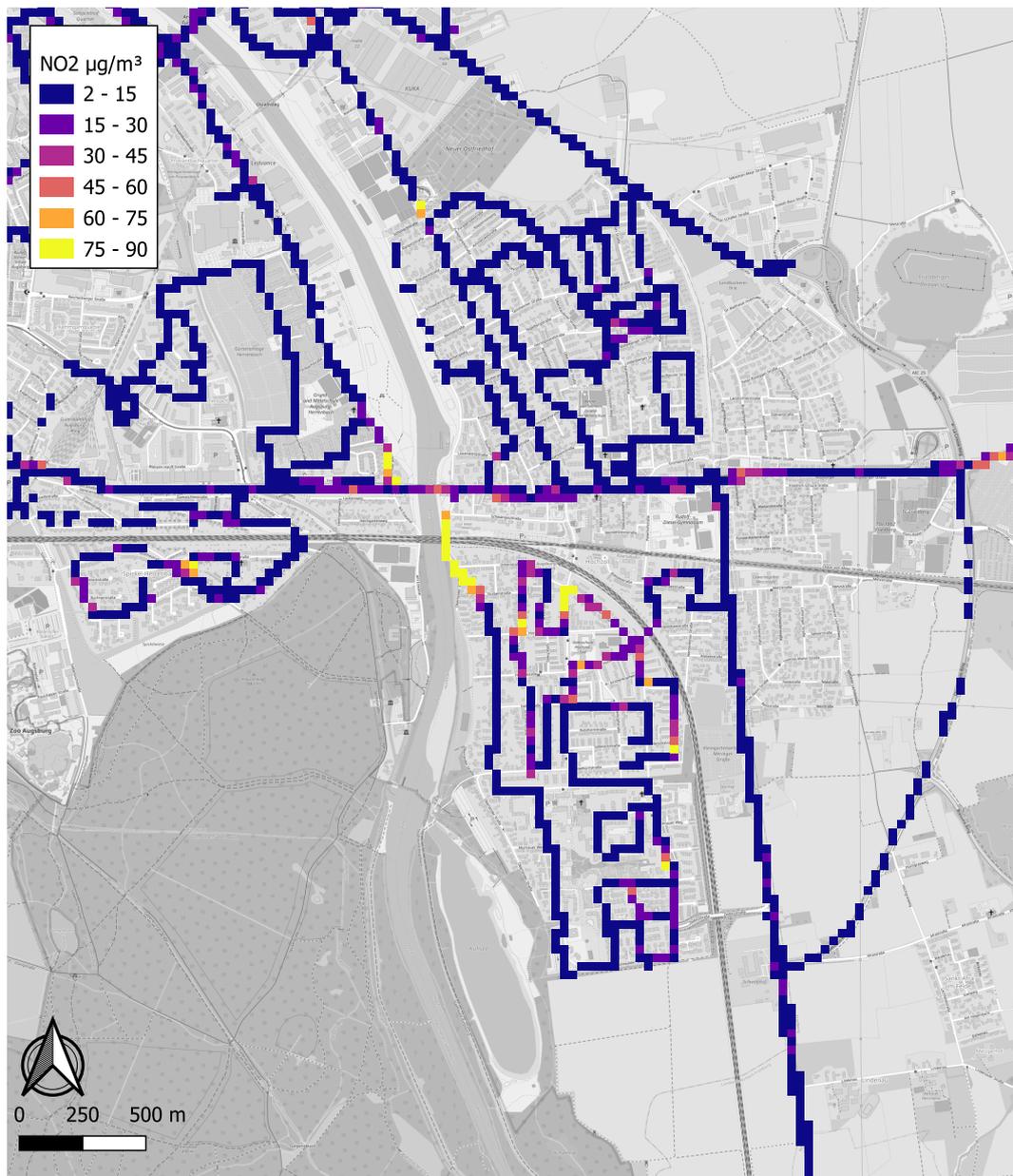


Figure E.2: Example of anomaly pattern.

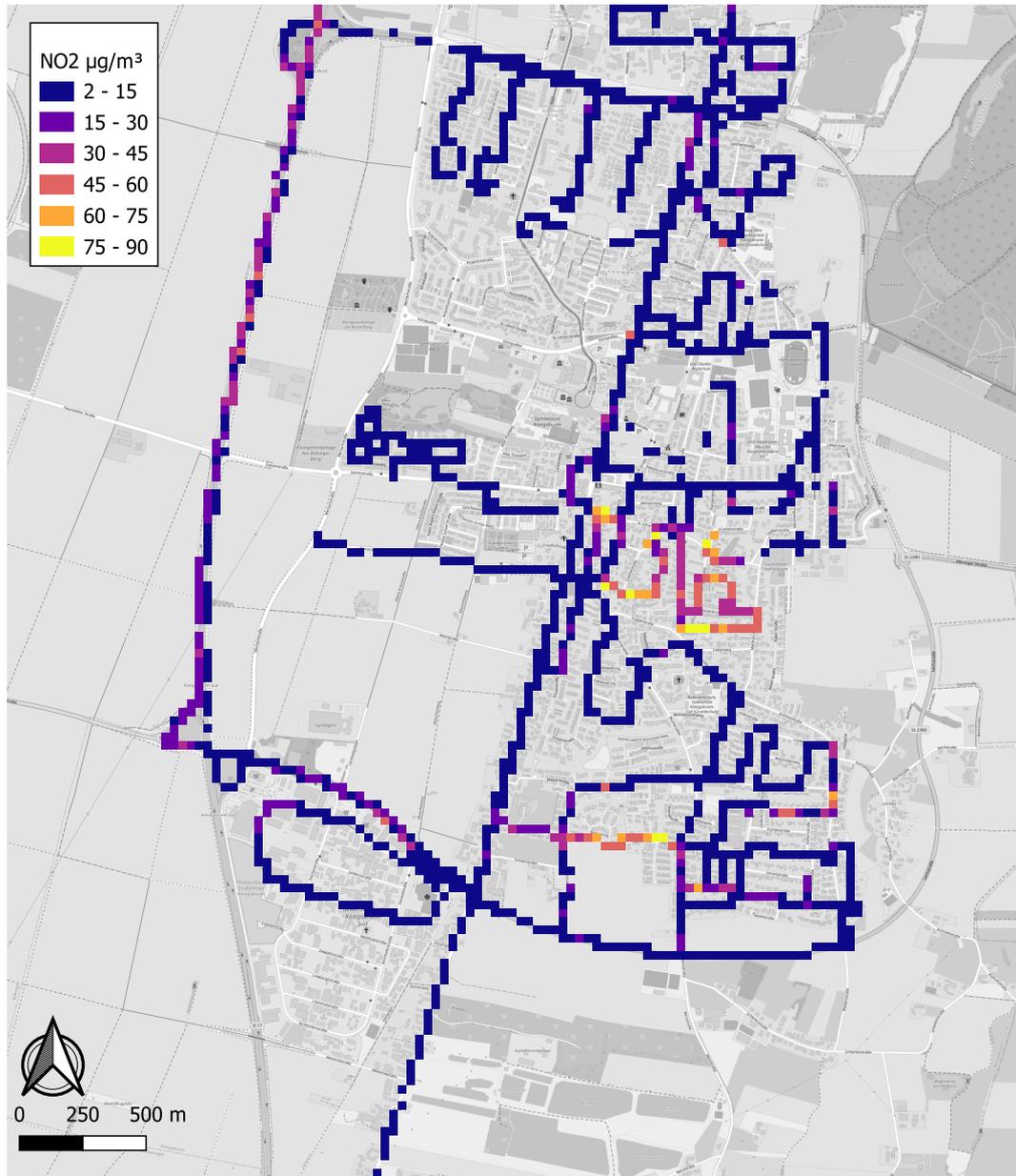


Figure E.3: Example of anomaly pattern.

F. Full Spatial Revisit Count Distribution

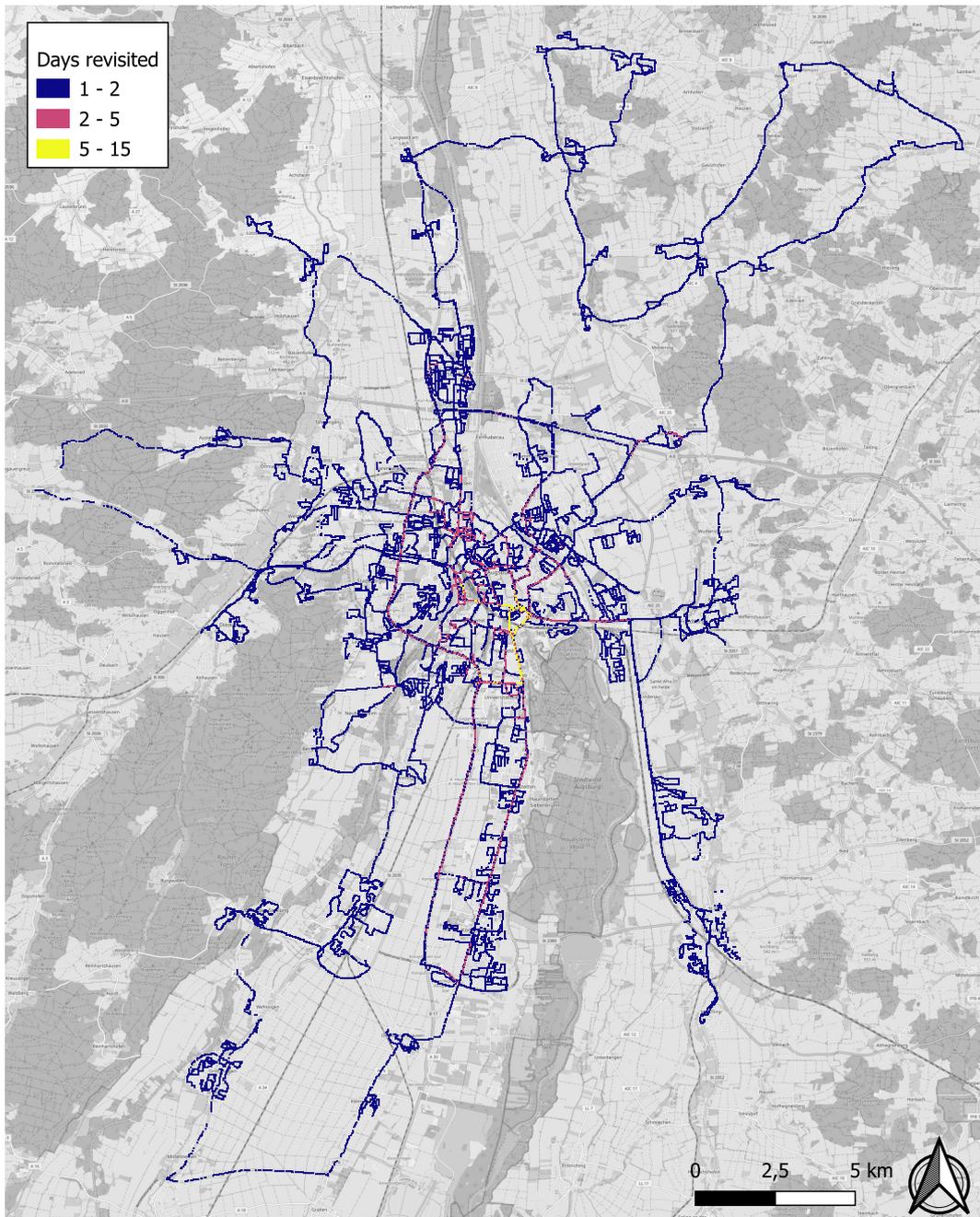


Figure F.1: Overview of revisit frequency.

G. Final Data Sample Spatial Overview of NO₂

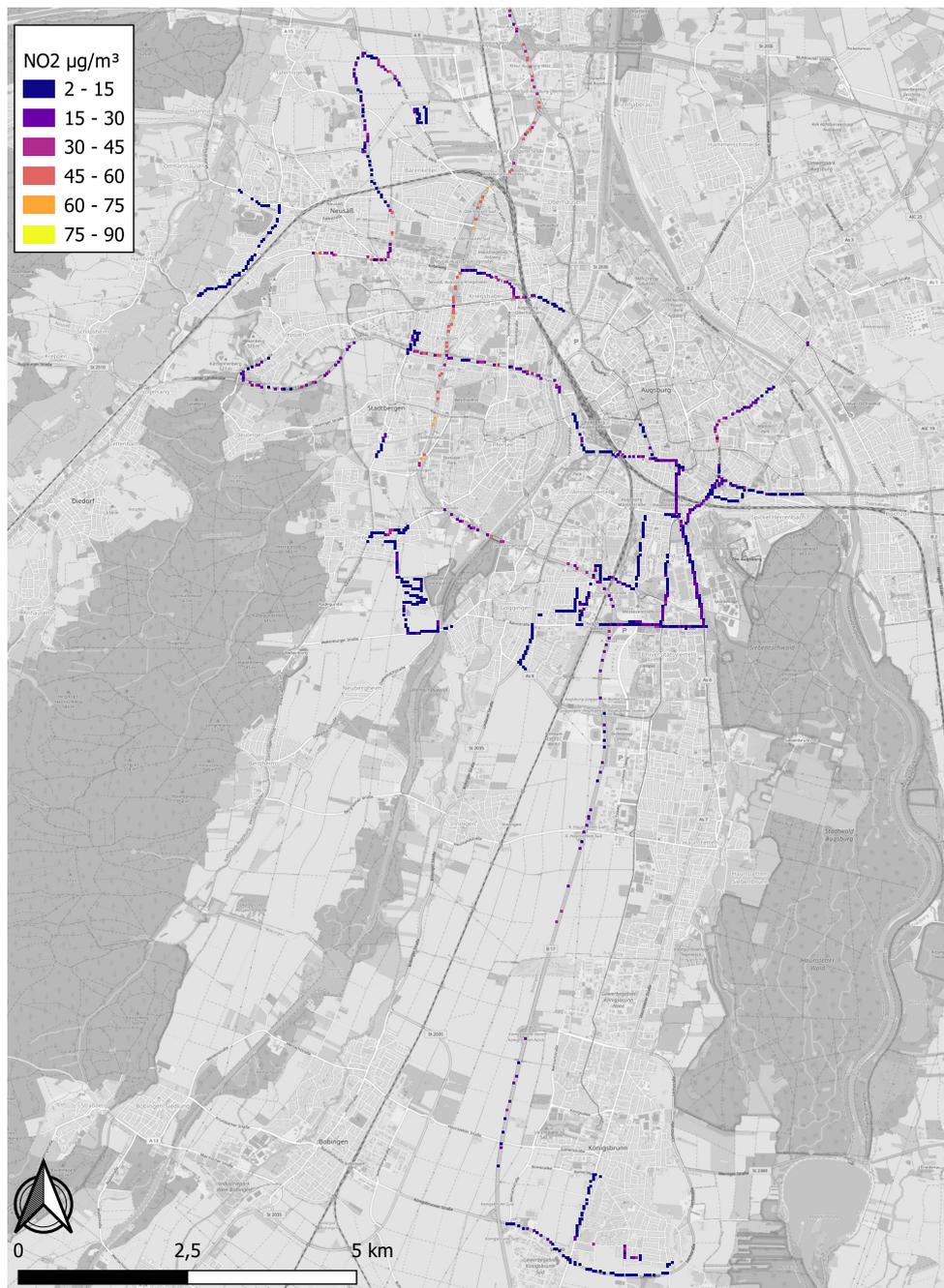


Figure G.1: Final sample NO₂ overview.

H. Final Data Sample Split

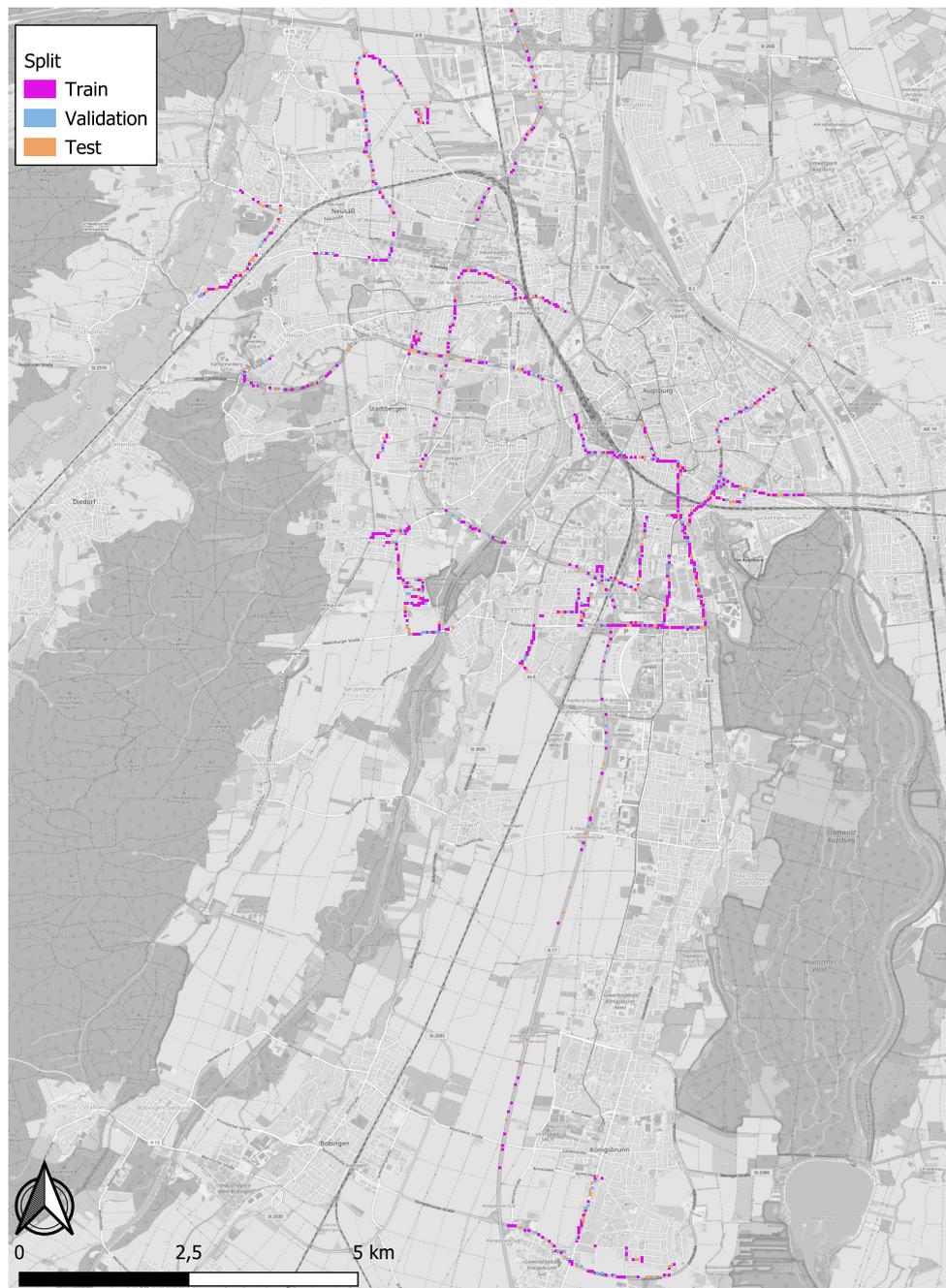


Figure H.1: Final data split.

I. Pre-trained Features

Table I.1: List of features and their frequencies used in the object detection model.

Feature	Frequency
Car	58669
Truck	9988
Person	5595
Traffic Light	9527
Bus	898
Bicycle	995
Stop Sign	935
Motorcycle	271

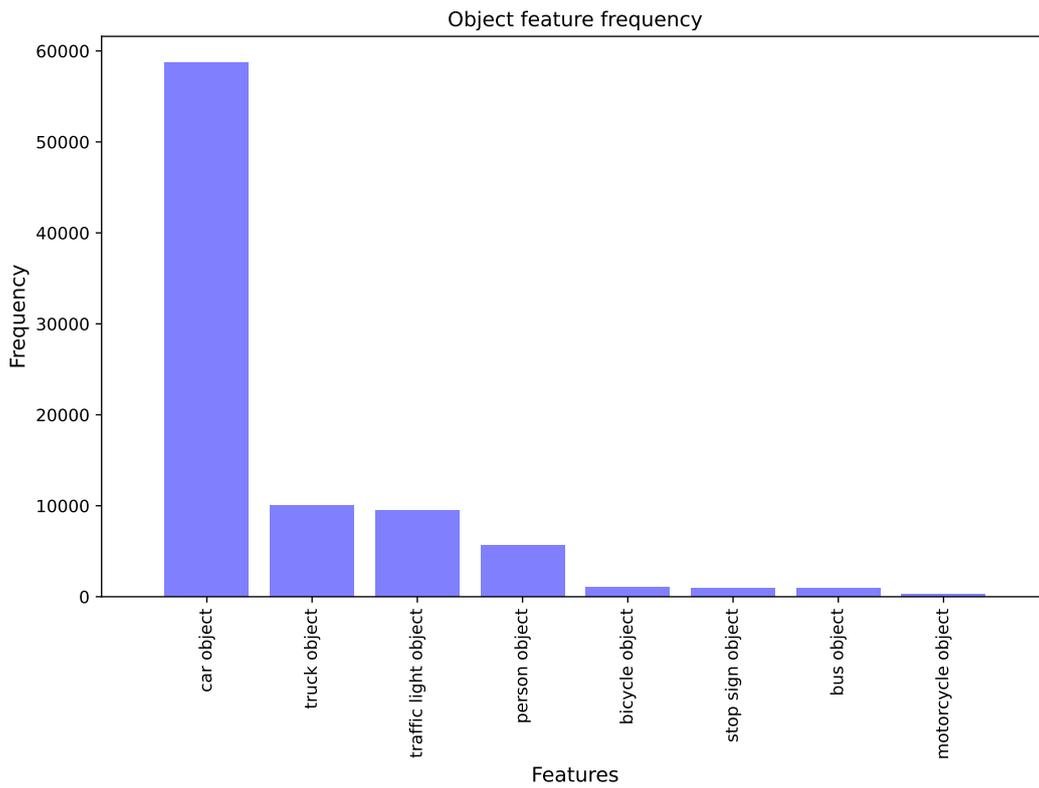


Figure I.1: Detected object frequency.

Table I.2: List of features and their frequencies used in the semantic segmentation model.

Feature	Frequency
Sky	263862.5
Vegetation	251589.2
Buildings	111503.1
Road	65379.6
Vehicle	43455.9
Terrain	25802.9
Sidewalk	17400.2
Pole	8261.1
Person	1400.2
Traffic Sign	1368.1
Bicycle	567.4
Traffic Light	389.2
Rider	205.0

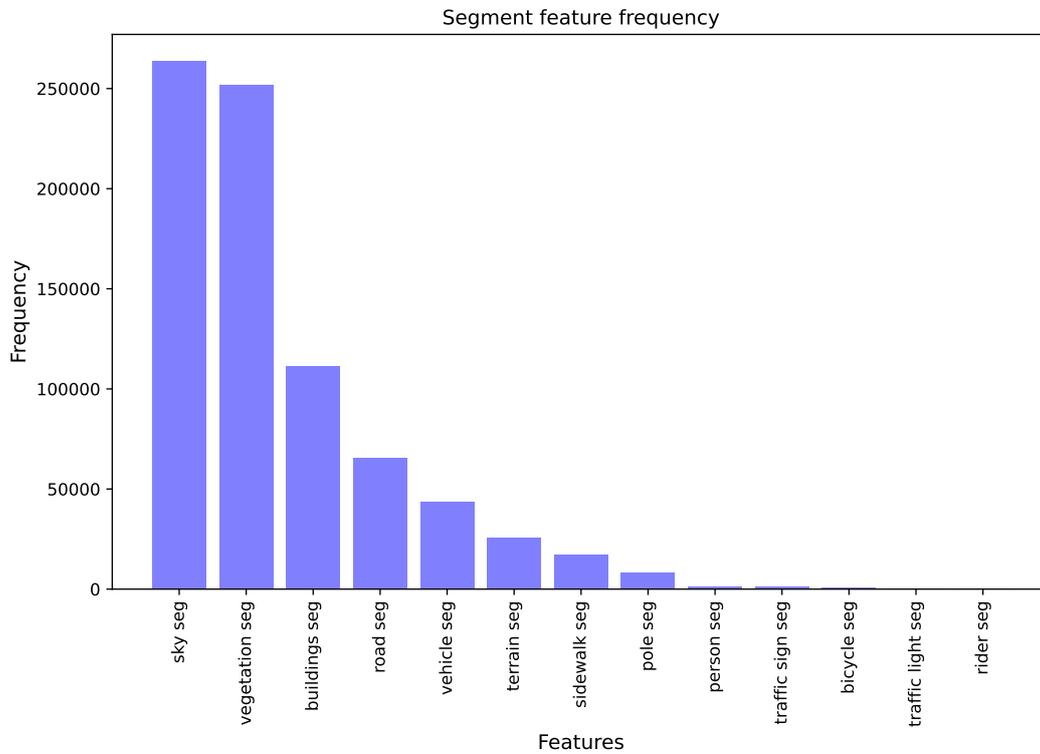


Figure I.2: Detected semantic frequency.

J. UFP Results

Table J.1: Comparison of Model Performance on Validation and Test Sets

Model	Validation Set		Test Set	
	R ² Score	RMSE	R ² Score	RMSE
Baseline	-1.10	13354.79	-0.02	12496.97
Linear Regression	0.08	12826.20	0.11	11681.01
SVR	0.03	13155.77	0.10	11762.88
XGBoost	0.16	12223.63	0.25	10700.29

Bibliography

- [1] 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". In: *International journal of hygiene and environmental health* 212.3 (May 2009), pp. 271–287. DOI: 10.1016/j.ijheh.2008.06.003. URL: <https://doi.org/10.1016/j.ijheh.2008.06.003>.
- [2] Renjie Chen et al. "Associations between short-term exposure to nitrogen dioxide and mortality in 17 Chinese cities: the China Air Pollution and Health Effects Study (CAPES)." In: *Environment international* 45 (2012), pp. 32–8. DOI: 10.1016/j.envint.2012.04.008.
- [3] United States Environmental Protection Agency. *Basic information about NO2* | US EPA. July 2023. URL: <https://www.epa.gov/no2-pollution/basic-information-about-no2>.
- [4] Shuo Huang et al. "Long-term exposure to nitrogen dioxide and mortality: A systematic review and meta-analysis". In: *Science of the total environment* 776 (July 2021), p. 145968. DOI: 10.1016/j.scitotenv.2021.145968. URL: <https://doi.org/10.1016/j.scitotenv.2021.145968>.
- [5] Jules Kerckhoffs et al. "Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen". In: *Environment international* 170 (Dec. 2022), p. 107575. DOI: 10.1016/j.envint.2022.107575. URL: <https://doi.org/10.1016/j.envint.2022.107575>.
- [6] P. Morten Hundt et al. "Mid-IR spectrometer for mobile, real-time urban NO2 measurements". In: *Atmospheric measurement techniques* 11.5 (May 2018), pp. 2669–2681. DOI: 10.5194/amt-11-2669-2018. URL: <https://doi.org/10.5194/amt-11-2669-2018>.
- [7] Maria Cerrato-Alvarez et al. "A portable, low-cost, smartphone assisted methodology for on-site measurement of NO2 levels in ambient air by selective chemical reactivity and digital image analysis". In: *Sensors and actuators. B, Chemical* 338 (July 2021), p. 129867. DOI: 10.1016/j.snb.2021.129867. URL: <https://doi.org/10.1016/j.snb.2021.129867>.
- [8] Rowell Hagemann et al. "Spatial variability of particle number concentrations and NOx in the Karlsruhe (Germany) area obtained with the mobile laboratory 'AERO-TRAM'". In: *Atmospheric environment* 94 (Sept. 2014), pp. 341–352. DOI: 10.1016/j.atmosenv.2014.05.051. URL: <https://www.sciencedirect.com/science/article/pii/S1352231014003987>.
- [9] A. Larkin et al. "Global Land Use Regression Model for Nitrogen Dioxide Air Pollution." In: *Environmental science & technology* 51 12 (2017), pp. 6957–6964. DOI: 10.1021/acs.est.7b01148.
- [10] Allison P. Patton et al. "Spatial and temporal differences in traffic-related air pollution in three urban neighborhoods near an interstate highway". In: *Atmospheric environment* 99 (Dec. 2014), pp. 309–321. DOI: 10.1016/j.

- atmosenv.2014.09.072. URL: <https://doi.org/10.1016/j.atmosenv.2014.09.072>.
- [11] Adyati P. Yudison and R. Driejana. "Development of indoor air pollution concentration prediction by geospatial analysis". In: *Journal of Engineering and Technological Sciences* 47.3 (July 2015), pp. 306–319. DOI: 10.5614/j.eng.technol.sci.2015.47.3.6. URL: <https://doi.org/10.5614/j.eng.technol.sci.2015.47.3.6>.
- [12] Jochem O. Klompmaker et al. "Spatial variation of ultrafine particles and black carbon in two cities: Results from a short-term measurement campaign". In: *Science of the total environment* 508 (Mar. 2015), pp. 266–275. DOI: 10.1016/j.scitotenv.2014.11.088. URL: <https://www.sciencedirect.com/science/article/pii/S0048969714016829?via%3Dihub>.
- [13] David C. Carslaw et al. "The diminishing importance of nitrogen dioxide emissions from road vehicle exhaust". In: *Atmospheric environment*. X 1 (Jan. 2019), p. 100002. DOI: 10.1016/j.aeaoa.2018.100002. URL: <https://www.sciencedirect.com/science/article/pii/S2590162118300029#:~:text=Abstract,2%20in%20many%20European%20cities>.
- [14] Nichole Baldwin et al. "Factors affecting pollutant concentrations in the near-road environment". In: *Atmospheric environment* 115 (Aug. 2015), pp. 223–235. DOI: 10.1016/j.atmosenv.2015.05.024. URL: <https://www.sciencedirect.com/science/article/pii/S1352231015300972?via%3Dihub>.
- [15] Wan Nurul Farah Wan Azmi et al. "Application of land use regression model to assess outdoor air pollution exposure: A review". In: *Environmental advances* 11 (Apr. 2023), p. 100353. DOI: 10.1016/j.envadv.2023.100353. URL: <https://www.sciencedirect.com/science/article/pii/S2666765723000133#bib0055>.
- [16] Rob Beelen et al. "Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project". In: *Atmospheric environment* 72 (June 2013), pp. 10–23. DOI: 10.1016/j.atmosenv.2013.02.037. URL: <https://www.sciencedirect.com/science/article/pii/S1352231013001386#tbl2>.
- [17] Ornella Luminati et al. "Land use regression modelling of NO₂ in São Paulo, Brazil". In: *Environmental pollution* 289 (Nov. 2021), p. 117832. DOI: 10.1016/j.envpol.2021.117832. URL: <https://www.sciencedirect.com/science/article/pii/S0269749121014147>.
- [18] Zhendong Yuan et al. "LUR modeling of long-term average hourly concentrations of NO₂ using hyperlocal mobile monitoring data". In: *Science of the total environment* (Feb. 2024), p. 171251. DOI: 10.1016/j.scitotenv.2024.171251. URL: <https://www.sciencedirect.com/science/article/pii/S0048969724013901?via%3Dihub>.
- [19] Markus Fritsch and Svenia Behm. "Agglomeration and infrastructure effects in land use regression models for air pollution – Specification, estimation, and interpretations". In: *Atmospheric environment* 253 (May 2021), p. 118337. DOI: 10.1016/j.atmosenv.2021.118337. URL: <https://www.sciencedirect.com/science/article/pii/S1352231021001552>.
- [20] Yahya Ghassoun and Marc-Oliver Löwner. "Land use regression models for total particle number concentrations using 2D, 3D and semantic parameters". In: *Atmospheric environment* 166 (Oct. 2017), pp. 362–373. DOI:

- 10.1016/j.atmosenv.2017.07.042. URL: <https://www.sciencedirect.com/science/article/pii/S1352231017304910>.
- [21] Bryan Hellack et al. "Land use regression modeling of oxidative potential of fine particles, NO₂, PM_{2.5} mass and association to type two diabetes mellitus". In: *Atmospheric environment* 171 (Dec. 2017), pp. 181–190. DOI: 10.1016/j.atmosenv.2017.10.017. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1352231017306787>.
- [22] Shali Tayebi et al. "Comparison of NO₂ and BC predictions estimated using Google Street View-Based and conventional European-Wide LUR models in Copenhagen, Denmark". In: *Atmosphere* 14.11 (Oct. 2023), p. 1602. DOI: 10.3390/atmos14111602. URL: <https://www.mdpi.com/2073-4433/14/11/1602>.
- [23] Tianhong Zhao et al. "Sensing urban soundscapes from street view imagery". In: *Computers, environment and urban systems* 99 (Jan. 2023), p. 101915. DOI: 10.1016/j.compenvurbsys.2022.101915. URL: <https://doi.org/10.1016/j.compenvurbsys.2022.101915>.
- [24] Esra Suel et al. "What you see is what you breathe? Estimating air pollution spatial variation using Street-Level imagery". In: *Remote sensing* 14.14 (July 2022), p. 3429. DOI: 10.3390/rs14143429. URL: <https://doi.org/10.3390/rs14143429>.
- [25] Jing Huang et al. "Estimating urban noise along road network from street view imagery". In: *International journal of geographical information science* 38.1 (Nov. 2023), pp. 128–155. DOI: 10.1080/13658816.2023.2274475. URL: <https://doi.org/10.1080/13658816.2023.2274475>.
- [26] Kris Y. Hong, Pedro O. Pinheiro, and Scott Weichenthal. "Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data". In: *Environment international* 144 (Nov. 2020), p. 106044. DOI: 10.1016/j.envint.2020.106044. URL: <https://www.sciencedirect.com/science/article/pii/S0160412020319991>.
- [27] Esra Suel et al. "Measuring social, environmental and health inequalities using deep learning and street imagery". In: *Scientific reports* 9.1 (Apr. 2019). DOI: 10.1038/s41598-019-42036-w. URL: <https://doi.org/10.1038/s41598-019-42036-w>.
- [28] Scott Weichenthal, Marianne Hatzopoulou, and Michael Brauer. "A picture tells a thousand... exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology". In: *Environment international* 122 (Jan. 2019), pp. 3–10. DOI: 10.1016/j.envint.2018.11.042. URL: <https://www.sciencedirect.com/science/article/pii/S0160412018322001>.
- [29] Joshua S. Apte et al. "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data". In: *Environmental science & technology* 51.12 (June 2017), pp. 6999–7008. DOI: 10.1021/acs.est.7b00891. URL: <https://doi.org/10.1021/acs.est.7b00891>.
- [30] Arman Ganji et al. "Predicting Traffic-Related Air Pollution Using Feature Extraction from Built Environment Images". In: *Environmental science & technology* 54.17 (Aug. 2020), pp. 10688–10699. DOI: 10.1021/acs.est.0c00412. URL: <https://doi.org/10.1021/acs.est.0c00412>.

- [31] Nicolas Carion et al. “End-to-End Object Detection with Transformers”. In: *CoRR* abs/2005.12872 (2020). arXiv: 2005.12872. URL: <https://arxiv.org/abs/2005.12872>.
- [32] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in context*. Jan. 2014, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_{_}48. URL: https://doi.org/10.1007/978-3-319-10602-1_48.
- [33] AI at Meta. *facebook/detr-resnet-101 · Hugging Face*. May 2020. URL: <https://huggingface.co/facebook/detr-resnet-101>.
- [34] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [35] Bowen Cheng et al. *Masked-attention mask transformer for universal image segmentation*. Dec. 2021. URL: <https://arxiv.org/abs/2112.01527>.
- [36] Bowen Cheng, Alexander G. Schwing, and Alexander Kirillov. *Per-Pixel Classification is Not All You Need for Semantic Segmentation*. July 2021. URL: <https://arxiv.org/abs/2107.06278>.
- [37] Marius Cordts et al. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. Apr. 2016. URL: <https://arxiv.org/abs/1604.01685>.
- [38] AI at Meta. *facebook/mask2former-swin-large-cityscapes-semantic · Hugging Face*. Dec. 2021. URL: <https://huggingface.co/facebook/mask2former-swin-large-cityscapes-semantic>.
- [39] André I. Khuri. “Introduction to Linear Regression Analysis, Fifth Edition by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining”. In: *International statistical review* 81.2 (Aug. 2013), pp. 318–319. DOI: 10.1111/insr.12020_{_}10. URL: https://doi.org/10.1111/insr.12020_10.
- [40] Alex J. Smola and Bernhard Schölkopf. “A tutorial on support vector regression”. In: *Statistics and computing* 14.3 (Aug. 2004), pp. 199–222. DOI: 10.1023/b:stco.0000035301.49549.88. URL: <https://doi.org/10.1023/b:stco.0000035301.49549.88>.
- [41] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. URL: <https://doi.org/10.1145/2939672.2939785>.
- [42] Hugging Face. *Pipelines*. URL: https://huggingface.co/docs/transformers/en/main_classes/pipelines.
- [43] Scikit-learn. *scikit-learn: machine learning in Python — scikit-learn 1.5.0 documentation*. URL: <https://scikit-learn.org/stable/>.
- [44] xgboost developers. *XGBoost Documentation — xgboost 2.0.3 documentation*. URL: <https://xgboost.readthedocs.io/en/stable/>.
- [45] Pandas. *pandas - Python Data Analysis Library*. URL: <https://pandas.pydata.org/>.
- [46] Numpy. *NumPy*. URL: <https://numpy.org/>.
- [47] Scikit-learn. *SVR*. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR>.
- [48] Landesamt für Umwelt (LfU). *Messwertarchiv - LfU Bayern*. May 2020. URL: <https://www.lfu.bayern.de/luft/immissionsmessungen/messwertarchiv/index.htm>.

- [49] Ying Zhu et al. "Spatial and temporal representativeness of point measurements for nitrogen dioxide pollution levels in cities". In: *Atmospheric chemistry and physics* 20.21 (Nov. 2020), pp. 13241–13251. DOI: 10.5194/acp-20-13241-2020. URL: <https://doi.org/10.5194/acp-20-13241-2020>.
- [50] Silvana Di Sabatino, Riccardo Buccolieri, and Prashant Kumar. *Spatial distribution of air pollutants in cities*. Sept. 2017, pp. 75–95. DOI: 10.1007/978-3-319-62731-1_{_}5. URL: https://doi.org/10.1007/978-3-319-62731-1_5.
- [51] Statista. *Entwicklung der Einwohnerzahl in Augsburg bis 2022*. May 2024. URL: <https://de.statista.com/statistik/daten/studie/375300/umfrage/entwicklung-der-gesamtbevoelkerung-in-augsburg>.
- [52] Statista. *Amsterdam: total population 2023 | Statista*. July 2023. URL: <https://www.statista.com/statistics/753235/total-population-of-amsterdam/>.
- [53] Statista. *Copenhagen: total population 2023 | Statista*. Aug. 2023. URL: <https://www.statista.com/statistics/1303909/population-copenhagen/>.
- [54] U.S. Census Bureau. *Explore Census data*. URL: <https://data.census.gov/table/DECENNIALPL2020.P1?q=Oakland%20city,%20California>.
- [55] Jules Kerckhoffs et al. "Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring". In: *Environment international* 154 (Sept. 2021), p. 106569. DOI: 10.1016/j.envint.2021.106569. URL: <https://doi.org/10.1016/j.envint.2021.106569>.
- [56] Jules Kerckhoffs et al. "Robustness of intra urban land-use regression models for ultrafine particles and black carbon based on mobile monitoring". In: *Environmental research* 159 (Nov. 2017), pp. 500–508. DOI: 10.1016/j.envres.2017.08.040. URL: <https://doi.org/10.1016/j.envres.2017.08.040>.