The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-6-2025
ISPRS, EARSeL & DGPF Joint Istanbul Workshop “Topographic Mapping from Space” dedicated to Dr. Karsten Jacobsen’s 80th Birthday
29–31 January 2025, Istanbul, Türkiye

Generating Training Data for Deep Learning-Based Segmentation Algorithms by Projecting Existing Labels onto Additional Aerial Images

Franz Kurz [1], Nina Merkle [1], Corentin Henry [1], Reza Bahmanyar [1], Felix Rauch [1], Jens Hellekes [1], Veronika Gstaiger [1], Dominik Rosenbaum [1], Peter Reinartz [1]

[1] DLR, Earth Observation Center, 82234 Wessling, Germany - (firstname.lastname)@dlr.de

Keywords: Segmentation, Deep Learning, Label Generation, Aerial Images, Traffic Infrastructure.

Abstract

Highly accurate manually-generated labels in aerial and satellite images are used for the training of deep learning-based segmentation algorithms and should be available in large numbers and cover many different scenarios to increase the accuracy and generalization capability of the underlying models. Existing labels can be efficiently reused by photogrammetric projections onto additional overlapping aerial or satellite images, enabling great variability in the appearance of the scenes based on differences in viewing angles and environmental conditions. In this work, we investigate whether the additionally generated training data can effectively lead to an increase in prediction accuracy. To this end, we collected aerial images overlapping with the already annotated Traffic Infrastructure and Surroundings (TIAS) dataset, taken from a large-scale historical database spanning 2011 to 2024, and generated new training data by means of photogrammetric projections of existing labels onto these additional images. Training a Dense-U-Net model on the whole TIAS dataset or a part therefore, with and without additional projected labels, showed that this technique could be beneficial to improve the performance of a model if only a small amount of annotations is available comparatively to a large amount of overlapping aerial images.

1. Introduction

Deep Learning (DL)-based semantic segmentation algorithms are crucial for generating thematic maps that support various applications such as urban planning, environmental monitoring, disaster management, and traffic analysis. These models rely on aerial images annotated with highly-accurate labels for training and evaluation. However, the process of manually annotating large-scale datasets is time-consuming and resource-intensive, limiting scalability and efficiency. As shown in (Zlateski et al., 2018), the most important factor for creating adequate training datasets is to allocate manpower to either annotate a large amount of data in a coarse manner, or a small amount of data in an accurate manner. In the end, the total amount of effort spent on a labeling task directly determines the maximum performance obtainable from a model, indicating that scaling up a dataset even with imperfect labels can be beneficial. This highlights the need for innovative approaches to reduce manual efforts, while maintaining or improving annotation accuracy, ultimately leading to more accurate and widely applicable segmentation models.

Several existing approaches have attempted to address the challenge of data generation with little to no manual labeling. One such approach, described in (Tian et al., 2023), involves projecting curbstone positions from a database onto various aerial images to generate training data. In (Chiciudean et al., 2024), the authors propose a data augmentation technique for UAV data where they propagate manually labeled images into a 3D mesh and generate images with new views. In addition, (Toker et al., 2024) explores the use of generative image diffusion to generate high-quality and diverse image labels for satellite imagery, achieving notable improvements in semantic segmentation performance.

In this paper, we investigate the possibility of reusing existing manually-generated annotations in aerial images through projections onto other overlapping aerial images from a large database. The main benefit is the reduction of the manual workload while increasing the variability in the appearance of the scenes. The DL-based algorithms are also expected to be able to train to become more accurate and robust, as the projected labels bring an extended diversity of visual appearance and environmental conditions for the individual object classes, such as different lighting situations, different brightness levels, surface conditions such as rain, resolution levels, etc. for the same scenes. At the same time, inaccuracies arise during the projection of the annotations, which can be roughly divided into geometric and semantic errors. Geometric errors are caused by inaccuracies in the georeferencing of the aerial images, by errors in the elevation model, and by shading. Semantic errors are mainly caused by the temporal distance between the aerial images and the resulting projected labels, which may as such no longer constitute an exact ground truth.

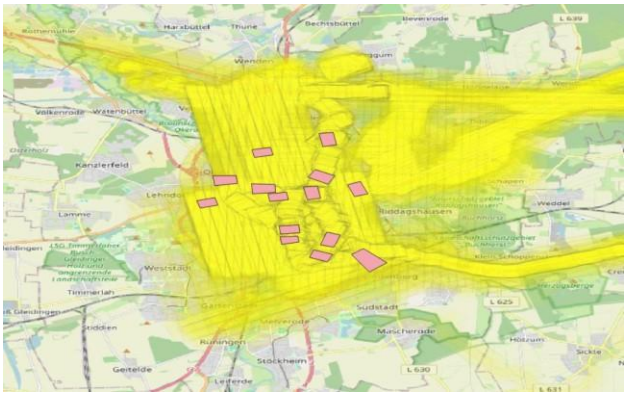


Figure 1. Coverage of our aerial image database (yellow) and labeled image footprints (pink) over Brunswick, Germany.

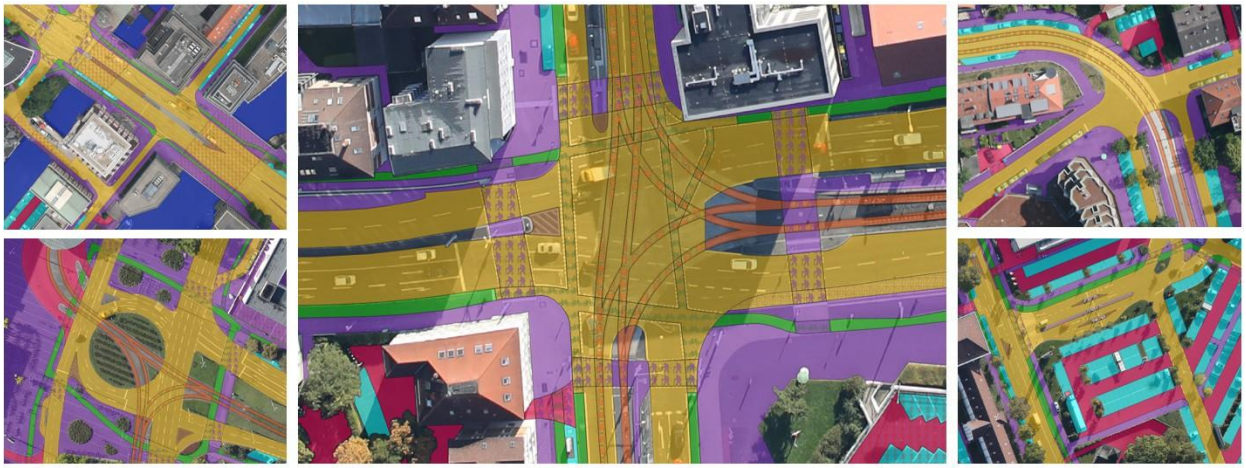


Figure 2. Example areas from the TIAS dataset with overlaid labels. Category colors are: parking area, road, access way, footway, bikeway, railroad bed, keep-out area, road shoulder, and water.

In the following sections, we will show ways of how to minimize these errors and investigate the benefits of the projected label for the quality of the predictions of DL-based segmentation methods through an experimental study. In order to do this, we will use the generated labels in combination with an existing dataset to train a neural network for the task of traffic area segmentation.

2. Methodology

In this section, we first provide an overview of our aerial image database and the Traffic Infrastructure and Surroundings (TIAS) dataset. Both form the basis of our study by providing a large amount of image data and high fidelity labels. We then discuss the generation of additional training data by projecting existing labels onto corresponding aerial images. In addition, we highlight the problems that arise during this process and offer solutions to overcome them. Finally, we present our neural network for segmenting traffic areas, with which we study the benefits of adding the images with projected labels to the training set.

2.1 Aerial Image Database

For the projection of existing labels onto new images, we rely on a unique database of 888,346 aerial images acquired between 2011 and 2024 across Central Europe. The Ground Sampling Distance (GSD) of these images range from 2 cm/pix to 30 cm/pix, covering different acquisition conditions, viewing angles, times of day and year, and camera settings. All images were captured using the 3K (Kurz et al., 2012) and 4K (Kurz et al., 2014) camera systems of the German Aerospace Center (DLR), with direct measurement of image positions and altitudes by a GNSS/inertial system. Figure 1 shows the image footprints over the city of Brunswick, overlaid with labeled aerial images.

2.2 The TIAS Dataset

The TIAS dataset (Merkle et al., 2024) is a novel dataset consisting of 57 aerial images with high-fidelity labels of traffic areas. This dataset accurately reflects urban scenarios from a transportation perspective by providing detailed, fine-grained labels. The dataset supports the reconstruction of traffic networks for motorized vehicles, bicycles, pedestrians, and rail traffic, enabling applications such as hazardous area identification (e.g., for automated vehicle and road safety analysis) and traffic area distribution analysis.

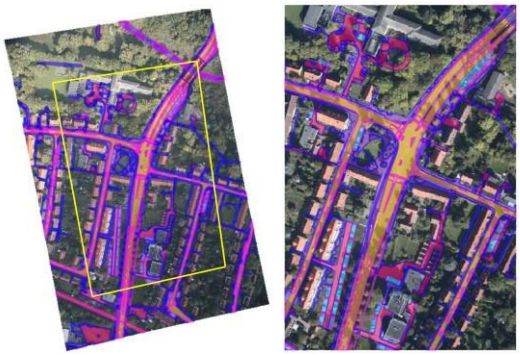


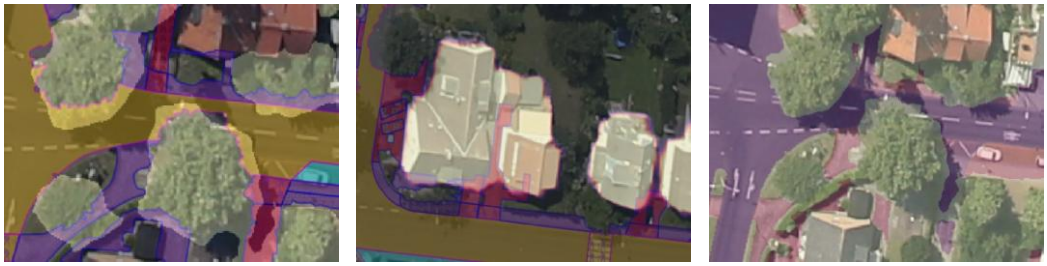
Figure 3. Left: Georeferenced labels in the source image, the yellow frame is the target image extent; Right: Labels projected onto the target image. Neither image is ortho-projected.

The images in the TIAS dataset were acquired over the German cities Berlin, Brunswick, Cologne, Garmisch-Partenkirchen, Hamburg, Landsberg, Kaufbeuren, Munich, Munster, Oldenburg, and Wolfenbuttel. Of the 57 images, 51 are from the aerial image database with GSD values ranging from 6–14 cm/pix, while the remaining 6 are ortho-projected images with a GSD of 10 cm/pix. Individual image sizes range from 17 to 22 Mpx.

The traffic areas within TIAS are classified into nine classes: parking area, road, access way, footway, bikeway, railroad bed, keep-out area, road shoulder, and water. To preserve the topological nature of the transportation network, attributes indicate whether the areas are: (1) shared by two or more traffic participants, and which they are, (2) elevated like bridges, (3) under construction, and (4) difficult to recognize for the annotator. Additionally, the attribute “unsure” provides a confidence with which an object of a given class is annotated. Figure 2 shows five sample areas with the corresponding labels of the TIAS dataset. A “background” class is assigned to all areas not included in the above such as trees and buildings, and more generally to all areas belonging to one of the 9 classes not visible in the images due to occlusion.

(a) (b) (c)

Figure 4. Visualization of different error types in the projected labels including the displacements of projected labels (a), missing parts due to different viewing angles and occlusions (b), and errors due to temporal offset (c).



(a) (b) (c)

Figure 5. Error elimination through automatic detection using tree and building segmentation methods: (a) and (d) highlight areas where labels are occluded by trees and buildings, while (f) shows the labels after automatic removal of these effects.

For all experiments performed in this paper, we only used three classes of the TIAS dataset: road, access way, and parking area. In addition, we extended the parking area class to include all areas with the attribute “shared with parking area” such as roads shared with parking area, typically used to annotate roadside parking areas lacking appropriate lane markings.

2.3 Photogrammetric Projection

The projection of the annotations is based on photogrammetric principles. The annotations are projected from a source aerial image onto a Digital Terrain Model (DTM) and projected back onto a target aerial image (see Figure 3). We use a highprecision laser DTM for the projection, in which buildings and other elevations such as trees are missing. Alternatively, we could have used a high-definition laser surface model, however, firstly, it is not available across the country and, secondly, height errors can still occur, especially around trees and buildings.
Using this method, the labels can be projected onto many overlapping aerial images from different viewing angles and scales. However, this also results in many errors, the minimization of which is described in the following sections. We would like to emphasize here that the proposed workflow also works for already ortho-projected labels. Nevertheless, for the sake of simplicity, we would like to limit the scope of the present study to projections from one non-orthoprojected image to another non-orthoprojected image. Thus, we excluded the 6 already orthoprojected labeled images from the projection process.

2.4 Projection Errors

Table 1 outlines the sources of errors that occur when projecting labels from a source image to a target image, along with the improvement methods tested in this paper. The proposed improvements are fully automated, eliminating the need for timeconsuming manual correction, and are described in the following sections. While this list covers common errors, it is not complete, and not all errors can be minimized. For example, in cases where there are large time gaps between aerial images, additional errors may occur, such as the presence of parked vehicles. Figure 4 shows examples of the errors in Table 1.

Type	*Error source*	*Improvements tested*
Geometric	Displacement due to errors in georef./DTM	Bundle adjustment, use of high-res DTM
Geometric	Errors due to occlusions in source and target images	building and tree segmentation and assignment of areas to background

Table 1. Classification of the errors identified in the projected labels and corresponding counter-measures.

A major source of error is the displacement of projected label boundaries, caused by geometric inaccuracies in both the georeferencing of the source and target images, as well as the terrain model. Since the database consists of aerial images with direct georeferencing measurements, their accuracy is often insufficient for pixel-accurate projection. To address this issue, we propose to improve the georeferencing of all aerial images using bundle block adjustment, as described in Section 2.5.

There are two additional sources of geometric error. First, areas that are visible in the target image but not in the source image, due to occlusions such as trees and buildings, or that are outside the image boundary. Second, areas that are visible in the source image but occluded in the target image. In both cases, the labels in the source image are not aligned with the labels in the target image.

Figure 6. Example of image pairs where it is difficult to find accurate tie points due to time offset, season, vehicles, shadows, viewing direction, and image scales.

In the first case, automatic correction is challenging because there are no source labels for these areas (see Figure 4(b)), so we treat them as label noise. In the second case, automatic correction is possible. In this work, we apply two deep learning methods (see Section 2.6) to detect tree and building boundaries in the target images and adjust the labels for the occluded areas accordingly, as shown in Figure 5.

To simplify further evaluation and to reduce other error sources, we reduce the number of classes and attributes to “roads”, “access ways” and “parking areas”, whereby the original class ”parking area” and the attribute ”shared with parking area” are combined for the latter.

2.5 Reducing the Geometrical Displacement

All aerial images in the database were acquired with the 3K/4k camera system with varying flight altitudes, focal lengths, camera models and installation configurations. The external orientation of the aerial images was always measured using a GNSS/inertial system, the accuracy of which, however, depends on various conditions, including the availability of correction signals.

In order to reduce the geometric displacements of all overlapping aerial images at one test site as far as possible, bundle block adjustment is necessary. A prerequisite for an automatic process is the automatic generation of tie points with e.g., SIFT or BRISK. Furthermore, during the bundle adjustment the parameters of the internal orientation must also be estimated for each camera and day of acquisition during the process.

Figure 6 shows a pair of images where it is difficult to find accurate tie points using standard methods. To solve the problem, we could in future use deep-learning based matching methods like superglue (Sarlin et al., 2019), but in this paper we have only used the conventional methods described above. We introduced the height of the terrain model as additional loose observations in the bundle adjustment, which helps to determine the focal lengths of the cameras and further increases the final accuracy.

2.6 Reducing the Errors due to Occlusions

Areas labeled in the source image and projected onto the target image may be assigned to trees and buildings because the height information is not accurate enough or is missing in terrain models. As described above, the label errors due to occlusion caused by elevated objects can be reduced e.g. by detecting trees and buildings in the target image. We use a deep learning approach to detect trees and buildings (Yuan et al., 2023) which is based on a Swin Transformer (Liu et al., 2021) trained on the ISPRS Potsdam benchmark dataset. We apply the trained model on all images and assign all detected tree and building areas in the target images to the background class (see Figure 5(c)).

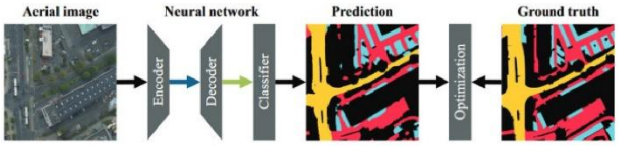


Figure 7. High-level overview of the model and training process.

2.7 Deep Learning-Based Segmentation of Traffic Areas

To extract the three traffic area classes ”roads”, ”parking areas”, and ”access ways” from aerial images, we used a Dense-U-Net121 model. This architecture is based on the established U-Net architecture as detailed in (Henry et al., 2021a) and incorporates the DenseNet-121 backbone in both the encoder and decoder, chosen for its ability to balance high accuracy with computational efficiency. Contrary to most other adaptations, it adheres closely to the original concept of U-Net, whereby the decoder is a mirror of the encoder with around as many layers and parameters. This way, it learns to extract more fine-grained details thanks to the low-resolution feature maps and the higher resolution skip-connections. Previous results showed that the resulting masks are generally smooth, homogeneous, and match object borders precisely (Henry et al., 2021b). A high-level overview of the model is provided in Figure 7.

3. Experiment

3.1 Generated Training Data

In order to generate additional training data, the footprints of the 51 non-ortho-projected labeled aerial images from the TIAS dataset are intersected with the images from our aerial image database. This resulted in 12433 aerial images, which have an overlap of at least 70% with the labeled TIAS images. To simplify matters, the number of projections was further reduced so that in the end five labeled images in two cities were projected onto 204 other overlapping images. When filtering the images, care was taken to ensure that the remaining images exhibit great variability in terms of the appearance of the individual classes, e.g. different days, positions of the sun, viewing directions.

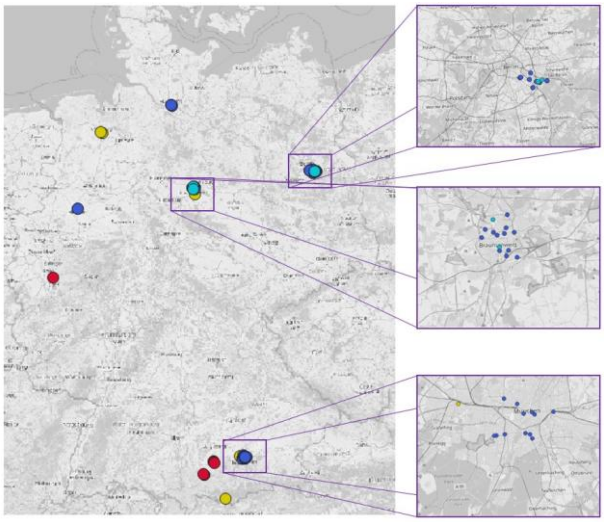


Figure 8. Positions of the images used for the ● training, ● validation and ● test sets of the deep-learning algorithms. An extra five images are added to the training set, which are used for label generation via projection, shown as ●.

Before filtering the images, a bundle block adjustment was performed for the overlapping images on the 51 TIAS positions in order to increase the georeferencing accuracy. As described above, it was not possible with standard algorithms to automatically generate tie points for some individual image pairs due to the large differences in scale, rotations and changes in viewing direction. Second, to further reduce processing time, a tree and building segmentation model was applied only on these images (see Section 2.5), where bundle adjustment was successful. The improved exterior and interior orientation was used together with the DTM to generate the 204 projected labels. Additionally, 204 segmentation maps for trees and buildings were generated with the algorithm from Section 2.6.

Experiment #	# TIAS training images	Projected labels?	Average [%] IoU	Precision	Recall	Class-wise IoU [%] Roads	Parking areas	Access ways
1	45	–	72.98	84.88	82.41	80.96	61.49	56.62
2	45	✓	72.02	85.94	80.37	80.80	59.95	54.76
3	21	–	65.76	83.84	73.96	73.56	54.39	43.86
4	21	✓	67.11	84.72	75.02	73.66	56.01	47.36

Table 2. Quantitative results comparison between the models from all four experiments on the test set of the TIAS dataset.

If we overlay the projected labels with the target image, we can qualitatively estimate the accuracy of the georeferencing. In most cases, the positional accuracy of the projected labels is accurate to within a few pixels. There are only exceptions if, for example, the scales between the source and target images are very different or the height changes due to differing road surfaces.

A prediction probability threshold of 40% was applied to the tree and building segmentation maps. The projected labels were then overlaid with the tree and building classes and were set to ”background” wherever tree or building pixels were detected.

3.2 Network Training

In order to investigate the influence of the projected labels on the accuracy of our Dense-U-Net model, we train the network on different extensions of the TIAS dataset, each exposing the model to different amounts of data and variety. Specifically, we perform the following experiments:

- Experiment #1: Baseline training on the TIAS dataset without the projected labels.
- Experiment #2: Training on the TIAS dataset with the projected labels for 5 chosen images. It constitutes the most optimistic scenarios, where a large amount of prior annotated data is available, with projected labels acting as convenient extra data.
- Experiment #3: Training on only 50% of the TIAS dataset, including all 5 images with associated projected labels. This constitutes a less optimistic training scenario, where less prior annotated data is available and the projected labels can make a significant difference.
- Experiment #4: Training on only 50% of the TIAS dataset, including all 5 images used for label projection, but without the projected labeled images.

The TIAS dataset is composed of 57 images, split into 45 training, 6 validation images, and 6 test images. While the training images span cities all across Germany, we chose validation and test images from clearly distinct regions in Germany to fairly evaluate the generalization capability of the models. The validation set therefore features the cities of Kaufbeuren, Landsberg am Lech and Cologne, and the test set features the cities of Oldenburg, Münster, Wolfsbüttel, Pasing-Obermenzing and Garmisch-Partenkirchen. In addition, images of pure background class, mostly fields and forests, taken from the area of Dortmund, Germany, are added to the training set only to reinforce the precision of the model, together with the corresponding empty label images. In total, the model sees the following amount of valid data pixels (i.e. excluding background no-data values in the projected labels), expressed in Megapixels (MP):

- Training set with Dortmund patches: 1415 MP
- Projected images alone: 2685 MP
- Training set with Dortmund patches and projected labels: 4100 MP
- Validation set: 123 MP
- Test set: 126 MP

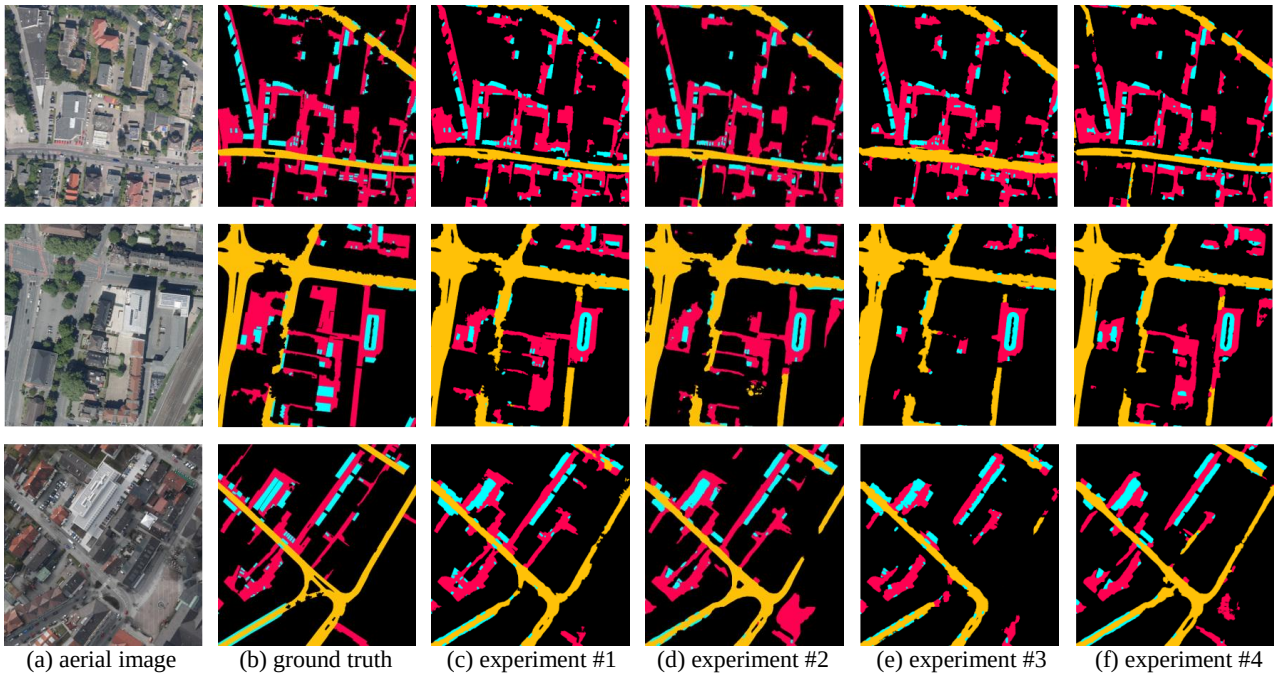
The model is trained over 50 epochs, i.e. the entirety of the data in the training set, using a cross-entropy loss, an AdamW optimizer, and a ”reduce on plateau” learning rate schedule starting at $4e-4$ and decaying by a factor of 0.90 after two epochs without any observed improvement greater than 0.10% in mean IoU score between the classes road, access way and parking area on the validation set. The images are sliced into 512×512 px non-overlapping, shuffled patches during training, and into 2528×2528 px patches during evaluation.

3.3 Results & Discussion

We report the results of each experiment on the test set in Table 2 and in Figure 9. We measured the performance in terms of mean IoU, precision and recall across all foreground classes (i.e. excluding the background class), as well as the binary IoU for each individual class. Our first set of experiments concerns the baseline model (#1) and its counterpart trained with the additional images with projected labels (#2). The baseline model actually outperformed the latter by a significant margin, around 1% mean IoU, a finding consistent across all classes, contrary to our initial expectations that it would be the other way around. This may be explained by the fact that with 45 training images, the TIAS dataset is already capable of providing good generalization capacities to other regions to a model, and that additional, noisy labels over the same training region leads not only to reduced benefits, but also to some degree of confusion. This is further confirmed by the increased precision score (+1%) at the cost of a reduced recall (-2%), showing that the noisy projected labels have caused the model to become more conservative in its predictions.

To reinforce this hypothesis, we observe the reverse trend in our second set of experiments, where one model is trained using only half of the TIAS training set (#3), while the other model has access to the projected data as well (#4). It appears that with less training data at its disposal, the third model struggles to attain results as high as the first model, and therefore benefits more from extra data, in the form of the projected labels: the fourth model indeed achieves a 1.6% increase in mean IoU, 0.9% in precision and 1.0% in recall. And whereas the performance on roads alone barely increases (0.1% IoU), the parking areas and especially the access ways see a large boost to their accuracy, with +1.7% and +2.5% IoU, respectively.

The qualitative comparison in Figure 9 confirms the quantitative results and shows that, overall, we achieve the best predictions with our baseline model #1, although model #2 achieves quite similar results. Each of these two models has advantages over the other in some areas and disadvantages in others, especially in terms of completeness. On the other hand, if we compare the results of models #3 and #4, we can see greater differences between the performance of the models. We can see a great improvement in the quality of the predictions when using the projected labels (#4). The road network is more complete and many more parking areas are also correctly classified.



(a) aerial image (b) ground truth (c) experiment #1 (d) experiment #2 (e) experiment #3 (f) experiment #4

Figure 9. Qualitative comparison between the models from all four experiments on the test set of the TIAS dataset. Category colors are: ■ parking area, ■ road, and ■ access way.

Our observations on the second set of experiments (#3-4) indicate that for smaller datasets, the use of projected labels could be an alternative way to increase the dataset size, and thus the quality of the predictions, without having to manually label more images. Furthermore, the results for our first set of experiments (#1-2) show that the TIAS dataset is suitable for training generalizable models for segmenting roads, parking areas and access ways for many urbanized regions in Germany and most likely in other areas with similarly looking cities.

4 Conclusions & Future Work

In this paper, we investigated the possibility of reusing existing manually-generated annotations in aerial images through projections onto additional aerial images from a large database. These databases, often collected during operational scenarios, represent real-world conditions but are typically underutilized due to the challenges of manual or semi-automatic labeling. By training on an existing annotated dataset, completed by projecting existing labels onto additional aerial images, we have demonstrated that our label generation approach effectively leverages large existing databases to improve the performance of segmentation algorithms. Our results show that this approach improves performance when the labeled training dataset is limited, reducing the reliance on extensive manual labeling efforts and underscoring its potential for resource-efficient dataset augmentation.

While the label projection method introduces some errors, many of these can be automatically mitigated. However, certain discrepancies, such as those caused by temporal inconsistencies such as moving vehicles, remain challenging and can be treated as label noise. These problems are particularly pronounced in sensitive classes such as parking lots, where the presence of moving objects can significantly distort the shape of the annotated regions.

In addition, the increased diversity achieved by projecting labels onto images captured under different conditions, such as different sensors, times, and environments, can improve model regularization. This diversity can help to make the trained models more robust and generalizable for real-world applications. Compared to synthetic datasets, the extended datasets generated by our approach can offer advantages by avoiding the domain gap challenges that often arise when transferring models trained on synthetic data to real-world scenarios.

References

Chiciudean, V., Florea, H., Blaga, B.-C.-Z., Beche, R., Oniga, F., Nedevschi, S., 2024. Data Augmentation for Environment Perception with Unmanned Aerial Vehicles. *IEEE Transactions on Intelligent Vehicles*, 1-15.

Henry, C., Fraundorfer, F., Vig, E., 2021a. Aerial road segmentation in the presence of topological label noise. *Proceeding of the International Conference on Pattern Recognition (ICPR)*.

Henry, C., Hellekes, J., Merkle, N., Azimi, S., Kurz, F., 2021b. Citywide estimation of parking space using aerial imagery and osm data fusion with deep learning and fine-grained annotation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII, 479– 485.

Kurz, F., Rosenbaum, D., Meynberg, O., Mattyus, G., Reinartz, P., 2014. Performance of a real-time sensor and processing system on a helicopter. C. Toth, T. Holm, B. Jutzi (eds), *ISPRS Archives*, ISPRS Technical Commission I Symposium, XL-1, ISPRS Archive, 189–193.

Kurz, F., Turmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P., Leitloff, J., 2012. Low-cost optical Camera System for real-time Mapping Applications. *Photogrammetrie Fernerkundung Geoinformation*, Jahrgang, (2), 159–176.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992–10002.

Merkle, N., Rauch, F., Henry, C., Hellekes, J., Kurz, F., 2024. TIAS: An aerial traffic infrastructure dataset to study transportation in urban environments. *GeoDPA - International Conference on Geoinformation Data, Processing and Applications*.

Sarlin, P., DeTone, D., Malisiewicz, T., Rabinovich, A., 2019. SuperGlue: Learning Feature Matching with Graph Neural Networks. *CoRR*, abs/1911.11763. http://arxiv.org/abs/1911.11763.

Tian, J., Zhuo, X., Auer, S., Kurz, F., Reinartz, P., 2023. Fusion of stereo aerial images and official surveying data for mapping curbstones using AI. *8th International Conference on Signal and Image Processing, ICSIP 2023*, IEEE, 1–5.

Toker, A., Eisenberger, M., Cremers, D., Leal-Taixe, L., 2024.´ Satsynth: Augmenting image-mask pairs through diffusion models for aerial semantic segmentation. *Proceedings - 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024*, 27685–27695.

Yuan, X., Tian, J., Krauß, T., Zhuo, X., Reinartz, P., 2023. Multi-layer thematic map representation for urban understanding. *2023 Joint Urban Remote Sensing Event, JURSE 2023*, 2023 Joint Urban Remote Sensing Event (JURSE), Institute of Electrical and Electronics Engineers, 1–4.

Zlateski, A., Jaroensri, R., Sharma, P., Durand, F., 2018. On the Importance of Label Quality for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, 1479–1487.