Influence of Augmentation of UAV Collected Data on Deep Learning Based Facade Segmentation Task

Patryk Balak
SkySnap sp. z o.o.
Warsaw, Poland *patryk.balak@skysnap.pl*

Agnieszka Łysak
Jan Kochanowski University; Institute of Physics; SkySnap sp. z o.o.
Kielce; Warsaw, Poland *agnieszka.lysak@ujk.edu.pl*

Kamil Choromański
SkySnap sp. z o.o.; Warsaw University of Technology; Faculty of Geodesy and Cartography
Warsaw, Poland *kamil.choromanski@skysnap.pl*

Marcin Luckner
Warsaw University of Technology; Faculty of Mathematics and Information Science
Warsaw, Poland *marcin.luckner@pw.edu.pl*

Abstract

Data augmentation is crucial for image segmentation, especially in transfer learning with limited data, however it can be costly. This study examines the cost-benefit of augmentation in facade segmentation using unmanned aerial vehicles (UAV) data. We analysed how dataset size and offline augmentation impact classification quality and computation using DeepLabV3+ architecture. Expanding the dataset from 5 to 480 thousand tiles improved segmentation efficiency by an average of 3.7%. Beyond a certain point, further dataset expansion yields minimal gains, in our case, just 1%, on average, after segmentation accuracy plateaued at around 76%. These findings help avoid the computational and time costs of ineffective data augmentation.

Keywords: unmanned aerial vehicles, computer vision, segmentation, augmentation.

1. Introduction

Data augmentation enhances deep learning (DL) models' generalisation while reducing data collection and labelling efforts, using various image modifications. This technique is crucial in architecture, engineering, and construction (AEC), where DL supports labor-intensive tasks like data collection, inspection, and management.

Augmentation addresses architectural variability and labelled data scarcity. While DL models require large datasets, AEC segmentation datasets remain small due to costly, specialized unmanned aerial vehicle (UAV) data acquisition. Despite its benefits across industries, a research gap exists in the AEC regarding cost-effective augmentation methods that improve the facade inspection process. Augmentation in the facade segmentation process presented in [6] is only a part of a broader segmentation approach. Our study fills the research gap by analysing how augmentation affects model efficiency and training costs in UAV-based facade segmentation. We showed that beyond a certain dataset size, returns diminish as model performance plateaus, providing strategic insight for resource allocation.

We explored limitations of public facade datasets in size, diversity, and content. Our medium-sized dataset (475-2,850 images, depending on the dataset variant) was collected for architectural variety. We tested key augmentation techniques, comparing models trained on augmented versus original data, to assess their economic impact.

2. Related Works

Various data augmentation strategies for video action recognition, presented in [8], demonstrated that augmentation can improve the model's performance. Utilising geometric transformations and colour variations to expand the dataset derived from UAV photogrammetry enhanced the DL model's ability to recognise windows under varying conditions and views [3]. The impact of augmentation on wall and roof damage classification using UAV images, focusing on vegetation on facades and roof dirt, was discussed in [1]. The tailored geometric transformations significantly improved classification accuracy.

While our research addresses similar data limitation challenges, we focus on more detailed facade semantic segmentation and employ more diverse augmentation techniques beyond geometric transforms to provide wider feature variability. The under-explored issue of augmentation cost is particularly relevant in today's context of resource-intensive computation versus expensive data collection. Our study addresses this gap by identifying efficient augmentation strategies that balance model performance with computational resource usage. It is worth noting our methodology's limitations—we use offline methods without adapting to class distribution.

3. Materials and Methods

3.1. Augmented Dataset Creation

The dataset comprised RGB photographs of facades from 26 buildings from four Polish cities, collected during UAV missions. Mission parameters were optimized to capture facade details, with 80% front and 60% side coverage. Drones flew vertically along facades at distances of 5-10 meters. All images were captured in daylight with good weather.

The dataset was split into six versions, with growing training and validation sets, keeping the test set constant. Selection maximized diversity in camera models, resolutions, architectural styles, and building functions. Images came from three drones and five cameras across four different resolutions. Three resolutions (5472×3648, 4000×3000, 8192×5460) were distributed across training, validation, and test sets, while the fourth resolution (5280×3956) appeared only in training and test sets. Of the 26 buildings, 19 were used for training, 3 for validation, and 4 for testing. Training data included all drone types, camera models, building styles, and functions, while validation and test datasets had more limited diversity.

The dataset was divided into six versions (V_1–V_6) to evaluate the impact of augmentation. Each subsequent version increased the number of training images per building, while 15% of images were reserved for validation. The test set remained fixed across all experiments, containing images from four buildings to ensure consistent performance comparison. This setup ensured varying training sizes while controlling evaluation conditions. A summary of the dataset versions is shown in Table 1.

We identified five classes: brick, metal, windows, plaster, and roofing. Distribution was uneven, with plaster dominating, followed by windows, especially in modern buildings. Metal, brick, and roofing were less common. Metal elements were rather small, while brick appeared mostly in historical buildings. Despite covering large areas, roofing was often hard to capture due to UAV angle and distance.

We applied a tailored data augmentation approach for UAV-based facade segmentation, guided by the characteristics of aerial imagery and facade analysis needs. Our research used the following augmentation methods: horizontal flip, random brightness and contrast shift, random shadow, and Gaussian noise. The augmentation techniques used in the research are shown in Fig. 1.

Table 1. Number of images and tiles in each data set.

	Training		Validation		Testing	
	Images	**Tiles**	**Images**	**Tiles**	**Images**	**Tiles**
V_1	19	5,086	3	767	180	50,613
V_2	95	25,366	15	4,018	180	50,613
V_3	190	53,425	30	7,909	180	50,613
V_4	285	80,594	45	12,618	180	50,613
V_5	380	103,579	60	17,459	180	50,613
V_6	475	133,675	75	21,936	180	50,613

Fig. 1. Augmentation techniques.

3.2. Comparison with other datasets

In the field of facade semantic segmentation utilising UAV-data, our dataset stands out for its high resolutions, varying from 4000×3000 to 8192×5460 pixels, and their corresponding tile count. In contrast to typical datasets in this field, characterised by lower-resolution photographs and smaller tile counts, our dataset features larger tiles: 512×512 pixels. This allows for a more global and contextual analysis. Our dataset is rich in classes, facilitating detailed facade segmentation. Regarding the image count, our dataset is relatively modest, consisting of 475 photos before and 2,850 after augmentation.

For the comparison, in [7], 1,035 ground-view images across five classes, including building, window, and door classes, with the addition of vegetation and ground were employed. In post-augmentation, the dataset grew to 3,924 tiles, each 300×300 pixels in size. However, the exact size of the test set was not specified. It was estimated to consist of approximately 2,757 images. Also, the exact resolution of the dataset was not specified. In contrast, when expanded to its largest version, our study's dataset had 133,675 tiles for training, 21,936 for validation, and 50,613 for testing. This highlights significant differences in both augmentation technique and methodology.

The dataset used in [1] consisted of 1458 photos from two distinct cameras (at 2464×1632 and 3840×2160 resolutions). Augmentation was not employed. Compared to our dataset, their dataset is more extensive regarding the number of photographs; however, it is characterised by a lower resolution than ours. Also, the datasets presented in [2, 3, 4, 5] are characterised lower resolution and smaller patch sizes than in our case.

4. Results

In our experiments, the DeepLabV3+ model was trained twelve times, six times with the datasets V_1 - V_6 without any augmentation applied, and six times with the augmented datasets V_1 aug - V_6 aug. Augmentation significantly increases training time. For version V_1, training without augmentation takes only 1.55 $[h]$, while with augmentation it rises to 15.32 $[h]$. This trend continues across all versions. Version V_2 requires 9.17 $[h]$ without augmentation and 38.51 $[h]$

with it. For V_3, the times increase to 14.84 and 44.44 $[h]$, respectively. The most substantial durations are observed for V_4, V_5, and V_6, with augmented training times reaching 72.06, 76.21, and 118.04 $[h]$, respectively. Those results underscores the computational demands associated with different dataset configurations and compares the training time for data with and without augmentation. Furthermore, the F1 score was evaluated across all classes. Given the class imbalance and semantic distinctions between the classes associated with different facade materials, assessing the model's segmentation efficacy for each class was crucial. Fig. 2 illustrates the model's varying performance across classes.

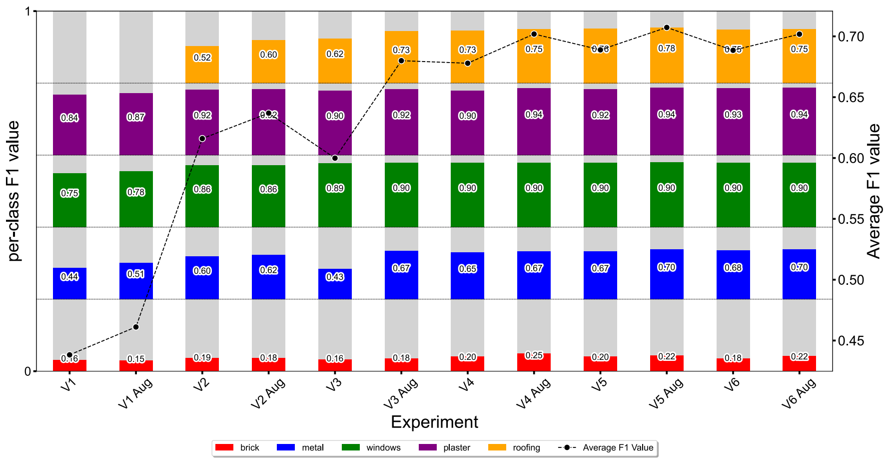

Fig. 2. F1 scores across classes

Analysing F1 scores revealed that data augmentation improves segmentation, especially in early dataset versions. Performance gains were most notable in versions V_1–V_3, where augmentation enhanced segmentation across all classes, particularly for challenging elements like metal. However, from V_4 onward, improvements plateaued, with only minor differences in plaster and background segmentation.

Metric gains (IOU, ACC, F1) dropped from 0.057 / 0.022 / 0.041 in V_1–V_3 to 0.029 / 0.015 / 0.018 in V_4–V_6, despite training time nearly tripling (from 24.24 to 63.10 hours). From V_4 to V_6, an extra 189.31 hours yielded minimal quality improvement. These results illustrate the diminishing returns of augmentation beyond a certain dataset size and support more efficient resource allocation.

4.1. Statistical analysis

To analyse the statistical importance of improvements by applying augmentation, for each quality measure, a vector of the obtained results was created for each recognised class, and each data set ($V_1 - V_6$). A similar vector was created for each augmented data set (V_1 aug $- V_6$ aug).

The vectors were compared separately for each dataset version using the one-tailed Wilcoxon paired signed-rank test – with a significance level of $\alpha = 0.05$ – to evaluate if the results obtained after augmentation were higher. Next, each vector was extended by adding the mean value of the measure calculated across all classes. This step stressed the importance of global change across all classes, not only a change in a single class. Finally, a vector containing all measures was created to perform the tests with the significance level $\alpha = 0.01$.

A statistically important improvement of all three measures by data augmentation was observed only for dataset V_5 ($p \leq 0.03$). However, for the vector containing all measures, improvement is observable for V_1 (but the reference level for the improvement is relatively low) and V_3–V_5 ($p \leq 0.01$). This observation implies that extending the most extensive dataset, V_6, is not rational from a statistical point of view.

5. Conclusions

Our research offers key insights into data collection and augmentation strategies in cost-intensive AEC sector. We identified a threshold of around 480,000 tiles where data augmentation's benefits plateaus. Below this point, computation hours yielded a 3.7% efficiency improvement; above it, 90.87 hours led to just a 1% gain, highlighting the limited benefit of expanding datasets past critical thresholds. These findings suggest that acquiring new data may be more effective than relying exclusively on augmentation.

Despite its contributions, this study has limitations. It focuses on offline augmentation, future work could explore adaptive, lightweight online techniques that are more efficient and context-aware. While DeepLabV3+ is widely used in AEC, newer models may offer better performance. Additionally, our dataset is also limited in size and diversity. Expanding it to include more building types, materials, and varied environmental conditions could yield deeper insights and potentially different results.

Acknowledgments

This research was made by the use of the UAV facade data, developed and provided by the Research and Development project. **INKUB - Inventory of technical conditions of cubature real estates based on UAV and AI POIR.01.01.01-00-0895/21** at **SkySnap sp. z o.o**.

References

[1] Cao, M.: Drone-assisted segmentation of tile peeling on building façades using a deep learning model. Journal of Building Engineering 80, pp. 108063 (2023)

[2] Chen, K., Reichard, G., Xu, X., Akanmu, A.: Automated crack segmentation in close-range building façade inspection images using deep learning techniques. Journal of Building Engineering 43, pp. 102913 (2021)

[3] Jiang, Y., Han, S., Bai, Y.: Scan4façade: Automated as-is façade modeling of historic high-rise buildings using drones and ai. Journal of Architectural Engineering 28(4), pp. 04022031 (2022)

[4] Lee, K., Lee, S., Kim, H.Y.: Bounding-box object augmentation with random transformations for automated defect detection in residential building façades. Automation in Construction 135, pp. 104138 (2022)

[5] Ottoni, A.C., de Amorim, R.M., Novo, M.S., Costa, D.B.: Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. International Journal of Machine Learning and Cybernetics 14(1), pp. 171–186 (2023)

[6] Zhang, L., Wang, G., Sun, W.: Automatic identification of building structure types using unmanned aerial vehicle oblique images and deep learning considering facade prior knowledge. International Journal of Digital Earth 16(1), pp. 3348–3367 (2023)

[7] Zhuo, X., Mönks, M., Esch, T., Reinartz, P.: Facade segmentation from oblique uav imagery. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) 2019 Joint Urban Remote Sensing Event (JURSE). pp. 1–4 (2019)

[8] Zou, Y., Choi, J., Wang, Q., Huang, J.B.: Learning representational invariances for data-efficient action recognition. Computer Vision and Image Understanding 227(November 2022), pp. 103597 (2023)