Buildings along Monge Street in Paris.

Leveraging deep learning segmentation techniques and connected component analysis to automate high-level cost estimates of facade retrofits using 2D images

María Escalada

Higher Technical School of Architecture of San Sebastián, Spain.
Email: mescaladamg@gmail.com

Abstract: Deep learning semantic segmentation techniques applied to 2D facade images hold a great promise in several domains that go far beyond model generation, mainly if the data used are front-parallel or orthonormal photographs. However, effective applications in the field of built heritage have not been adequately explored, largely due to the absence of multidisciplinary teams that include architecture professionals as early as the dataset creation stage. The aim of this research is to introduce a holistic view in order to demonstrate the practical usefulness of state-of-the-art segmentation models to automate high-level cost estimates of urban-scale residential building facade rehabilitations when combined with a connected component analysis. To achieve this, a scalable bottom-up approach is formulated in five simple phases, encompassing both data science and architecture expertise. This strategy seeks to improve the accuracy of analyses at early stages when limited information on constructions is available and there is a significant cost uncertainty, and therefore to optimise the strategies used by construction stakeholders involved in economic feasibility studies and decision-making processes.

Keywords: Deep learning; semantic segmentation; connected component analysis; high-level cost estimates; facade rehabilitations.

Received: 14/09/2024; **Accepted:** 03/12/2024; **Available online:** 13/12/2024; **Published:** 31/12/2024.

Cite as: Escalada, M. (2024). 'Leveraging deep learning segmentation techniques and connected component analysis to automate high-level cost estimates of facade retrofits using 2D images'. *VITRUVIO - International Journal of Architectural Technology and Sustainability*, *9*(2), 22421. https://doi.org/10.4995/vitruvio-ijats.2024.22421

1. Introduction

For the past decades, 2D facade parsing formulated as an image segmentation problem has been an important and challenging task in computer vision where the aim is to decompose an urban photograph into meaningful parts, after assigning a category label to each of the pixels. However, as Shapiro (1992: 1743) hints, it is not a trivial exercise but a process that must be guided by specific goals so as to be relevant in terms of application.

Thus, to grasp the untapped potential of segmentation techniques applied to facade images analysis and to propose useful implementations, it is essential to have a thorough understanding of both algorithms and datasets from the data science and the application domain point of view.

2. Literature review

2.1 Datasets

In architecture, publicly available benchmark databases of 2D exterior images include three main types of scenes, each associated with a distinct field of study:

- Aerial urban views, usually coming from large-scale orthophotographs collected by airborne sensors.
- Street scenes in foreshortened views, where buildings, outlined by their silhouette, are just another component of the complex urban environment.
- Street facades, which are broken down into the main constructive elements that contribute to their overall composition.

This last category includes both rectified and unrectified images, but as front-parallel views can be treated as simplified elevations, they are of greater utility for experimentation. In this regard, ECP (Teboul, 2011), ICG Graz50 (Riemenschneider et al., 2012), CMP (Tylecek, 2013) and ENPC2014 Paris Art-deco Dataset (Gadde et al., 2016) can be noted as examples of benchmark datasets (Figure 1). Both ECP and ENPC2014 are prepared to recognize and segment seven classes (door, shop, balcony, window, wall, sky and roof), Graz50 just four (door, window, wall and sky) and CMP eleven (facade, molding, cornice, pillar, window, door, sill, blind, balcony, shop, deco and background).

Databases are commonly linked to the early stages of 3D city modeling or virtual reconstructions as reported in the work of Teboul (2011), Riemenschneider et al. (2012), Musialski et al. (2013), Martinovic and Van Gool (2013), Mathias et al. (2016) and Varcity ETHZ (2017) project. Nevertheless, a mere glance is enough to reveal that the limited number of classes and the improper use of semantics fail to capture the full range of facade complexities, evidencing a shallow understanding of the selected architecture. This undoubtedly limits data usefulness in virtual urban modeling and other domains that remain unexplored.

2.2 Segmentation techniques

In order to tackle segmentation in the three above-mentioned scenarios, a range of strategies has been applied over time, often combined. Awareness of the full range of procedures is essential, but due to their abundance, this paper will focus on highlighting only a representative subset.

On the one hand, traditional methods that rely on manual feature extraction and rule-based systems (Zhuo et al., 2023) and machine learning approaches where classifiers are trained on labeled data to extract building features are common.

Many researchers explore the potentials of grammars for image-based urban 3D reconstruction using sets of rules that are mostly hand-crafted. Attention should be drawn to the early work of Ripperda and Brenner (2006), Müller et al. (2007) and Simon et al. (2011). Teboul (2011) for his part performs single-view image-based procedural modeling for haussmannian buildings, using Reinforcement Learning; Riemenschneider et al. (2012) show an example of a parse tree transformed into 3D model; Martinovic and Van Gool (2013) use a technique that learns 2D Attributed Stochastic Context-Free Grammars from labeled images in order to automate the creation of new instances of the same building styles and end up showing examples of new 3D building designs generated from a learned shape grammar after rendering them in CityEngine.

Different strategies within the broader framework of random forests are employed by Fröhlich et al. (2010) and Rahmani et al. (2017). Berg et al. (2007) present a process in stages, guided by a Conditional Random Field (CRF). Čech and Radim (2009) and Koziński et al. (2015) use Markov Random Field (MRF) strategies.

Mathias et al. (2016) use a combination of several classifiers in three layers including segmentation and classification techniques, a CRF framework, architectural principles for facade parsing and procedural modeling following CityEngine CGA rules. In the final step, they talk about the possibility of using the output for 3D reconstructions.

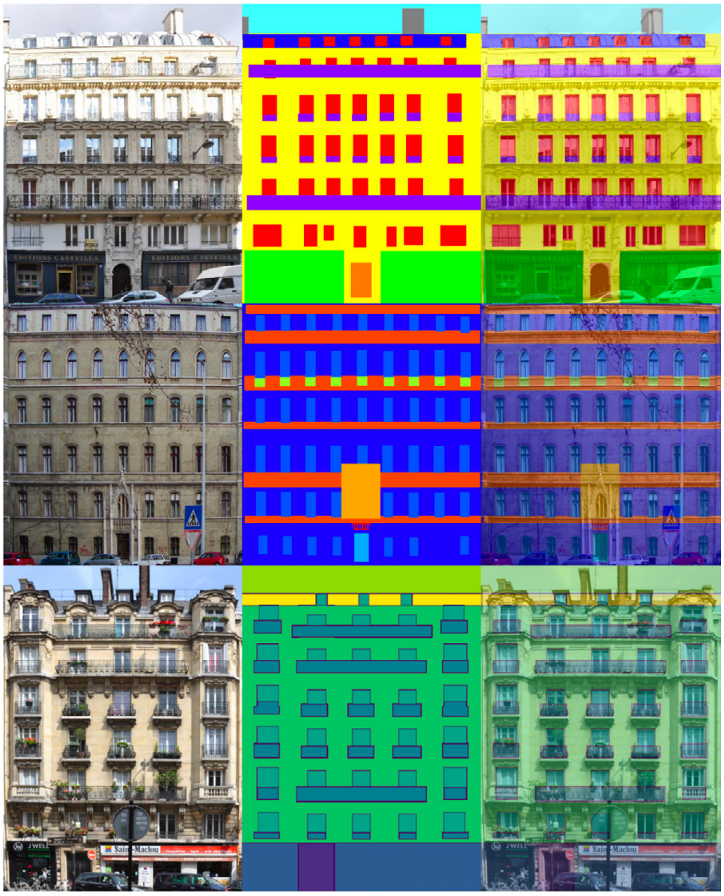

Figure 1 | Examples of ECP (row 1), CMP (row 2) and ENPC2014 Paris Art-deco datasets (row 3).

On the other hand, Deep Learning based techniques can be found, which have outperformed traditional vision approaches (Liu, et al. 2017). Hierarchical attributes are captured directly from the raw image data without needing explicit feature engineering.

In this context, Schmitz and Mayer (2016) present the first fully-convolutional approach for semantic interpretation of facade images based on a convolutional neural network inspired by Alexnet, in which they use parts of already-trained networks and fine-tune them.

Liu et al. (2017) develop a method called DeepFacade where they train deep convolutional neural networks with the constraints under man-made rules including a new symmetric regularizer applied to window, door and balcony classes. Their loss function penalizes segmentation regions that are not horizontally and vertically symmetric. The bounding boxes generated by object detection help locate and refine the shape of the predicted regions. At a later stage, they also apply region refinement.

Iglovikov and Shvets (2018) create TernausNet, a network that inherits classical U-Net architecture using VGG11 as the encoder, with 3 different ways of weight initialization. They defend that pre-trained networks substantially reduce training time that also helps to prevent overfitting.

Kelly et al. (2017) focus on the problem of procedurally creating structured models by leveraging data from multiple sources. They use segmentation for many purposes. On the one hand, to recognize those parts of the images that are unlikely or likely to have facade features using a Bayesian SEGNET CNN. On the other hand, to identify facade features and its boundaries. As for Femiani et al. (2018), they create four different refined models based on the SEGNET algorithm.

Pantoja et al. (2020) face the facade segmentation problem from two different approaches. In the first one, the output of the model is a facade mask. In the second one, the model detects the corners of the building facade and connects them to constitute a polygon. Both of the procedures take into account Iglovikov and Shvets (2018) work.

Dai et al. (2021) present FacMagNet for semantic segmentation of building facade components using a symmetric structure, dilated convolution and Faster R-CNN to handle class imbalance. Last but not least, they also include an ensemble learning strategy that integrates two models derived from the U-Net architecture adopting Inception ResNet-V2.

In recent years, a series of techniques such as the attention module and the vision transformer (ViTs) have been proposed to complement the CNNs. An example of it is the work of Zhang et al. (2022). They design a hieratical deep learning framework called DAN-PSPNet-Lsym where they include two types of attention modules (spatial and channel) and a novel loss function to integrate prior knowledge in order to force the algorithm to detect facade elements with a highly proportionate shape. This last idea was inspired by the work of (Liu et al., 2017).

2.3 General gap identification

It can be easily appreciated that research on segmentation techniques applied to facade images overemphasizes technical improvements that are highly specialized for architects and hard to follow because the field is self-contained. This issue can be explained by the absence of multidisciplinary teams that could reflect on the societal impact of the work and is evidenced by the fact that publications frequently proudly conclude by presenting metrics that demonstrate how state-of-the-art performance is surpassed. Consequently, the practical utility of these computer vision approaches in the field of built heritage is either disregarded, timidly reported (lacking clear demonstrations) or too narrowly focused on improving 3D model creation processes.

Therefore, efforts should be channeled into reflection on how to reverse the current situation and into showcasing the relevance of this technologies for building conservation purposes.

3. Research aim

The aim of this research is to demonstrate the practical usefulness of state-of-the-art segmentation models when combined with a connected component analysis, to automate high-level cost estimates of urban-scale residential building facade interventions.

The work developed is intended to be a modest step towards the creation of advanced tools that can genuinely assist urban stakeholders in their decision-making processes for an optimized allocation of resources in the case of retrofit projects (Figure 2). Owing to the limitations in human, material, and economic resources, as well as those inherent to the data used, the key contribution of this article is not the end product or result itself, but rather the reflective process and the bottom-up approach that encompasses both data science and architecture expertise.

4. Methodological approach

In architecture, facade retrofit projects go through a long lifecycle where cost calculation is not done in the same way. Before conceptual and detailed design projects start, there is a high level of ambiguity and uncertainty with respect to the expected budget because limited technical information is available, so often a rough order of magnitude is searched.

However, deep learning segmentation techniques can be of great help. Instead of making estimates based on broad simplifications such as cost per square meter (founded on reference projects), rectified facade photo analysis can be used to suggest a reasonable cost value.

Automated 2D parsing of its elements can replace manual measurement work, which is a labor-intensive task and it is usually not done at this early stage due to the lack of plans. Therefore, the implementation of AI techniques appears to be well-suited to perform urban-scale estimations.

Although the methodology is scalable and can be adapted for any segmentation algorithm, personal dataset or intervention scenario depending on the resources available (see future research possibilities of Figure 3), DeepLabv3+ trained on ECP benchmark database will be implemented due to its widespread use in state-of-the-art publications.

4.1 Phase 1: Dataset understanding from an architectural and computer vision point of view

In this case, the initial step involves reviewing an existing dataset to assess the opportunities it provides for experimentation within facade intervention projects. Analyzing the selected architecture and the annotations will provide insight into the tool's capabilities, specifically regarding the types of image information retrieval that could be performed.

4.1.1 Images and masks

As it was mentioned before, the ECP database was created by Teboul (2011). The benchmark from 2011 contains 104 manually rectified photographs of Monge street

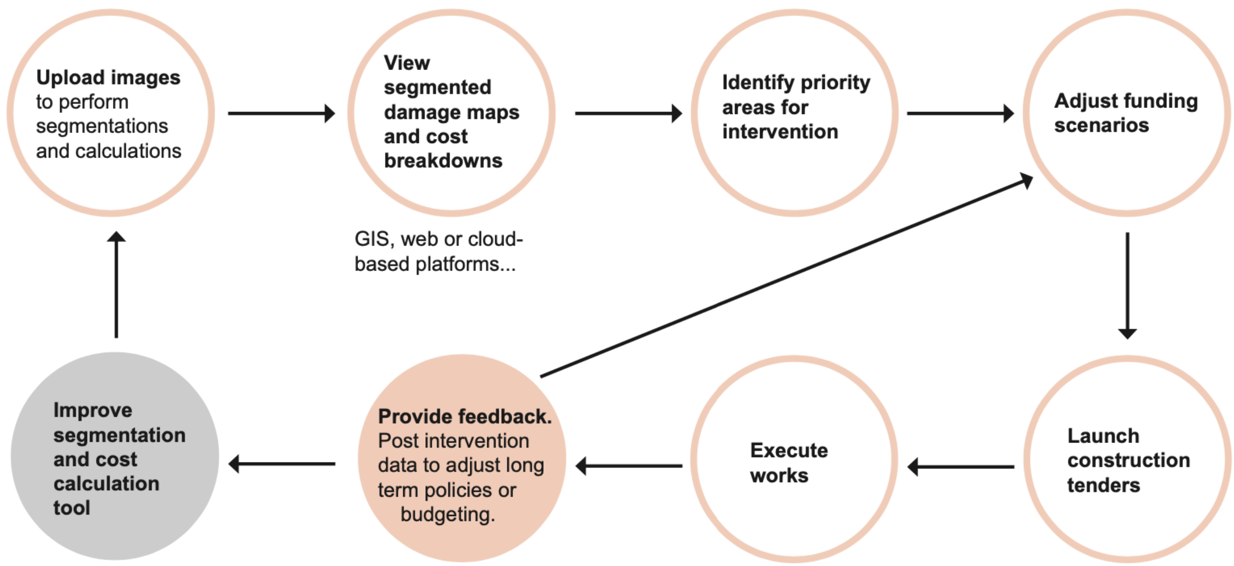

Figure 2 | A complex tool for Decision-Making.

buildings and their corresponding masks, where 7 classes are defined. For this experiment, the images were obtained from Jampani et al. (2015) and the improved annotations presented by Martinovic et al. (2012) were used. The provided multiclass semantic segmentation masks (Martinovic, n.d.) are in a RGB format, so in order to operate easily with regions, those values are transformed into integers: (0, 0, 255): 1 (Roof), (0, 255, 0): 2 (Shop), (128, 0, 255): 3 (Balcony), (128, 128, 128): 4 (Chimney), (128, 255, 255): 5 (Sky), (255, 0, 0): 6 (Window), (255, 128, 0): 7 (Door) and (255, 255, 0): 8 (Wall).

4.1.2 Construction period

Monge street is located in the Jardin des Plantes and Saint-Victor districts of Paris and it was opened by Decree on July 30 (Paris, n.d.). Although construction dates are not specified in any data science publication, it can be inferred that Teboul views the selected buildings as post-1860 and in compliance with Haussmann's regulations. In order to promptly verify this question without resorting to archival work, inventories based on topographical files and official records have been consulted (Dugast et al., 1990). The list of heritage protections of the 5th district (Mairie de Paris, 2023) and the inscriptions on the buildings have also been examined. As a result of all this brief analysis, the following data have been obtained: No. 3 (1865), No. 12 (1867-1868), No. 21 (1885), No. 56 (1878), No. 60 (1893), No. 63 (1872-1873), No. 68 (1882), No. 84 (1883-1888), No. 102 (1886), No. 111 (1878), No. 113 (1875), No. 115 (1878-1879), No. 117 (1880) and No. 119 (1878-1979).

It is therefore reductive to categorize all buildings from this ECP dataset as Haussmannian, since many were also raised in the later period, which theoretically started once the Prefect dimissioned in 1870. The above means that several regulations must be taken into account in order to deepen the insight into the selected facades.

4.1.3 Building individualization

The urban fabric in which Monge street is located consists of residential blocks (in many cases triangular) arranged in continuous alignment, featuring buildings with shared party walls and chamfered corners. When talking specifically about the dataset, parcel division is intended to coincide with the individualization of the built volume in the image. The vertical cuts are made without much precision near party walls and always end before the change of plane of the chamfer. Therefore, these issues must be taken into account when making any measurement or scaling the architectural elements.

4.1.4 Regulations

For the period under review, permitted building profiles were defined in three regulations: the Decree of July 27, 1859 (Laisney and Koltirine, 1988) during the 3rd Empire; and those of July 22, 1882 and July 23, 1884 during the Third Republic.

The first one set the maximum height of facades at 20 meters for streets over 20 m, with the proviso that in no case should more than 5 floors be built above the ground floor. Roofs had to be pitched at 45º and their height could not exceed half the depth of the building (around 25 m). For streets between 9.75 m and 20 m, the height was 22.40 m to 22.50 m. Emerging elements from the facade could have a maximum dimension of 10 cm, small balconies 22 cm and large ones 80 cm (Laisney et Koltirine, 1988: 77). The placement of continuous balconies was also standardized to align horizontally with neighboring buildings in the second and fifth floors.

For its part, the Decree of July 23, 1884 set the height at 28.50 m (20 m facade + 8.50 m attic space within an arc of a circle with a radius of 8.50 m). The dimensions of small balconies and large ones do not vary, except in the case that the latter could be situated at a height of less than 4 meters, being then limited to a depth of 50 cm. The cornices could have a 10 cm overhang if they were below 2.60 m and 50 cm if they were above (Laisney et Koltirine, 1988: 78).

Building peculiarities can therefore be mainly appreciated in the ornamental details and in the materiality of the facade rather than in the silhouette, as the result is normally as follows: Ground Floor + Intermediate floor + 3 + Attique (it can be set back or flush with the facade) + Mansard Roof with dormer windows.

Had all the aforementioned features been reflected in the masks, systematizing the calculations would have been straightforward, but this is not the case. To illustrate the above, calculating the total length of the railings of the balconies with the highest overhang would simply involve adding 1.60 m (0.80 cm x 2) meters to the true dimension estimate provided by the segmentation model.

4.1.5 Conclusions

One must be aware that Martinovic's masks present clear limitations for the stated purpose. There is an inconsistent region categorization and delimitation due to:

- Poorly rectified orthophotos that cause misalignment of building elements (windows and balconies). The primary structure of the facades, which is in fact a network of voids governed by level and plumb underlined by the enframements, the encompassing orders and the imposts lines, cannot always be respected due to the distortions.
- The inadequate choice of annotation shapes (rectangular instead of polygonal or freeform) that cannot be precisely adapted to the geometry of the constructive elements.
- Broad simplifications in the use of semantics and geometries, not encompassing the complete spectrum of facade complexities. On the one hand, no finishing is discernible no matter what constructive element is selected but the facade panel is the most obvious case. The architecture of ashlar and render facades is part of Paris' historic heritage for residential buildings and examples of both cases using a traditional load-bearing masonry system can be found in Monge street (eg. No. 1 uses large blocks of limestone, No. 28 and 29 have parts where a coating simulates stone by representing the joints, No. 8 has a smooth continuous render finish on upper floors…). On the other hand, isolated, linear or superficial ornamentation is ignored.

To sum up, the ECP dataset provides a useful framework for experimenting with the extraction of geometric properties of various constructive elements (quantitative and shape analysis). However, it is hindered by a significant limitation: masks lack information regarding the state of conservation of the facade and the stratigraphy of the walls, which will restrict the segmentation tool's potential. This highlights the future need to create new appropriate datasets.

4.2 Phase 2: Facade intervention goal definition, including budget lines (task description and unit prices)

The initial decision involves determining the type of project to be undertaken (restoration, energetic refurbishment…), followed by defining the specific tasks to be executed, from both a descriptive and economic standpoint.

In this sense, ECP dataset masks do not offer many options for consistent experimentation, but given that the primary aim is solely to demonstrate the applicability and effectiveness of computer vision techniques in the automated measurement extraction process, three budget items that could be congruent with a restoration of Monge street facades are listed as an example. For their correct drafting, cost estimation software and historical data should be taken into consideration and local market research should be conducted.

- Windows: Restoration of wooden balcony door by means of general consolidation, including the replacement of deteriorated elements, the covering of cracks and holes with epoxy resin and the removal of debris. Cost: 54,39 €/m^2
- Wall: Cleaning treatment of contaminants with water, ethyl alcohol and ammonia applied with soft bristle brushes. Cost: 17,45 €/m^2
- Balcony railings: Restoration of wrought iron metal balcony including mechanical repairs, review and change of deteriorated decorations of the railing, general cleaning, removal of oxides with metal brushes and painting. Cost: 99,93 €/m^2

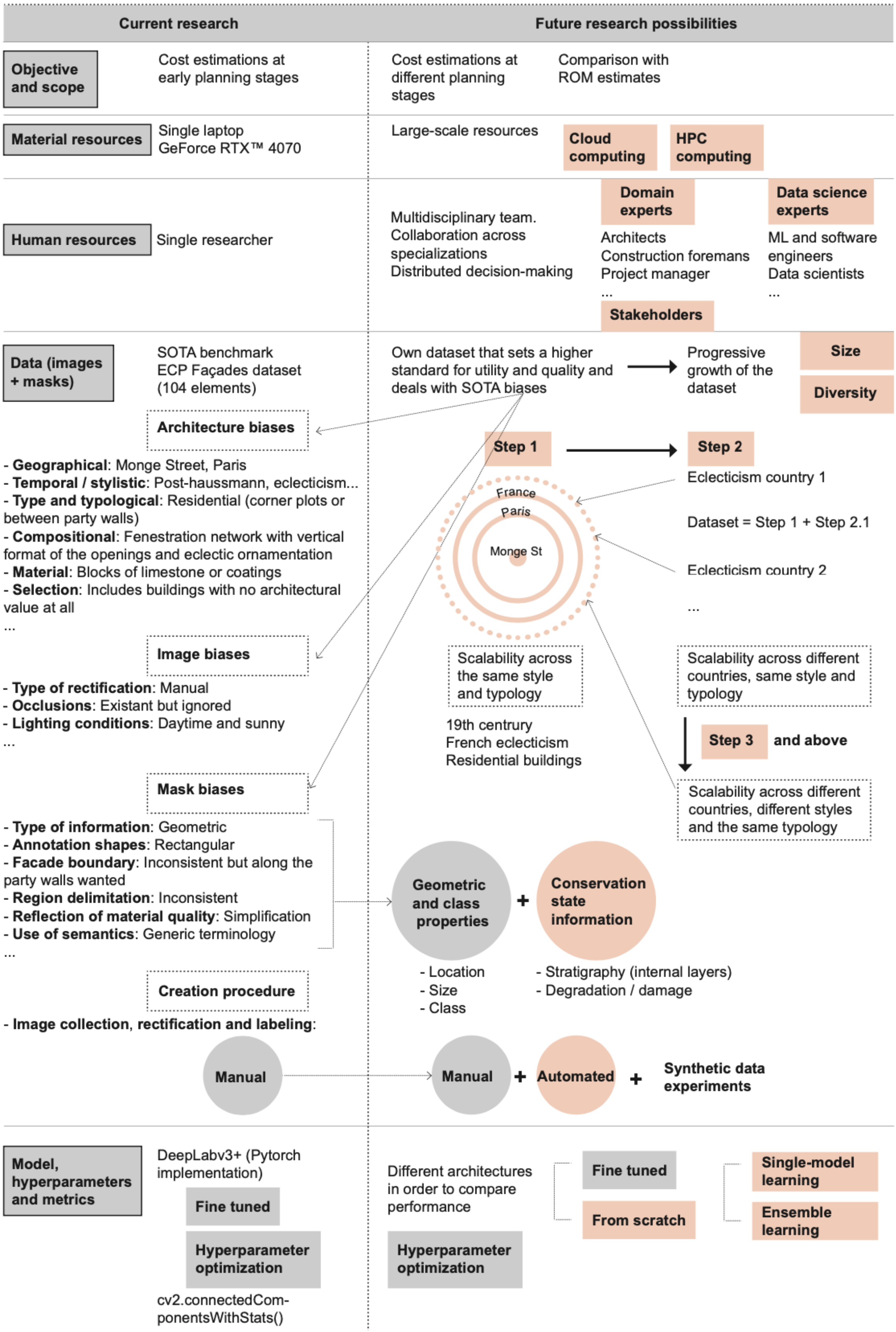

Figure 3 | Input scaling for impact.

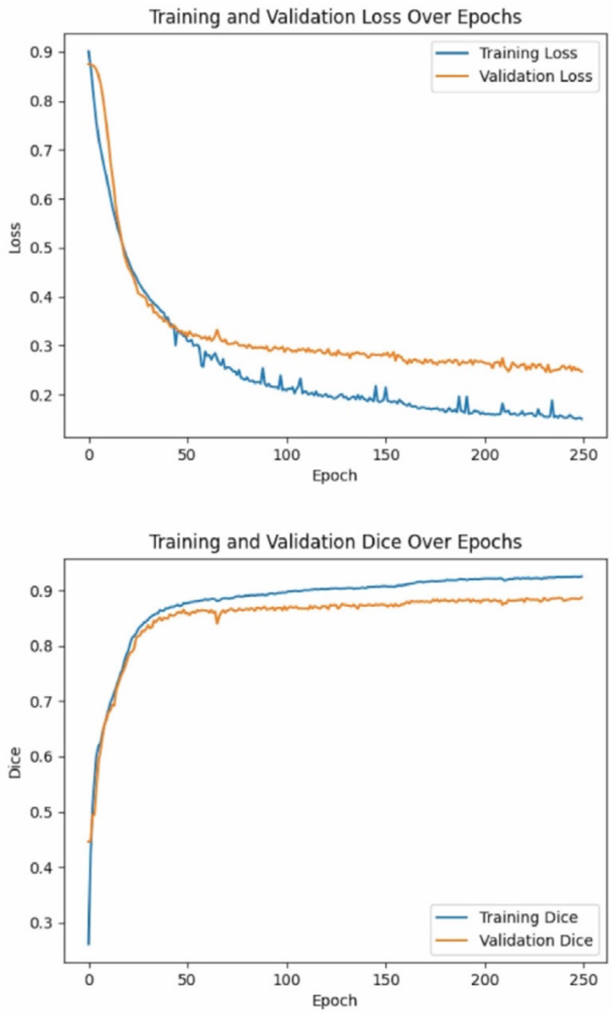

Figure 4 | Training and validation loss and dice curves after 250 epochs.

4.3 Phase 3: Available resources and environment preparation

For this project, an MSI Pulse 15 B13VGK workstation equipped with an Intel® Core™ i7 processor, 32 GB of RAM, and an NVIDIA® GPU GeForce RTX™ 4070 Laptop GPU with 12.1 CUDA support was used. A dedicated virtual environment was set up with Python 3.11.8 and Pytorch 2.2.1. Moreover, the configuration included a set of essential libraries in order to facilitate data preprocessing, analysis, and visualization.

To enhance workflow during the training process, additional packages were also installed such as SMP (Iakubovskii, 2019) for accessing pretrained models, torchmetrics (Detlefsen et al., 2022) for evaluating models performance and torchvision for transformations (TorchVision maintainers and contributors, 2016).

4.4 Phase 4: Semantic segmentation with deep learning

Although deep learning segmentation algorithms have demonstrated outstanding performance, in order to reduce overfitting problems due to the lack of labeled data, basic image augmentation techniques are implemented to artificially increase the train set. In this regard, torchvision.transforms.v2 module is used to apply color (gaussian blur, sharpening, autocontrast and jitter adjustment) and geometric (horizontal flipping) manipulations.

In this supervised experiment, a random splitting approach in 2 steps is proposed using Scikit learn train_test_split function to avoid contamination among subsets, hence over-optimistic model metrics. 83 samples are retained for training, 12 for validation and 9 for test.

After creating Pytorch datasets and dataloaders with their corresponding custom transformations, SMP library (Iakubovskii, 2019) is used, as it provides access to a wide range of encoder-decoder options to experiment with. DeepLabV3Plus (Chen et al., 2018) architecture with Resnet-34 as network backbone is implemented, including encoder pretrained weights on ImageNet.

The training procedure is carried out for 250 epochs with a learning rate of 0.0001, the Adam optimizer with $\beta 1 = 0.5$ and $\beta 2, = 0.999$ and the Diceloss function as the criterion. The performance is measured using the Dice coefficient averaged across the 9 classes. The resulting loss and dice curves are shown in Figure 4.

As stated at the beginning of the article, the objective is not to create a highly competitive model in this field that rivals the state-of-the-art technical publications, so the improvement of the dice score obtained should not be the main concern until the utility of this segmentation techniques is demonstrated. Furthermore, in the context of high-level cost estimates made in the early stages of any architectural project, there is a considerable tolerance for error. Nor should it be overlooked that the final budget will depend to a large extent on the price per unit fixed for each task. Therefore, the average dice value of 0.887 reached for validation is considered sufficient to continue with research, despite being lower than the one achieved by Liu et al. (2017) and others using different techniques.

The model is finally evaluated on 9 unseen images to assess its generalization ability, achieving an average Dice score of 0.865 on the test data. In qualitative terms, predicted masks do not show accurately delineated regions of interest. Nevertheless, although being noisy or blooby, they are sufficiently good when the annotations have been made with greater precision, the facades are not heavily sun-exposed and there is no vegetation on the balconies (Figure 5).

4.5 Phase 5: Feature measurement and facade intervention cost calculation

In order to operate with visually recognizable predicted segmentations, regions must be mathematically identified in the multi-class masks. For this purpose, cv2.connectedComponentsWithStats() from OpenCV (n.d.) is used, which performs a connected component analysis (CCA) on binary images. This technique assigns a unique label to each connected component and provides information about the number of regions identified, the class they belong to and some additional statistics (stats parameter) about the bounding box, such as x and y coordinates, width, height, area (number of pixels) and centroids.

Before proceeding with the calculation, two key considerations must be addressed:

- The classes of interest are Balcony (3), Window (6) and Wall (8).
- Balconies are in the foreground and cover part of the windows and the facade wall. Consequently, only measurements for the first elements can be extracted using Martinovic's masks, which are later automatically modified to exclude class 3 in order to enable calculations for numbers 6 and 8 (Figure 5).

After converting multi-class problems into binary ones, the number of pixels per class is calculated using both ground truth and predicted masks. The percentage of the image occupied by each class is then derived.

Subsequently, the images are scaled using the plot width value (obtained from the Paris cadastre) considering that theoretically, the party walls constitute the vertical limits of the photographs. It is acknowledged that this procedure lacks full automation and requires further consideration.

Next, percentages are applied to real measurements to obtain the number of square meters of each type of constructive element that seeks to be analyzed.

Lastly, it simply comes down to relating the considerations shared in phase 2 with the results of the previous section.

Figure 5 | Building No. 31 image and ground truth (GT) mask over image (first row), GT mask and prediction (PR) mask with balconies (second row), GT mask and PR mask without balconies (last row).

5. Results

After performing comparisons across different masks for each test example, the resulting deviations from the ground truth serve to verify the procedure's relevance. Results for building No. 31 are presented in Table 1; however, this process is fully automated for all cases. As a remark, all decimal places were considered in the calculations for the values in euros.

Table 1 | Building No. 31: results for ground truth and predicted masks with and without balconies.

C	No. of pixels	%	m^2	€
		Ground truth masks (GT)		
3	32343	14,79	57,62	5757,82 €
6	41405	18,93	73,76	3957,50 €
8	116144	53,11	206,91	3528,84 €
		Predicted masks (PR)		
3	29642	13,56	52,81	5277,03 €
6	40842	18,68	72,76	3957,50 €
8	113515	51,91	202,23	3528,84 €

Table 2 | Results across all examples of the test set.

Class	GT €	PR €
3	38172,49 €	39190,24 €
6	29298,28 €	27802,36 €
8	27315,59 €	24859,91 €
Total	94786,36 €	91852,51 €
Class	**Absolute Error (AE) €**	**Relative Error (RE) %**
3	1017,75 €	2,66618 %
6	1495,92 €	5,10582 %
8	2455,68 €	8,99003 %
Total	2933,85 €	3.09523 %

The result of the total budget for the 9 examples of the test set is shown in Table 2. Errors are derived from a different metric -price-, easily understandable for stakeholders that seek to avoid budget overruns right from the planning phase and to streamline resource allocation. Values vary from 2.67 to 8.99% and are tolerable in the early stages of any architecture project that inherently involves uncertainty due to incomplete data and evolving conditions.

6. Conclusions

In summary, this paper provides a holistic approach to semantic segmentation techniques and connected component analysis through an architectural project example that seeks to automate the calculation of high-level budgets for residential facade interventions. Regardless of CMP mask inconsistencies and limitations, the results obtained in section 5 validate the hypothesis that SOTA algorithms can retrieve useful information from facade images.

Nevertheless, despite showing promise, the proposed method is not yet mature enough to serve as a standalone solution for efficient resource allocation and aid management in the architectural heritage conservation field. It needs to be complemented with other diagnostic tools such as condition assessments, degradation monitoring, etc.

Additionally, for these particular tools to be integrated into the urban planning workflow (Figure 2) and thus achieve a real societal impact (Figure 6), the collaborative work of multiple professionals is required, who must attend to the following aspects:

- The creation of bigger and architecturally rigorous datasets tailored to the possible interventions (reactive, proactive or preventive) on buildings with good image quality and no distortions. To avoid leading to a superficial understanding of their condition, masks should capture both geometric and conservation state or damage data.
- The implementation of models that offer better performance in segmentation tasks and the use of mask refinement strategies.
- The consideration of a greater number of construction tasks.
- The synchronization with an up-to-date cost database that reflects market variations in the different locations where construction works should be executed.
- The design of user-friendly end-tools.

7. Acknowledgments

This PhD research is supported by the PIF contract from the University of the Basque Country (UPV/EHU) and the Higher Technical School of Architecture of San Sebastián.

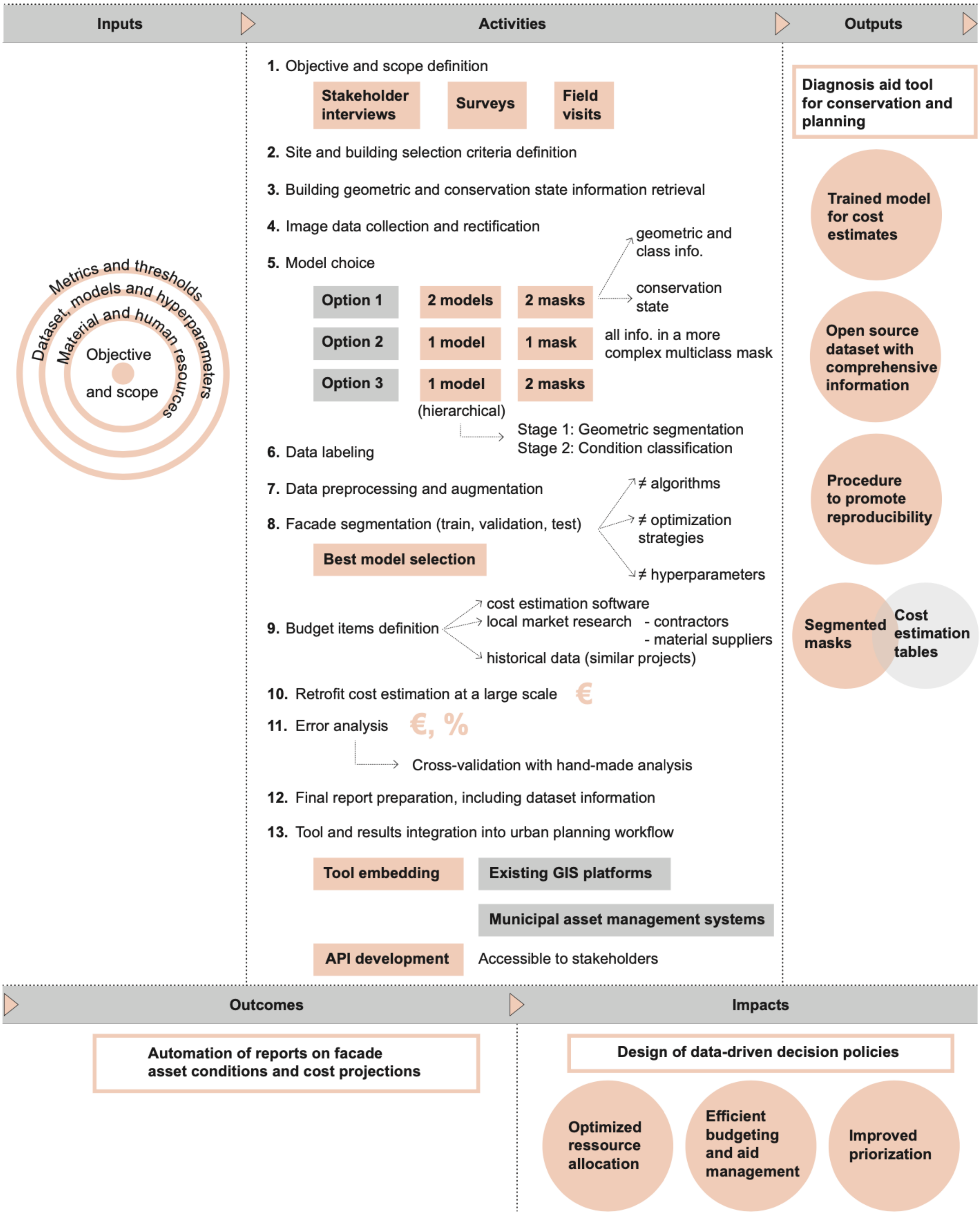

Figure 6 | Dream horizon: impact journey.

8. References

Berg, A., Grabler, F., & Malik, J. (2007). Parsing Images of Architectural Scenes. *2007 IEEE 11th International Conference on Computer Vision*. 1-8. https://doi.org/10.1109/ICCV.2007.4409091

Dai, M., Ward, W., Meyers, G., Densley, D., & Mayfield, M. (2021). Residential building facade segmentation in the urban environment. *Building and Environment, 199,* 107921. https://doi.org/10.1016/j.buildenv.2021.107921

Čech, J., & Radim, S. (2009). Languages for Constrained Binary Segmentation Based on Maximum A Posteriori Probability Labeling. *International Journal of Imaging Systems & Technology, 19*, 69-79. https://doi.org/10.1002/ima.20181

Chen, L-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* (Cornell University). https://doi.org/10.1007/978-3-030-01234-2_49

Detlefsen, N., Borovec, J., Schock, J., Harsh, A., Koker, T., Di Liello, L., Stancl, D., Quan, C., Grechkin, M. & Falcon, W. (2022). TorchMetrics - Measuring Reproducibility in PyTorch. *The Journal of Open Source Software, 7*, 70. https://doi.org/10.21105/joss.04101

Dugast, A., Parizet, I., & Fleury, M. (1990). Dictionnaire par noms d'architectes des constructions élevées à Paris aux XIXe et XXe siècles, période 1876-1899: notices 1 à 1340. Vol. I. *Bulletin monumental*. Institut d'Histoire de Paris.

Femiani, J., Para, W., Mitra, N., & Wonka, P. (2018). Facade Segmentation in the Wild. *arXiv* (Cornell University). https://doi.org/10.48550/arXiv.1805.08634

Fröhlich, B., Rodner, E., & Denzler, J. (2010). A Fast Approach for Pixelwise Labeling of Facade Images. *20th International Conference on Pattern Recognition*, 3029-3032. https://doi.org/10.1109/ICPR.2010.742

Gadde, R., Marlet, R., & Paragios, N. (2016). Learning grammars for architecture-specific facade parsing. *International Journal of Computer Vision, 117*(3), 290–316. https://doi.org/10.1007/s11263-016-0887-4

Iakubovskii, P. (2019). Segmentation Models Pytorch. *GitHub.* Available at https://github.com/qubvel/segmentation_models.pytorch (accessed 13 April 2024).

Iglovikov, V., & Shvets, A. (2018). TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *arXiv* (Cornell University). https://doi.org/10.48550/arXiv.1801.05746

Jampani, V., Gadde, R., & Gehler, P. (2015). *Facade segmentation*. Max Planck Institute for Intelligent Systems. http://ps-old.is.tue.mpg.de/project/Facade_Segmentation (accessed 14 May 2024).

Kelly, T., Femiani, J., Wonka, P., & Mitra, N. (2017). BigSUR: Large-scale Structured Urban Reconstruction. *Transactions on Graphics, 36*(6), 204. https://doi.org/10.1145/3130800.3130823

Koziński, M., Gadde, R., Zagoruyko, S., Obozinski, G., & Marlet, R. (2015). A MRF shape prior for facade parsing with occlusions. *2015 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR). 2820-2828. https://doi.org/10.1109/CVPR.2015.7298899

Laisney, F., & Koltirine, R. (1988). Règle et règlement. *La question du règlement dans l'évolution de l'urbanisme parisien, 1600-1902.* Research report [WWW document]. Ecole Nationale Supérieure d'Architecture de Paris-Belleville. Available at https://hal.science/hal-01903202 (accessed 20 May 2024).

Liu, H., Zhang, J., Zhu, J., & Hoi, S. (2017). DeepFacade: A Deep Learning Approach to Facade Parsing. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2017/320

Mairie de Paris. (2023). Protections patrimoniales: 5ème arrondissement. Plan Local D'Urbanisme De Paris [WWW document]. URL http://pluenligne.paris.fr/plu/sites-plu/site_statique_55/pages/page_1182.html (accessed 14 May 2024).

Martinovic, A. (n.d.). *Angelo Martinovic, Ph.D.* URL http://martinovi.ch/ (accessed 14 May 2024).

Martinovic, A., Mathias, M., Weissenberg, J., & Van Gool, L. (2012). A Three-Layered approach to facade parsing. *Lecture Notes in Computer Science, 7578*, 416–429. https://doi.org/10.1007/978-3-642-33786-4_31

Martinovic, A., & Van Gool, L. (2013). Bayesian grammar learning for inverse procedural modeling. *IEEE Conference on Computer Vision and Pattern Recognition*, 201-208. https://doi.org/10.1109/CVPR.2013.33

Mathias, M., Martinovic, A., & Van Gool, L. (2016). ATLAS: A three-layered approach to facade parsing. *International Journal of Computer Vision 118*(1), 22-48. https://doi.org//10.1007/s11263-015-0868-z

Müller, P., Zeng, G., Wonka, P., & Van Gool, L. (2007). Image-based procedural modeling of facades. *ACM Transactions on Graphics, 26*(3), 85. https://doi.org/10.1145/1275808.1276484

Musialski, P., Wonka, P., Aliaga, D., Wimmer, M., Van Gool, L., & Purgathofer, W. (2013) A survey of urban reconstruction. *Computer Graphics Forum, 32*(6), 146-177. https://doi.org/10.1111/cgf.12077

OpenCV (n.d.) *OpenCV: Structural Analysis and shape Descriptors.* Available at https://docs.opencv.org/3.4/d3/dc0/group__imgproc__shape.html (accessed 14 May 2024).

Pantoja, B., Swamy, V., & Sakota, M. (2020) Extracting Masonry Building Facades through Polygon Image Segmentation. EPFL. Available at https://www.epfl.ch/labs/mlo/wp-content/uploads/2021/05/crpmlcourse-paper782.pdf (accessed 14 May 2024).

Paris (n.d.) *Les voies de Paris : dénominations et numéros d'immeubles.* Available at https://www.paris.fr/pages/les-voies-de-paris-denominations-et-numeros-d-immeubles-7550 (accessed 20 May 2024)

Rahmani, K., Huang, H., & Mayer, H. (2017). Facade segmentation with a structured random forest. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV-1/W1, 175-181. https://doi.org/10.5194/isprs-annals-IV-1-W1-175-2017

Riemenschneider, H., Krispel, U., Thaller, W., Donoser, M., Havemann, S., Fellner, D., & Bischof, H. (2012). Irregular lattices for complex shape grammar facade parsing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1640-1647. https://doi.org/10.1109/CVPR.2012.6247857

Ripperda, N., & Brenner, C. (2006). Reconstruction of Facade Structures Using a Formal Grammar and RjMCMC. *Pattern Recognition. Vol. 4174*, 750-759. https://doi.org/10.1007/11861898_75

Simon, L., Teboul, O., Koutsourakis, P., & Paragios, N. (2011). Random Exploration of the Procedural Space for Single-View 3D Modeling of Buildings. *Int J Comput Vis., 93*, 253–271. https://doi.org/10.1007/s11263-010-0370-6

Shapiro, S.C. (ed.) (1992). *Encyclopedia of Artificial Intelligence*, 2nd edn., Vol. II. John Wiley & Sons, New York.

Schmitz, M. & Mayer, H. (2016). A convolutional network for semantic facade segmentation and interpretation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XLI-B3, 709–715. https://doi.org/10.5194/isprsarchives-xli-b3-709-2016

TorchVision maintainers and contributors (2016). *TorchVision: PyTorch's Computer Vision library. GitHub repository.* Available at https://github.com/pytorch/vision (accessed 14 May 2024).

Teboul, O. (2011). *Shape grammar parsing : application to image-based modeling.* PhD thesis [WWW document]. Ecole Centrale Paris. Available at https://theses.hal.science/tel-00628906 (accessed 14 May 2024).

Tylecek, R. (2013). *The CMP Facade Database (Version 1.1).* Research report [WWW document]. Available at https://cmp.felk.cvut.cz/~tylecr1/facade/CMP_facade_DB_2013.pdf (accessed 14 May 2024)

VarCity ETHZ (2017). VarCity - The Video - semantic and dynamic city modelling from images. Available at https://www.youtube.com/watch?v=6pjEs84DR6Q (accessed 14 May 2024).

Zhang, G., Pan, Y., & Zhang, L. (2022). Deep learning for detecting building façade elements from images considering prior knowledge. *Automation in Construction, 133*. https://doi.org/10.1016/j.autcon.2021.104016

Zhuo, X., Tian, J., & Fraundorfer, F. (2023). Cross field-based segmentation and learning-based vectorization for rectangular windows. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16*, 431-448. https://doi.org/10.1109/JSTARS.2022.3218767