2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

EXTRACTION OF ENERGY-INFLUENTIAL PARAMETERS FROM BUILDING FAÇADE IMAGES THROUGH GOOGLE STREET VIEW

Rafaela Orenga Panizza[1], and Mazdak Nik-Bakht[1]
[1]Concordia University, Montréal, QC, Canada

Abstract

Energy modeling is a crucial tool at the city level for city managers to take decisions related to the building stock. To achieve this, urban models need additional building information to ensure good-quality simulations. Automated image analysis has shown potential in many fields but has lacked to appear in works aiming to improve urban energy analysis. Thus, the objective of this study is to provide a methodology for the extraction of the window-to-wall ratio from building façade images. The methodology proposed in this study includes training a semantic segmentation model. Results of this study have shown that image segmentation models have great potential in extracting the window-to-wall ratio from façade images.

Introduction

While working towards being a North American leader in the fight against climate change, the Québec government has adopted demanding targets to be achieved by the year 2030 (Gouvernement de Quebec, 2016). Such targets include the reduction of non-renewable energy, as well as the enhancement of energy efficiency by 15%. Building energy modeling (BEM) – a process of creating computer simulations, with the help of software tools, to analyze the energy performance of a building - is a very important tool in the battle towards decreasing energy demand. In addition, BEM is also a great player in decreasing energy demand because of its capacity to help in achieving energy compliance, as well as assist in the decision-making of design parameters during building design (Nik-Bakht *et al.*, 2020; Allam *et al.*, 2022; Panizza and Nik-Bakht, 2022).

Since being major contributors to climate change, cities require a rapid adaptation towards more sustainable practices to become more energy efficient. In this context, energy modeling at a city level can be a crucial tool. With the help of city energy models, city managers and energy planners can perform more informed diagnostics on the existing building stock, as well as plan for future energy strategies to be implemented. To achieve this, however, urban-scale 3D models need to include enough building information to ensure good-quality predictions. CityGML, for instance, is an open standard data model that is used for storing digital models of cities (*CityGML | OGC*, no date). Providing enough building data, CityGML models can be used for the modeling of energy performance at a neighborhood or city level (Nouvel *et al.*, 2013).

The energy performance of buildings is highly dependent on building parameters such as their geometry, material, existence, and size of windows, among many other parameters (Allam *et al.*, 2020; Rafaela Orenga Panizza, 2020). The CityGML models currently available for many cities (including the city of Montréal) typically include data related to the building geometries as well as the form of their roofs (Level of Detail 2 - LoD2), excluding all information related to building façades (Nouvel *et al.*, 2013). Though the area and volume of a building are of extreme importance when it comes to calculating energy demand, existing research has shown that, in the cold climate of Québec, the window-to-wall ratio (WWR) parameter (i.e. the size of the windows relative to the size of the building façade) is one of the most significant parameters when modeling the energy performance of buildings (Panizza and Nik-bakht, 2020; Panizza and Nik-Bakht, 2022). CityGML models, however, do include this kind of information. So, the objective of this study is to propose a workflow method for the extraction of the window-to-wall ratio value of building façades for large-scale implementation.

The remainder of this paper is organized as follows. Section 2 is presenting an overview of the existing literature. In section 3, a brief overview of the dataset used throughout this study will be presented. In section 4, the proposed methodology will be explained, followed by an overview of the results and discussion, and finally, the concluding remarks.

Literature Review

In today's world, an incredibly large amount of data is already being collected in endless areas. Thus, with that, the potential of automating image analysis has been recognized in a variety of fields (Koch *et al.*, 2019). Building images, for instance, are a very rich source of building data. In the 3D urban modeling field, there is a significant amount of work utilizing laser scans and street view images for geo-localization, model reconstruction, etc. (Wang, 2013; Adegun *et al.*, 2018; Koch *et al.*, 2019). Laser scan-generated point-cloud images can simplify the detection of building façade elements but at the same time, are very expensive and time-consuming to obtain. Street view images, on the other hand, are mostly publicly

available and can be easily gathered (Neuhausen, Koch and König, 2016).

The existing literature in the field of urban image analysis is vast. A significant portion covers the use of image analysis for detecting street objects and/or building façade elements (Fathalla and Vogiatzis, 2017; Kang *et al.*, 2018). In the real estate field, for instance, works have been done to derive sociodemographic information from existing neighborhoods and residences (Gebru *et al.*, 2017). Works have also focused on the processing of street view images for the purpose of land-use classification (Adegun *et al.*, 2018), geo-localization purposes (Babahajiani *et al.*, 2017), as well as for 3D model reconstruction purposes (Wang *et al.*, 2017). The applications are endless and have shown great potential. Though, since the objective of this study is to automatically calculate the ratio between window and wall of building façades with the help of images, greater focus was put into image segmentation methods.

The pixel-wise feature of image segmentation techniques is what makes this method ideal for the calculation of window-to-wall ratio (Van Ackere *et al.*, 2019). Image segmentation, however, can be performed in two different ways: semantic and instance segmentation. Semantic segmentation is when the objects within the same class are treated as one entity, while instance segmentation identifies the different objects within the different classes. The extraction of the window-to-wall ratio parameter relies on the total area of windows in comparison to the area of the building façade. Therefore, in this study focused attention was given to semantic image segmentation methods.

While analyzing the existing literature with a focus on semantic segmentation methods, it has been noticed that a variety of deep learning models can be used to perform semantic segmentation tasks. To be able to perform semantic segmentation tasks, the convolutional neural network (CNN) models need to use an encoder-decoder architecture. Differently from a CNN architecture used for classification, the architecture used for segmentation has three main parts: the encoder, which is responsible for reading the input image and converting it into the appropriate format; the hidden state, which is the output of the encoder, or the coded message; and the decoder, which is where the coded message is converted into comprehensible language. Some encoder-decoder architectures used in the field include U-Net (Dai *et al.*, 2021), SegNet (Femiani *et al.*, 2018), PSPNet (Zhang, Pan and Zhang, 2022), and DeepLabv3 (Ali, Verstockt and Van De Weghe, 2021).

The different architectures found in the analyzed literature have shown to be successful in segmenting building facades for a variety of purposes. Dai et al., for instance, have used a U-Net architecture to train a building façade segmentation model for surveying purposes (Dai *et al.*, 2021). Other works have used PSPNet and DeepLabv3 for detecting defects and analyzing social changes, respectively (Ali, Verstockt and Van De Weghe, 2021; Zhang, Pan and Zhang, 2022). Though multiple works have been done to segment images of building façades, the existing literature has not yet used the help of image segmentation for collecting building parameters that are relevant to the energy performance analysis of buildings. Thus, the collection of building-related data at a large scale is a significant step towards the overarching goal of this research: to improve the quality of the existing city models by integrating building information details and therefore their ability to model energy demand.

Proposed Methodology

In order to accomplish the objective of this study, four main steps have been accomplished. During the first step ('Data preparation' phase), the labels from the selected dataset are manipulated to preserve only the classes of interest. Then, together with the original images, they form the final dataset that will be used during the next steps. The second step ('Model architecture testing' phase) is where the selected architecture is tested with different pre-trained layers to reach the best model architecture. Then, during the third step ('Training phase'), the final dataset goes through an augmentation process to then be fed as input to deep learning architecture (selected in the previous step) for training. The trained model is then tested and evaluated during the last step ('Evaluation phase') to validate the workability of this model with widely available images from Google Street View (GSV) as well as evaluate its performance. An overview of the methodology being proposed in this study can be found in Figure 1.

Dataset preparation

Two different datasets were used throughout this study: a training dataset and an evaluation dataset. The dataset used for training is also known as CMP dataset (Fritz, 2020). This manually labeled dataset contains 606 façade images of different architectural styles from different countries around the world, such as the Czech Republic, Slovakia, Argentina, Germany, Austria, England, Italy, Switzerland, Spain, Hungary, Greece, and the United States. The CMP dataset has labels for twelve different classes of objects and all object annotations have a rectangular shape. Classes of objects included in this dataset are the following: background, façade, window, blind, cornice, sill, door, balcony, deco, molding, pillar, and shop. The evaluation dataset consists of a sample of façade images of Montreal (Quebec, Canada) buildings extracted from GSV. The images include a wide variety of building types from different neighborhoods to enable the evaluation of the applicability of the proposed methodology at a large scale.

The CMP dataset contains 12 different classes, of which 9 of them are subclasses of either the 'façade' or the 'window' class. Thus, to know the ratio between window and wall and to facilitate the training and prediction processes of the deep learning model, the subclasses of 'façade' and 'window' were simplified at the pixel level. The 'window' and 'blind' classes from the original annotations are combined forming the new 'window'

class. And the remaining façade elements (e.g., 'doors', 'molding', 'sill') are combined forming the new 'façade' class. The pixels labeled with the 'background' class are kept the same. The processed annotations were then left with 3 classes: 'background', 'façade', and 'window'. An example image with its respective annotations (before and after processing) is showcased in Figure 1.

Different from the training dataset, the evaluation dataset was created specifically to be used in this project. This dataset is meant to be used for validating the workability of the model trained during this study, thus a wide variety of building façade types have been extracted to form the evaluation dataset. To be able to help in the evaluation of the model, this dataset needs to contain pixel-wise annotations that include the 3 relevant classes (just like the training dataset). The gathered images included ten different categories of buildings from the city of Montreal, Canada, i.e., low-, medium- and high-rise residential buildings, small and large office buildings, commercial, institutional/public, religious, industrial, and mixed-use buildings. These images were then manually annotated with the help of LabelMe (*LabelMe. The Open annotation tool*, no date). Lastly, the actual values for WWR of the façade shown in the photo were calculated manually by analyzing the 360° view of the façade from GSV.

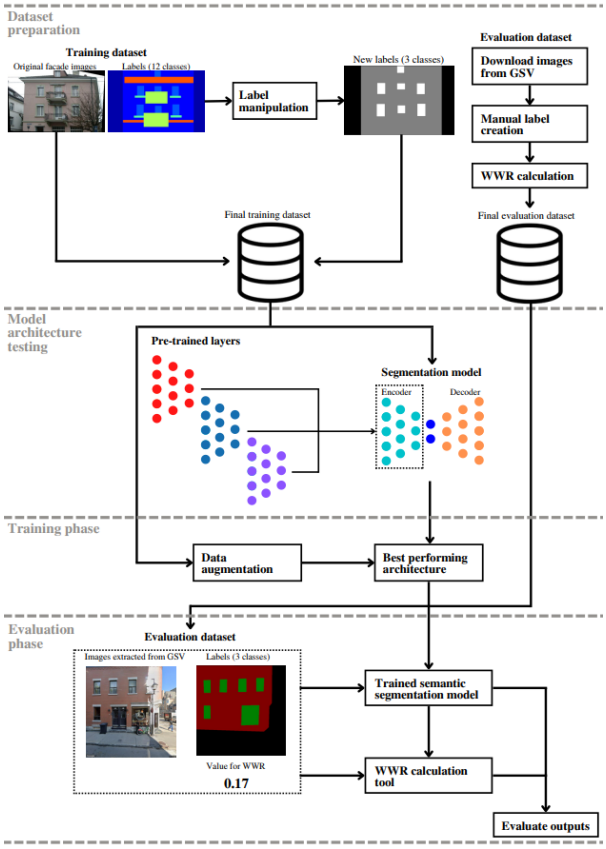

Figure 1. High-level methodology being followed by this study.

Model architecture

The architecture selected to be used for training of the image segmentation model in this study was the U-Net architecture (Ronneberger, Olaf; Fischer, Philipp; Brox, 2021). The U-Net architecture is built upon an architecture named "fully convolutional network" (Zhuang *et al.*, 2019) to be able to take fewer labeled images for training. The network architecture includes two paths: encoder and decoder. The encoder works like a convolutional network where the input images are downsampled into a compact and summarized representation of the input images based on the features recognized throughout the images (feature map). This compact representation is then the input to the decoder. The decoder path includes the upsampling of the feature map back to its original format. The most important part of this model is the process of encoding the input to provide a compact representation that is also complete and meaningful. Thus, in this study, pre-trained models have been used as the encoder to try to optimize the performance of the model.

Pre-trained models are layers of a network that have been previously trained with a large dataset. They can be used as is or they can be customized for a specific task, a practice known as transfer learning. The idea behind the use of pre-trained models is that, if they are trained with a large enough dataset (e.g. ImageNet (Jia Deng *et al.*, 2009)), this model can then be used as a generic model of the visual world. New models can then take advantage of the feature maps learned during the training of these models without having to start from scratch. The practice of using transfer learning is often favored over starting a model from scratch since it makes it so the model can be trained faster, generally provides better accuracies, and is an effective way to handle the challenge of training a model when the available dataset is not very large (Zhuang *et al.*, 2021).

There is a variety of existing pre-trained networks that can be customized for use in the training of an image segmentation model. These networks can take various architectures. They can differ on the depth of the network (i.e., the number of layers) which directly influences the number of features, the number of convolutional layers, the number of fully connected layers, the filter sizes for the convolutional layers, etc. Throughout this study, 32 different pre-trained models have been tested to select the most appropriate model architecture to achieve the objective of this work. The analyzed pre-trained networks included variations of the following convolutional neural networks: VGG (Vedaldi and Zisserman, 2015), ResNet (He *et al.*, 2016), Inception (Szegedy *et al.*, 2015), InceptionResNet (Szegedy *et al.*, 2017), SeResNet (Hu *et al.*, 2020), ResNext (Xie *et al.*, 2017), DenseNet (Huang *et al.*, 2017), EfficientNet (Tan and Le, 2019), SeResNext (Xie *et al.*, 2017; Hu *et al.*, 2020), SeNet (Hu *et al.*, 2020), and MoobileNet (Howard *et al.*, 2017).

Model training and evaluation

Based on the best-performing architecture selected in the previous section, the semantic segmentation model was trained. To avoid overfitting and ensure the best possible accuracy of the semantic segmentation model, the training

dataset was augmented before being fed into the model for training. Data augmentation is a great strategy to be applied to overcome the limitation of not having a large enough training dataset available (Laupheimer *et al.*, 2018) and improve the quality of the image segmentation model. Augmentation techniques utilized include random brightness contrasts, random rotations, grid distortion, and horizontal and vertical flips. Given the nature of the problem, the semantic segmentation model was trained with 50 epochs, SoftMax as its activation function, and Adam optimizer.

The output of this model is a pixel-wise prediction of the three classes: background, window, and wall. The quality of the produced output is then measured based on the intersection over union (IoU) metric. IoU is the primary accuracy measure for image segmentation. IoU for each class is calculated based on their true positives (TP), false positives (FP), and false negatives (FN) (Equation 1). With this predicted output (generated annotations), a pixel-wise analysis is performed in order to calculate the WWR of the façade. To do that, the size of both, the windows, and the walls of each building are collected by counting the pixels classified as 'window' and 'wall', the WWR of that façade can be calculated with Equation 2. Where p_{window} and p_{wall} are the numbers of pixels classified as window and the number of pixels classified as wall, respectively.

$$IoU = {TP}/{(\mathrm{TP} + \mathrm{FP} + \mathrm{FN})} \quad (1)$$

$$WWR = {p_{window}}/{(p_{window} + p_{wall})} \quad (2)$$

Results and Discussion

After going through all the steps of the above-explained methodology, the results are presented and discussed in two parts: the 'model architecture' and the 'model training and evaluation'. The first is focusing on the performance of the different architectures that have been tested for selecting the best-performing one to be selected for the following steps. And then the second part is focusing on the performance of the final trained model as well as its performance when applied to GSV data.

Model architecture

A total of 32 different backbone architectures were tested throughout this study. These models were trained with the original dataset and with 50 epochs, SoftMax as its activation function, and Adam optimizer. The selection of the best-performing model was done based on both, visual and metric evaluation of the trained models. The evaluation of the predictive results of the pool of models started with visualization. The visual evaluation is done initially to ensure that the models considered are providing reasonable predictions. Then, the remaining models are further evaluated based on the IOU metric. After the visual evaluation step, fourteen models that remained and their mean IOU throughout the three classes are showcased in Figure 2.

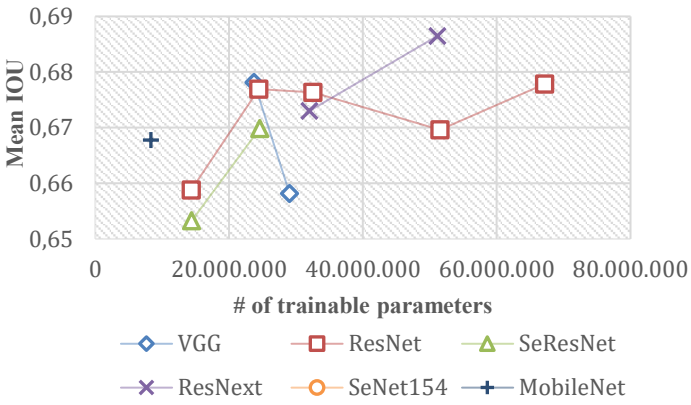

Figure 2. Analysis overview of the performance of validation set of pre-trained model architectures with their respective number of trainable parameters.

To help in the selection of the best architecture, the mean IOU of each architecture was analyzed alongside their number of trainable parameters. The number of trainable parameters in a model is generally a reflection of the depth of the architecture. More parameters to be trained may also result in longer training times due to a greater number of calculations needed, but that doesn't always mean better results. We can clearly see that with the VGG models in Figure 2: the architecture with fewer parameters (VGG16) achieves higher accuracy that the one with a higher number of parameters (VGG19). We can also see that with the ResNet models (ResNet18, ResNet34, ResNet50, ResNet101, ResNet152), it is probably due to the extraction of too many features which can cause overfitting. In order to keep the selected model within a reasonable number of trainable parameters as well as high IOU accuracy, the selected architecture to be used moving forward during this study is ResNext101. Throughout this preliminary step, ResNext101 has taken an approximate training time of 578 minutes (performed on a computer with Intel Core i7-6700 3.4 GHz CPU and 32 GB RAM, running a Windows 10 operating system) and achieved a mean IOU accuracy of 0.686.

Model training and evaluation

After the selection of the appropriate model backbone architecture (ResNext101), as discussed in the previous section, a more robust model was trained. This step aims to build a model of the best possible quality. Thus, the dataset used for model training, even though it was the same as the one used in the previous step, it was five times larger than the previous thanks to the data augmentation technique applied to the dataset before training. The training of this model at this phase is a lot more time-consuming due to the large number of images in the training set, but as expected better accuracy is also achieved. The retraining of the ResNext101 model could bring the semantic segmentation model to a training IOU of over 0.9 and a validation IOU of 0.85. The obtained IOU for window and wall classes were 0.89 and 0.96, respectively. In comparison to similar studies (such as

(Dai *et al.*, 2021) and (Zhang, Pan and Zhang, 2022)), the window class is maintaining a very similar to moderately increased accuracy, but the wall class has shown a moderate increase. A visual representation of the model prediction on a test set image can be seen in Figure 3.

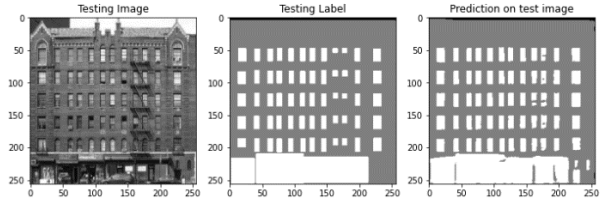

Figure 3. Semantic segmentation model visual result from image retrieved from the test set.

Now that the final model is already trained, it is time to start evaluating its applicability on a different dataset, in this study a dataset of GSV images. The GSV dataset is composed of 30 images, 3 of each building type. The GSV dataset, different from the dataset used for training, contains elements apart from the façade. The initial evaluation of the applicability of this methodology with GSV images has shown promising results. The model when applied to GSV images has provided a mean IOU of 0.5 and a mean square error (MSE) of 0.023 throughout the GSV dataset. Though it is not a very high accuracy, it is a very good preliminary result given that the noise that comes with the images was not handled during this study, which leads the authors to believe that this result has the potential to be improved provided some additional preprocessing of the GSV images in the future works. Some sample results from the GSV image analysis can be seen in Figure 4.

Though the dataset analyzed was limited, some preliminary findings could be highlighted from this analysis. While comparing the different building types during this analysis, it was noticed that this model produces a greater error in the WWR calculation of building categories such as commercial, high-rise residential, and mixed-used buildings in comparison to industrial, small offices, religious, low-and mid-rise residential, and public and institutional buildings. This happens mainly because, in high-rise buildings, the GSV image is not able to capture the building in its entirety. Also, the reason for the higher level of error in commercial and mixed-use buildings is related to the building noise that might be seen in the images, i.e., the neighboring buildings that can affect the results. Predominantly commercial areas (where commercial and mix-use buildings are generally found) are built in denser areas, which makes it very common for buildings to be very close to each other and, therefore, appear in unwanted GSV images and impact the estimated value of WWR. Residential buildings can also be found in denser areas as well, but these often have similar WWRs, which resulted in a low impact on the estimated WWR.

From these findings, it was noticed that in addition to the already mentioned challenges when dealing with these images, curtain-wall buildings are not easily recognized as fully glazed by the trained model. That is because the dataset used for training did not contain a significant amount of curtain wall images. Also, the dataset in general contains a lot of noise. For instance, vegetation (can be seen in the second row of Figure 4), vehicles, and adjacent buildings (can be seen in the first row of Figure 4), among other things. In addition, the images of the desired buildings are also found to be at an angle or not covering the entire building façade in one image alone. These differences between datasets are what make the large-scale implementation of this methodology a challenge.

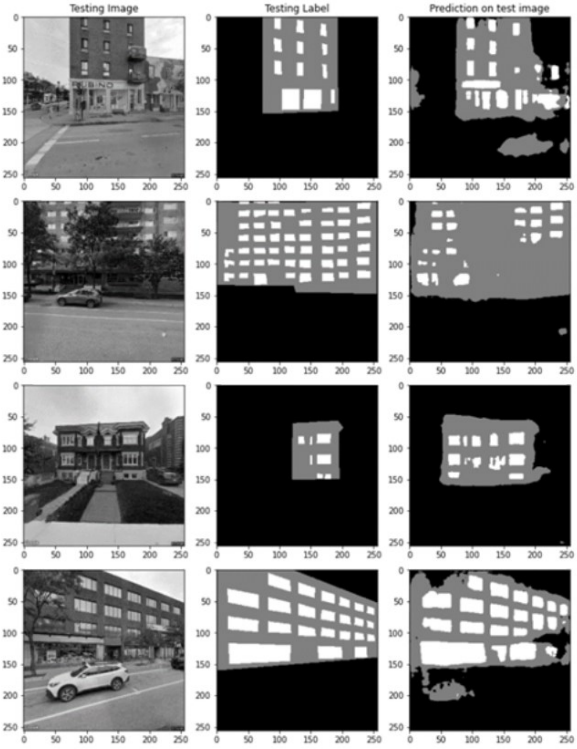

Figure 4. Model results when applied to GSV images (the building types from top to bottom are mixed-use, high-rise residential, low-rise residential, and small office).

Conclusion

This study has proposed a methodology for automatically extracting building information that can be used for the enrichment of urban-scale 3D models. The main idea is to provide urban models with building information that has great significance when it comes to the energy analysis of these buildings. The steps taken by this study towards this overarching goal include a methodology that makes use of semantic segmentation methods for training a model capable of segregating the building façade images into three different categories, which then allows the extraction of the window-to-wall ratio of façades. The model was trained with the help of an already existing dataset and used for testing on widely available GSV images.

The semantic segmentation model trained throughout this study achieved relatively high accuracies when compared with other studies (training IOU of over 0.9 and validation IOU of 0.85). However, when using this trained model on

images from a different dataset (in this case GSV), the model performance when it comes to its ability to segregate the images drops a bit due to the extra noise that the images contain. Nevertheless, it is still a promising performance of 0.5 IOU. From these predictions, the WWR was able to be calculated and has shown an MSE of 0.023 when compared to the actual WWR from the image. The noises found in the GSV dataset are what make the large-scale implementation of this methodology a challenge.

The technique used to gather WWR values from building façades at an urban scale is promising and can also be expanded to other types of available images as well. As it was noticed during the analysis of results when applied to GSV images, the handling of large-scale data may bring some added challenges, such as the angles in which the images are taken, objects that might be blocking the view of the façade, the shape, and height of the building itself, amongst others. Thus, in future works, these need to be addressed in order to be able to integrate this method with an urban model.

Acknowledgment

The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). This study was funded by NSERC through RGPIN/6697-2017.

References

Van Ackere, S. *et al.* (2019) 'Extracting dimensions and locations of doors, windows, and door thresholds out of mobile lidar data using object detection to estimate the impact of floods', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3/W8), pp. 429–436. Available at: https://doi.org/10.5194/isprs-archives-XLII-3-W8-429-2019.

Adegun, A.A. *et al.* (2018) 'Image segmentation and classification of large scale satellite imagery for land use: A review of the state of the arts', *International Journal of Civil Engineering and Technology*, 9(11), pp. 1534–1541.

Ali, D., Verstockt, S. and Van De Weghe, N. (2021) 'Single image façade segmentation and computational rephotography of house images using deep learning', *Journal on Computing and Cultural Heritage*, 14(4). Available at: https://doi.org/10.1145/3461014.

Allam, A.S. *et al.* (2020) 'Estimating the standardized regression coefficients of design variables in daylighting and energy performance of buildings in the face of multicollinearity', *Solar Energy*, 211, pp. 1184–1193. Available at: https://doi.org/10.1016/j.solener.2020.10.043.

Allam, A.S. *et al.* (2022) 'A framework for obtaining a BIM-compatible design solution based on quantitative decisions on building performance', *International Journal of Construction Management*, pp. 1–15. Available at: https://doi.org/10.1080/15623599.2022.2118103.

Babahajiani, P. *et al.* (2017) 'Urban 3D segmentation and modelling from street view images and LiDAR point clouds', *Machine Vision and Applications*, 28(7), pp. 679–694. Available at: https://doi.org/10.1007/s00138-017-0845-3.

CityGML | OGC (no date). Available at: https://www.ogc.org/standards/citygml (Accessed: 22 November 2022).

Dai, M. *et al.* (2021) 'Residential building facade segmentation in the urban environment', *Building and Environment*, 199(December 2020), p. 107921. Available at: https://doi.org/10.1016/j.buildenv.2021.107921.

Fathalla, R. and Vogiatzis, G. (2017) 'A Deep Learning Pipeline for Semantic Facade Segmentation', in *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 1–13. Available at: http://www.aast.edu/cv.php?ser=36825http://www.george-vogiatzis.org.

Femiani, J. *et al.* (2018) 'Facade Segmentation in the Wild'. Available at: http://arxiv.org/abs/1805.08634.

Fritz, K. (2020) *Instance Segmentation of Buildings in Satellite Images*. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1417200.

Gebru, T. *et al.* (2017) 'Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States', *Proceedings of the National Academy of Sciences of the United States of America*, 114(50), pp. 13108–13113. Available at: https://doi.org/10.1073/pnas.1700035114.

Gouvernement de Quebec (2016) *The 2030 Energy Policy*.

He, K. *et al.* (2016) 'Deep residual learning for image recognition', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, pp. 770–778. Available at: https://doi.org/10.1109/CVPR.2016.90.

Howard, A.G. *et al.* (2017) 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications'. Available at: http://arxiv.org/abs/1704.04861.

Hu, J. *et al.* (2020) 'Squeeze-and-Excitation Networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), pp. 2011–2023. Available at: https://doi.org/10.1109/TPAMI.2019.2913372.

Huang, G. *et al.* (2017) 'Densely connected convolutional networks', *Proceedings - 30th IEEE Conference on*

Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua, pp. 2261–2269. Available at: https://doi.org/10.1109/CVPR.2017.243.

Jia Deng *et al.* (2009) 'ImageNet: A large-scale hierarchical image database', pp. 248–255. Available at: https://doi.org/10.1109/cvprw.2009.5206848.

Kang, J. *et al.* (2018) 'Building instance classification using street view images', *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, pp. 44–59. Available at: https://doi.org/10.1016/j.isprsjprs.2018.02.006.

Koch, D. *et al.* (2019) 'Real estate image analysis: A literature review', *Journal of Real Estate Literature*, 27(2), pp. 271–300. Available at: https://doi.org/10.22300/0927-7544.27.2.269.

LabelMe. The Open annotation tool (no date). Available at: http://labelme.csail.mit.edu/Release3.0/ (Accessed: 14 November 2022).

Laupheimer, D. *et al.* (2018) 'Neural Networks for the Classification of Building Use from Street-View Imagery', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV(June), pp. 4–7.

Neuhausen, M., Koch, C. and König, M. (2016) 'Image-based Window Detection - An Overview', *23rd International Workshop of the European Group for Intelligent Computing in Engineering, EG-ICE 2016* [Preprint], (July).

Nik-Bakht, M. *et al.* (2020) 'Economy-energy trade off automation – A decision support system for building design development', *Journal of Building Engineering*, 30(January), p. 101222. Available at: https://doi.org/10.1016/j.jobe.2020.101222.

Nouvel, R. *et al.* (2013) 'CITYGML-based 3D city model for energy diagnostics and urban energy policy support', *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, pp. 218–225.

Panizza, R.O. and Nik-bakht, M. (2020) 'Ranking Energy Influential Parameters – How Building Type Affects the Parameters ' Influence', in *2020 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA*, pp. 416–422.

Panizza, R.O. and Nik-Bakht, M. (2022) 'On the invariance of energy influential design parameters in a cold climate–a meta-level sensitivity analysis based on the energy, economy, and building characteristics', *Advances in Building Energy Research*, 16(4), pp. 466–488. Available at: https://doi.org/10.1080/17512549.2021.1975559.

Rafaela Orenga Panizza (2020) *Correlation and sensitivity of building economy and energy consumption to design parameters*, *A Thesis in The Dept. of Building, Civil, and Environmental Eng*.

Ronneberger, Olaf; Fischer, Philipp; Brox, T. (2021) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', *IEEE Access*, 9, pp. 16591–16603. Available at: https://doi.org/10.1109/ACCESS.2021.3053408.

Szegedy, C. *et al.* (2015) 'Going deeper with convolutions', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, pp. 1–9. Available at: https://doi.org/10.1109/CVPR.2015.7298594.

Szegedy, C. *et al.* (2017) 'Inception-v4, inception-ResNet and the impact of residual connections on learning', *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pp. 4278–4284. Available at: https://doi.org/10.1609/aaai.v31i1.11231.

Tan, M. and Le, Q. V. (2019) 'EfficientNet: Rethinking model scaling for convolutional neural networks', *36th International Conference on Machine Learning, ICML 2019*, 2019-June, pp. 10691–10700.

Vedaldi, A. and Zisserman, A. (2015) 'VGG Convolutional Neural Networks Practical', pp. 1–28. Available at: http://www.robots.ox.ac.uk/~vgg/practicals/cnn/index.html.

Wang, R. (2013) '3D building modeling using images and LiDAR: a review', *International Journal of Image and Data Fusion*, 4(4), pp. 273–292. Available at: https://doi.org/10.1080/19479832.2013.811124.

Wang, S. *et al.* (2017) 'TorontoCity: Seeing the World with a Million Eyes', *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob, pp. 3028–3036. Available at: https://doi.org/10.1109/ICCV.2017.327.

Xie, S. *et al.* (2017) 'Aggregated residual transformations for deep neural networks', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, pp. 5987–5995. Available at: https://doi.org/10.1109/CVPR.2017.634.

Zhang, G., Pan, Y. and Zhang, L. (2022) 'Deep learning for detecting building façade elements from images considering prior knowledge', *Automation in Construction*, 133(May 2021), p. 104016. Available at: https://doi.org/10.1016/j.autcon.2021.104016.

Zhuang, F. *et al.* (2021) 'A Comprehensive Survey on Transfer Learning', *Proceedings of the IEEE*, 109(1), pp. 43–76. Available at: https://doi.org/10.1109/JPROC.2020.3004555.

Zhuang, J. *et al.* (2019) 'Fully Convolutional Networks for Semantic Segmentation', *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, pp. 847–856. Available at: https://doi.org/10.1109/ICCVW.2019.00113.