Politecnico di Torino
Degree Program: Data Science and Engineering

USING COMPUTER VISION FOR THE AUTOMATIC CLASSIFICATION OF BUILDING FACADES

Advisors:
Fabrizio Lamberti, Politecnico di Torino
Fabio Miranda, University of Illinois at Chicago

Candidate:
Davide Bartoletti
S296091

Anno Accademico 2022/2023

Contents

List of Tables

List of Figures

1 SUMMARY

The growth in data availability has led to an increase in the number of studies tackling different urban problems, including accessibility, walkability, and the impacts of climate change on communities. Despite this growth, however, certain studies are still limited by a lack of data that accurately describes the built environment. Such a data scarcity scenario creates opportunities for developing new computational frameworks that leverage and combine already collected data to extract new urban features. This thesis then presents an innovative framework called BuildingSurfaces that employs multi-scale training and semantic segmentation techniques to accurately identify building elements and classify their primary exterior material. We use labeled data from three major cities, combined with street-level imagery, to iteratively train a segmentation model that can achieve a classification accuracy of 92%. Our contributions can be summarized as follows:

1. We present a detailed survey on the availability of building data information across the US

2. We propose a computational framework for the integration of building data and street-level imagery

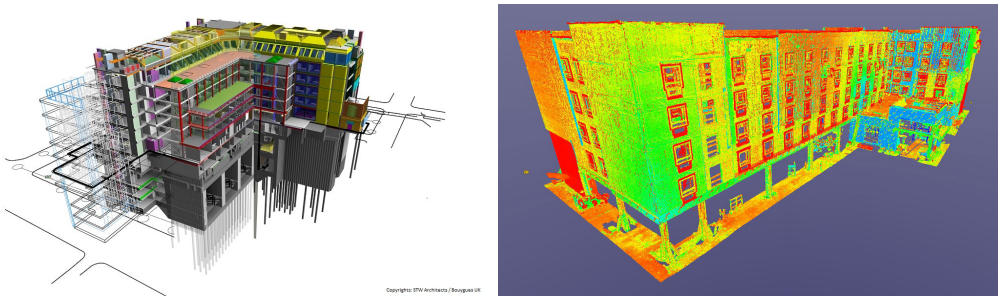
3. We present a detailed experimental evaluation of our segmentation model

4. We make our data available so that researchers can build on top of our efforts

2 INTRODUCTION

The rate at which urban areas are expanding is unprecedented, which is causing a great deal of strain on both resources and the environment. [1] [2]. Efficient use and reuse of resources are essential, which is why there is an urgent call for adopting circular economy principles. [3] In the realm of sustainable architecture, the integration of technology and design has become a transformative force. Building facades significantly impact a structure's energy efficiency, environmental footprint, and aesthetic appeal. A significant challenge is the lack of precise and dependable information on material resources in the constructed environment. [4].Assessment databases for buildings are data sources that hold vast information on properties in every country. They include building details such as identification numbers, positions, sizes, actual use types, apartment numbers, and owner information. However, they do not specify the exterior material of the building. This thesis explores the intersection of artificial intelligence (AI) and sustainable design principles, presenting a novel approach to recognize building facades' primary material through semantic segmentation's precision. Selecting appropriate facade materials is crucial in the face of escalating climate concerns and the imperative to curtail carbon emissions. Architects, engineers, and urban planners must embrace data-driven solutions that mitigate the environmental impacts of built environments. [5]. Automatically recognizing building facade materials can significantly benefit various fields, such as energy efficiency [6], city planning [7], historic preservation [8], and construction industry [9]. Experts can gain insights into material distribution, different buildings' energy efficiency, and historic structures' preservation needs by identifying the materials used in building exteriors. The material classification of building facades and the successive map of the results around cities has the potential to be applied in various fields

such as urban planning, architectural design, and disaster management. In urban planning, mapping the distribution of building materials can provide insights into the age, quality, and presence of hazardous materials in buildings. This information can then be used to make informed decisions about zoning, building codes, and infrastructure development. Moreover, mapping the distribution of different building materials in a city can be a valuable tool for disaster management purposes. Emergency responders can quickly identify areas with a high concentration of buildings susceptible to damage during natural disasters like earthquakes or hurricanes. This information can assist in developing evacuation plans, resource allocation, and post-disaster recovery efforts. [10] However, experts often perform this task manually, which is time-consuming, error-prone, and subject to subjective judgments. These possible applications and the highlighted lack of information about exterior building materials have brought about the need for an advanced AI architecture capable of effectively mapping out the intricate distribution of materials within a city without spending money and time. In the field of computer vision, there is a subfield called semantic segmentation, which can potentially revolutionize how facade materials are evaluated, selected, and used in construction projects. A training data set containing labeled building images is needed to apply this technique. Various digital techniques are available to gather building material data at the individual building level, such as imaging systems, building information modeling (BIM), Internet of Things (IoT), and laser scanning. BIM and IoT are newer technologies that require digital data and are not suitable for creating inventories of buildings built before the 2000s [11]. On the other hand, laser scanning can provide detailed information about construction materials, but it is a complex, labor-intensive, and expensive process that is difficult to scale across countries [12]. However, with the recent advancements in



(a) Building information modeling (b) Laser scanning technique

(c) Street View logo

Figure 1: Methodologies to extract building images

sensing techniques, a vast amount of imagery is now available for building exteriors, such as Google Street View [13], which has opened up new possibilities for cost-effective and scalable approaches to material data acquisition. This data is increasingly becoming openly accessible free of charge, making it a cost-effective source of information on existing buildings. Exploiting these data, we can train a neural network to accurately map the materials used in urban landscapes by identifying the primary material used in a building's facade. This would allow for precise data on material distribution patterns in a city and recognition of key features such as windows, doors, and roofs. This thesis introduces a novel framework, BuildingSurfaces, that employs the NVIDIA Hierarchical Multi-scale Attention architecture as the core for extracting features from input images and generating accurate predictions. The network obtained can then classify facade

materials according to eight different materials: brick, stucco, asphalt, asbestos, wood shake, glass, concrete, and vinyl. Figure 2 provides an overview of the framework, which consists of five blocks:

- Block A - Data extraction: This block uses the Google API to retrieve images of various buildings from the starting building assessment databases.
- Block B - WindRoof Network Training: This block trains a network to recognize important building elements such as windows, doors, and roofs. The training process includes 150 manual annotations and a subsequent pseudo-labeling step to enhance the performance and make the most of the large number of available images.
- Block C - Ground truth creation: After extracting the building images, two types of labels are created from them. The first label, which is specific to building recognition in an image, is obtained using the NVIDIA Hierarchical Multi-scale Attention architecture pre-trained on Mapillary. The second label, used for windows, doors, and roofs recognition, is obtained from the WindRoof network. The two labels are then merged together to create unique complete labels, on which it is possible to train the final architecture.
- Block D - BuildingSurfaces training: This block trains the final network achieving an average classification accuracy of 96% over all eight different materials and a mIoU of 86%.
- Block E - BuildingSurfaces implementations: Given the addresses of various buildings and an image dataset for a specific city, it is possible to map the distribution of materials across a specific place using the BuildingSurfaces trained network.

The resulting trained model has the potential to revolutionize the identification of building facade materials and provide valuable insights into energy efficiency and construction practices. The proposed architecture aims to develop a robust and accurate model for identifying critical elements in building facades and classifying their materials. The final section of the thesis presents interesting results and applies the network to different cities to map the building material distribution and provide a detailed evaluation of our segmentation model.

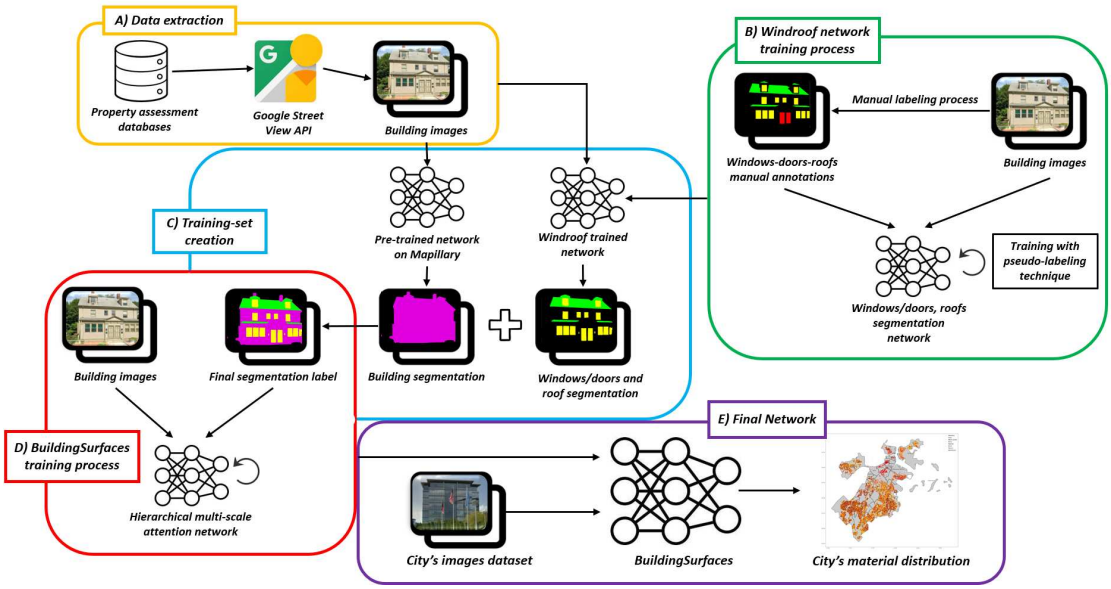


Figure 2: Framework diagram

2.1 Thesis structure

This dissertation is comprised of multiple chapters. The initial chapter provides the reader with general information to provide them an overview of the important topics covered in our research. Chapter 4 details all of the significant works that were analyzed to gain a thorough understanding of the current state of the art in the areas of material recognition, semantic segmentation, and energy efficiency. In Chapter 5, an evaluation of the available databases specific to buildings in US cities is presented, highlighting the lack of information regarding exterior facade materials. Chapter 6 provides a detailed overview of our project framework, BuildingSurfaces, explaining each block from image extraction to training processes and results. Chapters 7 and 8 present the experiments and the results of the AI architecture. The final chapters, Chapter 9 and Chapter 10, focus on potential future work and finish with some important conclusions about our work.

3 BACKGROUND

In this section, the reader has some general knowledge of computer vision, semantic segmentation, and the importance of materials in fields such as energy efficiency and climate impact to understand the proposed work.

3.1 Computer Vision

The field of computer vision is a subset of artificial intelligence (AI) that allows computers and systems to extract meaningful information from digital images, videos, and other visual inputs. With this information, computers can take action or make recommendations. If AI gives computers the ability to think, then computer vision allows them to see, observe, and comprehend. Computer vision works similarly to human vision, but humans have an advantage. Human sight has a lifetime of experience in learning how to differentiate objects, determine their distance, detect movement, and identify flaws in an image. Computer vision trains machines to perform these tasks with cameras, data, and algorithms rather than retinas, optic nerves, and visual cortex. A system trained to inspect products or monitor production can analyze thousands of items or processes per minute, detecting even the slightest imperfections or issues, surpassing human capabilities. Computer vision is utilized in industries such as energy and utilities, manufacturing, and automotive. The market for computer vision continues to grow.

3.2 Semantic segmentation

Semantic segmentation is a pixel-level classification task that assigns a semantic label to each pixel in an image. It provides a detailed understanding of the scene

by segmenting objects and regions based on their semantic meaning. Traditional approaches to semantic segmentation relied on handcrafted features and graphical models. However, with the advent of deep learning, significant progress has been made in this field, improving the performances in various tasks. Possible tasks are the following:

- Object Detection and Tracking: for object detection and tracking. You can locate and track objects more accurately by segmenting objects in an image.
- Autonomous Driving: semantic segmentation plays a crucial role in autonomous vehicles to identify and understand the surrounding environment. It helps in identifying road boundaries, pedestrians, vehicles, traffic signs, and other objects on the road.
- Scene Understanding: to understand and classify different objects and regions within a scene, aiding in scene analysis and comprehension.

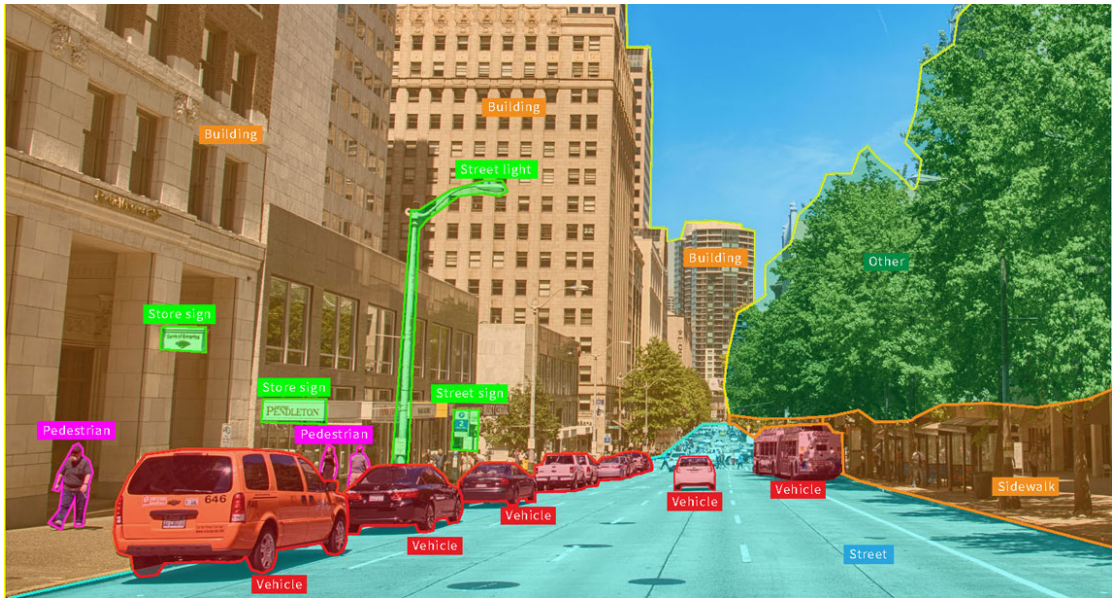


Figure 3: Example of semantic segmentation

3.3 Building materials

The choice of specific materials leads to differences in durability, maintenance, resistance, energy efficiency, climate impact, and sustainability of buildings. In this thesis the main attention is reserved for eight main categories of materials:

- *Brick*: highly valued building material due to its durability, timeless aesthetic appeal, and low maintenance requirements. It boasts excellent thermal mass properties, which effectively regulate indoor temperatures. It finds widespread use in both residential and commercial buildings, where it adds a classic and enduring look to facades and can be incorporated into various architectural styles.

- *Glass*: plays a crucial role in modern architecture due to its transparency and ability to allow natural light into buildings. This not only enhances indoor comfort but also provides a visual connection with the external environment. Glass is widely used in building construction, especially for windows, curtain walls, and glass facades. Its use creates visually appealing and energy-efficient building envelopes.

- *Concrete*: highly esteemed building material due to its immense strength, adaptability, and durability. It can be easily molded into a variety of shapes and textures, and is also resistant to fire and pests. Concrete is utilized in various types of structures, including residential, commercial, and industrial buildings. Typically, it is employed for structural components and exterior cladding, providing an enduring and robust facade.

- *Asbestos*: was once a popular material due to its fire resistance and insulation capabilities, but its use has declined due to health and safety

concerns. Inhaling asbestos fibers can pose significant health risks, leading to its phasing out in many countries.

- *Wood shakes*: are highly valued for their natural and rustic appearance. They are considered environmentally friendly when obtained sustainably and can provide excellent insulation. Typically, wood shake is utilized in residential buildings, especially in areas where a traditional or cabin-like aesthetic is preferred.

- *Vinyl* siding is highly regarded for its affordability, low maintenance, and diverse selection of colors and styles. It is resistant to decay and pests, making it a reliable choice for homeowners. Vinyl siding is commonly used in residential construction as a cost-effective option for cladding the exterior facade of a building.

- *Asphalt* is primarily used due to its excellent waterproofing and weather-proofing properties. However, it is more commonly associated with roofing applications instead of buildings facades.

- *Stucco* is a versatile material that is known for its ability to resist harsh weather conditions. It can be applied in various textures and finishes, which makes it a perfect choice for enhancing a building's appearance. Stucco is commonly used in residential and commercial buildings, particularly in dry regions, as it provides a durable and decorative exterior finish.

(a) Brick (b) Asbestos (c) Glass

(d) Stucco (e) Concrete (f) Asphalt

(g) Vinyl (h) Wood Shake

Figure 4: Building materials for buildings facades

3.4 Energy efficiency and Climate Impact

Distinct energy categories can be distinguished according to the proportion of materials used in building facades versus the area occupied by doors, windows, and roofs. A building's energy efficiency is primarily determined by the type of materials used, the amount of insulation incorporated, and the ratio of transparent to opaque surfaces on its envelope. Numerous international standards, notably LEED, BREEAM, and Passivhaus, have been developed to categorize buildings based on their energy efficiency. These standards factor in various criteria to determine the energy performance of a building, including its construction materials and the proportion of glazed and opaque surfaces. For example, the Passivhaus standard requires buildings to meet strict energy efficiency criteria, windows should not account for more than 25% of the total surface area of the building envelope. To promote energy efficiency and reduce the carbon footprint of buildings, it is possible to classify them into various energy classes based on the percentage of surface covered by materials compared to windows, doors, and roofs in the building facades. High-quality insulation and a more significant percentage of opaque surfaces typically result in better energy performance than poor insulation and more glazed surfaces.

(a) Breeam logo (b) Passivhaus Institute logo (c) Leed logo

Figure 5: International Standards for Energy Efficiency

The selection of materials used for constructing building facades is a critical determinant of a building's climate impact. Various factors impact this relationship, including insulation and thermal conductivity. Different materials possess varying levels of thermal conductivity, which affect the insulation properties of a building. Solar reflectance and absorption properties of facade materials significantly impact a building's energy consumption. Likewise, the production, transportation, and installation of facade materials contribute to their embodied carbon footprint. Material such as concrete has higher carbon emissions during manufacturing while using locally sourced, renewable, or recycled materials can reduce embodied carbon. Life cycle assessments provide a comprehensive view of the environmental impact of materials, including extraction, production, transportation, installation, and disposal/recycling.

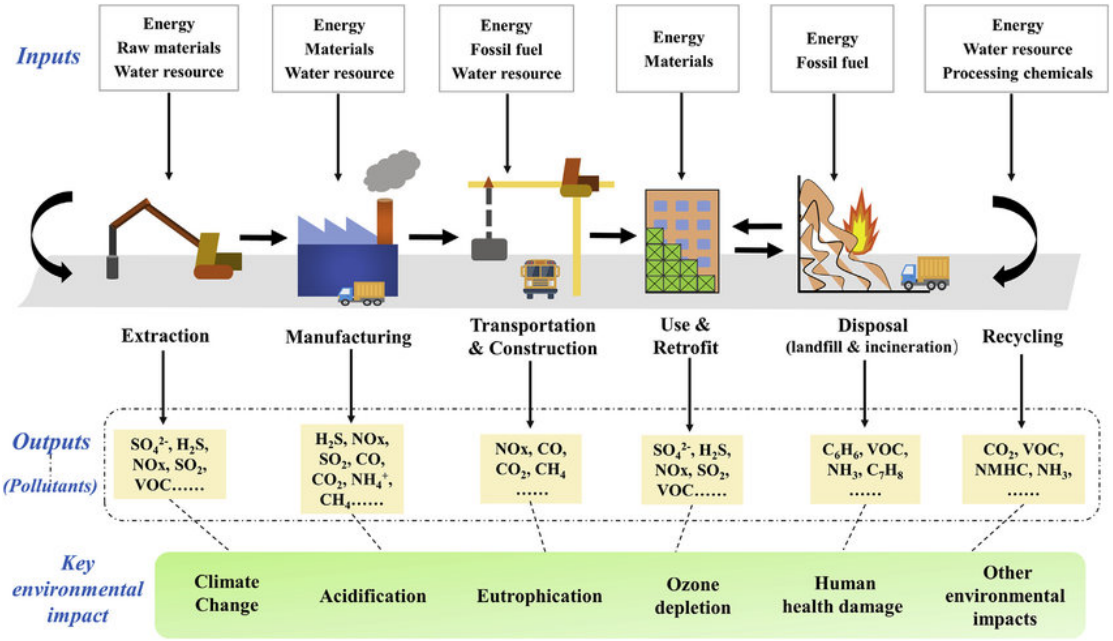


Figure 6: Key Environmental Impacts during the Life Cycle of-Building Materials

The connection between material, energy efficiency, and climate impact in building facades is increasingly important in sustainable design. Building facades are crucial in improving buildings' energy performance and indoor comfort condi-

tions(Yaman, 2021) [14]. As buildings become more complex devices that ensure the well-being of their occupants, there is a growing demand for new facade designs that comply with energy requirements (Yaman, 2021) [14]. Regarding facade design, the connections between facade panels also significantly impact energy performance. Abediniangerabi et al.(2020) [15] investigate the transient heat and moisture transfer in facade panel connections. The study highlights the importance of considering panel connections in energy performance analysis and suggests the need for novel connection designs and materials to improve building energy efficiency (Abediniangerabi et al., 2020) [15]. Guo & Liu (2020) [16] propose a new method for energy efficiency design and thermodynamic evaluation of building facades to support the design of energy-efficient facades. The paper analyzes factors that affect energy efficiency and provides calculation methods for parameters related to energy-saving performance.

Having the capability of recognizing the primarily used material of a building facade and cross-referencing this information with a database of thermal conductivity values or carbon emission values, it is possible to estimate the potential energy efficiency of the building based on the insulating properties of the materials and the embodied carbon of the building facade. Identifying the primarily used material of a building facade can allow architects, engineers, and builders to make informed decisions about facade materials, considering climate impact, energy efficiency, and environmental sustainability.

Figure 7: Envromental Impact of Building Material Pyramid

4 RELATED WORKS

This chapter showcases the related works analyzed to create an AI architecture for classifying materials on buildings' facades. It first analyzes the development of networks in the field of semantic segmentation until the arrival of multi-scale attention-based networks. Then, the overview focuses on the pseudo-labeling technique used to boost the accuracy of the network due to the absence of a large number of training labels. Finally, the last part covers papers related to building facade segmentation, climate impact, energy efficiency, and material recognition tasks.

4.1 Semantic Segmentation

Semantic segmentation is a fundamental computer vision task that involves assigning a label to each pixel in an image, facilitating the creation of detailed and informative image maps. There has been significant progress in semantic segmentation techniques in recent years, with many methods being developed and applied to a wide range of applications, ranging from traditional machine learning techniques to deep learning-based models.

One prevalent approach for semantic segmentation entails using fully convolu-

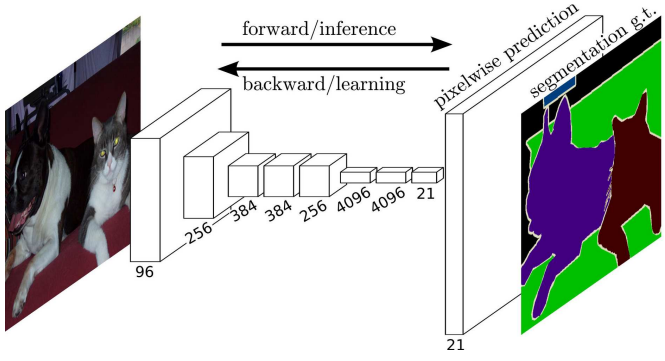


Figure 8: Fully Connencted Network

tional networks (FCNs) [17], which have exhibited impressive results in various image segmentation tasks. FCNs can learn representations of different scales and resolutions, effectively capturing the image's local and global features. Additionally, they can be trained end-to-end, facilitating a more efficient and unified training process. Another approach involves using encoder-decoder networks, which comprise an encoder to extract essential features and representations from input data, often using convolutional or recurrent layers and a decoder that generates desired output from the learned representations, such as segmented images or translated sequences. Variants like U-Net and SegNet have achieved impressive results in various benchmarks.[18] [19] Other approaches include using conditional random fields (CRFs) and generative adversarial networks (GANs) [20] to refine the segmentation maps and enhance their accuracy. Additionally, recent research has focused on utilizing multi-modal data, such as RGB images, LiDAR, and hyperspectral imaging, to improve semantic segmentation performance further.

Dilated Convolutional Networks (DCNs) [21] addressed the contextual challenge by integrating multi-scale context through dilated convolutions, allowing for more efficient information capture. Densely Connected Convolutional Networks (DenseNet) [22] utilized densely connected architectures to promote feature reuse.

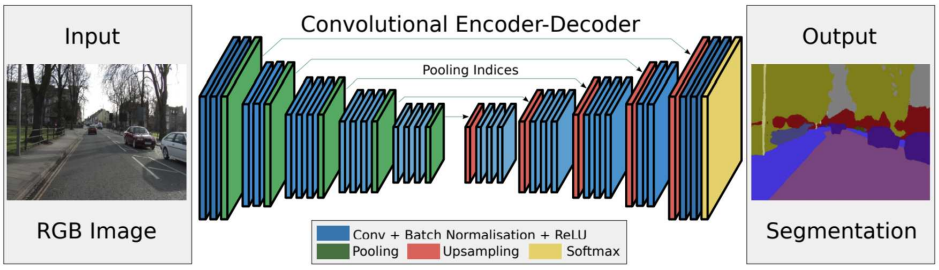


Figure 9: Encoder-Decoder architecture

Deep supervision methods introduced auxiliary supervision signals at multiple decoder stages, mitigating vanishing gradient issues and aiding gradient flow. Data augmentation strategies improved model generalization and robustness, including spatial and spectral transformations. [23] Transfer learning techniques, such as adapting models pre-trained on large datasets like ImageNet, significantly boosted segmentation performance, even with limited annotated data. Domain adaptation techniques have addressed domain shift challenges, enabling models trained on source domains to perform well on target domains with distinct characteristics. At the same time, recent approaches have involved advanced architectures like DeepLab [24], which integrated atrous spatial pyramid pooling for better context integration. Multi-scale context methods have gained importance due to their effectiveness in capturing information at different levels of granularity.

These methods have been applied to various data science tasks, including image

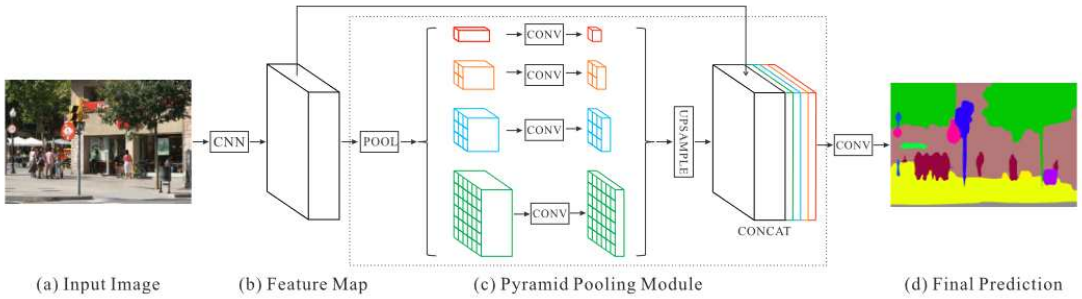


Figure 10: PSPNet Architecture

analysis, natural language processing, and graph-based analytics. State-of-the-art semantic segmentation networks utilize network trunks with low output stride, allowing for better resolution of fine details and resulting in a smaller receptive field, which can hinder predicting large objects in a scene. Pyramid pooling and relational context methods can address this issue by assembling multi-scale context and attending to the relationship between pixels. The PSPNet [25] model has a spatial pyramid pooling module that employs features from the last layer

of the network trunk. This module uses pooling and convolution operations to gather features at different scales.

4.1.1 Relational context methods

Pyramid pooling techniques often use fixed, square context regions due to the symmetrical application of pooling and dilation. However, these techniques are static and do not adapt to an image's specific features. Relational context methods offer a different approach by examining the relationships between pixels, allowing for context to be built without constraints to square regions. Additionally, the learned nature of relational context methods enables context to be built based on an image's composition. This results in a more appropriate context for non-square semantic regions like long trains or tall, thin lamp posts. OCRNet is an example of model that integrates relational context methods. The proposed architecture includes an Object-Contextual Representations (OCR) module that can be added to current convolutional neural network (CNN) backbones. This module gathers contextual information and long-range dependencies among objects and regions in an image, enabling the model to make more informed and context-sensitive segmentation choices.

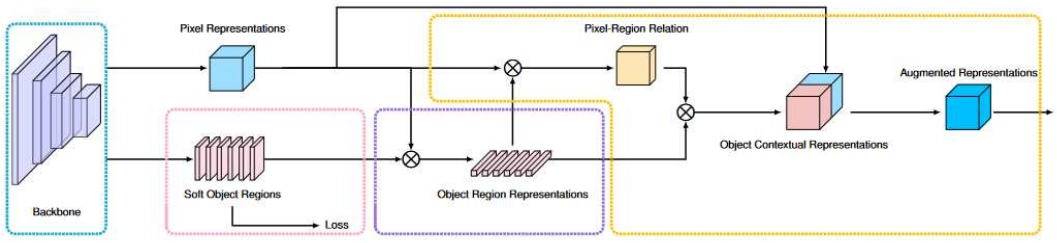


Figure 11: OCR Pipeline

4.1.2 Multi-scale inference

In semantic segmentation, both relation and multi-scale context methods have effectively achieved optimal results. Multi-scale inference is often used to combine network predictions at different scales, but traditional average pooling has limitations as it weights the output from each scale equally. To address this issue, attention mechanisms have been employed to combine predictions across multiple scales better. For example, Chen et al. (2018) [26] used attention heads trained across all scales simultaneously, while Yang et al. (2019) [27] combined features from different network layers to build better contextual information. However, these methods are limited as they are trained with a fixed set of scales and cannot adjust during runtime without re-training the network. To overcome this limitation, a novel hierarchical-based attention mechanism, the Hierarchical multi-scale attention architecture [28] was proposed, which is agnostic to the number of scales during inference time. This method improves performance over average pooling and allows visualizing the importance of different scales for various classes and scenes. Importantly, this approach is orthogonal to other attention or pyramid pooling methods that use a single-scale image and perform attention to combine multi-level features for generating high-resolution predictions. This network will be the trunk of our framework and will be used to generate predictions starting from input images.

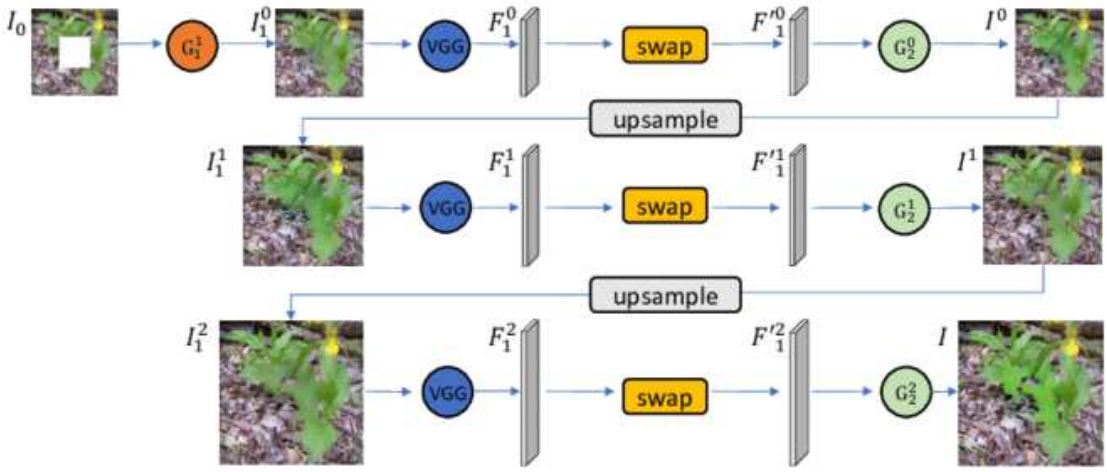


Figure 12: Multi-scale inference

4.2 Pseudo-labeling

Using pseudo-labeling as a semi-supervised learning technique has become very popular lately due to its effectiveness in situations with limited access to labeled data or challenges in obtaining such data. Self-training is one of the widely used and influential pseudo-labeling techniques. It involves an iterative process where a machine learning model is trained on a small set of labeled data and then used to predict labels for a larger pool of unlabeled data. These predictions are combined with the original set of labeled data, resulting in an enriched training dataset.

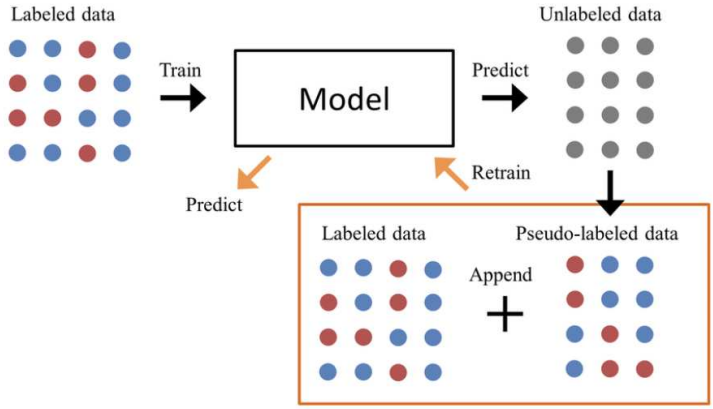


Figure 13: Pseudo-Labeling pipeline

The brilliance of self-training lies in its iterative nature. The model is re-trained continuously, refining its understanding of the underlying patterns within the dataset. The model gains new knowledge with each cycle from the expanding pool of labeled and pseudo-labeled data. This perpetual refinement process makes self-training a powerful and versatile semi-supervised learning technique. Self-training enables machine learning models to learn from themselves, drawing

upon the wealth of knowledge they have amassed with each iterative step. This iterative self-improvement process increases the training dataset's size and helps the model navigate real-world data, even when access to labeled data is limited. As a result, self-training is a crucial technique in the realm of semi-supervised learning, offering a compelling solution to the challenge of harnessing the potential of unlabeled data.

4.3 Facade Segmentation

Several works have been conducted in the field of material recognition and building facades segmentation. These works aim to accurately segment building facades for various applications such as 3D model reconstruction, architectural modeling, and geospatial mapping.

One of the early studies in this area was conducted by (Delmerico et al., 2011)

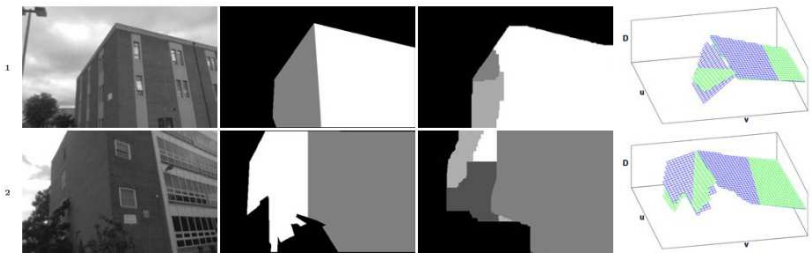


Figure 14: Delmerico paper - Building segmentation

[29]. They utilized texture and a priori knowledge to segment building facades and perform other facade-related tasks. Wendel et al. Delmerico et al. (2011) [29] used intensity profiles to identify repetitive structures in coherent image regions, enabling the segmentation and separation of different facades. Deep learning techniques have gained popularity in building facade segmentation in recent years. Sezen(2022) [30] proposed a deep learning-based approach for door and window

detection from building facades. Their method employed an object detection branch and a bounding box localization branch to detect windows. They also used a partition mask to isolate window samples, enabling segmentation and classification for recognizing windows in an image.

Figure 15: Sezen paper - Windows and Doors recognition

Building Information Modeling (BIM) has also been utilized in facade segmentation studies. BIM allows for the digital representation of various aspects of building information, including geometric and non-geometric aspects. Façade segmentation was initially studied in the 1970s using hand-crafted expertise, and later, detection and segmentation studies emerged based on object shapes and parametric rules. Furthermore, studies have been on efficient building facade structure extraction using image-based laser point cloud (Wang et al., 2023) [31]. Machine and deep learning methods, such as hierarchical clustering, random forests, and adversarial networks, have been introduced for building facade extraction from 3D point cloud data (Wang et al., 2023) [31]. These methods rely on training data to build structure descriptors and classify facade structures from the point cloud (Wang et al., 2023) [31]. Another approach to facade segmentation is through the use of oblique UAV imagery. Zhuo et al. (2019) [32]investigated building segmentation on full-tile UAV imagery and found that

deep neural network-based segmentation methods have demonstrated dominant performance compared to traditional approaches (Zhuo et al., 2019) [32]. In addition, developing a city-scale approach for facade color measurement and building functional classification has been explored using deep learning and street view images (Zhang et al., 2021) [33].

4.4 Material Recognition

Material classification networks are a type of deep learning model that can automatically extract discriminative features from haptic and visual information to classify different types of materials (Zheng et al., 2016) [34]. These networks have been shown to achieve state-of-the-art classification accuracy in various domains. One approach to material classification uses convolutional neural networks (CNNs). CNNs have been successfully applied in image classification tasks, including waste classification, musculoskeletal image classification, fabric fiber material classification, and metallic material classification. These CNN models are trained on large datasets of images and can accurately classify different materials based on their visual features. An essential study in the field involves Citysurfaces (Hosseini et al., 2022) [35], which utilized an active learning-based framework to classify sidewalk materials using commonly available street-level images. This network identified eight distinct types of materials and demonstrated its ability to generalize to different cities, indicating that it was not limited by its training set.

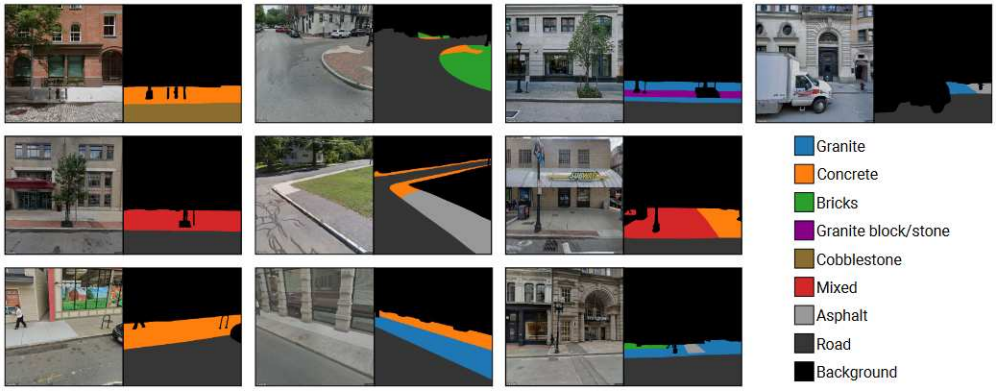


Figure 16: Citysurfaces sidewalk materials recognition

5 DATA SOURCE ASSESSMENT

Myriem's work [35] on semantic segmentation of sidewalks in urban areas inspired the application of this architecture to more complex structures, such as building facades. The first step was to conduct extensive research to find suitable databases that provide information about the addresses and exterior materials of buildings, which could be used as ground truth for our work. After conducting an extensive search for potential data sources, we discovered that each country in the United States possesses assessor parcel data, also known as property parcel data or cadastral data. This type of data is generated by the county assessor's office and serves various purposes, including property taxation, planning, and emergency response. These databases typically provide information about property and land attributes such as its land area, address, use, and zoning (most prevalent features in Table1). However, information regarding the exterior material of buildings is only available in some cases. This is due to four main reasons [36]:

1. **It can be challenging to collect this information**: The exterior materials of buildings can vary widely, and it can be difficult and time-consuming to accurately identify them. This is especially true for buildings with complex or non-standard exteriors.

2. **It can be expensive to collect this information**: The cost of collecting information about the exterior materials of buildings can be prohibitive for some assessor's offices. This is especially true for large counties or cities with many buildings.

3. **It is not always necessary to collect this information**: For some purposes, such as property taxation, buildings' exterior materials are not essential. Collecting this information may not be worth the time and expense

in these cases.

4. **Privacy concerns**: Some people may be concerned about the privacy implications of collecting information about the exterior materials of their buildings. Exterior material data may inadvertently reveal additional sensitive information about a property, such as the economic status of its owner or the presence of specific amenities. [37]

To address the limited amount of information about the exterior material of buildings, we developed a network that employs an AI architecture capable of classifying exterior materials based solely on a set of provided addresses. This network enabled us better to understand the distribution of materials within a city using the vast amount of data available in the assessor's data. To reach this goal, we needed a training set of labeled images, but first of all, reference datasets from which we could extract information about exterior materials.

Parcel number	A unique identifier for each parcel.
Property address	The street address of the property.
Legal description	A legal description of the property's boundaries.
Owner name	The name of the property owner.
Assessed value	The value of the property for tax purposes.
Land area	The size of the land parcel.
Building area	The size of the buildings on the property.
Use	The current use of the property.
Zoning	The zoning classification of the property.
Tax year	The year for which the assessed value is calculated.
Tax rate	The tax rate that is applied to the assessed value to calculate the property tax bill.
Exemptions	Any exemptions that apply to the property, such as homestead exemptions or senior citizen exemptions.
Lien information	Any liens that are attached to the property, such as mortgages or unpaid taxes.
Transaction history	A history of all the sales and transfers of the property.

Table 1: Common characteristics found in assessors' parcel databases

5.1 Building material datasets

The first step was to find databases containing useful information about the exterior material of building facades. A deep search revealed that many cities have building assessment databases about the location and the surface covered by each building within a city but no information about the used material. Only four datasets had information that could be used as a starting point to create the training and test set for our AI architecture:

- Boston Buildings Inventory
- San Francisco Tall Building Inventory
- New York City Building Assessment
- Wake County Property Assessment

5.1.1 Boston Buildings Inventory

The availability of an adequate and reliable dataset is crucial for the success of any machine learning algorithm. In this thesis, the Boston Inventory Data set was identified as a valuable starting point to gather information about building facade materials. This data set was compiled from various sources, including the city's assessing database, the Boston Redevelopment Authority's urban renewal plans, and the city's zoning map. It contains valuable data and assumptions from building experts, making it an ideal resource to identify individual building characteristics of all buildings in Boston. The City of Boston released this data set on 2020-05-05, and it is updated annually. The Boston Inventory Data set is rich in information, with 107 columns. However, this study focused on the location of various buildings and the exterior finishing material specified in the 'ext_fin'

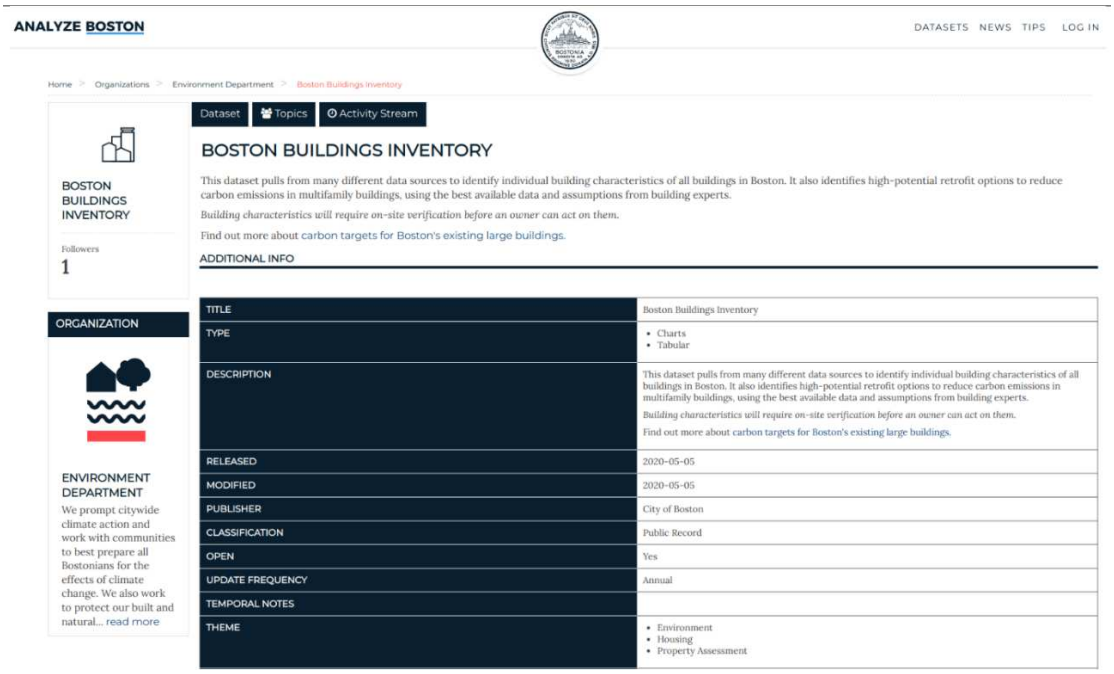


Figure 17: Boston Inventory Database Website

column. After dropping the Nan rows, the cardinality for each material was calculated, and materials with few data, such as other, glass, and concrete, were excluded. Brick material, that as shown in Table 1 is divided in two subcategories, is merged into a single classification. This decision was made because the difference between the three was solely due to the material used in the basement, which was not relevant to this study. The material labels from this data set will serve as the ground truth for training and validating the proposed semantic segmentation model, which can distinguish building facade materials, windows, doors, and roofs.

It was fundamental to identify which areas and districts of Boston were included in the study and how representative the dataset was of the city's entirety. To achieve this, we generated a map of Boston, featuring a dot for each data point in the database. This enabled us to visualize the spatial distribution of the data and identify any geographical biases in the dataset. By analyzing

Material	Dataset label	Cardinality
Vynil	M	32374
Wood shake	W	12121
Brick	B	5506
Asbestos	A	3506
Asphalt	P	1003
Brick/Concrete	C	974
Stucco	S	656
Aluminium	U	560
Brick/Stone Veneer	V	116
Other	O	10
Glass	G	5
Concrete	K	2

Table 2: Cardinality Boston Inventory Dataset

the data distribution on the map, we gained insights into which neighborhoods and districts were underrepresented or overrepresented in the dataset.

Acquiring this information was crucial in ensuring that any conclusions drawn from the analysis were reliable and representative of Boston. It also allowed us to pinpoint any potential limitations in the dataset. As is possible to see in Figure 18 there are data from each district, except for the urban area, where we have few data available or none, as in West End and South Boston Waterfront districts. After thoroughly analyzing material distribution in Boston, we discovered a significant shortage of materials commonly found in tall urban structures, such as glass and concrete. We incorporated additional data sources, specifically the San Francisco Tall Buildings Database and the City of New York Assessment Database, to address this issue. These resources successfully incorporated the previously missing elements into the database and rectified the imbalance. This procedure aimed to create a complete dataset for an adaptable architecture suitable for cities where materials like concrete and glass are commonly used.

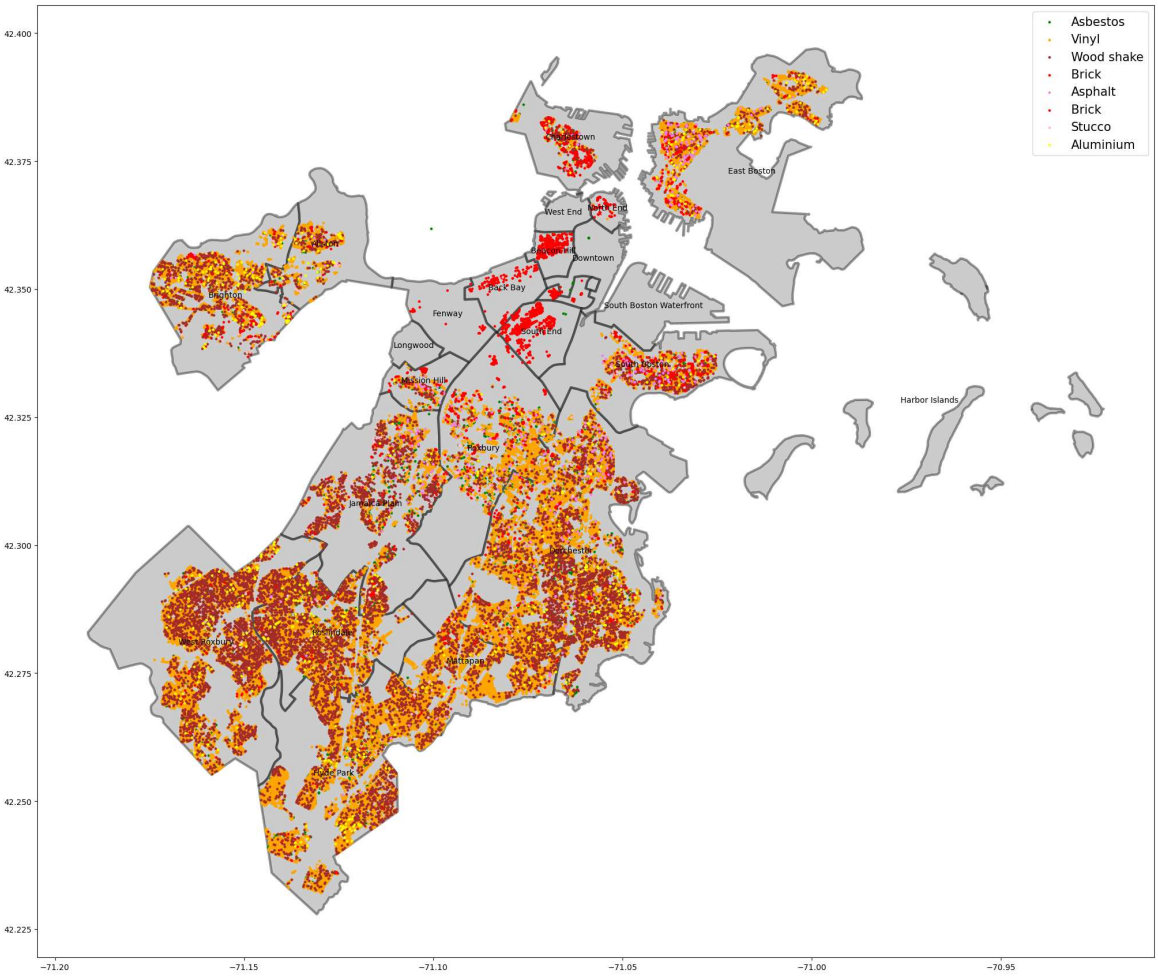


Figure 18: Distribution of buildings materials according to Boston Inventory Database

5.1.2 San Francisco Database

The San Francisco Tall Building Inventory is a database that catalogs and provides technical information about tall buildings in San Francisco, California. This database is typically compiled and maintained by local authorities, urban planning departments, architectural firms, or research institutions interested in urban development and building construction. The inventory aims to capture a wide range of technical details about these tall buildings, which often significantly shape a city's skyline and urban landscape. Some of the key technical details that might be included in the San Francisco Tall Building Inventory are:

- Building Information: The building's name, address, coordinates, and identification number.
- Architectural Details: The architectural design, style, and features that distinguish the building.
- Physical Characteristics: Information about the building's height, number of floors, and other relevant dimensions.
- Materials Used: Details about the construction materials used, including glass, concrete, steel, and others.
- Year of Construction: The year the building was completed or opened to the public.
- Purpose and Usage: Whether the building is for residential, commercial, mixed-use, or other purposes.
- Ownership and Management: The owner or managing entity of the building.

- Historical Significance: Any historical or cultural significance associated with the building.
- Photographs and Visuals: Images and visual representations of the building, both exterior and interior shots.
- Data Source: Information about the data source, including the organization responsible for compiling and updating the inventory.

Similar to the approach taken with Boston Database, we used the same reasoning here and focused solely on the pertinent information about the location and materials used, disregarding extraneous features.

Material	**Dataset label**	**Cardinality**
Concrete	C	93
Glass	G	51
Stone	-	7
Precast	-	5
Terra	-	5
Aluminium	-	2
Granite	-	2

Table 3: Cardinality of San Francisco Tall Building Inventory

Upon examining the cardinality table of materials in the database, it appears that the most commonly used materials are glass and concrete. The main problem is the lack of a significant number of buildings for these two categories. To assess this issue, the New York City Building Assessment was used to increase the number of samples for these two materials.

5.1.3 New York City Building Assessment

The New York City Building Assessment is a comprehensive process that involves the assessment and documentation of various attributes of buildings located within the city. This assessment is typically carried out by city agencies, real estate professionals, research institutions, or organizations involved in urban development and property management. The primary purpose of building assessment in New York City is to gather accurate and up-to-date information about the city's diverse building stock. This information is critical for several purposes, including urban planning, property taxation, infrastructure management, and regulatory compliance. The New York City Building Assessment covers various aspects of buildings, such as structural integrity, electrical and mechanical systems, plumbing, fire safety, and environmental sustainability. The cardinality of this database is about 600k rows. The issue, in this case, is the lack of a feature specific to the facade material. Then, to increase the number of samples for glass and concrete buildings, it was decided to utilize the main streets in the Manhattan area abundant with these materials. The Google API was employed to gather images of such buildings, and with the guidance of experts, the target structures were identified and added to the image training set. The main streets considered are located in the area of Manhattan, as shown in Figure 19.

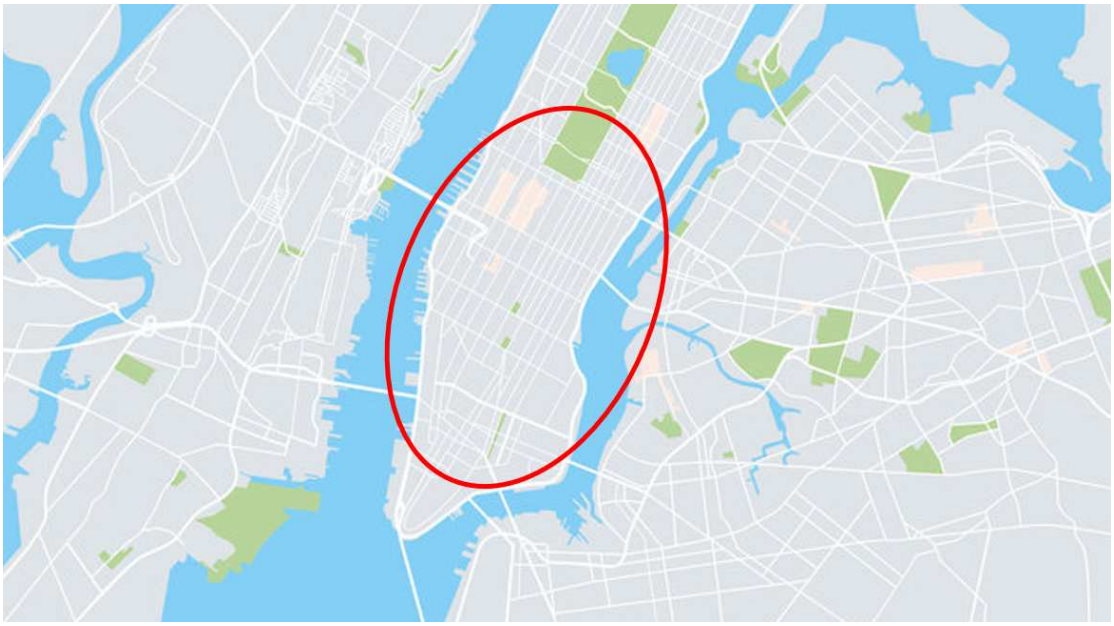


Figure 19: Area of selected buildings in Manhattan

5.1.4 Wake County Property Assessment

Material	Dataset label	Cardinality
Vinyl	M	45217
Brick	B	34475
Stucco	S	1311
Concrete	C	270
Glass	G	90

Table 4: Cardinality of Wake County Property Inventory

Wake County, North Carolina, maintains a property assessment database to manage and track property values and related information within the county. It assesses property values for tax purposes, calculates property tax rates for individual parcels of land and real estate within the county, and determines the amount of property tax owed by each property owner. The database contains detailed information about each property in the county, including property owner information, parcel numbers, physical addresses, and other identifying details. It also includes details about the property's characteristics, such as size, type, construction details, any improvements made to the property, and the exterior material. The property assessment values are determined through periodic assessments and consider factors such as property condition, location, and recent property sales in the area. Property tax records are linked to the assessment database, indicating the amount of property tax owed by each property owner and including payment history. For our study, we retrieved information about the address of each property and the exterior material. In this case, brick, stucco, vinyl, concrete, and glass are mainly used materials. Table 4 shows the cardinality of the database for each single material after removing null and duplicated rows. This dataset will be used as an evaluation data source to test the results of the trained architecture in images from a different geographic area.

6 BuildingSurfaces

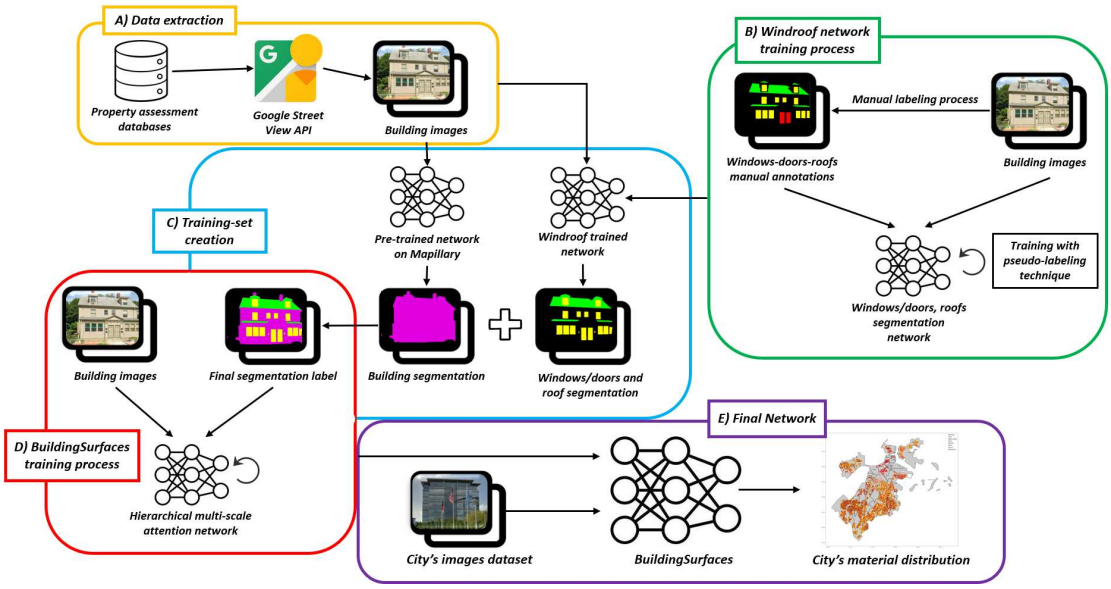


Figure 20: Framework diagram

In this section, we will discuss the process involved in creating BuildingSurfaces, a scalable approach for building material classification. This framework aims to recognize the primary material used in buildings' facades, as well as essential elements like doors, windows, and roofs. Picture 17 provides an overview of the different steps we took, while the following sections give a detailed explanation. Here are the main points:

- Data Extraction: We started by extracting building images from assessment databases and using the Google Street View API. (Block A in Picture 17)
- Windroof Network Training: We trained a network to recognize windows/-doors and roofs using manual annotation, then implemented a pseudo-

labeling technique due to a large amount of unlabeled data. (Block B in Picture 17)

- Training Set Creation: We created unique building labels by combining two labeled images: one for building recognition obtained from the pre-trained multi-scale attention network on Mapillary and one for windows/doors and roofs from the trained windroof network. (Block C in Picture 17)

- BuildingSurfaces Training: We trained a multi-scale attention architecture using the obtained labels, achieving a 95% classification accuracy and an mIoU of 86% (Block D in Picture 17).

Successively to these sections, the evaluation section will display interesting applications of the network to different cities in order to show the network capabilities (Block E in Picture 17).

6.1 Data Extraction

To train our system, we needed to gather accurate data. Our first step was to extract information about buildings (such as address and exterior materials) from databases and then obtain corresponding images using the Google Street View API. We were able to adjust camera parameters to capture images from different angles and heights, focusing on building structures rather than other elements. To achieve this goal, the main tool used is the Google Street View API that, given an address, allows obtaining the corresponding image.

6.1.1 Google API and Dataset creation

The Google Street View Static API enables embedding a static Street View panorama or thumbnail into a web page, with viewport parameters specified through a URL. Upon receiving an HTTP request with these parameters, the API returns a static image. For each HTTP request, the following parameters were included, in order to change the point of view, zoom, and other settings:

- **location**: Specifies the address of each building in the format: *st_num st_name city zip_code*.
- **size**: Determines the dimensions of the image returned (in this case, 640x640).
- **fov**: Represents the horizontal field of view of the image, expressed in degrees with a maximum allowed value of 120. Essentially, the field of view corresponds to zoom, with smaller values indicating higher levels of zoom.
- **pitch**: Dictates the up or down angle of the camera about the Street View vehicle. Although often flat horizontal, positive values angle the camera upwards. In this case, low positive values were employed to obtain images

encompassing the entire building while minimizing the focus on irrelevant sections, such as sidewalks and roads. This parameter is mainly adjusted when high buildings are considered, to capture all the structure.

- **heading**: indicates the orientation of the camera. Accepted values are between 0 and 360 (both values indicate north, 90 indicates east, and 180 indicates south). If no heading is specified, a value will be calculated that directs the camera to the specified location value from the point where the nearest photograph was taken. In our case, this parameter was not set to direct the camera on the address and on the building as a consequence.

(a) fov parameter set to 50 (b) fov parameter set to 70

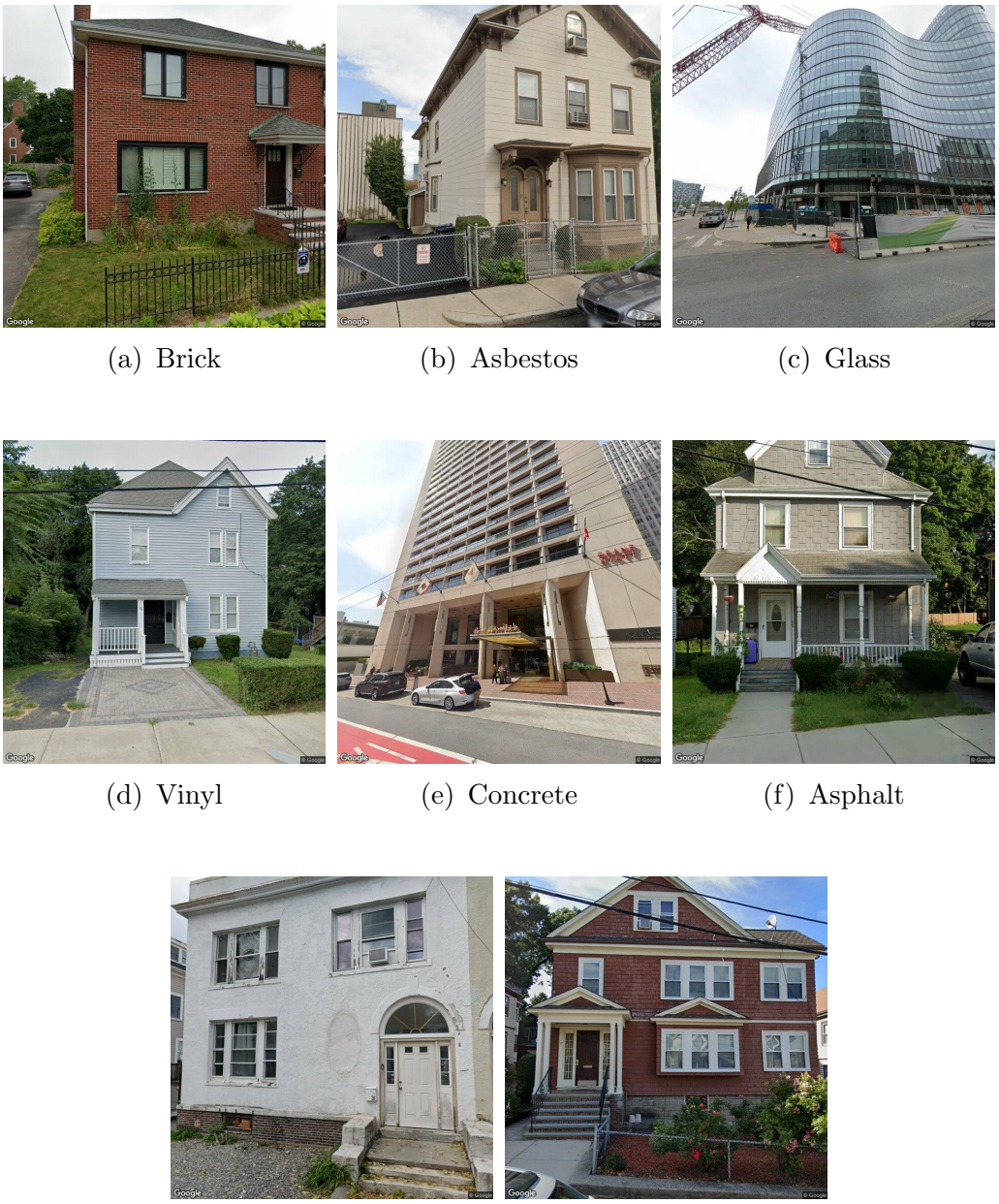
Figure 21: Images of the same building, but with different fov parameter values

This tool allowed us to obtain many buildings pictures for each material present in the starting databases. We started with 500 images for each material, but after a data cleaning procedure, we ended up with 250 images for each label. The considered materials are eight: brick, vinyl, wood shake, stucco, concrete, glass, asbestos, and asphalt. The data cleaning procedure was done according to two main decisions:

- deleting wrong images, in which the building material was different from the correspondent label
- deleting images in which were present more than one building and with different materials (example in Figure 21). In this case, the zoom was adjusted to include only one structure and with the right material.

(a) wrong iamge (b) adjusted image

Figure 22: On the left the wrong image in which two buildings of different materials, on the right the adjusted one where only the label material building is present (brick)



(a) Brick (b) Asbestos (c) Glass

(d) Vinyl (e) Concrete (f) Asphalt

(g) Stucco (h) Wood Shake

Figure 23: Images of the different materials in buildings facade

6.2 Ground truth creation

After creating the image set, we needed to obtain the segmentation of these pictures to use them as ground truth. We first needed an architecture capable of recognizing and delineating buildings in an image, then a network for recognizing elements such as windows/doors and roofs. The final step is to merge the two labels to create a unique one containing all this information.

6.2.1 Building segmentation

Taking inspiration from the work of Maryem [35], to obtain the segmentation of buildings in an image, we started from the pre-trained Hierarchical multi-scale attention network [28]. This network was trained on two primary datasets: Cityscapes and Mapillary, two prominent databases used in machine learning, specifically for object recognition and semantic segmentation tasks in urban environments. Cityscapes is a comprehensive dataset with high-quality pixel-level annotations of urban street scenes from 50 cities. The dataset consists of images taken from a car-mounted camera, providing a perpendicular view of streets with buildings along the sides. Researchers often use Cityscapes to train and evaluate computer vision algorithms for urban scene understanding and autonomous driving applications.

Figure 24: Cityscapes Database samples

In contrast, Mapillary is a diverse dataset from a global community of contrib-

utors capturing street-level imagery from various perspectives, including images with buildings as the focal point. The dataset's extensive geographical coverage and varying conditions, such as lighting and weather, make it ideal for training robust machine learning models that can generalize across different scenarios. An analysis of the results from the network trained on Cityscapes and Mapillary

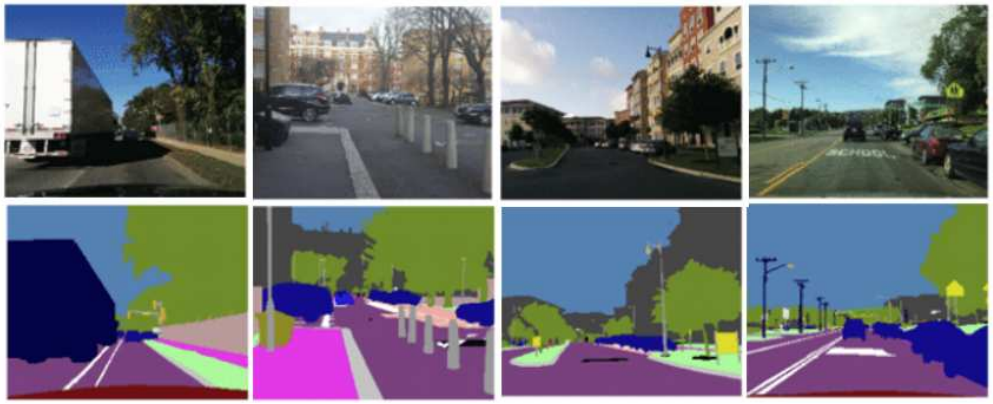
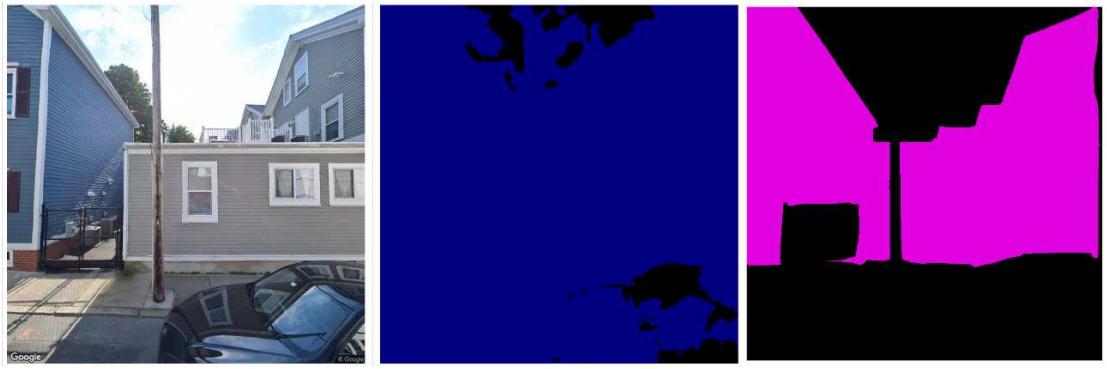


Figure 25: Mapillary Database samples

databases was needed to understand which pre-trained model could be used to find buildings in the images set. Each one of the datasets has labels containing different elements, but our focus was on the recognition of buildings; for this reason, everything that was not labeled with this name was set as background and colored in black.

Then, a comparative analysis was done on the results of the two datasets to determine the best training method. Results reported in Figure 8 show that the Cityscapes dataset is unsuitable for our intended goal, possibly because the training images capture different perspectives with respect to the images present in our dataset. Cityscapes mainly show perpendicular street images with buildings on the edges, while our pictures focus on the building itself. As a result, the network is not recognizing the imagery correctly. Mapillary, on the other hand, showed a big accuracy in delineating building surfaces, as shown



(a) Example1

(b) Example2

Figure 26: Comparison building segmentation of NVIDIA net trained on Cityscapes (center images) and Mapillary (right ones)

in Figure 26. After choosing the proper pre-trained network, the one trained on Mapillary, a different color for every single material, was used. However, this approach did not allow us to extract and include doors, windows, and roofs in our consideration, so we needed to incorporate these elements into our analysis. This aim was born from the need to understand the percentage of facade surface covered by the primary building material, with respect to the one covered by glass surfaces prevalent in doors and windows; in addition, we wanted to exclude elements that could interfere with the prediction and not belonging to the exterior surface material. Figure 28 shows the building segmentation results for each single different material.

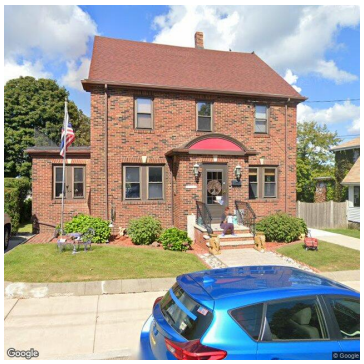
6.2.2 Scales selection

The best possible ground truth was essential to have the best building segmentation. To reach these results, different scale combinations were used. After analyzing the results, it was possible to notice that the best results were reached by the set of scales [0.5, 1, 2]. The possibility to choose between different scales is due to the implementation of the hierarchical multi-scale attention architecture, in which the network has no need to be retrained for each new set of scales. In this case, it was useless to consider small scales since the elements that we want to label are usually characterized by big dimensions.

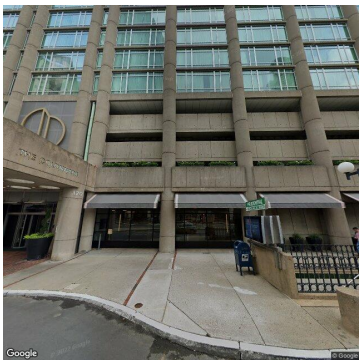
(a) Example1 (b) Example2

Figure 27: Segmentation of buildings using a different set of scales

(a) Asbestos building image



(b) Brick building image



(c) Concrete building image

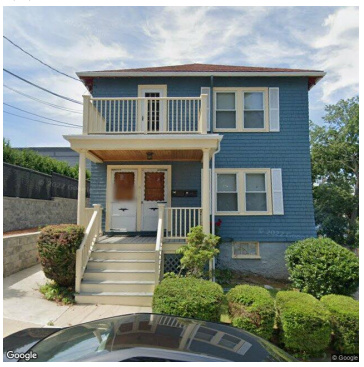
(d) Asbestos building label

(e) Brick building label

(f) Concrete building label

(g) Vinyl building image

(h) Stucco building image

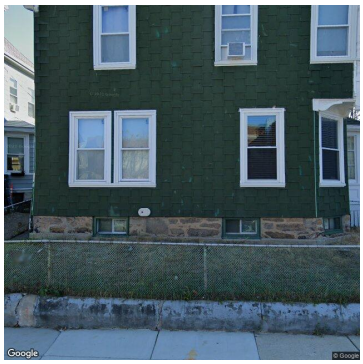


(i) Wood building image

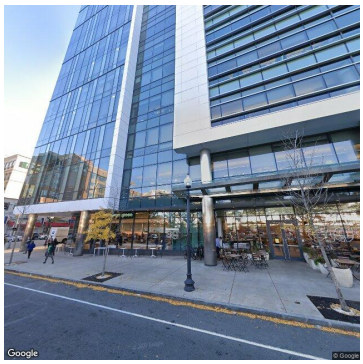
(j) Vinyl building label

(k) Stucco building label

(l) Wood building label



(a) Asphalt building image



(b) Glass building image

(c) Asphalt building label

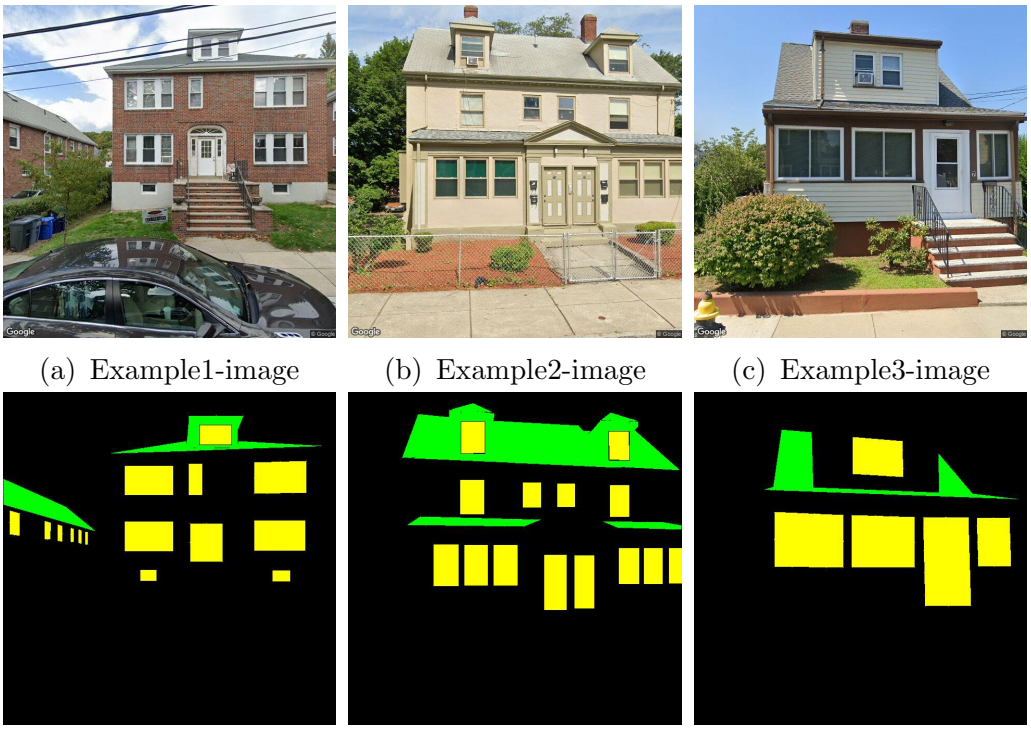
(d) Glass building label

Figure 29: Building segmentation results

6.3 Windows Roofs Segmentation

Accurate segmenting roofs, windows, and doors for building facade analysis can provide critical information for various applications. In this study, we propose a novel approach to improving the segmentation of windows and roofs by training a new neural network. Only images of buildings not present in urban areas were considered for this step. This choice was driven by the difficulty in creating labels for high buildings, where we can find hundreds of windows. Additionally, information regarding energy efficiency is already present for this kind of structure, as opposed to rural areas. To accomplish this, we manually labeled as many images as possible to create a suitable data set for the training. We annotated 150 images and divided them into a training set of 120, a validation set of 10, and a test set of 10. Initially, we used different colors to label windows and doors in order to distinguish between the two categories. However, we found that we didn't require this distinction for our purposes. As a result, we treated both labels in the same way and colored them both yellow (Figure 26).

We used the NVIDIA backbone to train our architecture, the hierarchical multi-scale attention network [28]. Our training resulted in a mean intersection-over-union (mIoU) of 86%, indicating that the network can accurately recognize two distinct labels: windows/doors and roofs. The remaining pixels are classified as background. Our approach effectively utilizes a newly trained neural network to precisely segment windows and roofs. Successful implementation of this network has the potential to advance building analysis and can be applied to various real-world applications. To further enhance our neural network's accuracy, we turned to semi-supervised learning techniques, explicitly pseudo-labeling, which capitalizes on the abundance of unlabeled data. We extracted the best model performance epoch and used it to obtain pseudo-labels for the unlabeled data.



(a) Example1-image (b) Example2-image (c) Example3-image

(d) Example1-label (e) Example2-label (f) Example3-label

Figure 30: Images and labels from manual annotations of three different buildings

After meticulously selecting the top pseudo-labels, we obtained approximately 450 labeled images. We then initiated a new training process using this enlarged dataset, which significantly improved segmentation accuracy. Our mean intersection-over-union (mIoU) skyrocketed to 94%, indicating a remarkable accuracy increase compared to the initial training phase. These results underscore the potential benefits of incorporating semi-supervised learning techniques, such as pseudo-labeling, to boost neural network accuracy, especially when limited labeled data.

(a) 15 HINCKLEY ST (b) Network prediction (c) Image and prediction

(d) 2 PURITAN AV (e) Network prediction (f) Image and prediction

(g) 40 VICTORY RD (h) Network prediction (i) Image and prediction

(j) 47 49 FRANCONIA ST (k) Network prediction (l) Image and prediction

Figure 31: WindRoof network segmentation results on Boston images

6.4 Merging operations and final ground truth

After successfully obtaining two kinds of labels for the same image, one for the building recognition and the second for windows/doors and roofs, we aimed to merge the two predictions to generate an all-inclusive label for each image, through a masking procedure. We wanted a combined label to provide more details about the building facades' materials. The resulting label employed a unique color for each building facade material, while yellow represented windows and doors, and green indicated the roof. Despite the neural network predictions' imprecision, these labels still significantly improved over using the building or windows and roof segmentation models alone. Consequently, we created a dataset containing 250 images for each material and corresponding labels. The labels we created would be utilized to train the ultimate unified network capable of precisely recognizing building materials, roofs, and windows/doors.

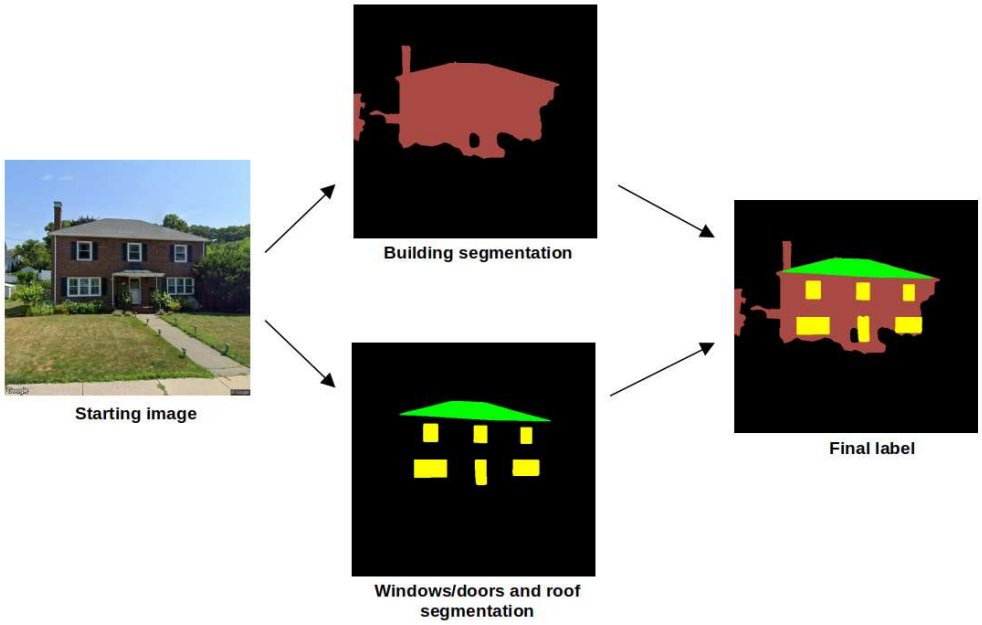
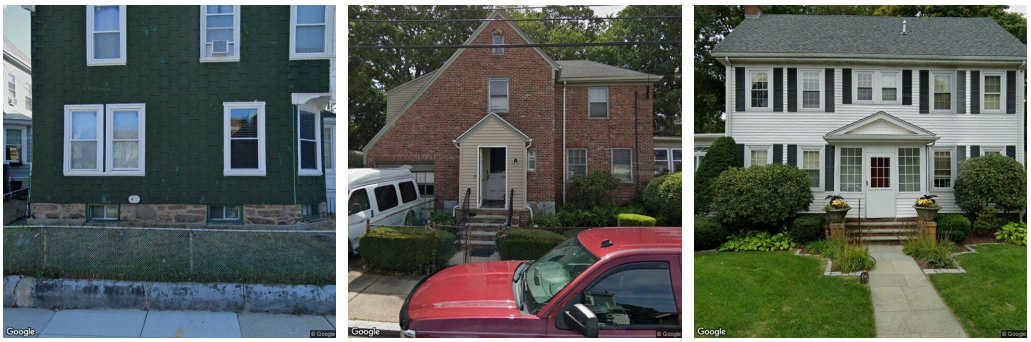


Figure 32: Steps to reach the final training labels

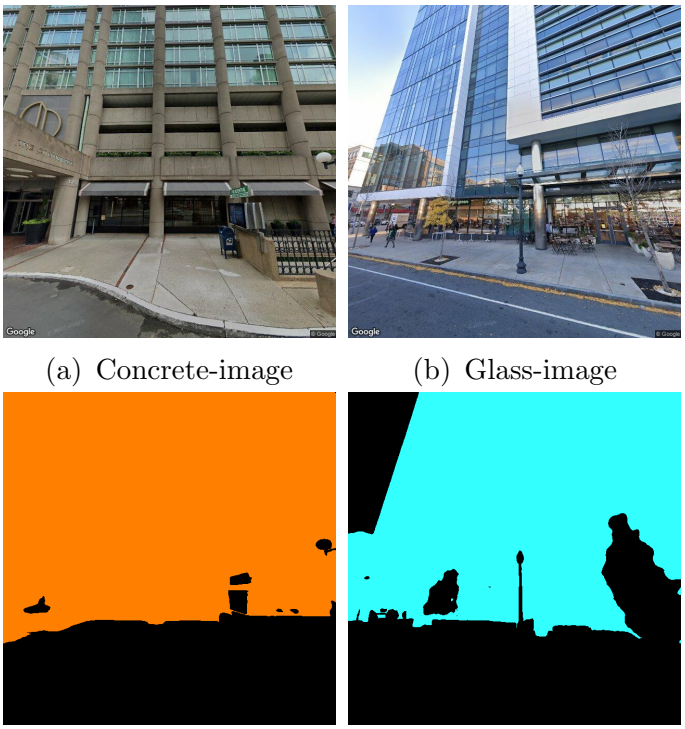


(a) Asphalt-image (b) Brick-image (c) Vinyl-image

(d) Asphalt-label (e) Brick-label (f) Vinyl-label

(g) Stucco-image (h) Wood-image (i) Asbestos-image

(j) Stucco-label (k) Wood-label (l) Asbestos-label



(a) Concrete-image (b) Glass-image

(c) Concrete-label (d) Glass-label

Figure 34: Images and labels, one for each different material

6.5 Backbone network

This chapter provides a detailed description of the AI architecture utilized as the backbone for both creating the training dataset and achieving the final goal. The network employed in this experiment is NVIDIA's, which utilizes a hierarchical multi-scale attention architecture (Hierarchical MSA). Multi-scale inference is used to achieve optimal results, improving the ones of semantic segmentation, with attention being a common technique for combining network predictions at multiple scales. However, existing attention methods are trained with a fixed set of scales and use average or max pooling to combine the different scales. The proposed hierarchical attention mechanism is agnostic to the number of scales during inference and improves performance over average pooling.

6.5.1 Hierarchical multi-scale attention

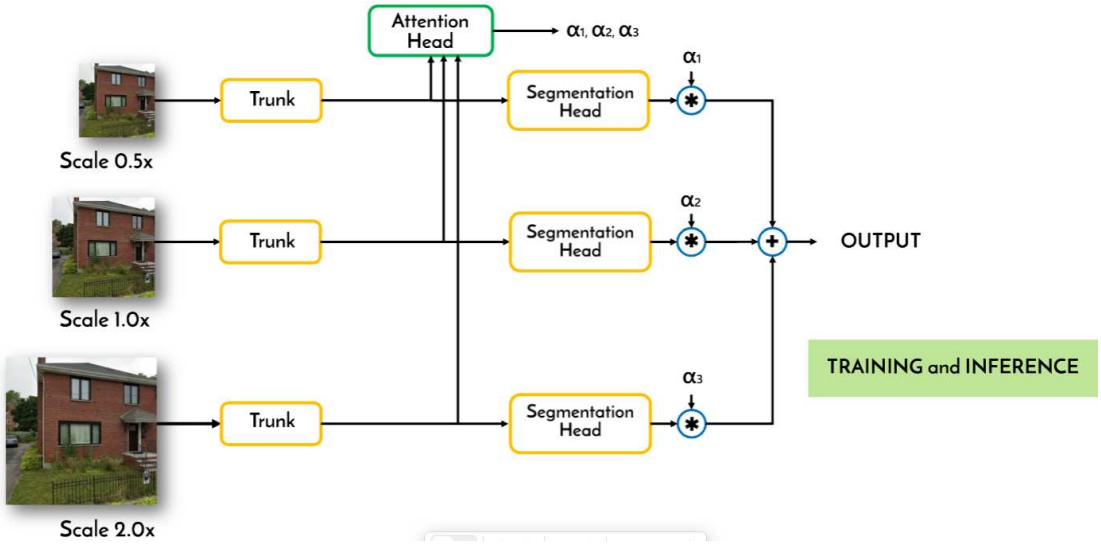


Figure 35: Multi-scale training and inference steps

This approach is very similar to that of [26], where a dense mask is learned for each scale, and the multi-scale predictions are combined by pixel-wise multipli-

cation between masks and predictions, followed by pixel-wise summation across scales to produce the final results (see Figure 6). The only difference is that this hierarchical approach involves learning a relative attention mask between adjacent scales rather than one for each fixed scale. During network training, are only considered adjacent scale pairs. As illustrated in Figure 6, the network predicts the dense pixel-wise relative attention between two image scales given a set of features from a single lower scale. To obtain the pair of scaled images, a scale-down on a single input image by a factor of 2 is performed, resulting in a 1x scale input and a 0.5x scaled input, although any scale-down ratio could be used. It's essential to note that the network input is a re-scaled version of the original training images because image scale augmentation is used during training. This enables the network to predict relative attention for various image scales. During inference, the learned attention hierarchically is applied to combine N scales of predictions in a chain of computations, as shown in Figure and described by the equation below. We prioritize lower scales and work our way up to higher ones, assuming they have more global context and can determine where higher-scale predictions should refine predictions. During the training process, an input image undergoes scaling by a factor r, where r = 0.5 indicates down-sampling by a factor of 2, r = 2.0 denotes up-sampling by a factor of 2, and r = 1 indicates no operation. Our training process utilizes r values of 0.5 and 1.0. The shared network trunk processes the two images with r values of 1 and 0.5, resulting in semantic logits L and an attention mask(α) for each scale. These masks are used to combine the logits L between scales. For two-scale training and inference, the bilinear upsampling operation U is utilized, and pixel-wise multiplication and addition (+) are performed. The equation can be formalized as follows:

$\mathcal{L}_{(r=1)} = \mathcal{U}\left(\mathcal{L}_{(r=0.5)} * \alpha_{(r=0.5)}\right) + \left(\left(1 - \mathcal{U}\left(\alpha_{(r=0.5)}\right)\right) * \mathcal{L}_{(r=1)}\right)$ There are two ad-

vantages to using the proposed strategy that lead to improved performances with respect to other architectures:

- During inference, the proposed attention mechanism allows for a flexible selection of scales, including adding new scales such as 0.25x or 2.0x to a model trained with 0.5x and 1.0x. This differs from previous methods that were limited to using only the same scales as those used during model training.

- The hierarchical structure offers improved training efficiency over the explicit method, as demonstrated by the reduced training cost. For example, using scales 0.5, 1.0, and 2.0 with the explicit method results in a training cost of $0.5^2 + 1.0^2 + 2.0^2 = 5.25$ relative to single-scale training. With our hierarchical method, however, the training cost is only $0.5^2 + 1.0^2 = 1.25$.

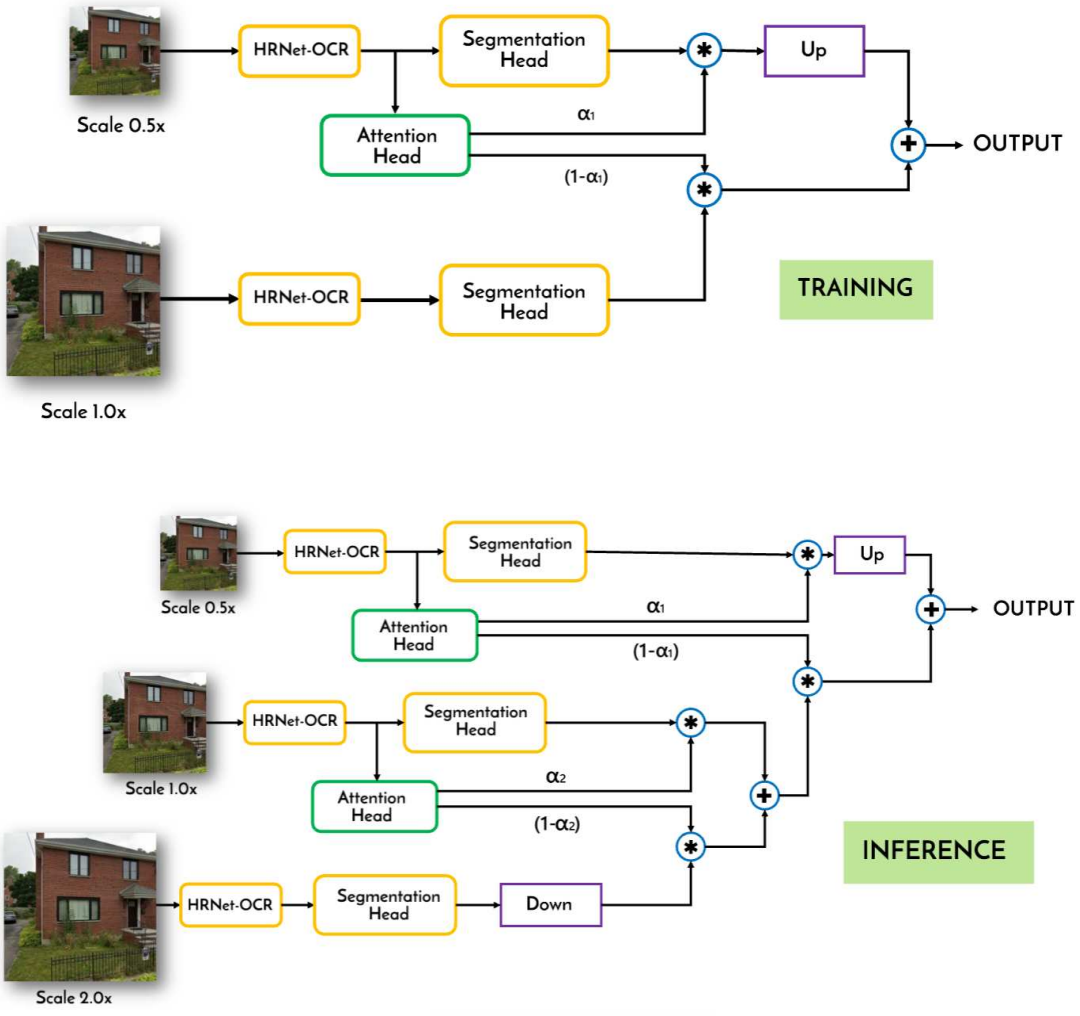


Figure 36: Hierarchical multi-scale attention architecture during training and inference steps

6.5.2 Architecture

- **Backbone** For the ablation studies in this section, HRNet-OCR is used as the backbone of the network, configured with an output stride of 8.

- **Semantic Head** To perform semantic predictions is used a dedicated fully convolutional head that consists of a 3x3 convolution, batch normalization (BN), rectified linear unit (ReLU), another 3x3 convolution, another BN, another ReLU, and a 1x1 convolution. The final convolution generates *num_classes* channels and is responsible for combining the predictions from multiple network scales.

- **Attention Head** For attention predictions, a separate head is implemented structurally identical to the semantic head, except for the final convolutional output, which generates a single channel. Semantic and attention heads are fed with features from the OCR block, and an auxiliary semantic head takes its features directly from the HRNet trunk before OCR. This auxiliary head consists of a 1x1 convolution, BN, ReLU, and another 1x1 convolution.

After applying attention to the semantic logits, the predictions are upsampled to the target image size with bilinear upsampling. (Figure 17)

6.5.3 HRNet-OCR

6.5.4 HRNet

State-of-the-art frameworks utilize a subnetwork that connects high-to-low-resolution convolutions in series (e.g., ResNet, VGGNet) to encode an input image into a low-resolution representation. This low-resolution representation is then used to recover the high-resolution representation of the image. However, High-Resolution Network (HRNet) [38] maintains high-resolution representations throughout the entire process. These representations are not only semantically strong but also spatially precise. This comes from two aspects:

- The connection between high and low-resolution convolution streams is in parallel rather than in series. Thus, this approach can maintain the high resolution instead of recovering high resolution from low resolution, and accordingly, the learned representation is potentially spatially more precise.
- Most existing fusion schemes aggregate high-resolution low-level and high-level representations obtained by upsampling low-resolution representations. Instead, repeated multi-resolution fusions are applied here to boost the high-resolution representations with the help of the low-resolution representations and vice versa. As a result, all the high-to-low-resolution representations are semantically strong.

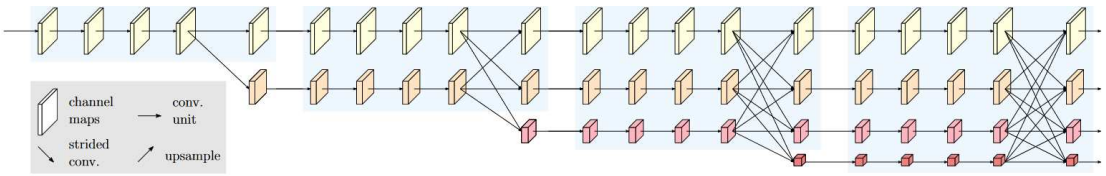


Figure 37: High Resolution Network

6.5.5 OCR

OCR is a method created to generate object-contextual representations for pixels by utilizing the representation of the corresponding object class. This approach involves three steps:

- The contextual pixels are separated into various soft object regions corresponding to different classes. This can be achieved through coarse soft segmentation that is computed using a deep network like ResNet or HRNet. The division is learned with the guidance of ground-truth segmentation.
- Estimate the representation for each object region by aggregating the representations of the pixels in the corresponding object region.
- Each pixel's representation is enhanced with object-contextual representation (OCR), which is a weighted aggregation of all object region representations based on pixel-object region relations.

OCR differs from other contextual methods because, exploiting object regions as structures for the contextual pixels, it can differentiate between the same-object-class contextual pixels and the different-object-class contextual ones. (Figure 19)

Figure 38: Illustrating the multi-scale context with the ASPP as an example and the OCR context for the pixel marked with ■. (a) ASPP: The context is a set of sparsely sampled pixels marked with ■, ■. The pixels with different colors correspond to different dilation rates. Those pixels are distributed in both the object region and the background region. (b) Our OCR: The context is expected to be a set of pixels lying in the object (marked with color blue). The image is chosen from ADE20K.

6.5.6 HRNet-OCR

The introduction of the OCR in the final stage of the HRNet enhances the network's ability to focus on essential image regions while maintaining a holistic understanding of the entire image. This is achieved by integrating the previously described attention mechanisms into the network architecture, which allocates more resources to relevant regions and suppresses irrelevant or redundant information. The attention module is beneficial in accurately recognizing fine details and small objects in high-resolution images. This makes the overall architecture more powerful. Here are shown some key aspects of the network:

- *Multi-Scale Feature Fusion*: HRNet is well-known for its capability to uphold various scales of information across the network. The attention module is vital in merging features from different resolutions or scales. It guarantees that details from low-resolution features are correctly merged with high-resolution features, enabling the network to capture both global context and fine-grained details.

- *Spatial and Channel Attention*: the attention module can incorporate both spatial and channel-wise attention to enhance performance. Spatial attention emphasizes the spatial relationships between pixels or locations in an image, while channel attention focuses on the relationships between feature channels. These attention mechanisms assist the network in highlighting important spatial regions and channels while suppressing noise.

- *Adaptability*: can be customized to suit different tasks and requirements. It has the flexibility to use different attention mechanisms, like self-attention or non-local attention, to capture long-range dependencies in the image; it can be designed in various ways.

- *Improved Performance*: the OCR is used to improve performance in tasks that demand accurate localization and segmentation, such as object detection and semantic segmentation. It effectively manages objects of varying sizes and scales, making it ideal for numerous computer vision applications. figure

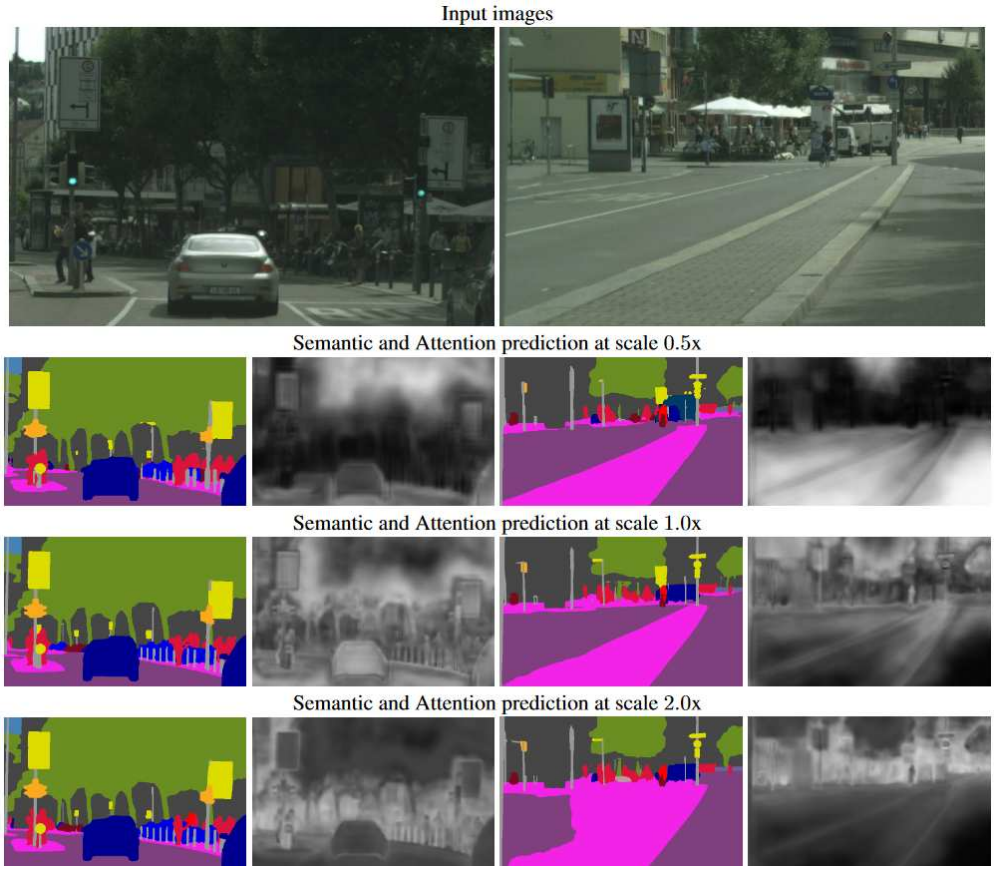


Figure 39: Two different predictions, semantic and attention ones, made at two different scale levels. One scene displays a problem with fine details, while the other scene illustrates a problem with large region segmentation. High attention values are represented by a white color, with the attention values for each pixel summing up to 1.0 across all scales. On the left side, the thin posts on the roadside are best resolved at a 2x scale, and the attention effectively prioritizes that scale compared to others. This is evident in the white color for the posts in the 2x attention image. On the right side, the large road/divider region is most accurately predicted at a 0.5x scale, with the attention focusing primarily on that scale for that region.

7 EXPERIMENTS

We started the training process once we had obtained the training, test, and validation sets. The training set was based on 250 images for each material, while test and validation sets comprised 40 images for each category. This section will provide a comprehensive outline of our implementation process. We trained our models using PyTorch on UIC servers with distributed data-parallel training and synchronous batch normalization. Stochastic Gradient Descent (SGD) was used for training with a batch size of 1 per GPU, a momentum of 0.9, and a weight decay of 0.0005. We applied the "polynomial" learning rate policy. The primary loss function was Cross Entropy Loss for the first experiment and RMI (Table 4) loss for the second one, both with default settings. As an optimizer, ADAM and RADAM were implemented for the first and the second training. For the Facade dataset, we trained for 175 epochs on 2 DGX nodes with a poly exponent of 2.0, an initial learning rate of 0.01. To augment the dataset during training, we used various techniques such as Gaussian blur, color augmentation, random horizontal flip, and random scaling (0.5x - 2.0x) on the input images. We used a crop size of 2048x1024. The best results are obtained with RMI Loss and RADAM Optimizer and are shown in the following section.

7.1 RMI Loss

In Semantic segmentation, a crucial computer vision problem that is commonly approached is that most segmentation models use a pixel-wise loss as their optimization criterion; this approach fails to account for the dependencies between pixels in an image. To address this issue, researchers have explored methods such as conditional random fields (CRF) and pixel affinity-based techniques, which

require additional model branches, memory, or inference time. The region mutual information (RMI) loss models pixel dependencies more simply and efficiently. Unlike the pixel-wise loss, which treats pixels as independent samples, RMI uses one pixel and its neighboring pixels to represent that pixel. For each pixel in an image, is obtained a multi-dimensional point that encodes the relationship between pixels. This results in the image being represented as a multi-dimensional distribution of these high-dimensional points. Maximizing the mutual information (MI) between the prediction and ground truth's multi-dimensional distributions makes achieving high-order consistency possible. RMI only requires a few extra computational resources during the training stage and does not impose any overhead during testing. Therefore, RMI offers a more efficient and effective means of modeling pixel dependencies in semantic segmentation.

7.2 RADAM Optimizer

The learning rate warmup heuristic successfully stabilizes training, accelerates convergence, and improves generalization for adaptive stochastic optimization algorithms like RMSprop and Adam. Adaptive learning rate has a problematically large variance in the early stage; RAdam was proposed to solve this problem by introducing a term to rectify the variance of the adaptive learning rate. Extensive experimental results on image classification, language modeling, and neural machine translation verify this aspect and demonstrate the effectiveness and robustness of the selected optimizer. (Figure 39)

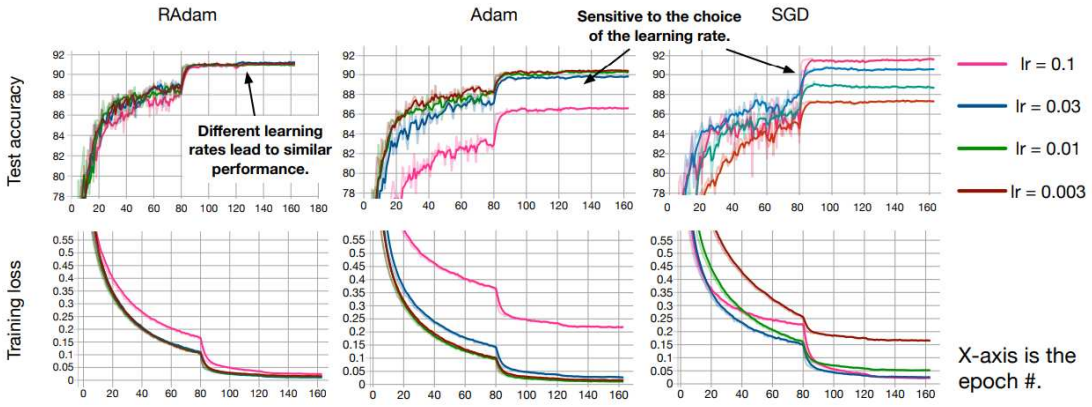


Figure 40: RADAM optimizer comparison

7.3 Results

This section shows the results of our best training made with the implementation of RMI Loss and RAdam optimizer on the Hierarchical multi-scale attention backbone and the training set produced in the previous stages of our framework. Table 5 highlights the good performances of our network, with a mIoU of 86.76 and an average classification accuracy of 96.05 over the entire validation set.

Labels	**IoU**
Background	95.37
Glass	80.05
Concrete	88.40
Vynil	90.48
Wood shake	70.23
Brick	86.75
Asbestos	83.43
Asphalt	82.84
Stucco	92.06
Windows/Doors	86.75
Roofs	92.96
mIoU	**86.76**

(a) Semantic Segmentation IoU

Labels	**Classification**
Glass	96.00
Concrete	94.00
Vynil	100.00
Wood shake	93.70
Brick	100
Asbestos	95
Asphalt	91.00
Stucco	100
Averall accuracy	**96.05**

(b) Facades classification results

Table 5: Best training results

During the training, we obtained the confusion matrix which excluded the background pixels. The purpose was to concentrate on the incorrect predictions made by the network regarding various materials. Figure 40 indicates that there is no significant discrepancy in the recognition of materials. This is a positive outcome that helps us achieve our end goal of reaching the material distribution of materials within a city. Figure 42 displays the results for each material to help readers understand network performance.

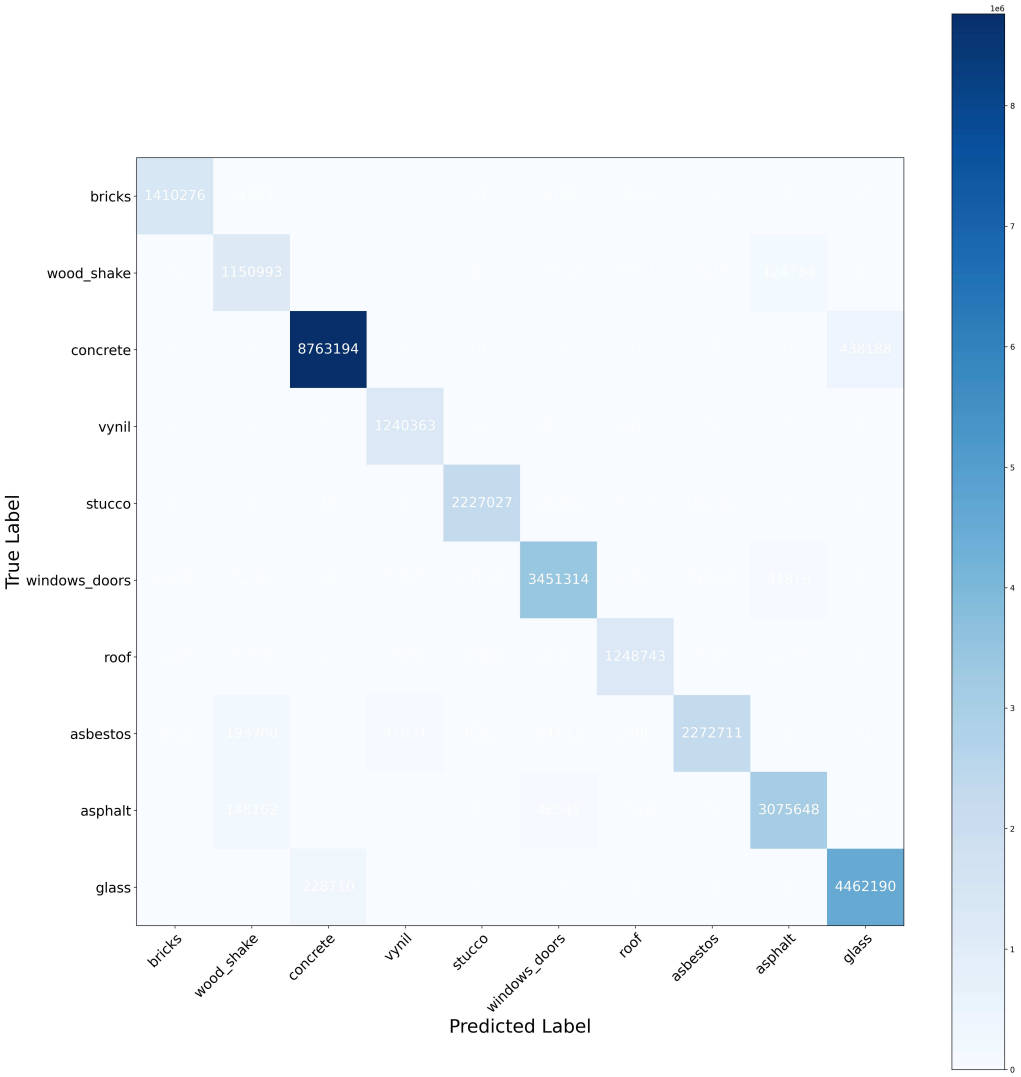


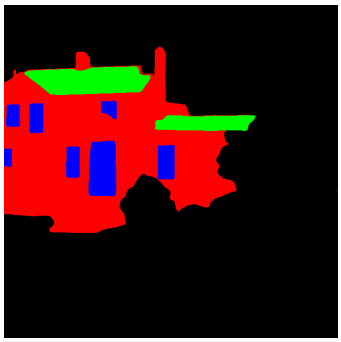
Figure 41: Segmentation confusion matrix

(a) Asbestos-image

(b) Prediction

(c) Image and prediction

(d) Brick-image



(e) Predictione

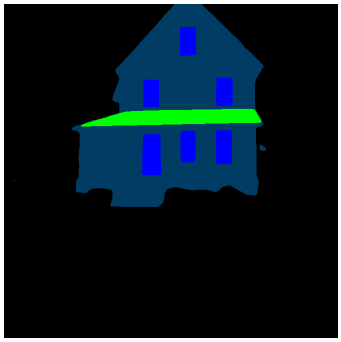
(f) Image and prediction

(g) Wood shake-image

(h) Prediction

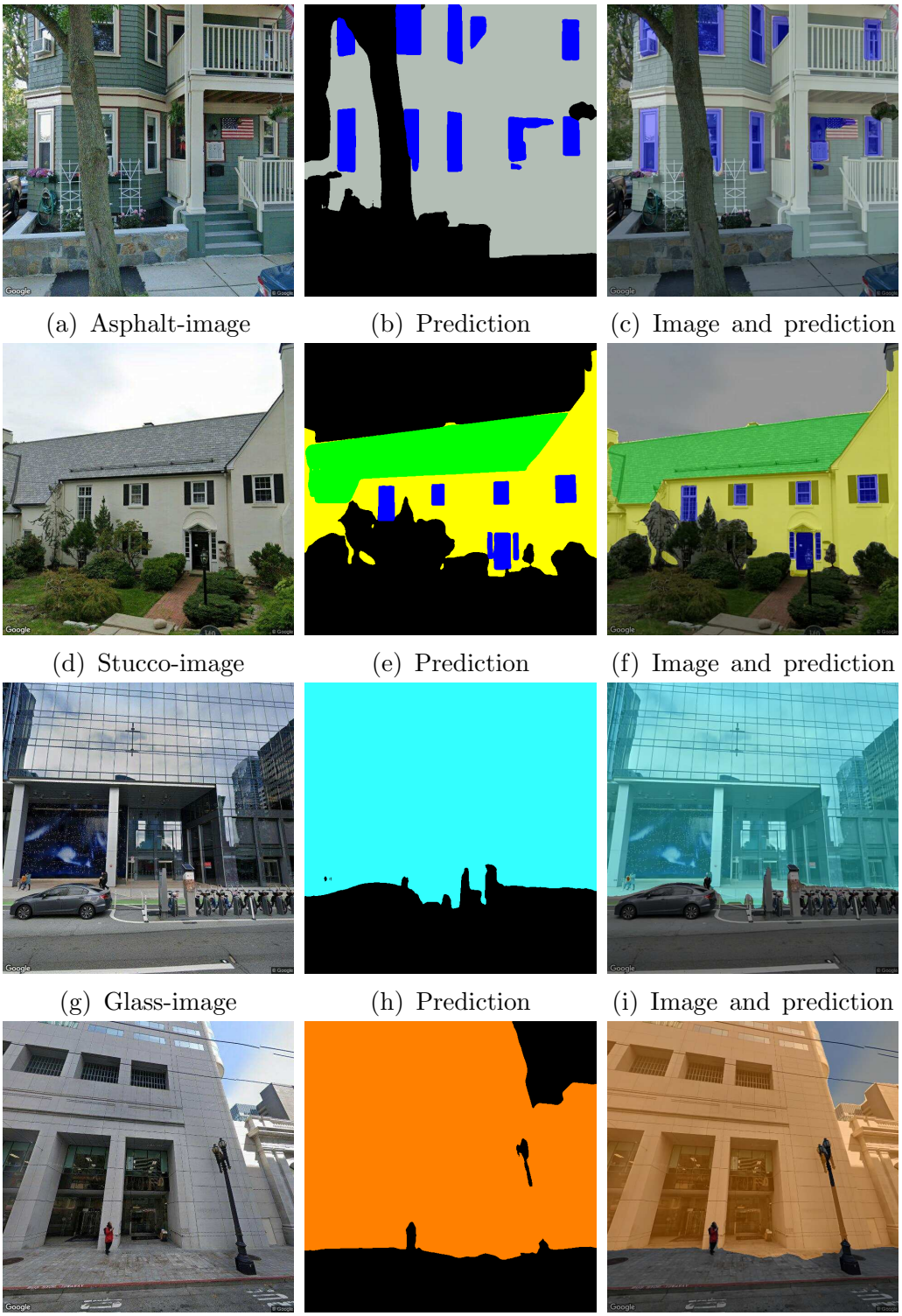
(i) Image and prediction

(j) Vinyl-image



(k) Prediction

(l) Image and prediction



(a) Asphalt-image (b) Prediction (c) Image and prediction

(d) Stucco-image (e) Prediction (f) Image and prediction

(g) Glass-image (h) Prediction (i) Image and prediction

(j) Concrete-image (k) Prediction (l) Image and prediction

Figure 43: Validation segmentation results examples

7.4 Distribution of materials in a city

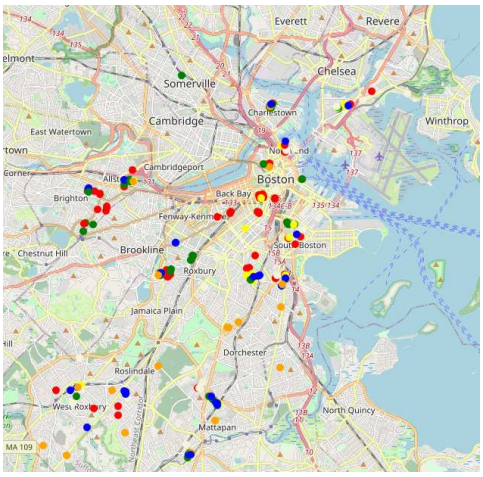
With the help of the trained neural network, it is now possible to identify and track the different buildings in a city based on the materials used in their facades by analyzing the corresponding image dataset. This analysis can then generate a map of the city, highlighting the distribution of these different materials.

Two types of maps can be generated using this information:

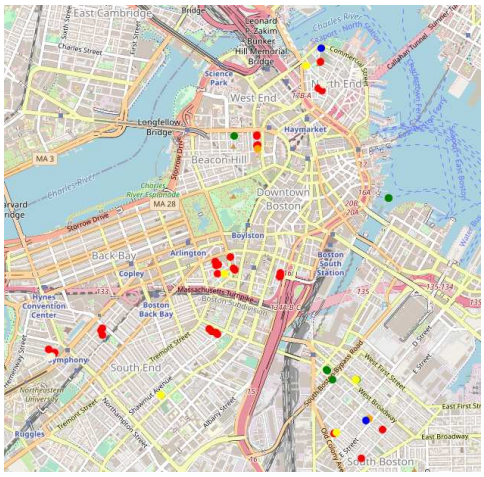
- *Interactive map*: using the Folium library users can zoom in and out of specific areas of interest, providing a detailed exploration of the distribution of buildings in that area. This feature is particularly useful for researchers and urban planners looking to study specific city regions in detail. (Figure 43)

- *Static map*: can be created using geopandas, which provides an overview of the entire city and its building material distribution. This map can help identify patterns or trends that may exist in the distribution of materials across different areas, offering a broader perspective of the city. (Figure 18)

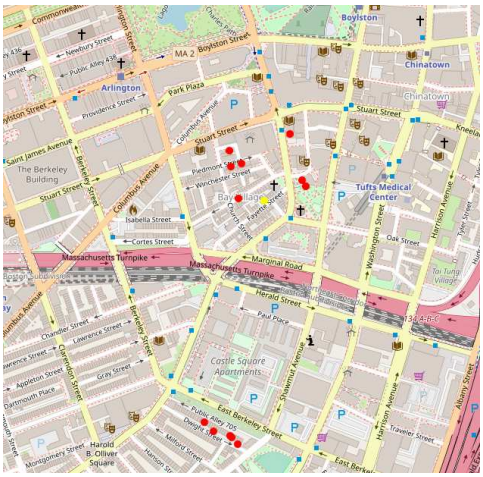
In summary, by utilizing a trained neural network and geographical information, it is possible to create detailed maps of a city's building material distribution, which can be used for various purposes, from urban planning to research and analysis.



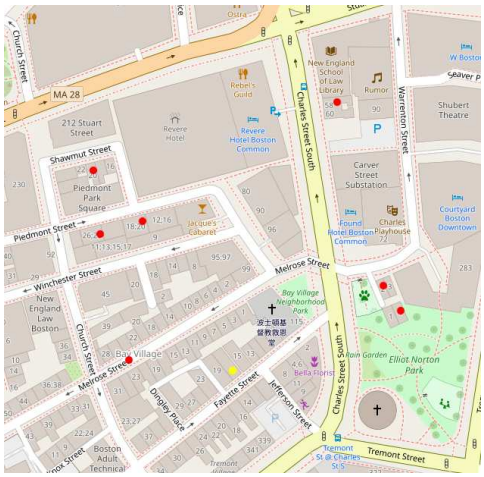
(a) Zoom1



(b) Zoom2



(c) Zoom3



(d) Zoom4

Figure 44: Different levels of zoom of a specific area in Folium map

8 EVALUATION RESULTS

8.1 Testing on Wake County

We conducted a study to assess the accuracy of the architecture in various cities. To achieve this, we used a property assessment database from Wake County in North Carolina. Although this county has several cities, we only evaluated the most significant ones, including Raleigh, Cary, Apex, and Fuquay Varina. The dataset we used contained similar information to the Boston dataset utilized during the training phase, which included data on the location and exterior materials of buildings in the area. The process of obtaining images through the Google Street View API was the same as what was done during the training phase. We cleaned the images of outliers before feeding them into the network. Table 6 below shows the results obtained from the images. The main issue here is that the dataset considered did not contain all materials recognizable by the architecture, but only five of them: Glass, Concrete, Brick, Vinyl, and Stucco. Anyway the results show also in this case a good average classification accuracy, reflecting the outcome obtained during the validation phase.

Material	**Accuracy**
Concrete	84
Glass	86
Brick	91
Stucco	87
Vinyl	83
Average Accuracy	86.2

Table 6: Classification accuracy in Wake County dataset

(a) Brick building segmentation

(b) Vinyl building segmentation

(c) Stucco building segmentation

(d) Concrete building segmentation

(e) Glass building segmentation

Figure 45: Examples of segmentation on Wake County images for five different materials

8.2 Chicago material distribution

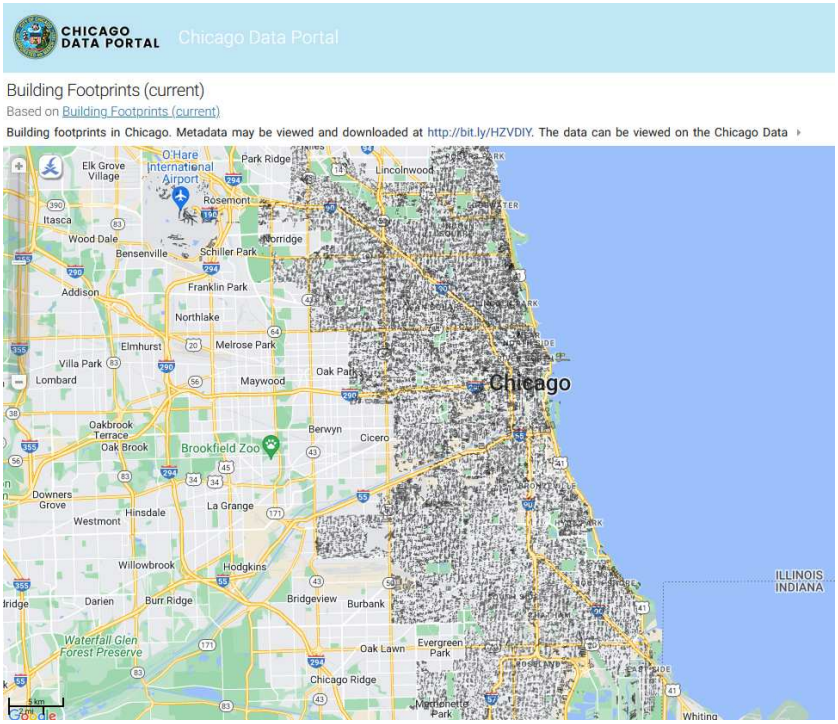


Figure 46: Chicago building distribution

In this section, we will be showcasing the outcomes of our architecture in the city of Chicago. Initially, we utilized the Building Footprints database available on the Chicago Data Portal, which contains geospatial data about the outlines or footprints of buildings in Chicago. When working with building footprint datasets, it is important to keep in mind several common attributes. Geometry refers to the shape of the building footprint, usually represented as polygons or multipolygons. The building ID or Identifier is a unique identifier for each building within the dataset, while the address provides the street address and any additional location details of the building. We used this information to create a dataset of 100 images for each single neighborhood in the Building Footprints database. These images were then fed into BuildingSurfaces to classify each building, resulting in a CSV file where each address is associated with a facade material, as shown in the

picture. The final step was to plot the results on the map of Chicago, resulting in the picture shown in Figure 44. Upon analyzing the picture, it is possible to notice a greater number of concrete and glass constructions in areas near the urban part of the city that are rich in high buildings. Conversely, brick constructions are more prevalent in areas further away from downtown, indicating that high buildings are no longer present, and independent houses are the norm. For the remaining materials, the distribution is quite uniform in all the city. The pie chart in Figure 45 illustrates the distribution of materials. Brick accounts for the majority of the materials at 46%, while stucco, concrete, and asphalt each account for around 10%.

It is also possible to evaluate at a different aggregation level. For instance, if the objective is to analyze the distribution in a particular neighborhood. Figure 46 presents the outcomes for four significant Chicago neighborhoods: West Loop, Lake View, Wicker Park, and River North.

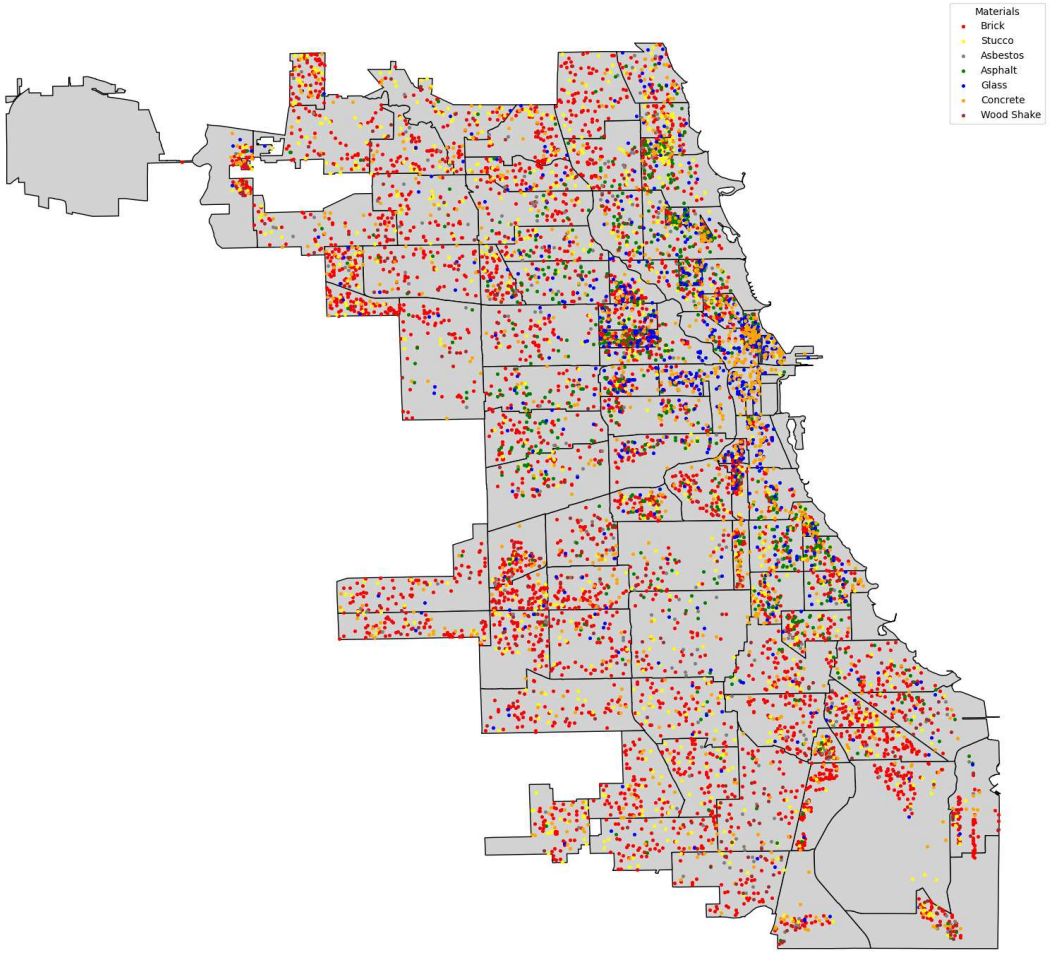


Figure 47: Chicago buildings material distribution

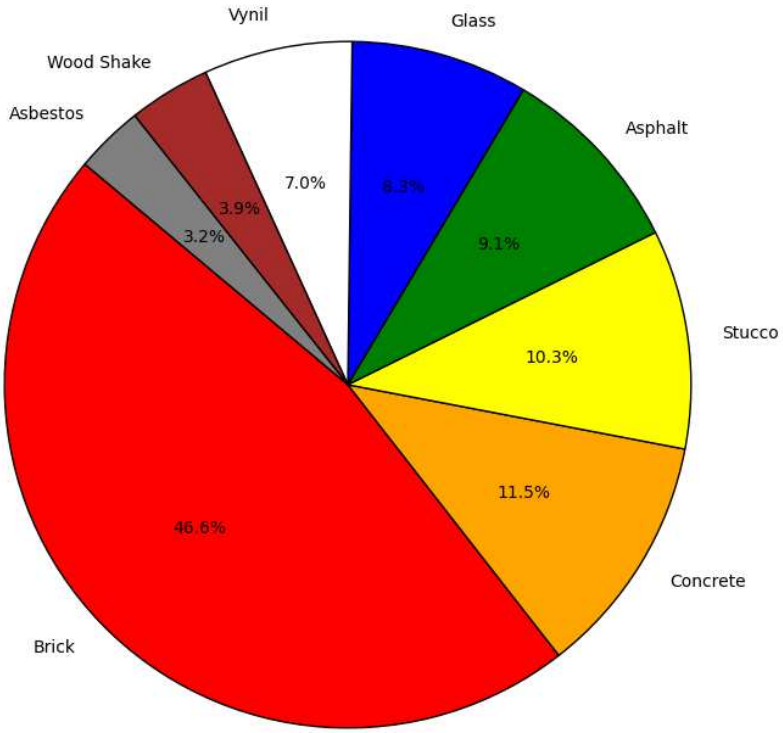
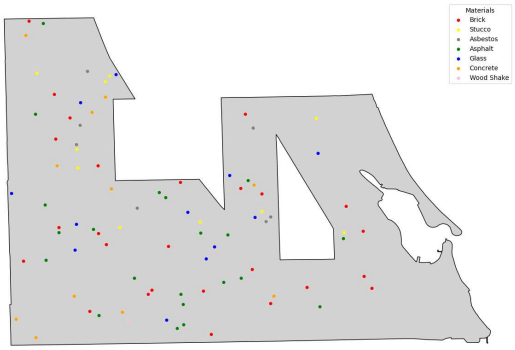
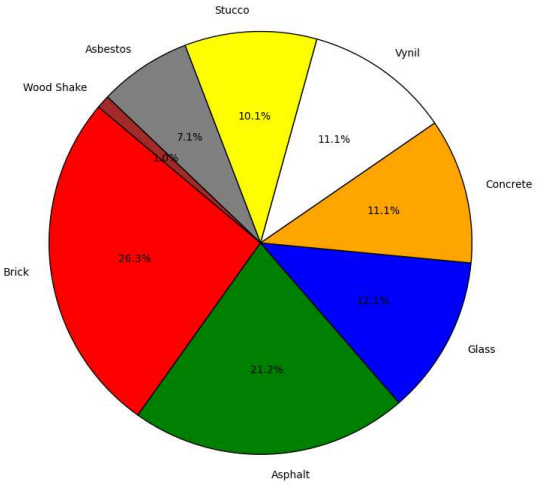


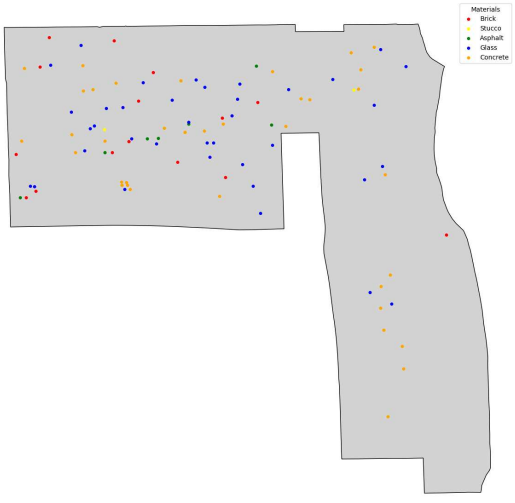
Figure 48: Pie chart on Chicago buildings materials



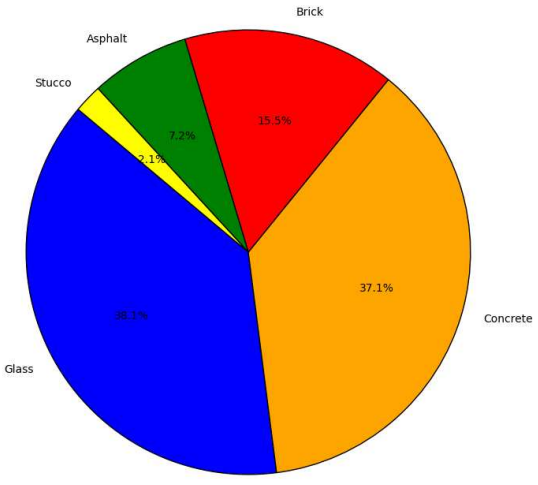
(a) Lake View distribution



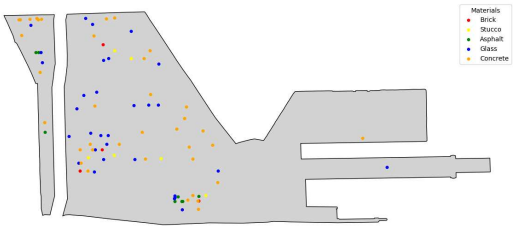
(b) Lake View pie-chart



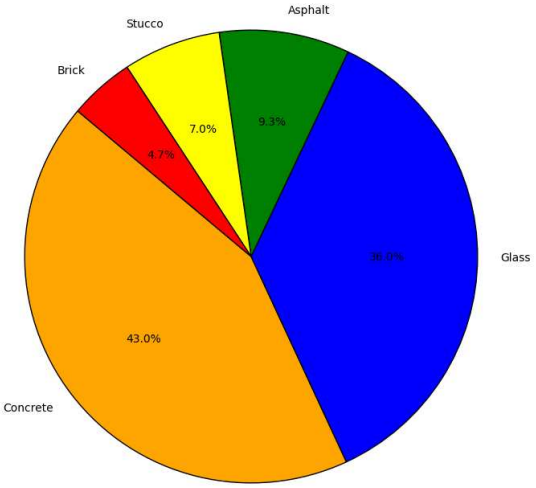
(c) West Loop distribution



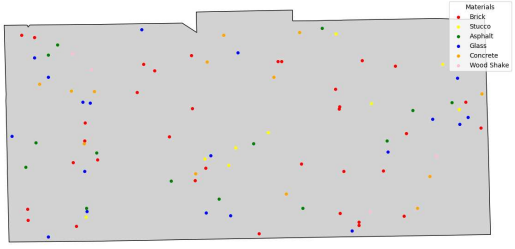
(d) West Loop pie-chart



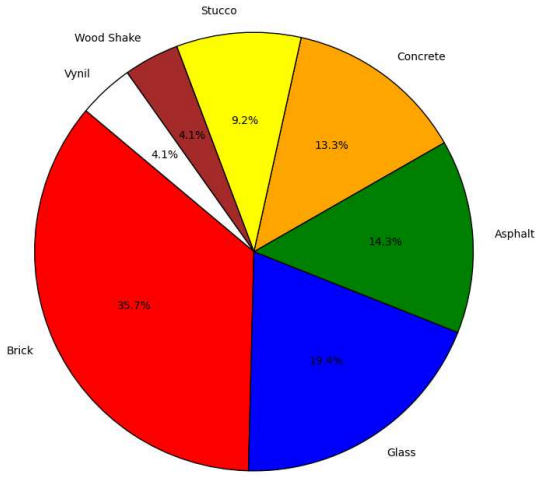
(a) Streeterville distribution



(b) Streeterville pie-chart



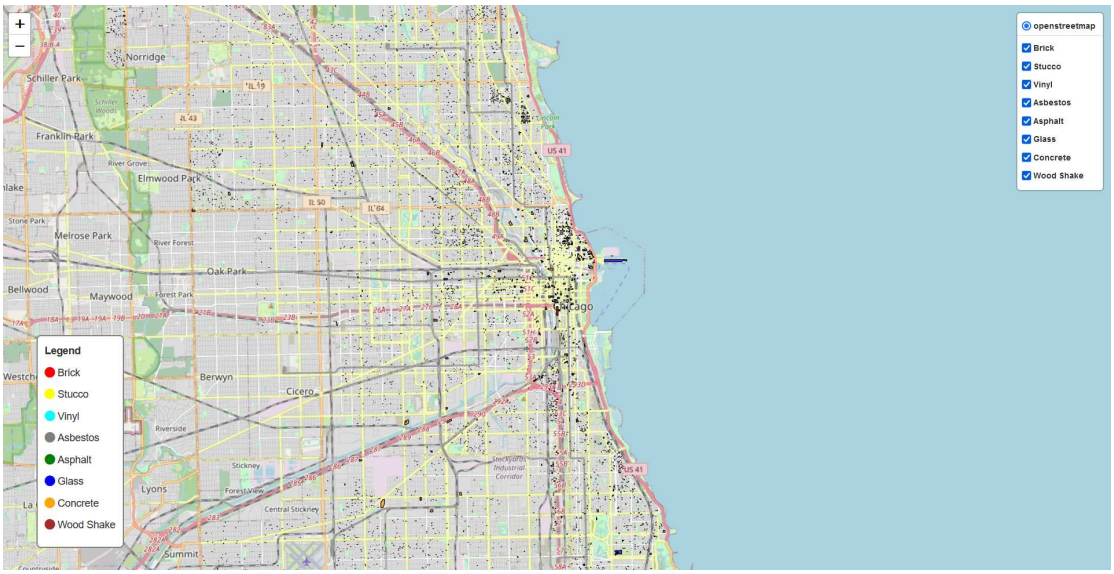
(c) Wicker Park distribution



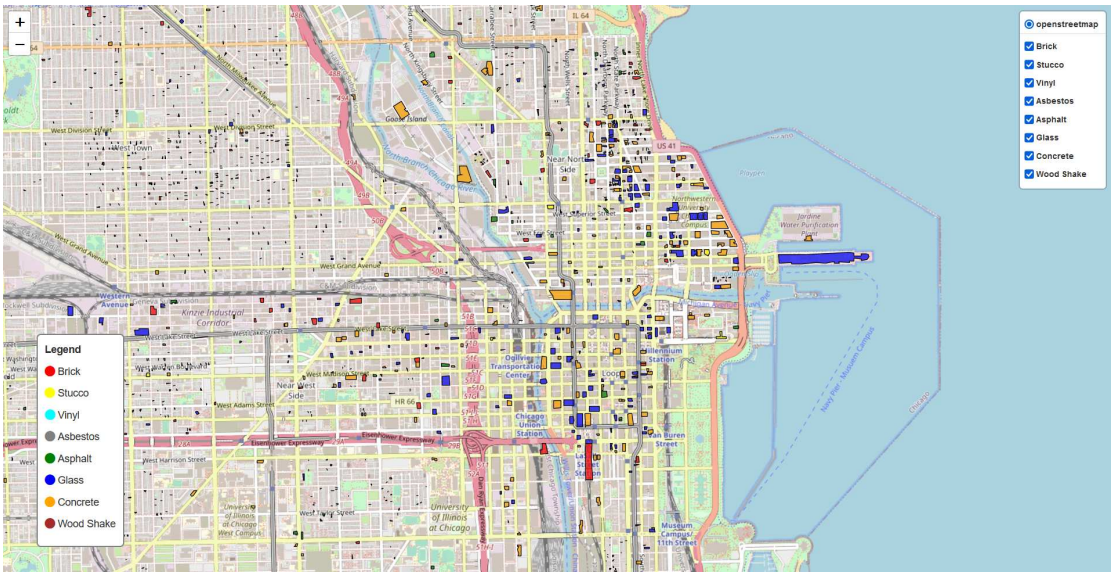
(d) Wicker Park pie-chart

Figure 50: Material distribution in some of the most important Chicago neighbours

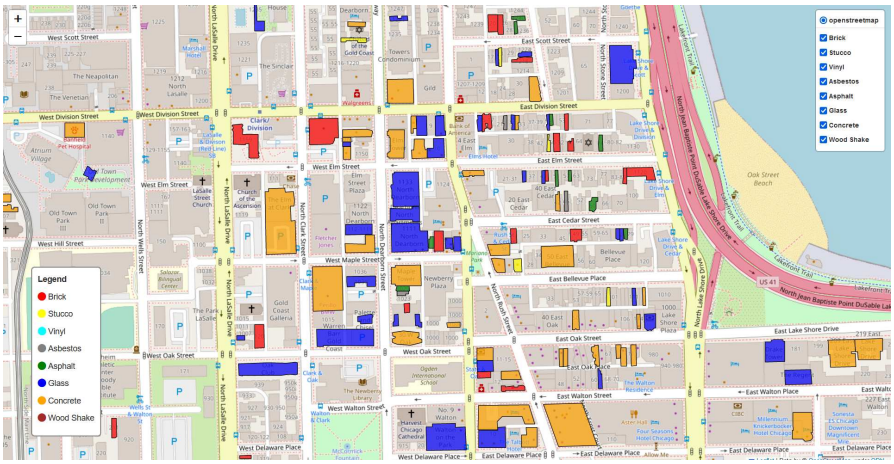
It is possible to create an interactive map using the Folium library, which allows for detailed analysis. You can zoom in and out of the map to focus on specific areas of interest. Filter operations can be applied in the top right corner to select one or more materials and hide the rest. On the left side, a legend displays the colors of the materials used to color the buildings. You can also see its address by hovering over a building with the mouse. Figure 48 shows some examples.



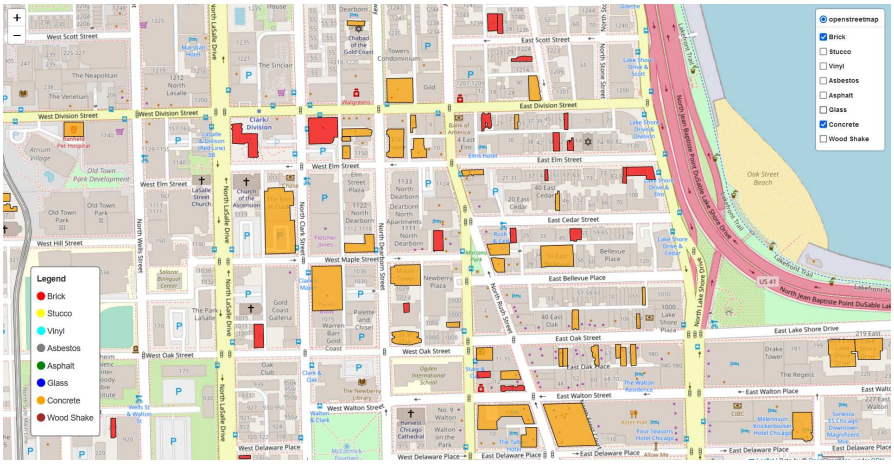
(a) Folium map zoom 0



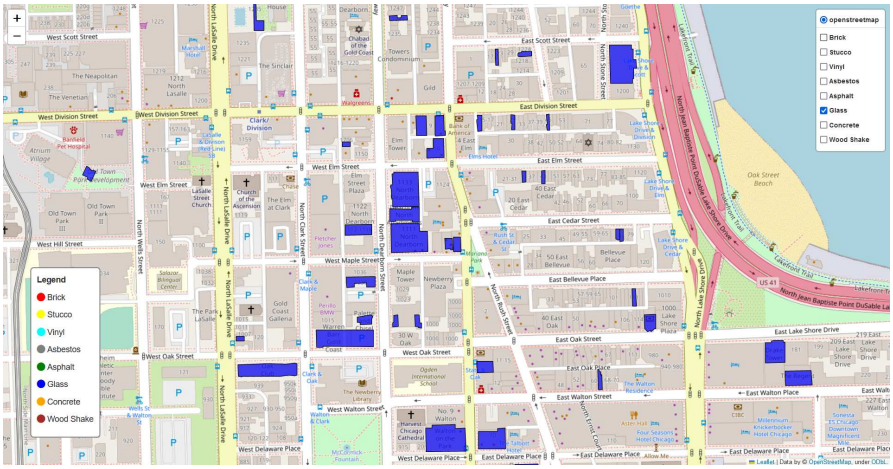
(b) Folium map zoom 1



(a) Folium map with all materials



(b) Folium map with brick and concrete buildings



(c) Folium map with glass buildings

Figure 52: Examples of Folium map application

9 FUTURE WORK

The results yielded by the architecture are fascinating and have the potential to map material distribution in numerous cities. Further research could concentrate on integrating a wider range of materials to cover a greater number of countries that vary in their usage of building materials and architectural styles. Additionally, creating more intricate labels that showcase not just the primary material, but also the other materials present in the buildings could lead to more precise and effective evaluations. The current system oversimplifies materials by categorizing them as "brick", "wood", or "concrete". A more sophisticated approach would involve creating a taxonomy that distinguishes between different types and qualities of these materials. For example, identifying different varieties of stone, such as granite, limestone, and sandstone, or different brick patterns, such as bond patterns, would yield more precise and valuable information. It would be beneficial to enhance the model's capabilities to identify materials and deduce the purpose and utilization of buildings. This would enable identifying whether a building is intended for residential, commercial, or industrial use, which is crucial for urban planning and policy-making. It is recommended to incorporate sustainability metrics into the analysis to further sustainable urban planning and construction practices. An ecological footprint classification system can be developed to assess the environmental impact of various building materials, allowing for the classification of buildings based on their ecological footprint. We plan to integrate BuildingSurfaces with other data sources, such as shadows [39], and investigate their use for pedestrian comfort. Moreover, we plan to integrate BuildingSurfaces with urban frameworks [40], such as the Urban Toolkit [41], A high-level grammar designed for common urban visualizations, enabling easy web-based authoring with flexibility and extensibility.

10 CONCLUSIONS

This thesis argues that the excellent performance of AI architecture can provide valuable insights into material distribution within cities. The insights gained can be used by politicians to make environmentally friendly decisions, resulting in a more sustainable and green approach to urban planning and resource allocation.

The following key points support this thesis:

- **Data-Driven Decision Making**: The AI architecture can analyze vast amounts of data related to material distribution in cities. This data-driven approach enables a more comprehensive understanding of how resources are allocated and utilized.
- **Efficiency and Resource Optimization**: AI can optimize the distribution of materials and resources, leading to reduced waste, improved efficiency, and cost savings. Politicians can use these insights to support policies and initiatives that promote resource efficiency.
- **Environmental Impact Reduction**: Making environmentally friendly decisions based on AI-driven insights can reduce a city's environmental impact. This may include reducing emissions from transportation, minimizing resource depletion, and promoting sustainable practices.
- **Long-Term Sustainability**: Politicians and city planners can use the AI architecture's outcomes to develop strategies that promote long-term sustainability. This may involve investments in renewable energy, sustainable transportation infrastructure, and waste reduction initiatives.

In conclusion, the thesis argues that the AI architecture can provide valuable insights into material distribution, empowering politicians to make environmentally friendly decisions. This can lead to a more sustainable and green approach

to urban planning and resource management, bringing tangible benefits in terms of efficiency, environmental impact reduction, and long-term sustainability.

References

[1] European Environment Agency (EEA). *Urban sustainability issues—what is a resource-efficient city?* 2015.

[2] L. Sun, J. Chen, Q. Li, and D. Huang. Dramatic uneven urbanization of large cities throughout the world in recent decades. 2020.

[3] W.R. Stahel. The circular economy. *Nature*, 531:435–438, 2016.

[4] B.C. Guerra and F. Leite. Circular economy in the construction industry: an overview of united states stakeholders' awareness, major challenges, and enablers. *Resources, Conservation and Recycling*, 170:Article 105617, 2021.

[5] Leilei Zhang, Guoxin Wang, and Weijian Sun. Automatic identification of building structure types using unmanned aerial vehicle oblique images and deep learning considering facade prior knowledge. *[Insert Journal Name]*, 2021.

[6] Khadraoui Mohamed Amine, Leila Sriti, and Besbas Soumaya. The impact of facade materials on the thermal comfort and energy efficiency of office buildings. *Journal of Building Materials and Structures*, 5(1):55–64, 2018.

[7] Jane Jacobs. *The Death and Life of Great American Cities*. Random House, 1961.

[8] Jane Jacobs. *The Death and Life of Great American Cities*. Random House, 1961.

[9] Construction and building materials.

[10] Federal Emergency Management Agency (FEMA). National response framework, 2017.

[11] D.A. Larson and K.A. Golden. Entering the brave, new world: an introduction to contracting for building information modeling. *Wm. Mitchell L. Rev.*, 34:75, 2007.

[12] S.M. Sepasgozar, S. Shirowzhan, et al. Challenges and opportunities for implementation of laser scanners in building construction. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, page 1. IAARC Publications, 2016.

[13] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. Google street view: Capturing the world at street level. *Computer*, 43:32–38, 2010.

[14] M. Yaman. Different façade types and building integration in energy efficient building design strategies. *International Journal of Built Environment and Sustainability*, 8(2):49–61, 2021.

[15] B. Abediniangerabi, M. Shahandashti, and A. Makhmalbaf. Coupled transient heat and moisture transfer investigation of facade panel connections. *Journal of Engineering, Design and Technology*, 19(3):758–777, 2020.

[16] A.C. Guo and Z. Liu. A new method for energy efficiency design of building facade and its thermodynamic evaluation. *International Journal of Heat and Technology*, 38(4):903–913, 2020.

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.

[19] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2481–2490, 2017.

[20] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[21] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *International Conference on Learning Representations (ICLR)*, 2016.

[22] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[23] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[24] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(4):834–848, 2018.

[25] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[26] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L. Yuille. Attention to scale: Scale-aware semantic image segmentation. *arXiv preprint arXiv:1511.03339*, 2015.

[27] Shiqi Yang and Gang Peng. Attention to refine through multi scales for semantic segmentation. In *Pacific Rim Conference on Multimedia*, pages 232–241. Springer, 2018.

[28] Catanzaro Tao, Sapra. Hierarchical multi-scale attention for semantic segmentation. *arXiv preprint arXiv:1511.03339*, 2020.

[29] Jeffrey Delmerico, Pierre David, and Jason Corso. Building facade detection, segmentation, and parameter estimation for mobile robot localization and guidance. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[30] G. Sezen, M. Cakir, M. Atik, and Z. Duran. Deep learning-based door and window detection from building façade. *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, XLIII-B4-2022:315–320, 2022.

[31] Yu Wang, Xin Hu, Tao Zhou, Ying Ma, and Zhen Li. Efficient building facade structure extraction method using image-based laser point cloud. *Transactions in GIS*, 27(4):1145–1163, 2023.

[32] Xinyu Zhuo, Matthias Mönks, Thomas Esch, and Peter Reinartz. Facade segmentation from oblique uav imagery. In *Proceedings of the Joint Urban Remote Sensing Event (JURSE)*, 2019.

[33] Jiaqi Zhang, Tomohiro Fukuda, and Nobuyoshi Yabuki. Development of a city-scale approach for façade color measurement with building functional classification using deep learning and street view images. *ISPRS International Journal of Geo-Information*, 10(8):551, 2021.

[34] H. Zheng, L. Fang, M. Ji, M. Strese, Y. Ozer, and E. Steinbach. Deep learning for surface material classification using haptic and visual information. *IEEE Transactions on Multimedia*, 18(12):2407–2416, 2016.

[35] Maryam Hosseini, Fabio Miranda, Jianzhe Lin, and Claudio Silva. CitySurfaces: City-scale semantic segmentation of sidewalk materials. *Sustainable Cities and Society*, 79:103630, April 2022.

[36] ... ... In *2010 Residential Buildings Energy Efficiency Meeting*, Denver, Colorado, July 20–22 2010.

[37] Alessandro Acquisti and Ralph Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. *Privacy Enhancing Technologies*, pages 36–58, 2006.

[38] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1, 2020.

[39] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Luc Wilson, Mondrian Hsieh, and Cláudio T. Silva. Shadow Accrual Maps: Efficient accumulation of city-scale shadows over time. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1559–1574, 2019.

[40] Harish Doraiswamy, Juliana Freire, Marcos Lage, Fabio Miranda, and Claudio Silva. Spatio-temporal urban data analysis: A visual analytics perspective. *IEEE Computer Graphics and Applications*, 38(5):26–35, 2018.

[41] Gustavo Moreira, Maryam Hosseini, Md Nafiul Alam Nipu, Marcos Lage, Nivan Ferreira, and Fabio Miranda. The Urban Toolkit: A grammar-based framework for urban visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 2024. Accepted, to appear.