ESTIMATION OF COMPLETE SOLAR POTENTIAL IN URBAN AREAS BASED ON THREE-DIMENSIONAL BUILDING FAÇADE RECOGNITION

FAN XU

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

The Department of Land Surveying and Geo-Informatics

Estimation of complete solar potential in urban areas based on three-dimensional building façade recognition

Fan XU

A thesis submitted in partial fulfillment of

the requirements for the degree of Doctor of Philosophy

January 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ _____ (Signed)

_____XU Fan_____ (Name of student)

Abstract

Solar photovoltaic (PV) harvesting is a significant force leading to the rapid expansion of renewable energy. To facilitate the installation of PV modules at solar-abundant locations, an accurate estimation of solar PV spatial potential is indispensable. Solar energy could be reflected on high-albedo building surfaces inside the urban canyon. However, using constant albedos to represent the urban vertical surfaces or ignoring the indirect components in estimating received irradiation is the typical solution in current research, which leads to inaccuracies in final results. Using conventional ways to construct albedo datasets for different building surfaces is extremely labor-intensive.

In this study, we address these challenges by proposing a novel framework that integrates façade material identification using street-view images. This framework incorporates the effects of multi-reflection, enabling both qualitative and quantitative analysis of the impact of façade albedo on solar energy distribution. To achieve this, we built a façade material dataset from street views and developed an segmentation model to effectively identify façade materials from street view images. Furthermore, this study provides the first accurate estimation of solar energy potential in complex metropolitan environments and elucidates how metropolitan environments with different albedo characteristics affect solar potential distribution.

Due to the distinguishable features between materials in terms of the subtle texture and patterns rather than just their shapes and colors, identification requires more details from images, which makes a multi-scale inference structure a promising solution. Compared with existing methods combining scale features at the pixel level, we proposed a novel Multi-Scale Contextual Attention Network (MSCA) using a Multi-Scale Object-Contextual Representation (OCR) block to exploit and combine contextual information from different scales in high dimensional layers. The experimental results show that the proposed model significantly outperforms the existing models, achieving a mean Intersection over Union

(mIOU) of 70.23%. The results indicate that the MSCA can effectively obtain the materials information from street views and can be a reliable solution to providing urban albedo information for solar estimation.

The segmentation results of the façade materials are further projected onto a 3D GIS model, which allows precise albedo values to be assigned to each urban surface. This enables the accurate simulation of solar potential, incorporating both direct and reflected solar radiation, as well as capturing the complex multi-reflection effects occurring in dense urban environments. By simulating how solar radiation interacts with various building surfaces, we provide a realistic estimation of solar potential distribution and comprehensively discuss the effects that the sophisticated albedo environment might bring to the solar potential. The experimental results show that the discrepancies in albedo significantly affect the overall solar potential by 8.0% to 9.1%. If multiple reflections among buildings are disregarded, the impact can reach 11.9% to 17.8%.

The findings of this study offer valuable insights for urban planning by providing a scalable method for more precise solar potential assessment. The integration of real-world material data with 3D GIS enhances the decision-making process for optimizing photovoltaic (PV) deployment in urban areas, thereby contributing to more sustainable urban energy planning and efficient use of renewable resources.

Publications Arising from the Thesis

Xu, F., Wong, M.S., Zhu, R., Heo, J., Shi, G., 2023. Semantic segmentation of urban building surface materials using multi-scale contextual attention network. *ISPRS Journal of Photogrammetry and Remote Sensing* 202, 158–168.

Xu, F., M.S., Wong, 2024. Solar Cities: multiple-reflection within urban canyons. (under review)

Xu, F., M.S., Wong, 2025. Evaluation of the importance of urban morphological parameters in albedo-based three-dimensional solar potential estimation (ready to submit).

Acknowledgements

I am deeply grateful for the opportunity to pursue my Ph.D. at The Hong Kong Polytechnic University. This journey has been an unforgettable experience, and I owe a great deal of my success to the incredible support and guidance I have received along the way.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Wong. His unwavering support, insightful guidance, and belief in my potential have been instrumental in my academic journey. Professor Wong has not only provided me with invaluable advice and mentorship but has also generously supported my research financially, enabling me to conduct experiments that were crucial to my thesis. His dedication and encouragement have been a constant source of motivation for me.

I would also like to extend my sincere thanks to my co-supervisor, Dr. Zhu. His expertise and constructive feedback have significantly shaped my doctoral thesis. Dr. Zhu's meticulous attention to detail and his willingness to help at every stage of my research have been immensely beneficial.

I am grateful to my fellow lab members and colleagues, Yuan Meng, Coco Yin Tung Kwok, Jing Li, Keru Lu, Xuan Liao, Songyang Li, Shaolin Wu, Meilian Wang, Qian Peng, Tony, Xindi, Kang Zou, Ziyi Huang. Their camaraderie, support, and collaborative spirit have made this journey enjoyable and enriching.

Additionally, I would like to express my deepest appreciation to my best friends, Ying Dang, Zhihang XU, Chen Yang, Yezhou HU, Zhiming Li. Their unwavering emotional support, encouragement, and understanding have been a pillar of strength for me throughout this challenging journey. Their constant presence and belief in me have helped me overcome numerous obstacles and stay focused on my goals. I am truly blessed to have such amazing friends who have stood by me through thick and thin. Thank you all for your invaluable contributions to my academic and personal growth.

Table of Contents

List of Figures

List of Tables

List of abbreviation

ANN	Artificial Neural Network
BIPV	Building Integrated Photovoltaics
BRDF	Bidirectional Reflectance Distribution Function
CNN	Convolutional neural networks
CO2	Carbon dioxide
ERF	Effective receptive field
FCNs	Fully Convolutional Networks
FiT	Feed-in Tariff scheme
FLOPS	floating-point operations per second
GIS	Geographic information system
GNSS	Global Navigation Satellite System
GW	Gigawatts
IGDB	International Glazing Database
IMU	Inertial Measurement Unit
JS	Jensen – Shannon
KL	Kullback – Leibler

KSA	Kingdom of Saudi Arabia
LiDAR	Light Detection and Ranging
LoD	Level-of-Detail
MHA	Multi-Head Attention
MINC	Materials in Context Database
mIOU	mean Intersection over Union
MMS	Mobile Mapping Systems
MSA	Multi-Scale Attention
MSCA	Multi-scale contextual attention network
OCR	Object-Contextual Representation
PV	Photovoltaic
RGB	Red Green Blue
RMS	Root Mean Square
RMSE	The root mean square error
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
SWIR	Short-wave infrared
TJ	Terajoules

TRF	Theoretical receptive field
VNIR	Visible and near-infrared
WEKA	Waikato Environment for Knowledge Analysis
3D GIS	Three-dimensional Geographic Information Systems

Chapter 1 Introduction

1.1 Background and Research Gaps

Electricity is the lifeblood of modern societies and economies, providing the energy needed to power homes, businesses, and industries. However, the production of electricity is also a significant contributor to carbon dioxide (CO_2) emissions worldwide, which are a major cause of climate change. In order to mitigate the impact of these emissions, there has been a rapid expansion of renewable energy sources, such as solar, wind power, and hydroelectricity, which is leading the transition to net zero emissions while the world's total renewable electricity capacity is predicted to rise to 4500 gigawatts (GW) in 2024 (IEA, 2023).

1.1.1 Renewable energy in Hong Kong

In this context, solar photovoltaic (PV) systems are becoming increasingly popular in metropolitan cities. These systems use solar panels to convert sunlight into electricity, providing a clean and renewable source of energy. In Hong Kong, for example, the solar energy produced increased from 47 terajoules (TJ) in 2018 to 74 TJ in 2019, representing a significant increase of 57% in just two years (Electrical and Mechanical Services Department, 2021). This growth is part of Hong Kong's broader push towards carbon neutrality by 2050, as outlined in the "Hong Kong's Climate Action Plan 2050". The plan sets a target of reducing Hong Kong's total carbon emissions by 50% before 2035 compared to the 2005 levels, with renewable energy, including solar, expected to play a crucial role in achieving these goals. Although renewable energy currently accounts for less than 1% of the city's total electricity generation, there is a strong focus on increasing its share in the energy mix to between 7.5%-10% by 2030 and to 15% gradually thereafter (Hong Kong Government, 2021).

In addition, another factor driving the growth of solar PV systems is that many governments and organizations are offering incentives and subsidies to encourage the adoption of solar energy, making it an attractive option for both individuals and businesses. Solar PV systems provide a way for individuals and businesses to take action and reduce their carbon footprint while also saving money on their energy bills. In Hong Kong, the government has implemented the Feed-in Tariff (FiT) scheme, which incentivizes individuals and businesses to invest in solar PV systems by allowing them to sell excess electricity generated back to the grid at attractive rates. Under the FiT scheme, participants can earn between HKD 2.5 and HKD 4 per kilowatt-hour of electricity exported, depending on the capacity of the system installed. As of 2021, over 16,000 applications for renewable energy installations, representing a total capacity of around 265MW, mainly solar PV, have been approved, further encouraging the development of solar infrastructure in both residential and commercial sectors (China Light and Power Company, 2021). The increasing popularity of solar PV systems in metropolitan cities is a positive trend that is helping to drive the transition to a cleaner and more sustainable energy future. As the awareness of the need to reduce greenhouse gas emissions grows, it is likely that the use of solar PV systems will continue to expand and play an increasingly important role in the global energy mix.

1.1.2 2D Solar potential estimation

To enhance the efficiency of PV equipment deployments, some studies have estimated the urban PV potentials in different cities (Gassar et al., 2021; Choi et al., 2019). These studies provide an essential basis for energy policy decision-making and panel deployment in metropolitan cities. By understanding the potential for solar energy production in a given city, policymakers and energy companies can make informed decisions about where to install solar panels and how to optimize their use (Zhu et al., 2023). For example, Zhu et al., (2019) estimated the urban PV potential in Hong Kong, while a study by Dehwah et al., (2018) estimated the potential in the residential sector of the Kingdom of Saudi Arabia (KSA).

Izquierdo et al., (2011) build on an existing geo-referenced method for determining available roof area for solar facilities in Spain to produce a quantitative picture of the likely limits of roof-top solar energy. These studies provide valuable information about the amount of solar energy that could be produced in these cities, as well as the potential economic and environmental benefits of deploying solar PV systems. Overall, the estimation of urban PV potentials is an important tool for enhancing the efficiency of PV equipment deployments in metropolitan cities. By providing information about the potential for solar energy production and the challenges and barriers to deployment, these studies can help policymakers and energy companies make informed decisions about the use of solar PV systems and contribute to the transition to a cleaner and more sustainable energy future.

In recent years, there has been a growing interest in improving the efficiency and accuracy of the land surface solar irradiation estimation (Gassar et al., 2021). This is an important task that is essential for understanding the potential for solar energy production in different areas and making informed decisions about the deployment of solar PV systems. In empirical models, multiple meteorological parameters, e.g., temperature, relative humidity, and precipitation, are utilized to describe the long- or short-term distribution of solar potential in large-scale areas (Chen et al., 2019). While these models can provide a general understanding of the solar potential in a given area, they have certain limitations. With the increasing variables, traditional models are not capable of reflecting complex and nonlinear relationships (Ağbulut et al., 2021). To overcome the limitation, machine learning methods have been increasingly used in recent years. These methods are more effective in explaining complex and nonlinear relationships and can provide more accurate estimates of the solar potential in a given area (Meenal et al., 2018; Jiang et al., 2017; Ibrahim et al., 2017). However, these approaches still have certain limitations. One of the main limitations is the lack of incorporation of urban microclimate conditions and the interaction among buildings, such as shadowing and solar multi-reflections. This can lead to inaccurate results, as these factors can significantly affect the amount of solar energy that is available for use (Freitas et al., 2015;

Zhu et al., 2023). Another limitation of these approaches is the neglect of solar irradiation received by urban vertical surfaces, such as building façades. This can vastly underestimate the solar PV potential in cities, as the façades can generate more energy than rooftops due to their larger area, despite the less vertical irradiation and high shadow covering rate. For example, a study by Redweik et al., (2013) found that if the experiments only considered the roof area of the Campus of the University of Lisbon, the total potential of solar energy production would be about 34 GW h/year. However, if the potential of the building façades is also taken into account, the total potential would be almost double, at about 53 GW h/year. This is because, although the average annual irradiation per unit area on the façades is lower than that of the roofs, the much larger area of the façades means that a significant amount of solar energy reaches them throughout the year. In this case, the façades receive about 19 GW h/year of solar energy, which significantly contributes to the campus's total solar potential.

Especially in metropolitan cities, due to the high density and large ratio of façade area/roof area, the urban morphology conclusively determines the solar photovoltaic distribution, which is difficult to incorporate in two-dimensional estimation (Walch et al., 2020; Park et al., 2021; Assouline et al., 2015). However, most studies only consider urban areas as a two-dimensional plane, using satellite images and cloud coverage data to estimate the solar radiation received by rooftops (Walch et al., 2020; Park et al., 2021; Assouline, Mohajeri et al., 2017). This approach neglects the solar irradiation received by urban vertical surfaces, which can significantly affect the solar PV potential in cities. For example, a study by Walch et al., (2020) used Machine Learning to incorporate high spatial resolution building and environmental information in parts of Switzerland but did not consider the solar irradiation received by building façades. Similarly, a study by Park et al., (2021) used satellite images and cloud coverage data to estimate the solar radiation received by rooftops in a dense urban area but did not consider the solar irradiation received by building façades. These studies have omitted multi-reflected solar radiation made by façades in the estimation of the solar

potential in city-wide studies, which should be deemed as a significant component of received solar radiation (Sánchez et al., 2015; Boccalatte et al., 2020).

1.1.3 3D Solar potential estimation

The 3D geographic information system (GIS) model has proven to be a promising approach for accurate solar analysis at the building scale (Zhu et al., 2020; Zhu et al., 2019; Li et al., 2016; Erdélyi et al., 2014; Zhu et al.,2022). This approach involves creating a 3D model of a building or urban area, which can be used to quantify the building occlusion and the solar PV potential on the vertical surface. No matter the Level-of-Detail (LoD), 3D models can provide valuable information about the solar potential in a given area and have been used in a number of studies to estimate the solar potential in different cities. However, despite the potential of 3D GIS models for accurate solar analysis, there are still certain limitations that need to be addressed. One of the main limitations is the lack of consideration of radiation reflections in the calculation. While some studies have proposed 3D models for the estimation of the solar potential, radiation reflections are not incorporated in the calculation (Calcabrini et al., 2019; Jakubiec et al., 2013; Li et al., 2016). For example, a study by Redweik et al., (2013) used a 3D model to estimate the solar potential in a dense urban area but did not consider the effect of radiation reflections on solar irradiation. Due to the lack of information on urban envelopes albedo, the inter-building reflection, which should be deemed as a significant component of solar irradiation, has been omitted in most studies. This is a significant limitation, as the inter-building reflection can significantly affect solar irradiation in densely populated urban areas. On the contrary, a study by Zhu et al., (2020) used a 3D GIS model to estimate the solar potential in ten cities, including Athens, Honolulu, Lisbon, Hong Kong, Los Angeles, Mandalay, New York, Paris, Singapore, and Toronto. The study indicates that the solar reflection simulation provides a more accurate estimation of solar distribution in urban environments, as the high albedo in cities can significantly alter the distribution. The study also suggests that, under the same conditions (latitude and clouds),

cities with a higher density of tall buildings and an irregular fluctuation of building heights tend to have a larger solar capacity. In contrast, some studies apply a constant value to represent the albedo of all urban façades. For example, a study by Zhu et al., (2020), they incorporated reflection into the 3D model by applying a constant value to represent the albedo of all urban façades. However, this approach has certain limitations, as the albedo of urban façades can vary significantly depending on the architectural style and materials used. In metropolitan cities, where there is a wide range of architectural styles and materials, the irradiation of reflected solar light could vary significantly in different districts. For instance, a rooftop near a commercial building with a high albedo glass façade is likely to receive higher irradiation than a rooftop near an old residential building with a mosaic tile façade.

1.1.4 Research Gaps

It is challenging to accurately estimate the solar potential incorporating albedo-based multi-reflection and quantify the effect of façade albedo on solar PV distribution, as the albedo collection is exceptionally labor-intensive at a city scale. In this study, we propose a multi-scale contextual attention network to extract the material information from street views, which can link the results with 3D models and hence improve the accuracy of solar estimation. However, there are still some difficulties that need to be addressed before extracting the material information. Since the major objective of this study is obtaining facade material information on a large scale, using street view images as a source of information has become a very intuitive, economical, and effective method. The significant challenge when utilizing street-level images is the unstable image qualities, with massive occlusion, over-exposure, and unclear perspective. The other reason that makes the use of street-level images for material identification a challenging task is that the colors and shapes of different materials may have similar visible spectral characteristics, making it difficult to distinguish between them. In addition, there is a lack of specific datasets dedicated to façade material identification, which can further complicate the task. Although there are some

datasets for material identification, such as the Materials in Context Database (MINC) (Bell et al., 2015) and the dataset from MIT (Sharan et al., 2009), no specific dataset is dedicated to façade material identification. This can make it difficult to train and evaluate the performance of material identification models. In the past few years, previous works (Liu et al., 2017; Ma et al., 2020; Dai et al., 2019) have applied convolutional neural networks (CNN) to façade parsing, which have achieved better performances than traditional models (Gadde et al., 2016). However, the same challenges still exist in current research. Most façade related research was conducted by utilizing non-street-level images, which simplifies complicated environments and obstacles and thus reduces the generalization of methods. This can limit the applicability of these methods to real-world scenarios, where street-level images are often the only available data source. Therefore, this study proposes a novel street-level semantic segmentation dataset. The dataset collected street view images of different streets and regions in Hong Kong, including ShekMun, LaiChiKok, WestKowloon, Kowloon Bay, NorthPoint, and Central, under different weather and time conditions. This aims to enhance the robustness of the model in identifying materials in different complex urban environments, such as new and old residential areas, commercial areas, government land, etc.

Furthermore, another difficulty is that the colors and shapes of different materials may have similar visible spectral characteristics in street-view images. This makes identifying materials much more challenging than identifying façade components, like windows or balconies. The model needs to have strong semantic understanding ability to be competent in this task. In current research, there are two main methods for identifying materials: semantic segmentation and object recognition. Semantic segmentation involves dividing an image into multiple regions and assigning a label to each region or pixel, indicating the type of material present in that region. Object recognition, on the other hand, involves identifying specific objects within an image and classifying them with bounding boxes. Both semantic segmentation and object recognition have their own advantages and disadvantages, and the choice of method depends on the specific requirements of the task. Object recognition is

useful for identifying specific objects within an image, such as windows or doors, and classifying them based on their characteristics. This can be useful for tasks such as building facade parsing, where it is important to identify specific building components and their properties. However, object recognition may not be as effective at identifying materials in complex scenes, as it may not be able to distinguish between different materials that are present in the same region. This study chooses to use a segmentation model to extract material information from street views. This approach allows us to identify materials in complex scenes and makes it easier to map the identified results of different building parts into a 3D GIS model. In semantic segmentation models, it is critical to balance the network dimensions (i.e., width, depth, and resolution) (Tan et al., 2019). Some studies use low-resolution images as input to cover a relatively larger receptive field and lower floating-point operations per second (FLOPS) (Richter et al., 2021). Façade materials identification has a high demand on pixel-level details to differentiate specific materials, which means maintaining the image original resolution is crucial. In addition, finer-resolution inputs usually mean better performance in detecting small objects while coarser is good for large ones. Using multi-scale inference is a popular way to handle the trade-off. Lin et al., (2017) used average pooling to combine the features between scales. Tao et al., (2020) proposed the attention mechanism to determine the weighted mask and then use the mask to trade off the information of different scales. However, it is still a pixel-level operation, which cannot fully exploit the contextual information of the output tensor.

In this study, a multi-scale contextual attention network (MSCA) is proposed for the façade segmentation. Compared with previous works (Chen et al., 2016; Tao et al., 2020), MSCA combines contextual information in high-dimension feature space to achieve feature-level fusion between scales. The main idea of our work is to use the attention layers to fuse hierarchical information in a revised Object-Contextual Representation (OCR) module (Yuan et al., 2019) rather than after the segmentation head of the network. This can significantly improve the contextual comprehending ability of the network and thus could better handle the

trade-offs between high demand on details and contextual comprehension ability on large objects.

Based on the extensive material information obtained from the MSCA, this research aims to develop a comprehensive evaluation framework for investigating the effect of albedo on solar PV potential distribution. The proposed framework is designed to provide a systematic approach for evaluating the influence of albedo on solar PV potential distribution, taking into account various factors such as building function, façade materials, and inter-building reflections. After MSCA, the segmentation results are projected to the 3D GIS model, which allowed for the visualization and further analysis of the study area in a spatial context. Based on the identified materials, each building in the study area was assigned one of three different albedos: constant albedo, simulation albedo, or segmentation-based albedo. The 3D building models with different albedo strategies were then used to evaluate the effect on solar potential distribution. At the end of the study, the quantitative results were compared and discussed to assess the impact of different albedos in urban areas on the solar PV potential distribution.

1.2 Research Objectives

Solar photovoltaic (PV) systems are increasingly becoming popular in metropolitan cities. To provide an essential basis for energy policy decision-making and panel deployment in metropolitan cities, this study proposes to realize an accurate estimation of solar potential by incorporating the effect of the multi-reflective solar radiation made by façades. To handle this challenge, this study has the following major objectives:

1) To develop a semantic segmentation dataset for façade material recognition.

2) To develop a deep learning model for façade material recognition.

3) To convert the 2D material segmentation results into albedo information and project it into the 3D model.

4) To develop a solar potential estimation model that incorporates the effect of the multi-reflective solar radiation made by façades.

1.3 Thesis outline

The remaining parts of this thesis were organized as follows. The first chapter introduces the background, research gaps, and objectives of this study. Chapter 2 reviews related work and current technology, including façade imagery analysis, existing material Bidirectional Reflectance Distribution Function (BRDF) library, and solar potential estimation solutions. Chapter 3 presents the study area and data collection. In Chapter 4, we describe the methodology of this study, including the framework of this study, assumption, data annotation process, structure model, and detail formulas. In Chapter 5, we evaluate the performance of the proposed method and the quantitative results were compared and discussed to assess the impact of different albedos in urban areas on the solar PV potential distribution. Chapter 6 summarized the conclusion of this study and followed by future work for this study.

Chapter 2 Literature Review

This section first reviews the current approaches for solar potential estimation, including traditional approaches and machine learning-based approaches. The existing limitations are also summarized in the chapter. It then reviews the research on material recognition, façade imagery analysis, and multi-scale segmentation, identifying the current gaps in accessing large-scale façade information.

2.1 Solar potential estimation

2.1.1 Overview of Estimation Methods

Accurate solar potential estimation is critical for designing and optimizing photovoltaic (PV) systems, especially in urban environments. Various methods have been developed to estimate solar potential, which can be broadly classified into empirical, physical, and machine learning-based models. As shown in Table 2.1, these methods range from simple statistical approaches to advanced AI-driven models, and their application depends on the complexity of the study area, available data, and the required accuracy.

Table 2.1 Comparison of solar potential estimation methods

Methods	Description	Study Area	Accuracy	Representative works
Empirical Models	Based on statistical analysis of historical data	Large regions	Medium	Chen et al. (2019); Makade et al. (2019)
Physical Models	Simulates solar radiation processes based on principles of physics	Urban, Street	High	Nguyen et al. (2013)
	3D Ray-Tracing	Urban, Street	Very High	Zhu et al. (2020); Li et al. (2016)
Machine Learning Models	Uses algorithms like SVM and ANN to predict	Various	High (data-dependent)	Assouline et al. (2015); Nwokolo et al. (2023)

Empirical models, as documented by Chen et al. (2019) and Makade et al. (2019), are derived from statistical analysis of historical data and are often used for their simplicity and ease of application. Physical models, on the other hand, as described by Nguyen et al. (2013) and Zhu et al. (2020), are grounded in the fundamental principles of physics and attempt to simulate the actual processes affecting solar radiation. Machine learning models, as explored by Assouline et al. (2015, 2017) and Nwokolo et al. (2023), employ algorithms that can learn from data and identify complex patterns, thereby providing a more dynamic approach to solar potential estimation.

2.1.2 Horizontal solar potential estimation

Models for estimating solar potential using two-dimensional input parameters have been extensively studied. The conventional meteorological parameters such as cloud cover, ambient temperature, and relative humidity, which have been widely acknowledged by researchers including Besharat et al. (2013), there is a growing body of literature that suggests the incorporation of additional data points like sunshine duration or rainfall. These parameters have been shown to optimize the estimation process, as demonstrated by Quej et al. (2016) and Meenal et al. (2016). The rationale behind this is to utilize a more comprehensive set of variables that influence solar irradiance, thereby enhancing the predictive capability of the models. However, as the number of parameters that are found to be relevant to solar distribution increases, traditional models often exhibit a lack of sufficient generalization ability to accurately map the relationships between independent (input) and dependent (output) variables. This limitation has prompted researchers like Meenal et al. (2018) to employ data mining tools such as the Waikato Environment for Knowledge Analysis (WEKA) to identify the most influential factors. They evaluated the accuracy of advanced computational techniques such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) in capturing the complex interdependencies between the variables.

2.1.3 Vertical solar potential estimation

To further refine the accuracy of solar potential estimations, three-dimensional modelling techniques have been introduced. Li et al. (2016) proposed a 3D model that calculates the shadow effects on rooftops, which is critical for understanding the availability of solar radiation in urban environments. Similarly, the SORAM model, developed by Erdélyi et al. (2014), employs a sophisticated ray-tracing algorithm to assess whether a 3D ray vector intersects with a voxel, thereby enabling the calculation of dynamic shading effects from urban structures. Moreover, the assessment of solar potential on vertical surfaces has been increasingly recognized as an important factor in comprehensive solar resource evaluation.

This aspect has been incorporated into models by researchers such as Xu et al. (2019), An, Chen et al. (2023), and Willenborg et al. (2018b). Meanwhile, researchers like Bill et al. (2016) and Willenborg et al. (2018a) leverage advanced technologies such as Light Detection and Ranging (LiDAR) data, semantic 3D city models, and 3D mesh models to achieve higher levels of detail (LoD) and accuracy in 3D model.

Despite these advancements, only a handful of studies, such as those by Zhu et al. (2020, 2022), have included the complex phenomenon of inter-building reflection in their estimations. In these studies, the albedo, which represents the reflectivity of urban surfaces, is often assigned a constant value. This simplification has led to less precise results, as the variability of albedo in urban environments can significantly influence the distribution of reflected solar radiation. The challenge lies in quantifying the effect of this variability on solar potential estimation, which remains an area for further research and refinement.

The integration of photovoltaic systems into building façades, termed Building Integrated Photovoltaics (BIPV), is increasingly recognized for its potential to enhance the sustainability of urban structures. One of the critical factors influencing the performance of BIPV systems is their sensitivity to the variations in solar indirect components, notably multi-reflection phenomena, which are significantly influenced by the reflectivity of surrounding surfaces (Fouad et al., 2017). These reflections can substantially alter the amount of solar energy received by photovoltaic systems. Kotak et al. (2015) conducted a detailed investigation into this phenomenon and utilized the empirical equation formulated by Liu et al. (1963) to estimate the contributions of reflected energy. Their findings indicated a substantial increase in energy gain, specifically recording a 48% enhancement when factoring in the reflective properties of nearby structures. Zhu et al. (2020) further contributed to this body of knowledge by selecting an empirical parameter of 0.4 for their estimations based on observations that the average albedo of commonly used building materials ranged between

0.32 and 0.38. Despite these advances, the challenge of accurately acquiring reliable albedo information for building materials in specific study areas persists.

Traditionally, the acquisition of albedo data relied heavily on conducting in situ investigations, which can be both time-consuming and resource-intensive. In earlier research efforts, such as those by Dana et al. (1999), and subsequent studies by Fritz et al. (2004) and Mallikarjuna et al. (2006), the focus was primarily on material recognition through the analysis of textures in close-up photographs devoid of any contextual background. This method saw significant development over time, shifting towards the recognition of materials on real-world objects beyond mere texture distinctions, as evidenced in the studies by Sharan et al. (2014) and Bell et al. (2015).

However, these techniques predominantly targeted indoor objects and did not address the unique challenges presented by urban envelope materials, which must often be identified from greater distances, with more complex environmental conditions, and with less distinct shapes.

2.2 Material Recognition

2.2.1 Close-up images identification

In the domain of computer vision, distinguishing materials from images represents a notably complex challenge when compared to tasks such as object detection or scene understanding. The intrinsic difficulty lies in the fact that materials can manifest a highly diverse range of appearances, complicating the identification of consistent image features that can reliably categorize them. Initial efforts in material recognition primarily focused on analyzing textures captured in close-up photographs without any contextual background. This approach is exemplified by the work of Dana et al. (1999), who introduced the CUReT dataset, comprising 61 material surfaces photographed under more than 200 different illumination

and viewing conditions. However, a significant limitation of CUReT is that each class includes only a single instance of each material, which severely restricts the dataset's ability to generalize across varied real-world scenarios.

2.2.2 Real-world objects identification

To address these limitations, subsequent datasets such as KTH-TIPS (Fritz et al. 2004) and KTH-TIPS2 (Mallikarjuna et al., 2006) expanded the variety of samples for about ten different materials, including textiles like corduroy, linen, and cotton. These datasets included multiple instances of each material, which helped improve the robustness and generalization of material recognition algorithms. Nevertheless, the challenge of recognizing materials on real-world objects remains substantially greater than identifying close-up textures due to the complexity and variability of objects in natural settings.

This recognition challenge was further addressed by Sharan et al. (2014), who created the FMD, a dataset consisting of 1,000 images of complete objects categorized into ten different types. This dataset was designed to minimize irrelevant background interference, thereby focusing on the materials themselves. Following this, the OpenSurfaces dataset (Bell et al., 2013) provided over 25,000 images captured in more typical indoor settings rather than close-ups, offering a more realistic view of materials within everyday environments. Building on this, Bell et al. (2015) developed the MINC dataset, which included 7,061 labelled material segmentations across 23 categories, with a more uniform representation of materials aimed at enhancing the performance of automated classification systems.

2.2.3 Façade materials in Hong Kong

A survey was conducted during 2004, to analyse the materials used for external wall finishes in 111 private residential buildings in Hong Kong (Ho et al., 2004). The study identified seven categories of commonly used materials: Paints, Mosaic Tiles, Ceramic Tiles, Mosaic and Ceramic Tiles, Mosaic and Others, Ceramic and Others, and Others (including granite

and marble cladding). The frequency distribution revealed that Ceramic Tiles were the most prevalent, accounting for 40.5% of the sample, followed by Mosaic Tiles at 27%, and Paints at 13.5%. Notably, the use of Paints and Mosaic Tiles has significantly declined in recent developments. Most buildings with paint finishes were over 20 years old, aligning with findings from Shui On et al. (1984). However, Paints has seen a resurgence in public rental housing due to advancements in paint technology, offering better durability and a wider range of colours. Additionally, the adoption of metal formwork has facilitated the use of Paints without compromising the aesthetic appeal of the buildings.

Despite these advancements, based on the surveys on façade materials in Hong Kong, a significant gap remains in the datasets, which primarily focus on indoor objects and settings. In contrast, materials featured on building façades in urban street views present unique challenges. These materials are often captured from greater distances, resulting in images with blurrier details and more complex backgrounds. Furthermore, the less distinct shapes and the dynamic lighting conditions typical of outdoor environments add additional layers of complexity to the task of material recognition. This distinction underscores the need for developing specialized methodologies and datasets that are tailored for the identification and classification of materials in outdoor urban environments, specifically those that can address the challenges posed by variable distances, complex scenes, and diverse lighting conditions.

2.3 Façade imagery analysis

2.3.1 Façade parsing

The utilization of street view imagery for façade analysis presents numerous advantages, primarily due to the ease of access and the potential for crowdsourced data collection. This approach enables the compilation of extensive datasets that can be employed to enhance the understanding and classification of building façades. In this context, Gadde et al. (2016) introduced an innovative approach by employing an unsupervised clustering method. This

method focused on aggregating simple rules into complex patterns, offering an alternative to traditional methods that rely on handcrafted grammars specifically designed for parsing distinct classes of building façades. Such methodologies demonstrate a shift from rigid, predefined models to more flexible and adaptive learning frameworks capable of understanding the variability in urban architecture.

The utilization of street view imagery for façade analysis presents numerous advantages, primarily due to the ease of access and the potential for crowdsourced data collection. This approach enables the compilation of extensive datasets that can be employed to enhance the understanding and classification of building façades. In this context, Gadde et al. (2016) introduced an innovative approach by employing an unsupervised clustering method. This method focused on aggregating simple rules into complex patterns, offering an alternative to traditional methods that rely on handcrafted grammars specifically designed for parsing distinct classes of building façades. Such methodologies demonstrate a shift from rigid, predefined models to more flexible and adaptive learning frameworks capable of understanding the variability in urban architecture.

Further advancing the field, Liu et al. (2017) proposed a novel neural network architecture equipped with a symmetric regularizer. This design leverages the inherent structural symmetry found in man-made architectures, particularly in building façades, which are often characterized by highly regularized shape priors and fine-grained details. The researchers highlighted a significant challenge associated with the direct application of standard deep learning techniques, such as those discussed by Schmitz et al. (2016), which do not consistently achieve optimal results in façade imagery analysis. This discrepancy is largely due to the unique architectural features that are not adequately captured by conventional segmentation models primarily designed for more generic applications.

In response to the limitations observed in both grammar-based approaches, which heavily depend on prior architectural knowledge, and the underperformance of standard segmentation

models, Kong et al. (2020) introduced a novel convolutional neural network (CNN) pipeline. Their approach synergistically combines pixel-wise segmentation with global object detection to address the complexities inherent in street-level datasets. This methodology aims to improve the generalization capabilities of façade analysis tools by effectively integrating detailed local texture analysis with broader structural recognition.

However, the task of extracting material categories from façade imagery introduces additional complexities that surpass those encountered in the identification of façade components such as windows or balconies. The primary challenge arises from the subtle differences in shape and spectral characteristics among various building materials, such as ceramic tiles and marble, which are often much less pronounced than the distinctions between different architectural components. This subtlety makes the task of accurately categorizing façade materials using image analysis techniques more arduous than simply identifying different structural elements of a façade. As such, developing robust methods that can discern these subtle differences is critical for advancing the capabilities of automated façade analysis models, which are essential for applications ranging from urban planning to heritage conservation.

2.3.2 Data collection for façade recognition

Accurate façade recognition relies heavily on the quality and source of the data collected. Various methods exist for gathering façade imagery, each with its own set of advantages and disadvantages. Common data sources include Mobile Mapping Systems (MMS), handheld devices, Baidu, Google, panorama images, and cell phone images. As shown in Table 2.2, these methods differ in terms of resolution, availability of open-source databases, and suitability for specific study areas.

Table 2.2 Comparison of different data collection methods

Data Collection Method	Advantages	Disadvantages	Resolution	Open-Source	Area
Mobile Mapping Systems (MMS)	High resolution, with high-precision geographic coordinates	Expensive, limited availability	High	Limited	Urban
Baidu Street View	Wide coverage, relatively accessible	Inadequate resolution, privacy concerns	Medium	Yes	Primarily China
Google Street View	Wide coverage, relatively accessible	Inadequate resolution, privacy concerns	Medium	Yes	Global
Panorama images	Comprehensive spatial context	Difficult to process for model	Medium	No	Urban
Cell Phone Images	Highly accessible, easy to collect	Varies greatly by phone quality	Medium to high	No	Highly localized

While these methods provide valuable data for general façade recognition tasks, they have significant limitations when applied to the specific task of identifying façade materials. The primary issue lies in resolution and accuracy. High-resolution images are necessary for distinguishing subtle differences in building materials, such as the textures of ceramic tiles

versus marble. However, many widely available sources, such as Google Street View and Baidu Street View, lack the resolution required for such fine-grained material segmentation. Additionally, panorama images, while offering high resolution and comprehensive spatial context, are difficult for typical segmentation models to process.

Moreover, the inconsistency in resolution and accuracy across different platforms makes it challenging to apply segmentation models to datasets from multiple sources. In the context of this study, which aims to develop robust material segmentation methods, the limitations in available datasets underscore the need for controlled, high-quality image data. Therefore, freely available resources like Google Street View, while useful for general façade analysis, are unsuitable for our research, which requires high precision in identifying façade materials through semantic segmentation.

2.4 Segmentation method

Semantic segmentation is a significant area of computer vision that involves partitioning given images into segments by assigning a class label to each pixel. The goal of semantic segmentation is to label regions of an image with their corresponding object categories. It plays an essential role in tasks like autonomous driving, medical image analysis, and urban mapping, where pixel-level classification is required to understand the scene.

2.4.1 Semantic segmentation

In semantic segmentation, deep learning architectures such as Fully Convolutional Networks (FCNs) (Long et al., 2015) have been pivotal in the evolution of this field. These models replace the fully connected layers of traditional convolutional neural networks (CNNs) with convolutional layers, thus enabling pixel-wise classification for end-to-end image segmentation. This approach has become foundational for many applications where high spatial resolution is crucial for identifying small objects and details within an image.

Typical semantic segmentation architectures often rely on fixed-size receptive fields, which can limit the model's ability to capture contextual information. While they excel in detecting fine-grained details, such as edges or small objects, their lack of ability to scale effectively can cause underperformance when identifying larger or more complex objects. For instance, objects like buildings or landscapes require a larger context to be accurately classified. As a result, semantic segmentation models frequently encounter difficulties balancing high-resolution predictions with the understanding of broader contextual features (Chen et al., 2017; Zhao et al., 2017).

To mitigate these challenges, various improvements have been introduced, including encoder-decoder architectures and skip connections. These additions allow for the preservation of high-level semantic information while maintaining fine-grained spatial resolution. Models such as UNet (Ronneberger et al., 2015), which use these techniques, have shown considerable improvements in achieving pixel-level accuracy across different scales, especially in medical imaging and remote sensing applications.

Despite these advancements, typical semantic segmentation methods often struggle with capturing multi-scale features, particularly in scenarios where objects of varying sizes coexist within the same image. This limitation has motivated the development of multi-scale segmentation techniques, which extend the ability of segmentation models to recognize objects across different resolutions.

2.4.2 Multi-scale segmentation

In the domain of pixel semantic segmentation within deep learning frameworks, recent advancements have predominantly employed low output stride backbones to enhance the resolution of predictions, thereby facilitating the identification of fine-grained details in images. However, as noted by Wei et al. (2017), these architectures typically suffer from small receptive fields, which, although beneficial for capturing minute details, tend to

underperform when tasked with identifying larger objects. This limitation underscores a significant challenge in balancing the need for detail with the ability to comprehend larger contextual elements within an image.

To address this inherent trade-off, Zhao et al. (2017) developed the Pyramid Scene Parsing Network (PSPNet), which incorporates a pyramid pooling module to aggregate contextual information across multiple scales effectively. This innovative approach allows the network to maintain high-resolution insights while also capturing broader contextual data necessary for understanding larger image segments. Simultaneously, other studies have explored the use of encoder-decoder structures complemented by skip connections, as exemplified by the UNet architecture proposed by Ronneberger et al. (2015). These designs facilitate the seamless transfer of contextual information between layers of varying depth within the network, thereby enhancing the feature integration across different scales.

Nevertheless, Chen et al. (2016) identified a critical issue with traditional pooling operations, such as average or max pooling. They observed that these operations often render the features extracted at each scale either uniformly important or selectively sparse. To mitigate this issue, they introduced an attention mechanism designed to dynamically fuse features across scales, thereby enabling the model to adaptively focus on the most pertinent scales for any given task.

Expanding upon this foundation, Tao et al. (2020) further optimized the model by reducing the computational cost associated with scale integration. They achieved this by predicting relative weightings between adjacent scales, thus obviating the need for introducing additional scales into the network. However, despite these advancements, the specific challenge of material segmentation at the pixel level, as highlighted by Schwartz et al. (2016), remains a daunting task. This is primarily due to the difficulty in precisely distinguishing the material origins of patterns within an image. This study proposes to handle the contextual

information at the feature level by applying the attention module before the segmentation head instead of employing it at the end of the network.

To more effectively tackle this challenge, this study proposes a novel approach that applies the attention module prior to the segmentation head rather than at the end of the network. By manipulating the contextual information at the feature level earlier in the processing pipeline, this methodology aims to enhance the precision of material categorization, ensuring that the segmentation network can more accurately and reliably determine the material composition of various objects within an image. This adjustment not only refines the process of feature integration but also aligns the network's focus more closely with the nuanced requirements of material segmentation.

Chapter 3 Study area and data collection

3.1 Study area

This study primarily aims to evaluate the solar potential in metropolitan cities, taking into account the significant impact of façade reflections. Hong Kong, a densely populated metropolis in southern China known for its economic development and dense architectural landscape, serves as the ideal research site for this investigation. The city's diverse building styles, ranging from skyscrapers and museums to aged residential structures, contribute to a wide variety of façade materials, making the albedos of these façades a critical factor in assessing the indirect part of solar potential.

First, to address the challenge of material identification on building façades, this study constructs a segmentation dataset. The data was collected by a mobile mapping system. The street view images cover various districts, ensuring the façade types of Hong Kong are comprehensively collected, which includes Shek Mun, Lai Chi Kok, West Kowloon, Kowloon Bay, North Point, and Central. The specific routes are illustrated in Figure 3.1.

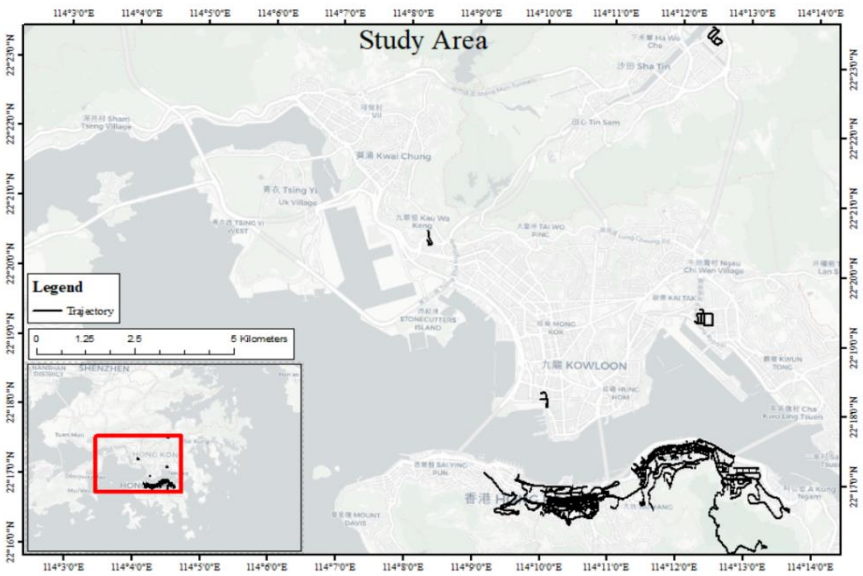

Figure 3.1 The data collection routes of constructed dataset, including ShekMun, LaiChiKok, WestKowloon, Kowloon Bay, NorthPoint, and Central in Hong Kong

As shown in Figure 3.2, in the subsequent phase of solar potential analysis, the research narrows its focus to the North Point area, which belongs to the Eastern District, located in the northeastern part of Hong Kong Island. The high population density, more than 25000 people per square kilometre (Hong Kong Government, 2021), in North Point results in complex land utilization and heterogeneous façades. As shown in Figure 3.3, according to the land utilization data (Planning Department, 2023), the eastern land of North Point is mostly used for residential purposes. According to on-site investigations, the residential buildings in this area were typically built in the last century. The façades of this kind of building are usually mosaic tiles and painted. In contrast, the west and south parts of North Point are more multifunctional, consisting of residential, commercial, industrial, and a large percentage of institutional land use. Therefore, the corresponding architectural materials are more diverse, like glass, ceramic, and metals, which form a sophisticated environment for estimating solar reflection in this district. Moreover, the southern region of North Point, in comparison with the bustling, high-density downtown areas adjacent to the seaside, is bordered by expansive

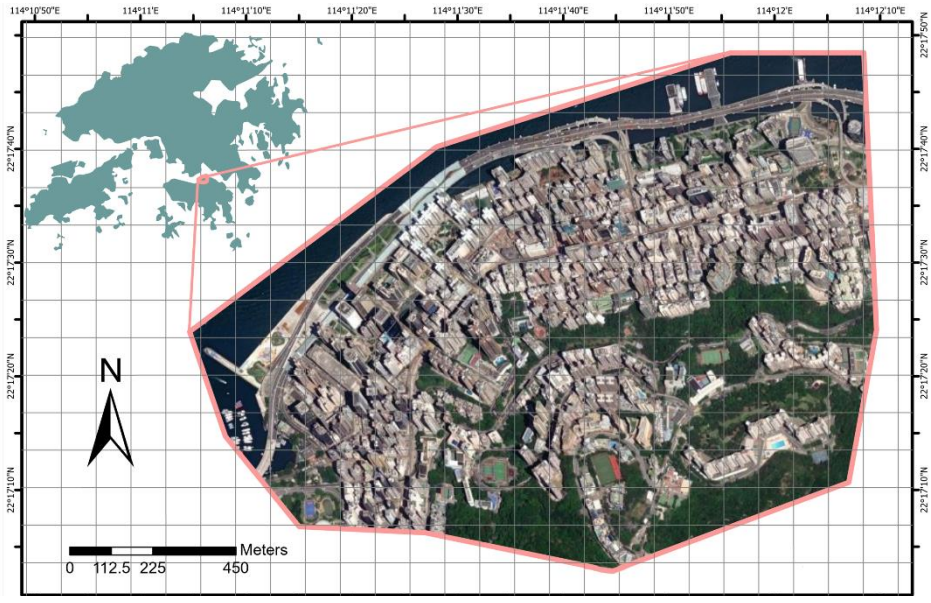

Figure 3.2 Study area (North Point, Eastern District) for the estimation of solar potential.

woodlands. This natural barrier fosters a setting where residential buildings are relatively

spaced apart, minimizing the phenomenon of inter-building reflection. This spatial arrangement provides an excellent baseline for comparison with the high-density urban districts, allowing for more controlled observations of solar reflection dynamics. The isolation and reduced density of the buildings in this part of North Point make it an ideal control group for studies focused on understanding the impact of urban density on solar reflection. This nuanced understanding of the varying land use and architectural materials across different parts of North Point is crucial for developing accurate models of solar energy potential across urban landscapes.

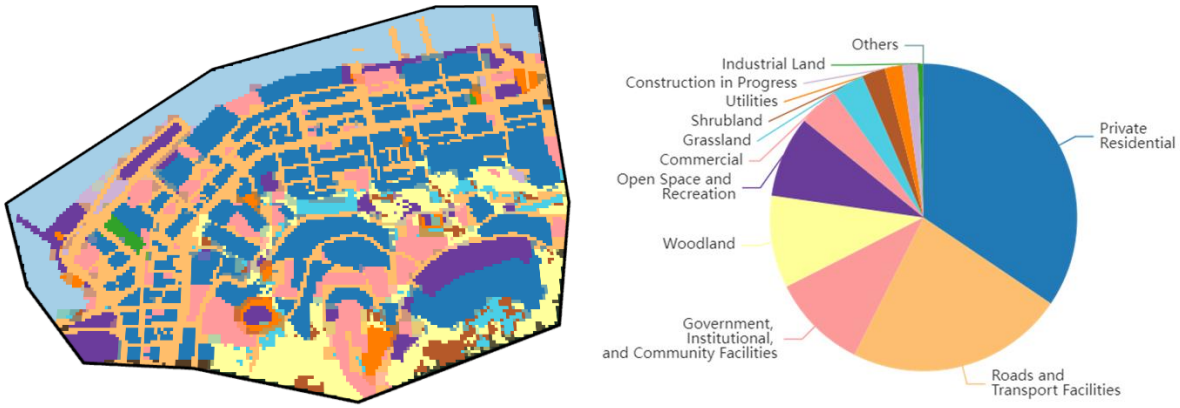

Figure 3.3 Land utilization and corresponding percentages in the North Point.

3.2 Data collection

This research, conducted in partnership with the Hong Kong Highways Department, developed a façade segmentation dataset aimed at enhancing the understanding of urban envelope materials environments. The data acquisition spanned from 2017 to 2019, during which the Hong Kong Highways Department systematically collected street view imagery. These images were captured under a variety of weather conditions and solar illuminations, as per the designated collection routes detailed in Figure 3.1.

To optimize the quality of the dataset, the research team implemented a rigorous selection process post-data collection, where images compromised by duplication, overexposure, or underexposure were discarded. Despite these exclusions, the dataset maintained a rich

diversity in architectural styles, encompassing over 10,000 distinct buildings, thereby providing a solid foundation for subsequent experiments.

Figure 3.4 The mobile mapping system.

The street view images were collected by the vehicle-based mobile mapping system, Leica Pegasus: Two, as shown in Figure 3.4 (Leica Geosystems, 2024). Leica Pegasus: Two contains 8 cameras, which means that we can obtain street view images in eight directions along the data acquisition track. The detailed parameters are shown in Table 3.1. Furthermore, the instrument also contains navigation systems including low noise FOG IMU and the triple band – L-Band, SBAS, and QZSS for GPS, GLONASS, Galileo, and BeiDou constellations. The precision of these navigational aids is documented in Table 3.2. The high accuracy of the positioning system, with horizontal and vertical Root Mean Square (RMS) errors below 0.020 meters in open sky conditions, facilitates the precise georeferencing of images within the Hong Kong 1980 (HK80) coordinate system. Additionally, the accurate coordinates significantly streamline the integration of semantic segmentation results into three-dimensional urban models. This high level of precision in data collection critically supports the project's aim to create detailed and reliable façade segmentation models, which are pivotal for the subsequent solar potential estimation.

Table 3.1 Camera sensors parameters of the mobile mapping system.

Item	Configuration
Number of cameras	**8**
CCD size	2000 x 2000
Pixel size	5.5 x 5.5 microns
Lens	8.0 mm focal, ruggedized; 2.7 mm focal, top
Coverage	360° x 270° excluding rear down facing camera

Table 3.2 GNSS/IMU/SPAN sensor parameters of the mobile mapping system.

Item	Configuration
Frequency	200 Hz
Mean Time Between Failures	35,000 hour
Gyro bias in-run stability (±deg/hr)	0.75
Gyro bias offset (deg/hr)	0.75
Gyro angular rand. walk (deg/√hr)	0.1
Gyro scale factor (ppm)	300
Gyro range (±deg/s)	450
Accelerometer bias (mg)	1
Accelerometer scale factor (ppm)	300
Accelerometer range (±g)	5
Position accuracy after 10 sec of outage duration	0.020 m RMS horizontal, 0.020 m RMS vertical, 0.008 degrees RMS pitch/roll, 0.013 degrees RMS heading

Chapter 4 Methodology

In this chapter, a novel Multi-Scale Contextual Attention Network (MSCA) is proposed to identify façade materials from street view images, bridging the gap of lack of large-scale façade information in existing studies. The proposed model uses a Multi-Scale Object-Contextual Representation (OCR) block to exploit and combine contextual information from different scales in high-dimensional layers. To validate the effectiveness of the proposed model, comparative analyses against baseline models, including DeepLabv3, DeepLabv3+, OCR, and Hierarchical MSA, across diverse datasets are also designed in this chapter. Moreover, this chapter delineates a methodology for integrating segmentation outcomes into a 3D GIS model, subsequently leveraging these results to allocate albedo values to each surface within the designated study area. Based on the identified materials, an evaluation method of quantifying the effect of façade albedo on Solar potential distribution is also presented in this chapter.

4.1 Research framework

The comprehensive evaluation framework developed for this study is depicted in Figure 4.1. The framework is designed to quantitatively assess the impact of façade materials on solar potential within urban environments, leveraging advanced geospatial technologies and image processing techniques.

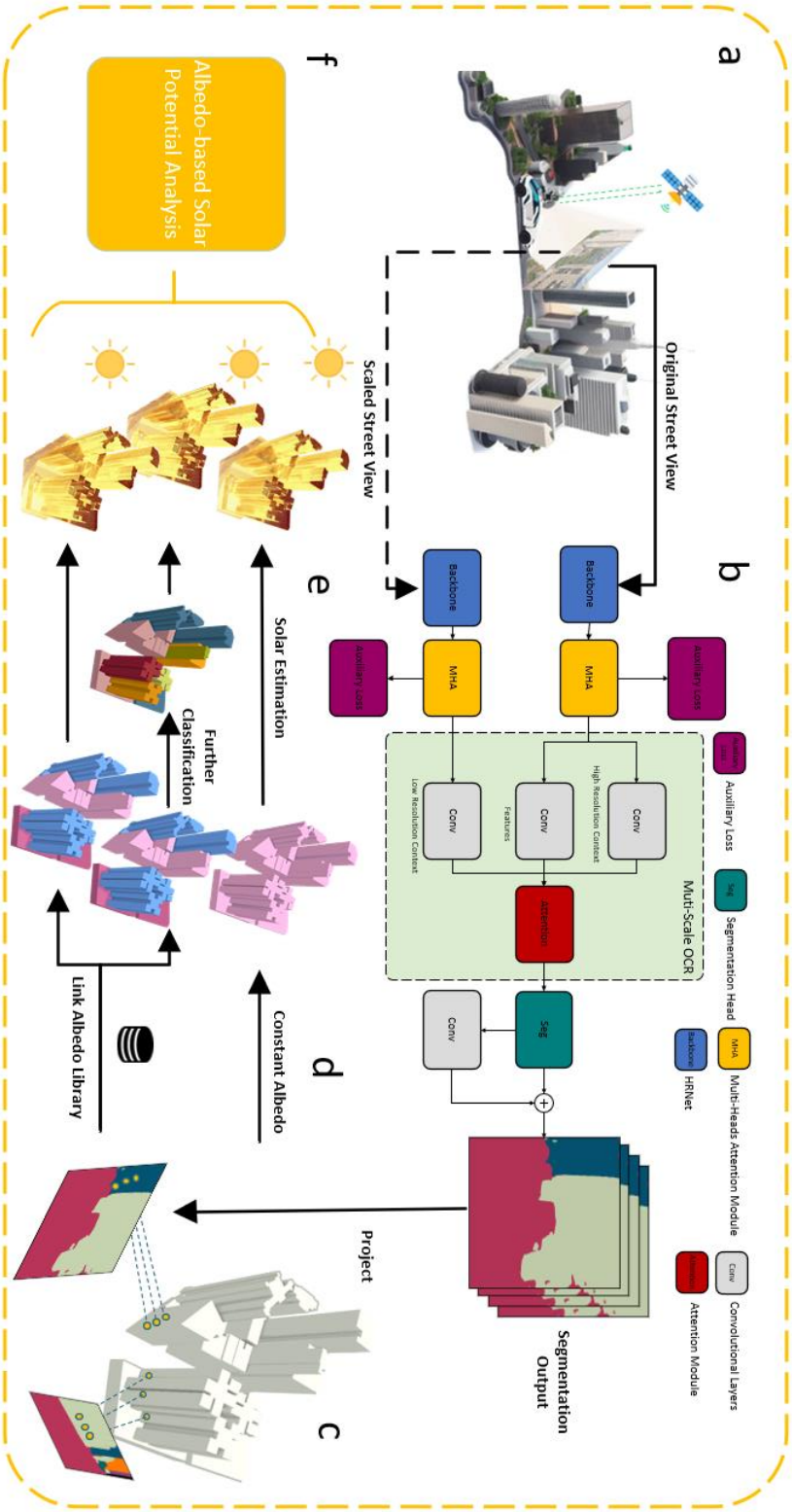

Figure 4.1 The evaluation framework. (a) Using the vehicle-based mobile mapping system to collect street view images with geographical coordinates. (b) MSCA is utilized to identify the

material categories from street views. (c) Projecting the 2D segmentation results to the 3D urban GIS model. (d) Based on three different albedo schemes, mapping the identified materials to distinct albedo: constant albedo, simulation albedo, and segmentation-based albedo. (e) Solar potential estimation based on three albedo schemes. (f) Analysing the solar potential distributions with different albedo schemes and spatial characteristics.

As delineated in Figure 4.1 (a), the initial phase of data acquisition employs a vehicle-based mobile mapping system, which is outfitted with the Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) apparatus. This sophisticated setup enables the capture of street view imagery, which is not only high-resolution but also precisely annotated with high-precision geographical coordinates. The integration of GNSS and IMU ensures that each image is accurately geo-referenced, providing a robust foundation for subsequent image analysis and material identification processes.

Subsequently, as illustrated in Figure 4.1 (b), the acquired street view images serve as inputs to the Multi-Scale Contextual Attention Network (MSCA). The MSCA is developed to identify a variety of materials from street-level imagery on a large scale. In this step, we introduce a multi-scale attention structure designed to capture and interpret contextual information contained within high-level features of images. This approach is a pivotal component of the entire research, as it enables the extraction of specific material information from the urban envelopes, which is essential for the accurate assessment of solar potential.

In the subsequent step, depicted in Figure 4.1 (c), the study employs the Collinearity equation to transform the pixel coordinates from the image coordinate system to the Hong Kong 1980 Grid System. This transformation is a critical step in associating the segmentation results with the corresponding 3D model, thereby facilitating a seamless integration of two-dimensional image data with three-dimensional spatial models.

Once all buildings within the study area have been accurately identified, the study proposes the application of various albedo schemes to evaluate the influence of these materials on solar potential. As demonstrated in Figure 4.1 (d), the framework contemplates multiple albedo assignment strategies. The upper row of (d) illustrates a simplified scheme where all surfaces, irrespective of material categories, are assigned a uniform albedo value. This approach provides a baseline for understanding the general impact of albedo on solar potential. In contrast, the bottom row of (d) presents a more nuanced scheme that allocates distinct albedo values to each material type based on the segmentation results obtained from the MSCA. This scheme acknowledges the heterogeneity of urban surfaces and aims to reflect the varied reflective properties of different materials. However, recognizing the limitations in the number of material categories that the MSCA can currently identify, the study introduces a third scheme, as shown in the middle row of (d). This scheme involves a further classification of the initial categories identified by the MSCA into more detailed subclasses, thereby simulating a more intricate urban environment. The intent behind this additional classification is to create a more comprehensive simulation that can better capture the complex interplay of multi-reflections in an environment with diverse albedo characteristics.

Furthermore, this study primarily focuses on façade material albedo contributions, which are inherently more complex to obtain and quantify than rooftop. For rooftop albedo estimation, we derived simplified classifications based on color intensity (e.g., dark vs. light surfaces) through satellite imagery, assigning albedo values to each category.

Finally, in step (f), the study quantifies and discusses the effects of albedo and other spatial factors, such as land utilization, building height, density, and function, on solar potential. This comprehensive analysis not only sheds light on the direct impact of material albedo on solar potential but also considers the broader urban context in which these materials are situated. By integrating spatial factors into the evaluation, the study provides a holistic

understanding of the interdependencies between urban morphology, material properties, and solar potential distribution.

4.2 Hong Kong Façades Materials Segmentation Dataset

This section delineates the development of the proposed dataset designed to enhance façade analysis in metropolitan environments. Metropolitan cities are characterized by intricate street configurations and densely packed building structures, presenting unique challenges for urban façade segmentation. Traditional datasets in this domain, such as those reported by Korc et al. (2009), Teboul et al. (2011), and Riemenschneider et al. (2012), predominantly feature single-view images of façades with relatively sparse building distributions and minimal obstructions. Such datasets often fail to capture the complexity and density typical of metropolitan cityscapes, thereby limiting their practical applicability for detailed urban façade analysis. In response to this gap, our research collaboration with the Hong Kong Highways Department has led to the creation of a specialized façade segmentation dataset. This dataset comprises 1,823 street-level images, specifically curated to reflect the dense and intricate urban fabric of Hong Kong. The choice of Hong Kong as the research site is strategic, given its status as a densely populated metropolis with a diverse range of building styles and façade configurations. This setting provides an ideal backdrop for developing a dataset representing the complexities encountered in large urban centers.

The data routes were strategically designed to ensure comprehensive coverage of Hong Kong's urban diversity. Routes were selected to span districts with varying building densities (e.g., high-rise clusters in Central vs. mixed-use areas in Lai Chi Kok), façade styles (modern glass towers in West Kowloon vs. traditional structures in North Point), and street geometries (narrow alleys vs. arterial roads). Data collection utilized a vehicle-mounted mobile mapping system equipped with calibrated cameras and the IMU systems, operating during daylight hours under diverse lighting conditions (sunny, overcast) to enhance robustness. Regarding dataset size, while 1,823 images are large compared to current datasets, it is still not sufficient for deep-learning. So, the model is first pre-trained on Cityscapes to learn urban scene priors (e.g., object boundaries, occlusion patterns), then fine-tuned on our dataset to specialize in façade material features. This strategy further expand effective training samples, ensuring generalizability across Hong Kong's heterogeneous urban fabric.

However, the construction of this dataset also necessitates some methodological concessions. To adapt the façade segmentation model to the complex styles of modern urban façades and the variable solar lighting conditions typical in such environments, we introduced several assumptions. These assumptions are essential for simplifying the model to a manageable complexity while still maintaining a reasonable approximation of real-world conditions. Such compromises aim to balance the trade-offs between model accuracy, generalizability, and computational efficiency.

4.2.1 Data specifications

The Highways Department used a vehicle-based mobile mapping system to collect data. The geographical scope of this data gathering included Shek Mun, Lai Chi Kok, West Kowloon, North Point, and Central. These locations were selected due to their varied urban compositions, ranging from older residential zones to dynamic business districts, thereby ensuring a broad scope of façade types and architectural styles.

The mobile mapping system was equipped with cameras oriented in eight different directions, allowing for comprehensive data capture from multiple perspectives. Following the initial data collection phase, a data cleaning process was implemented. This involved the exclusion of images that were overexposed, repetitive, blurred, or illegible, ensuring that only high-quality images were retained for dataset compilation. Ultimately, from the extensive collection of captured images, 1,823 were selected to form the dataset. Regarding dataset utilization, approximately 1,463 images (80.2% of the total dataset) were designated for the training set, while the remaining 360 images (equivalent to 19.7%) were allocated to the validation and testing sets. Labeling of the dataset was carried out using the "Labelme" tool (Russell et al., 2008), a well-regarded annotation software that facilitates precise and efficient manual annotation of images. Following the initial labeling, a rigorous cross-checking process was instituted. This quality control measure was essential to verify the consistency and accuracy of the annotations, ensuring that the labels correctly represent the diverse characteristics of the urban façades captured in the dataset.

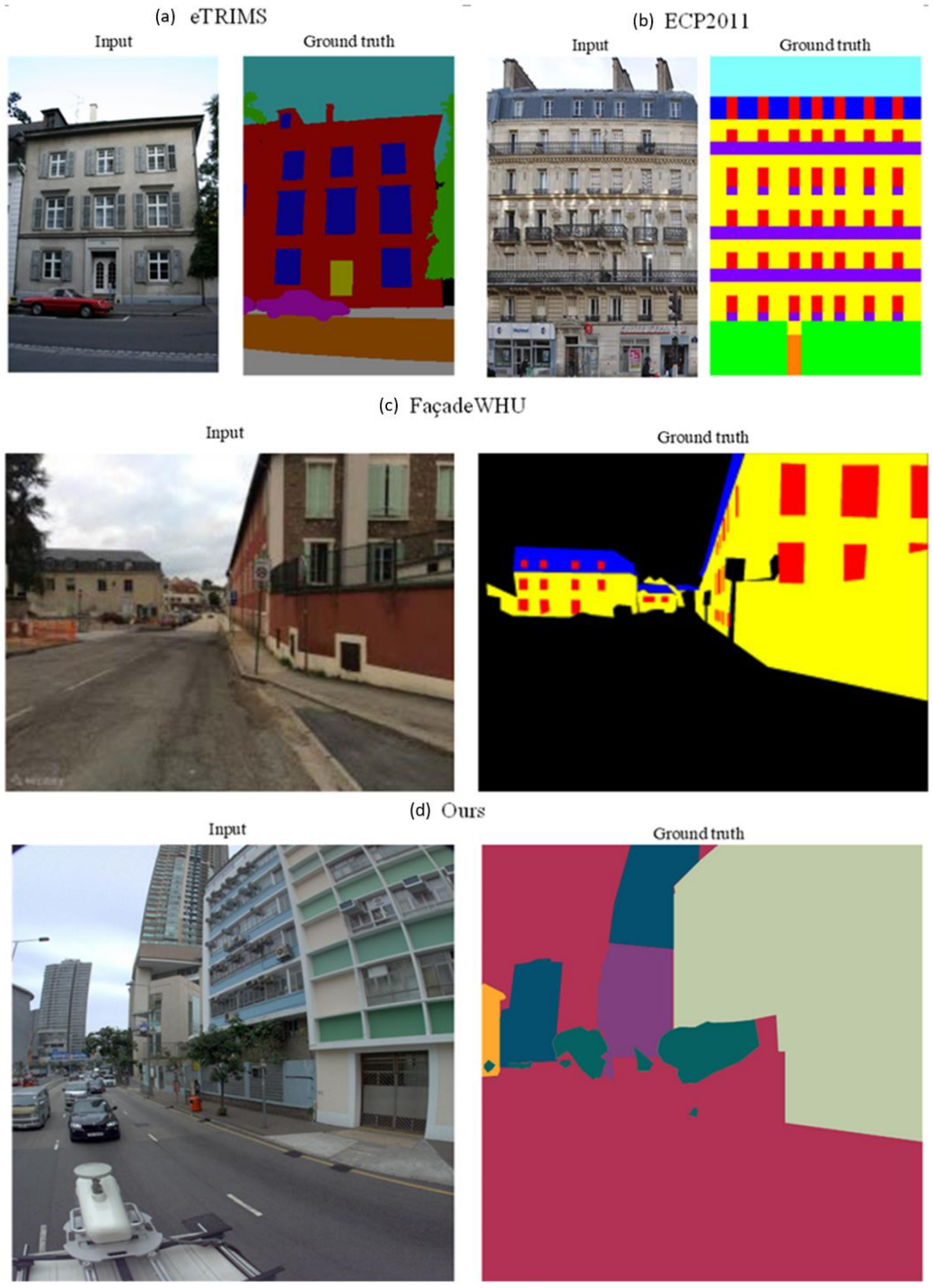

Figure 4.2 Comparison of existing façade-related datasets and the proposed dataset. The images on the left are the original images. The annotated images on the right side are used as the ground truth. In (a)(b)(c), different colors represent different façade opponents (i.e.

windows, balconies, doors and so on). In (d), different colors represent different materials of façade.

Compared to existing façade datasets, such as those reported by Korc et al. (2009), Teboul et al. (2011), Riemenschneider et al. (2012), and more recently by Kong et al. (2020), our dataset offers several distinctive advantages, which are stated as follow:

- Enhanced Façade Styles and High-Resolution Imaging: Existing datasets in the domain of façade recognition are generally limited both in the diversity of architectural styles and the number of images, which constrains the training and testing effectiveness of deep learning models. For instance, the eTRIMS dataset (Korc et al., 2009) comprises merely 60 annotated images, ECP2011 (Teboul et al., 2011) includes 104, and Graz2012 (Riemenschneider et al., 2012) contains only 50 images. Although FaçadeWHU (Kong et al., 2020) significantly increases this number to 900, it still falls short of the variety required to train models capable of accurately identifying façade materials in the diverse architectural environments of modern cities. Furthermore, these datasets predominantly feature monotonous building types, limiting their generalizability to different urban settings. In contrast, our dataset is comprised of 1,823 manually annotated images encompassing over 10,000 buildings, thereby offering a rich diversity of façade styles. Additionally, our images are captured at a resolution of 2046×2046 pixels, which surpasses most existing datasets, providing finer detail and supporting more accurate segmentation results.

- Accounting for Complex Urban Environments: Accounting for Complex Urban Environments: typical façade-related datasets often portray buildings with regular façade shapes taken from frontal views, devoid of occluding objects to simplify façade segmentation. For instance, as shown in Figure 4.2, ECP2011 focuses on close-up photos of building façades, meticulously cropped to minimize occlusion interference. Similarly, eTRIMS contains minimal background information, optimizing clarity but sacrificing environmental realism. By comparison, FaçadeWHU incorporates street-level images, enhancing generalization capabilities.

However, it still contains minimal occlusions such as pedestrians and billboards, which do not adequately reflect the cluttered environments of metropolitan areas.

- Inclusion of Dynamic Urban Elements and Variable Lighting Conditions: Our dataset distinctly includes street-level images capturing complex foreground occlusions such as trees, commercial signage, and dense traffic, significantly challenging the model's capability to isolate and identify façades accurately. Moreover, the dataset encompasses images under varied lighting conditions, affecting the hue, brightness, and saturation of façades, which introduces additional challenges but also opportunities for improving the robustness and adaptability of façade segmentation models. We posit that the heterogeneous quality and the realistic urban complexity represented in our images will substantially enhance the model's generalization across different urban settings and lighting conditions.

The greater diversity in building types and architectural styles and more comprehensive coverage of various urban environments make our dataset a valuable resource for advancing research in façade segmentation, particularly in the context of complex urban landscapes.

4.2.2 Assumption

In addressing the inherent challenges in identifying materials from Red Green Blue (RGB) images, this study introduces two critical assumptions that significantly reduce the complexities associated with labeling and minimize the associated labor costs:

- Each building has at most two components.

This research posits that each building can be segmented into at most two major components, each potentially composed of different materials. This assumption is based on observations and analyses of prevalent building types within Hong Kong. As depicted in Figure 4.3, most buildings in Hong Kong can be roughly divided into three categories. 1) Free-standing buildings, which are primarily for residential functions. The entire façade of this type of building is usually made of consistent material. 2) Complexes,

which typically consist of two parts, lower for commercial function while upper is a residential zone. The façade material of each component can be independent. 3) Special buildings, such as museums or cultural institutions, where façades may incorporate innovative and irregular material applications to convey unique aesthetic or thematic expressions.

- Each component consists of only one primary material.

Despite the observable fact that many building façades incorporate multiple material types, ranging from glass in windows to metal elements and decorative coatings, it remains technically challenging to accurately label every material type due to their complexity and the fine granularity required. Furthermore, buildings often feature alternating patterns and non-uniform distributions of materials that complicate the annotation process. To address these issues, this study focuses on identifying only the primary material of each component, strategically ignoring less dominant materials. Considering the first assumption, it is assumed that the façade of an individual building can be segmented into at most two major materials.

Figure 4.3 Common building structures in our dataset.

4.2.3 Classes and annotations

In the research conducted by Ho et al. (2004), it was identified that mosaic and ceramic tiles are predominantly used as façade materials in domestic buildings within Hong Kong. This

preference is primarily attributed to their self-cleaning properties and the relatively low cost associated with their maintenance. In stark contrast, the architectural landscape of commercial buildings, particularly in central business districts, shows a preference for high-rise structures where glass and glass-mixed façades are favored. These materials are chosen for their safety features, aesthetic appeal, and ability to confer a modern and prestigious appearance to the buildings.

For the purposes of this study, a comprehensive dataset has been developed, within which nine distinct annotation classes have been defined. These include background, ceramic tile, glass, hybrid, metal, mosaic tile, paint, tree, and unidentified materials. The selection of these classes was strategically informed by several key factors: the reflectivity of the materials, their visual distinguishability, and the relative effort required for accurate annotation. It is important to note that classes comprising materials that were exceedingly rare in the dataset were intentionally omitted to maintain a focus on prevalent and representatively significant materials. Furthermore, including trees as a distinct category helps isolate structural materials from common occlusions in urban imagery, preventing the misclassification of vegetation as part of the building facade. Besides, separating trees enables potential downstream applications like quantifying urban greenery coverage while ensuring material-focused models prioritize architectural elements. The primary challenge in material identification from images lies in the visual distinguishability of the materials, which is particularly problematic in images captured from a distance. Reflectivity, although an important characteristic, does not always guarantee high visual distinguishability. This necessitated some compromise in terms of differentiating materials based solely on their reflectivity.

Specifically, as illustrated in Figure 4.4 (e), façades composed of small, closely packed tiles are common in residential areas. While these tiles might be made from a variety of materials such as ceramic, wood, or brick, their similar visual appearances make them difficult to differentiate and label individually. Consequently, for pragmatic reasons related to visual distinguishability, all façades featuring small tiles were collectively categorized under the *Mosaic tile* class.

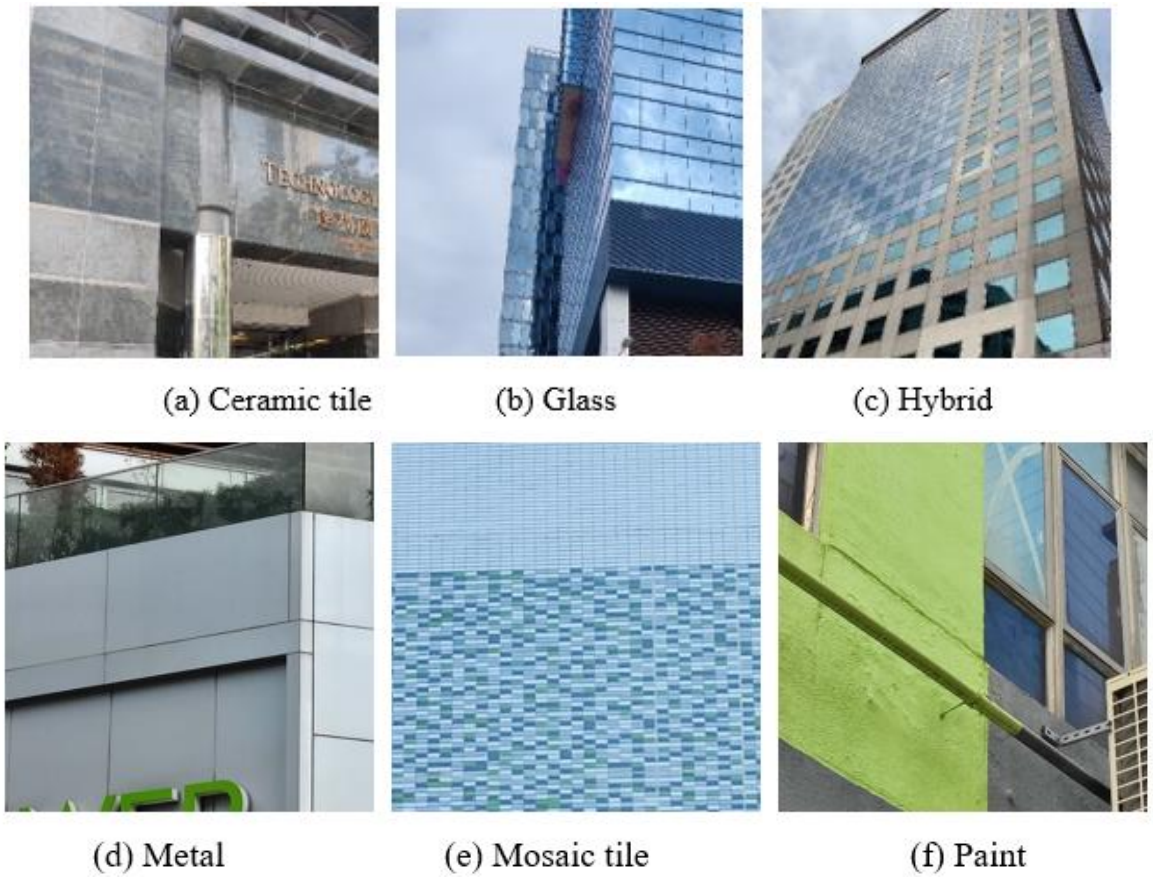

Figure 4.4 Examples of façade material pictures for each class.

Similarly, distinguishing between façades made from genuine marble and those made from marble-like ceramic materials poses significant challenges due to their similar appearances. As shown in Figure 4.4 (a), in this study, the classification '*Ceramic tile*' has been assigned to façades consisting predominantly of large tiles, regardless of whether they mimic marble textures. Furthermore, the *'Hybrid'* class, depicted in Figure 4.4 (c), includes façades that generally consist of a combination of glass and another material. The *'Metal'* category encompasses façades made from metal materials such as aluminum plates or other metal alloys, which are typically seen in commercial settings. *'Paint'* refers to façades that are primarily covered with ordinary paint coatings, commonly found in older residential communities. *'Glass'*, as shown in Figure 4.4 (b), is designated for typical glass-dominated office buildings.

In the annotation process, multiple resources, including field investigations and Google Street View, were utilized to provide a variety of perspectives, thereby aiding in the verification and

classification of façade materials. Images where material identification was particularly challenging due to visual ambiguities were labeled as '*Unidentified materials*.' This approach aims to ensure that the dataset not only reflects a high level of accuracy in material classification but also accommodates the inherent complexities in urban architectural environments.

4.3 Semantic Segmentation of Urban Building Surface Materials using Multi-Scale Contextual Attention Network

4.3.1 Architecture

In this study, we introduce a sophisticated multi-scale attention structure designed to capture and interpret contextual information contained within high-level features of images. This approach is instrumental in understanding the nuanced variations and general information in street views, which is critical for effective semantic segmentation in complex imaging environments. The proposed model, which incorporates innovative modifications to the existing Hierarchical Multi-Scale Attention (Hierarchical MSA) framework as described by Tao et al. (2020), includes the integration of Multi-Head Attention (MHA) mechanisms following the HRNet architecture, as well as enhanced attention mechanisms within OCRNet. These refinements are targeted at bolstering the model's capability to process and interpret complex feature sets for better segmentation accuracy.

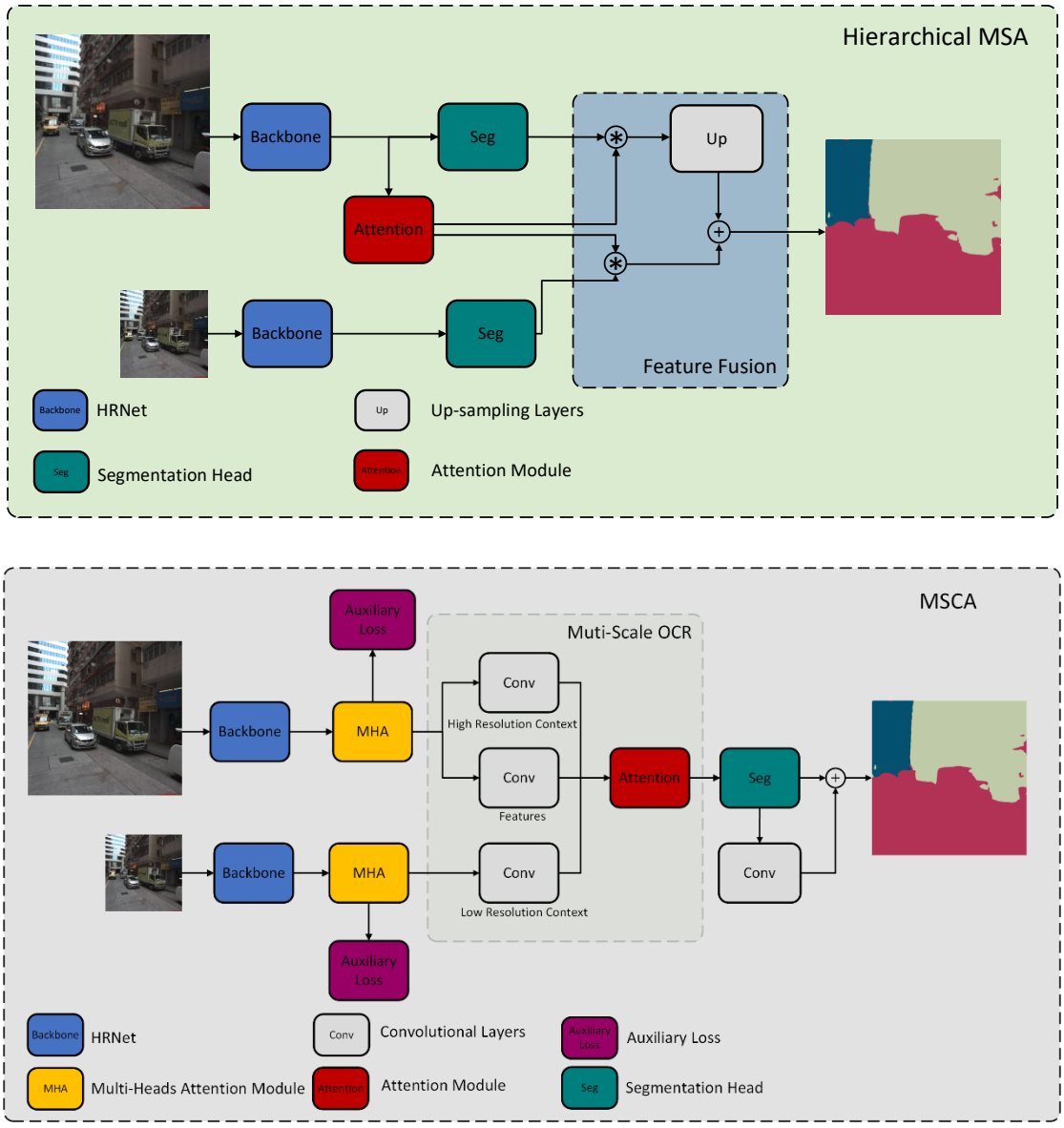

Figure 4.5 Network Architecture: Up and Down panels show Hierarchical MSA vs. MSCA (Ours) architectures, respectively.

To elucidate, the former modification embeds the extracted features into diverse representation subspaces. This embedding facilitates a broader and more detailed perspective on the semantic classes being analyzed, allowing the model to capture a more comprehensive set of feature interactions and dependencies. The subsequent enhancement, as illustrated in Figure 4.5, introduces a Multi-Scale OCR mechanism that empowers the network to adaptively concentrate on the salient semantic information at the feature level. Compared to

the Hierarchical MSA framework, which operates by combining features from different scales after the segmentation head and utilizes a weighted tensor derived from an attention head to modulate the feature fusion process, the proposed model optimizes the fusion of features before arriving at the segmentation head. This mechanism allows for a more integrated processing of contextual information across multiple scales.

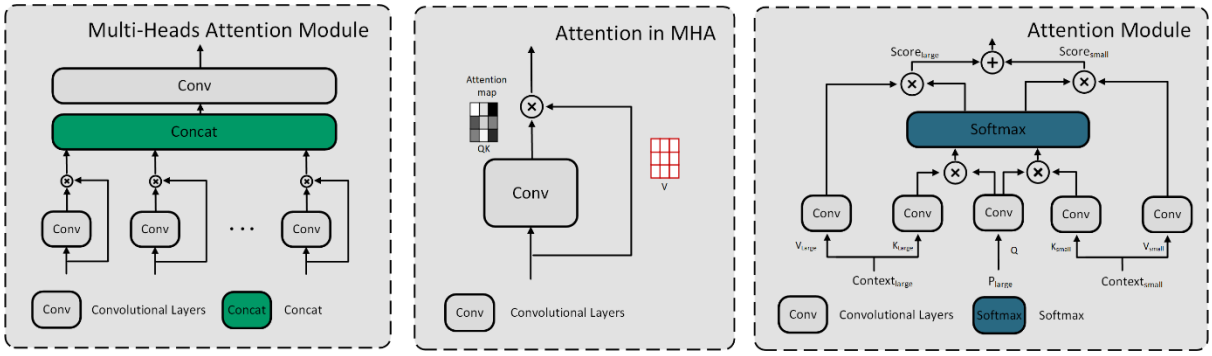

Figure 4.6 Attention Module: The details of the Multi-Heads Attention Module and Attention Module. Specifically, two different attention blocks are used in distinct modules.

Further detailing the model's architecture, as shown in Figure 4.5, the HRNet-W48 (Sun et al., 2019) serves as the backbone of our network. This decision is predicated on the exceptional ability of HRNet-W48 to retain fine-grained details from high-resolution imagery, which confers a significant advantage over other prevalent backbone architectures such as ResNet-101 (He et al., 2016). Next, the integration of MHA with HRNet further enhances the feature processing capabilities of the proposed model. In this step, MHA projects the backbone-derived features into different representation subspaces, with each 'head' in the MHA module interpreting semantic information in an individual manner. Preliminary experimental results have led us to configure the number of heads in the MHA to correspond with the number of semantic classes, thereby achieving optimal performance. The output generated by MHA is then utilized to compute the auxiliary loss, serving as a preliminary result.

Following the feature enhancement through MHA, the refined features from two distinct scales are transformed into query and value matrices. Concurrently, the high-resolution feature set is employed as the key vector, which is instrumental in calculating the attention scores that delineate the interdependencies of contextual information across various scales. The multi-scale attention mechanism within OCRNet is then employed to amalgamate crucial contextual information based on these scores, channelling it toward the segmentation head. In an effort to refine the segmentation outputs, we explored the possibility of deepening the segmentation head. However, this approach encountered certain limitations, such as exacerbated degradation issues and diminished accuracy. To circumvent these challenges, our study adopts a segmentation head that incorporates a residual block in lieu of a more profound structure. At the end of the network, the model yields an output tensor that represents a detailed probability map for each semantic class.

4.3.2 Multi-head attention

In this research, the methodology diverges from the conventional approach of directly leveraging the outcomes derived from the backbone architecture. Instead, it introduces an innovative mechanism that employs multiple attention functions to execute linear projections. This strategic modification enables the model to concurrently interpret and analyse the features extracted from various representations. The underlying principle of this approach is to enhance the model's ability to discern and prioritize relevant information from a complex dataset. The computational process is delineated as follows:

$$head_i = Attention_i(F_{backbone}) \tag{4-1}$$

$$MultiHead(F_{backbone}) = \rho\big(Concat(head_1, \dots, head_N)\big) \tag{4-2}$$

for each attention head, denoted as $head_i$, the operation $Attention_i(\cdot)$ is performed, where $F_{backbone}$ represents the feature matrix obtained from the backbone, characterized by its dimensions [batch size, 720, 256, 256]. As shown in the left part of Figure 4.5, multiple attention blocks are used to process $F_{backbone}$. The number of attention blocks, N, is 8, which has been tested to achieve optimal results in the preliminary stage of experiments. Then the

information from all attention heads is packed together by convolutional layers $\rho(\cdot)$ and trained jointly. Furthermore, the experimentation phase involved the utilization of two distinct types of attention blocks across different modules. The first type, employed within the Multi-Heads Attention Module, draws conceptual parallels to the attention mechanism proposed by Tao et al. (2020). Unlike traditional methods that rely on query and key matrices, this approach generates a dense mask directly from the features, which is then subjected to pixel-wise multiplication to derive the final output, as elucidated in the middle plate of Figure 4.5. By contrast, the attention in Multi-scale OCR is a self-attention module. This module is specifically designed to evaluate and integrate inputs across different scales. It accomplishes this by calculating attention scores for inputs of varying scales, subsequently amalgamating them following a softmax operation.

4.3.3 Multi-scale OCR

The Multi-scale Object-Contextual Representations (OCR) methodology is a sophisticated approach that systematically integrates contextual information at the feature level, derived from image features across multiple scales. This integration is crucial for enhancing the semantic segmentation capabilities of the network by providing a comprehensive understanding of the scene at different resolutions. The process begins with the extraction of features from a pair of input images, denoted as F_{large} and F_{small}, which are the outputs of a Multi-Head Attention (MHA) mechanism. These features encapsulate the visual information from the images at varying scales, with F_{large} representing the features from the larger scale image and F_{small} from the smaller scale.

The initial phase of the multi-scale OCR involves the computation of pixel representations from the large-scale features, F_{large}, using a convolutional operation defined by the function $g(\cdot)$. This operation is expressed mathematically as:

$$P_{large} = g\left(F_{large}\right) \tag{4-3}$$

Here, P_{large} is the pixel representation corresponding to the large-scale image features. The function $g(\cdot)$ is composed of a convolutional layer with a 3x3 kernel size, followed by a batch normalization layer and a ReLU activation layer. This sequence of operations is designed to capture the detailed textural information present in the large-scale features. Following the acquisition of pixel representations, the contextual information for the large scale, $Context_{large}$, is aggregated using the following equation:

$$\mathrm{Context_{large}} = \mathrm{f}\left(\mathrm{F_{large}}, \mathrm{P_{large}}\right) \tag{4-4}$$

In this context, $Context_{large}$ represents the object region representation within the OCRNet framework. The function $f(\cdot)$ employs a softmax operation to compute the probability distribution of each pixel representation P belonging to various object regions. This probability distribution is then element-wise multiplied with the feature matrix F, resulting in a contextually enriched feature representation that emphasizes the relevant object regions. The subsequent stage involves the application of an Attention Module, as depicted in Figure 4.5, to calculate attention scores and fuse features from different scales. The attention scores for the large and small scales are computed as follows:

$$Score_{large} = Attention\left(P_{large}, Context_{large}\right) \tag{4-5}$$

$$\mathrm{Score_{small}} = \mathrm{Attention}\left(\mathrm{P_{large}}, \mathrm{Context_{small}}\right) \tag{4-6}$$

In formulations, the network utilizes a self-attention mechanism to calculate the interaction scores between the pixel representations P and their respective contextual information $Context$. Given that P_{large} contains more detailed information compared to P_{small}, it is strategically used to compute both $Score_{large}$ and $Score_{small}$ with their corresponding contexts. As illustrated in the right section of Figure 4.5, P_{large} is employed to generate the query matrix, which is then used to estimate the attention scores by interacting with the key

and value matrices from both scales. The $MultiScaleFeature$ is then constructed by linearly combining these attention scores, with the equation:

$$\text{MultiScaleFeature} = \text{r} \cdot \text{Score}_{\text{small}} + \text{Score}_{\text{large}} \quad (4\text{-}7)$$

In this formulation, to maximize the utilization of fine-grained details, the network prioritizes $Score_{large}$ as the primary contributor to the $MultiScaleFeature$. In contrast, $Score_{small}$, which benefits from larger receptive fields, acts as an enhancement mask within this architecture, providing additional contextual breadth. A dropout function is introduced on the branch corresponding to $Score_{small}$ to prevent overfitting and to reduce the network's dependence on less distinct features from the smaller scale. In this context, r is a vector of independent Bernoulli random variables with a probability of 0.5, which serves as a regularization mechanism to promote model robustness.

The multi-scale OCR framework can be characterized as a process that generates contextual information from features at multiple scales and utilizes this information to compute a weighted output. The weights are determined by the relationships between the pixel representations and the region representations from the multi-scale contexts. This weighting mechanism ensures that the network assigns appropriate importance to different regions of the image, based on their relevance to the object categories of interest. By doing so, the multi-scale OCR effectively captures both the fine details and the broader contextual information, leading to a more accurate and contextually aware semantic segmentation.

4.3.4 Loss function

In this study, the cross-entropy loss function is employed as a fundamental component of the total loss. In the domain of machine learning and in tasks that involve classification, the cross-entropy loss function is a cornerstone metric for evaluating the performance of a predictive model. This loss function, also known as log loss, measures the dissimilarity between two probability distributions: the true distribution, as defined by the ground truth

labels, and the predicted distribution, as output by the model. Cross-entropy loss is particularly favoured in scenarios where the outputs can be interpreted as probabilities, as it is inherently designed to quantify the degree of uncertainty in these probabilistic predictions.

The cross-entropy loss function is mathematically articulated for a multi-class classification problem in following Equation:

$$loss = -\sum_{i=1}^{N} y_i \, log(p_i) \tag{4-8}$$

Here, (N) represents the total number of possible classes. The variable y_i is a binary indicator, which is set to 1 if the true class label corresponds to class (i), and 0 otherwise. The term p_i denotes the predicted probability that the given input is classified as belonging to class i. The cross-entropy loss function is adept at capturing the penalty for incorrect predictions, with the penalty escalating as the predicted probability deviates from the actual label. The logarithmic term in the equation ensures that the loss is sensitive to the confidence of the predictions, with highly confident but incorrect predictions incurring a greater penalty. This characteristic of the cross-entropy loss function is instrumental in guiding the model towards more calibrated and accurate probability estimates.

The total loss for the network is an amalgamation of the main loss and auxiliary losses, as specified by Equation:

$$loss_{total} = \alpha \cdot loss_{aux}^{s} + \beta \cdot loss_{aux}^{l} + loss_{main} \tag{4-9}$$

In this composite loss function, $loss_{main}$ signifies the cross-entropy loss computed on the final output of the network, serving as the principal training signal. The auxiliary losses, $loss_{aux}^{s}$ and $loss_{aux}^{l}$, are derived from the preliminary results S_s and S_l, respectively. These auxiliary components are introduced to provide additional gradient signals during the training process, which can be particularly beneficial in stabilizing the learning trajectory and enhancing the convergence of deep or complex networks. The gradient signal from the main

loss may attenuate as it propagates through multiple layers, and the auxiliary losses can help to mitigate this issue by reinforcing the gradient flow.

The coefficients α and β are hyperparameters that act as weighting factors for the auxiliary losses and are set to 0.5 in this study, in alignment with the methodology adopted by the Hierarchical MSA framework (Tao et al., 2020). The selection of these hyperparameters is pivotal, as they modulate the relative impact of the auxiliary losses in comparison to the main loss. By setting equal weights for both auxiliary losses, the study posits that the contributions of the small-scale and large-scale preliminary results to the learning process are of commensurate significance.

The auxiliary losses are computed using the same cross-entropy loss function as the main loss, ensuring a consistent optimization objective across the various components of the network. Employing the cross-entropy loss for both the main and auxiliary losses facilitate a harmonized approach to penalizing incorrect predictions and fosters the learning of precise class probabilities at all scales of the network's output.

In summary, the total loss function in this study is a linear combination of the main cross-entropy loss and two auxiliary cross-entropy losses. The auxiliary losses act as supplementary training signals, enhancing the robustness of the learning process across the network's hierarchy. This composite loss structure is meticulously crafted to optimize the network's performance by capitalizing on multi-scale information and ensuring comprehensive learning at all hierarchical levels of the network's architecture.

4.3.5 Experiments setups

In this section, we delve into the specifics of the training regimen and present a comprehensive analysis of the experimental outcomes. This study evaluates the performance of the proposed model by comparing it against the latest state-of-the-art algorithms using our Hong Kong street view dataset and the FaçadeWHU dataset. The comparative results

unequivocally demonstrate the superior performance of our model on both datasets. Additionally, ablation studies are conducted to dissect and scrutinize the contributions of each sub-module within our model. The findings from these studies substantiate the effectiveness and integral value of the proposed modules.

4.3.5.1 Training details

The development and experimentation of the proposed model were carried out using the PyTorch framework (Paszke et al., 2019). The computational experiments were performed on a high-performance server equipped with two TITAN RTX GPUs, enabling efficient processing of the large-scale data involved in this study.

The input images for the pipeline were processed at two distinct scales: 1.0x to capture finer details and 0.5x to provide a larger receptive field. This dual-scale approach ensures that the model can leverage both high-resolution details and broader contextual information. Due to the high computational cost associated with processing these images, the experiments involved cropping the images to dimensions of 896×896 pixels. The batch size was set to 2 per GPU to optimize the balance between computational efficiency and memory constraints.

Table 4.1 Details of experiment configuration.

Item	Configuration
Image scale	{1, 0.5}
Crop size	896x896
Batch size	2 per GPU
Learning rate	0.02
Optimizer	SGD
Learning rate scheduler	Polynomial
Power of learning rate scheduler	1
Minimum learning rate	0.0001
Loss function	Cross-Entropy Loss

For optimization during training, we employed the Stochastic Gradient Descent (SGD) algorithm with a momentum of 0.9 and a weight decay of 0.0001. This choice of optimizer is well-suited for our task, as it helps effectively navigate the complex loss landscape of deep neural networks. We conducted a comparative analysis of polynomial decay with power parameters of 1.0 and 2.0 for the learning rate scheduler. Based on this analysis, we selected a linear decay with power 1.0, which aligns with the methodology used by Tao et al. (2020). The specific configurations and hyperparameters used in our experiments are detailed in Table 4.1.

To rigorously validate the effectiveness of the proposed Multi-Scale Contextual Attention (MSCA) network, we conducted a comparative analysis with several renowned algorithms: DeepLabv3 (Chen et al., 2017), DeepLabv3+ (Chen et al., 2018), OCR (Yuan et al., 2019), and Hierarchical MSA (Tao et al., 2020). It is noteworthy that Hierarchical MSA achieved optimal results on the Cityscapes validation set, thus serving as a critical benchmark for our comparisons. In our experimental setup, both the baseline models and the backbone of the MSCA were pre-trained on the Cityscapes dataset and subsequently fine-tuned on our proposed dataset under similar configurations to ensure consistency and comparability.

4.3.5.2 Evaluation Metrics

The evaluation of the model's performance was conducted using several key metrics that are widely recognized in the field of semantic segmentation. Specifically, we selected mean Intersection over Union (mIOU), precision, recall, and F1-score to assess the experimental results quantitatively. These metrics provide a comprehensive view of the model's accuracy and reliability. Among them, the mIOU is calculated using the formula:

$$mIOU = \frac{1}{N+1}\sum_{i=0}^{N}\frac{TP_i}{FN_i+FP_i+TP_i} \tag{4-10}$$

In this equation, N denotes the number of classes, with $N+1$ including the 'background' class. The term TP_i stands for the true positives for class i, representing the number of

pixels correctly identified as belonging to class i. FN_i represents the false negatives, which are the pixels that belong to class i but were incorrectly identified as another class. FP_i denotes the false positives, which are the pixels incorrectly identified as class i when they actually belong to another class. Then, the precision for the multi-class scenario is computed as:

$$Precision_m = \frac{1}{N}\sum_{i=0}^{N}\frac{TP_i}{FP_i+TP_i} \tag{4-11}$$

Here, precision measures the proportion of true positive predictions among all pixels predicted to belong to class i. It evaluates the accuracy of the positive predictions. Recall for the multi-class scenario is given by:

$$Recall_m = \frac{1}{N}\sum_{i=0}^{N}\frac{TP_i}{FN_i+TP_i} \tag{4-12}$$

Recall measures the proportion of true positives among all actual pixels of class i, providing an indication of the model's ability to capture all relevant instances of the class. The F1-score, which is the harmonic mean of precision and recall, is calculated as:

$$F1-score_m = 2\cdot\frac{Precision_m Recall_m}{Precision_m+Recall_m} \tag{4-13}$$

The F1-score combines both precision and recall into a single metric, balancing the trade-off between the two to provide a more comprehensive evaluation of the model's performance for each class.

To ensure a balanced evaluation across all classes, we utilized macro-averaging (Sokolova et al., 2009) to calculate the mean values of these metrics. This approach aggregates the metrics by giving equal weight to each class, thereby avoiding biases that could arise from class imbalance. In most of the evaluations presented in this paper, background and unidentified materials were excluded from consideration to focus specifically on the model's ability to accurately segment and identify meaningful categories.

Furthermore, to explore the relationship between receptive field and model performance, we calculated the theoretical receptive field (TRF) of pixels at the network's output layer. The TRF represents the maximum area in the input image that can influence a pixel in a specific layer and is computed using the following equation:

$$r = \sum_{l=1}^{L} \left((k_l - 1) \prod_{i=1}^{l-1} s_i \right) + 1 \quad (4\text{-}14)$$

where r is the receptive field size of the network. k_l is the kernel size of layer l. s is stride. The actual impacted area, known as the effective receptive field (ERF), is typically smaller than TRF (Gu et al., 2021). The specific ERF depends on the information utilization ability by different networks.

4.4 Effect of Façade Albedo on Solar Potential Distribution in Different Urban Districts: A Case Study of Hong Kong

After obtaining extensive material information from the MSCA, this section introduces the proposed framework for comprehensively investigating the effect of albedo on solar PV potential distribution. The proposed framework is designed to provide a systematic approach for evaluating the influence of albedo on solar PV potential distribution, taking into account various factors such as building function, façade materials, and inter-building reflections.

4.4.1 Materials and methods

This section introduces the details of the evaluation framework that investigates the effect of albedo on solar PV potential distribution. The framework includes the deep learning pipeline which acquires real-world building reflectance information at a large scale, a methodology that converts segmentation output into reflectance and connects them to a 3D model, and the solar irradiation estimation that incorporates multi-reflection.

4.4.1.1 Solar Potential Estimation in Street Canyon

The procedure for evaluating the effect of albedo on solar potential estimation is comprehensively illustrated in Figure 4.7. The process begins with the segmentation outcomes obtained from previous work using the Multi-Scale Contextual Attention network, which are initially presented as 2D images. However, these 2D images cannot be directly utilized in a 3D GIS model, necessitating a crucial step of converting the pixel data from the image coordinate system to the Hong Kong 1980 Grid System. This transformation is essential for associating the segmentation results with the 3D model accurately. In the first step, image correction is performed to minimize distortion effects and ensure the accuracy of this conversion. Without this correction, pixels located at the periphery of the image would suffer from significant projection errors when mapped over distances extending tens to hundreds of meters. The corrected pixel coordinates, along with the viewpoint location in the HK80 System, are then used to construct 3D rays that extend from the viewpoint to infinity, intersecting the buildings in the urban environment. The next step involves identifying the first intersection point of these 3D rays with the 3D GIS model of the buildings. This intersection point represents the projection of the corresponding pixels onto the 3D model. Once all segmentation results are accurately projected onto the 3D buildings, the material category and RGB distribution of each building can be established.

Based on this material and RGB distribution data, the study proposes several albedo schemes to evaluate the impact of different materials on solar potential. The lower section of Figure 4.7 delineates three distinct albedo schemes. The first scheme, depicted in the left column, directly applies the identified material categories to assign different albedo values based on the segmentation results. However, recognizing the limited material categories that MSCA can determine, the second scheme, illustrated in the middle column, introduces a more nuanced classification. This approach further categorizes the initial material types into detailed subclasses, simulating a more complex urban environment. This refined classification aims to emulate the multi-reflections in diverse albedos, providing a deeper understanding of how materials with varied reflective properties interact in an urban context.

The third scheme, shown in the right column of Figure 4.7, takes a different strategy by assigning a constant albedo value to all surfaces within the study area, regardless of the material categories. This uniform albedo application serves as a control scenario, offering a baseline against which the other schemes can be compared.

After applying these three albedo schemes to the 3D models in parallel, the study proceeds to estimate the annual solar potential distribution, incorporating the effects of multi-reflection. This comprehensive evaluation not only enhances the precision of solar potential estimates but also allows for a detailed analysis of how different materials and their reflective properties influence the overall solar potential in an urban environment.

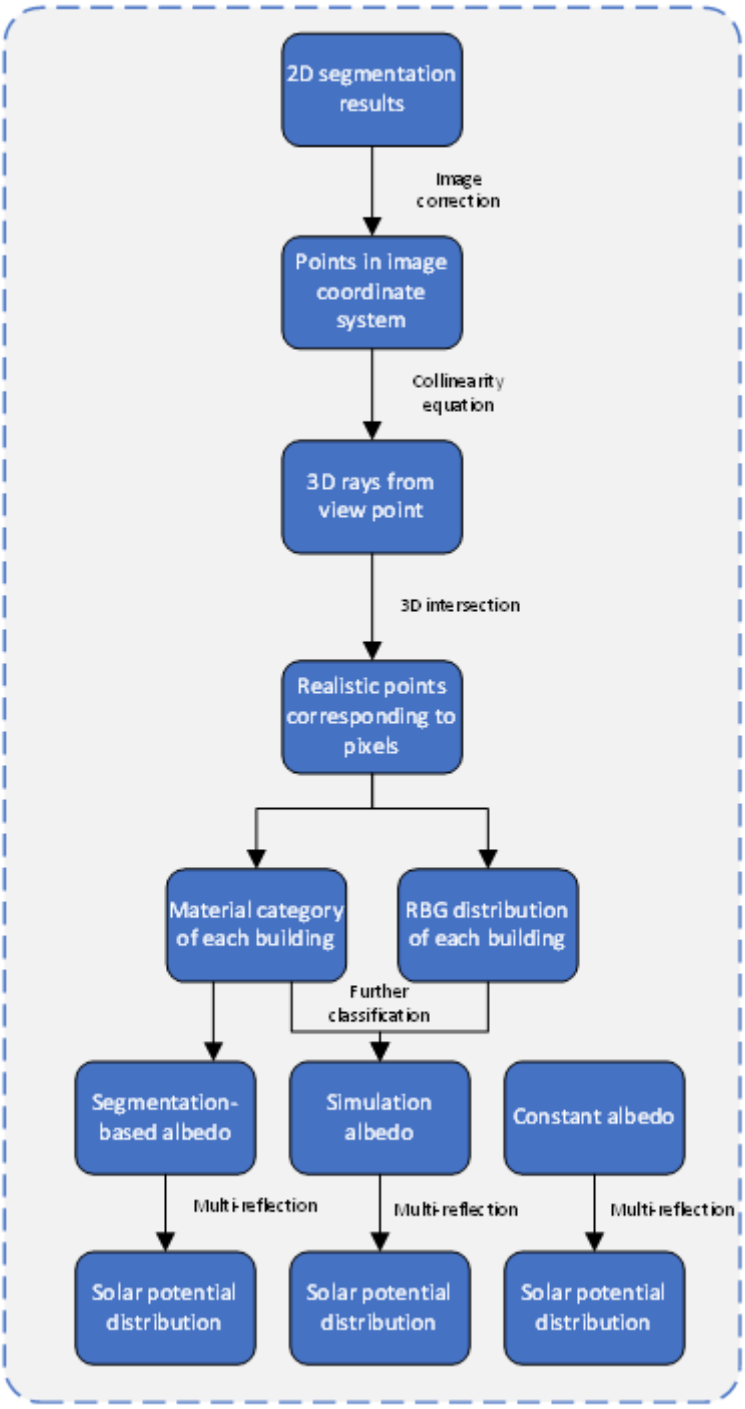

Figure 4.7 The procedure after segmentation.

4.4.1.2 Façade materials acquiring

To efficiently acquire material information for urban façades, this study utilizes the MSCA to identify the materials of building façades from street view images. In prior research, two foundational assumptions were established to streamline the segmentation process: 1) Each building is composed of no more than two primary components. 2) Each component consists predominantly of a single material. These assumptions are essential for simplifying the model to a manageable complexity while still maintaining a reasonable approximation of real-world conditions. Besides, this also enhances annotation efficiency at the expense of some accuracy. This approach is considered reasonable for the context of Hong Kong, where buildings typically feature a mix of upper residential sections and lower commercial sections, justifying the assumptions as a practical compromise.

Furthermore, since the original MSCA is trained on the dataset of Hong Kong street views, which is also the study area in this study, this research uses the same pretrained model and segmentation strategies to acquire the façade materials. Based on the reflectivity and visual distinguishability of different materials, façades are divided into ceramic tile, glass, hybrid, metal, mosaic tile, and paint. However, due to the complex reflection characteristics of materials and the restriction of visual manner, the original classification does not completely accord with the material albedos. Furthermore, the limited number of identified categories does not fully capture the diversity of albedo environments in urban settings.

Table 4.2 Further classification of facade materials and the characteristics of each subclass.

Class	No.	Color	Description
Ceramic	C1	brown	glazed
	C2	red	glazed
	C3	black	smooth
	C4	grey	uneven,new
	C5	grey	granite-like, polished
	C6	grey	granite-like, glazed
	C7	red	granite-like, glazed
	C8	red	granite-like, weathered
Metal	M1	grey	paint-sprayed, smooth
	M2	green	painted, smooth
	M3	grey	aluminium, shiny
Paint	P1	grey	concrete
	P2	grey	concrete, porous
	P3	grey	concrete, fine roughness
	P4	grey	concrete, smooth
	P5	grey	painted, smooth
	P6	brown	concrete
	P7	indigo	concrete
	P8	light grey	concrete
	P9	bronze	concrete
	P10	cedar	concrete
	P11	dark red	concrete
Hybrid	H1	grey	glass-ceramic hybrid, polished
	H2	grey	glass-ceramic hybrid, glazed
	H3	red	glass-ceramic hybrid, glazed
	H4	red	glass-ceramic hybrid, smooth
	H5	grey	glass-paint hybrid, concrete
	H6	grey	glass-paint hybrid, concrete
	H7	grey	glass-paint hybrid

Therefore, according to the three albedo schemes, which are applied to evaluate the effects of materials on solar potential, this study makes some adjustments and conducts a further classification on the segmentation result of MSCA. For the segmentation-based scheme, we merge the mosaic tile and ceramic tile because they have only minor distinctions in materials,

apart from significant differences in appearance. As a result, there are only five classes in this scheme (ceramic tile, glass, hybrid, metal, and paint). The classification method is basically based on the segmentation results of MSCA, which makes the materials identification and albedo assignment in this scheme have relatively high reliability, with 0.80 precision, 0.84 recall, and 0.82 F1-score.

Besides the segmentation-based scheme, this study proposes to use a further classification to simulate the complex albedo environment in urban areas. As shown in Table 4.2, excluding glass, the original four materials are further divided into 29 more detailed categories. Each subcategory in the table is visually distinguishable from each other, including differences in color or roughness. After obtaining the material from MSCA, the RGB information and segmentation results of each pixel in street views are projected to the 3D model. Based on the preliminary results, the RGB distribution of each building is collected and used to calculate similarity between other subclasses under the identified category (like ceramic tile, glass, hybrid, metal, and paint). This study measures the similarities between material distributions on building façades by calculating the Jensen–Shannon (JS) divergence, an important metric in information theory. The JS divergence is used to quantify the relative entropy or the difference in information content between two probability distributions. It is formulated as follows:

$$JS(P_1 \,||P_2) = \frac{1}{2}\, KL(P_1 \,||\frac{(P_1+P_2)}{2}) + \frac{1}{2}\, KL(P_2 \,||\frac{(P_1+P_2)}{2}) \tag{4-15}$$

where $KL(\cdot)$ is the Kullback–Leibler divergence. P_1 is the distribution of the building while P_2 is the distribution of each subclass from albedo library.

In this formula, the Kullback–Leibler (KL) divergence is a crucial component. P_1 is the distribution of the building while P_2 is the distribution of each subclass from albedo library. Furthermore, the $KL(\cdot)$ from a distribution p to a distribution q is defined by:

$$KL(p||q) = \sum_{i=1}^{N}[p(x_i)\log p(x_i) - p(x_i)\log q(x_i)]] \quad (4\text{-}16)$$

where $KL(\cdot)$ is the Kullback–Leibler divergence. P_1 is the distribution of the building while P_2 is the distribution of each subclass from albedo library. The KL divergence measures the relative entropy between two distributions, indicating how much one distribution diverges from a second, expected distribution. However, KL divergence is asymmetric and can yield infinite values under certain circumstances. The JS divergence addresses these limitations by averaging the KL divergences between each distribution and their midpoint distribution $\frac{(P_1+P_2)}{2}$, ensuring the measure is symmetric and finite. This makes it particularly useful for comparing the empirical distributions derived from the RGB values of building façades P_1 with the theoretical distributions from the albedo library P_2. By calculating the JS divergence, this study can effectively measure and compare the reflective properties of various materials, aiding in the accurate classification and analysis of urban façades.

4.4.1.3 Image correction

In urban façade analysis, particularly when using street-level imagery to identify and classify building materials, image correction is a crucial preprocessing step. Image distortion, often caused by the camera lens, can significantly impact the accuracy of projecting 2D pixel data onto a 3D model. The greater the distortion, the lower the accuracy, which can lead to substantial errors when estimating the locations and properties of building materials.

Image distortion is typically caused by imperfections in the camera lens, which can result in nonlinear mapping of the scene. The two primary types of distortion are radial and tangential. Radial distortion occurs when light rays bend more near the edges of the lens than at the

center, causing straight lines to appear curved. This effect is more pronounced in wide-angle lenses. Tangential distortion arises when the lens and the image plane are not perfectly parallel, causing some areas of the image to be shifted. The OpenCV library provides a powerful function, cv2.undistort, to correct these distortions. The function requires a camera matrix and distortion coefficients, which can be obtained through a camera calibration process. The undistort function works by transforming the distorted pixel coordinates back to their original, undistorted positions. The basic formula underlying this process is as follows:

$$img_{undistored} = undistort(img_{distored}, K, distCoeffs) \tag{4-17}$$

Where $img_{distored}$ and $img_{undistored}$ are the coordinates of the undistorted and distorted pixels, respectively. K is the input camera matrix, which can be formula as follow:

$$K = \begin{vmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{vmatrix} \tag{4-18}$$

$distCoeffs$ are the distortion coefficients (k_1, k_2, p_1, p_2, k_3). Image correction is a fundamental preprocessing step in urban façade analysis, especially when using street-level imagery. f_x and f_y represent the focal lengths of the camera in the x and y directions (measured in pixels). c_x and c_y are the coordinates of the principal point (optical center) in the image plane, defining where the optical axis intersects the imaging sensor. The undistort function is instrumental in reducing image distortion, thereby enhancing the accuracy of projecting 2D pixel data onto 3D models. By applying the image correction, this study ensures that the material classifications obtained from the MSCA network are precisely aligned with the 3D representations of buildings. This approach enables the RGB distribution and material identification result from multiple angles and distances street views to be accurately and reliably projected on corresponding buildings.

4.4.1.4 Data projection

After the image correction process, the segmentation results become qualified for projection onto the 3D GIS model. Image correction ensures that distortions caused by the camera lens or environmental factors are minimized, leading to more accurate segmentation outputs. These corrected images are essential for reliable projections because distortions can significantly skew the data, leading to errors in the final geographic information system (GIS) model. By correcting the images first, we establish a solid foundation for the subsequent steps.

The typical method to recover the actual coordinates from street views involves multi-image space intersection. This method relies on capturing multiple images of the same object from different angles to triangulate its position accurately. Each image contributes to a more precise calculation of the object's coordinates by intersecting the lines of sight from various viewpoints. This technique, widely used in photogrammetry, helps in creating detailed and accurate 3D reconstructions from 2D images.

However, the original street view images often encounter issues such as overexposure and occlusion. Overexposure can wash out critical details, making it challenging to identify features accurately, while occlusion occurs when objects in the foreground block parts of the scene. These issues necessitate screening out most unqualified images to maintain the integrity and accuracy of the data used for reconstruction. Only high-quality images that provide clear and unobstructed views are used for further processing.

This necessity for screening makes it difficult to ensure that there are sufficient consecutive photos for each filmed object to identify its location accurately. In practical scenarios, especially in dynamic urban environments, it is rare to capture perfectly consecutive and unshaded images of an object from all necessary angles. Consequently, the geographic coordinates of real-world points often need to be calculated using a single image situation, rather than relying on multiple images.

In this study, the Collinearity equation and 3D building models are utilized to achieve this objective. The Collinearity equation provides a mathematical relationship between the image coordinates and the corresponding real-world coordinates, making it possible to project points from a single image onto a 3D model. 3D building models, which contain detailed information about the geometry and position of structures, assist in this projection by offering a reference framework for calculating real-world coordinates.

Before applying the Collinearity equation, the pixels in pixel coordinates must be transformed to image coordinates. This step involves converting the pixel positions, which are typically given in terms of row and column numbers, into spatial coordinates (x, y) in the displayed image. This transformation is crucial because the Collinearity equation operates within the image coordinate system, not the pixel grid. The relationship between pixel coordinates and image coordinates can be seen in Figure 4.8:

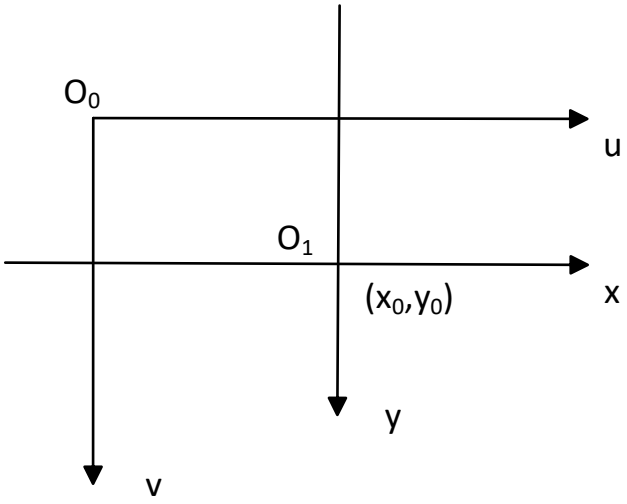

Figure 4.8 The relationship between pixel coordinates and image coordinates.

The image coordinate (x, y) of each pixel can be obtain by following equation:

$$x = (u + 1) * \left(\frac{du}{U}\right) - x_0 * \left(\frac{du}{U}\right) \quad (4\text{-}19)$$

$$y = -(v + 1) * \left(\frac{dv}{V}\right) + y_0 * \left(\frac{dv}{V}\right) \quad (4\text{-}20)$$

where u and v are the corresponding column and row of each pixel, respectively. (x_0, y_0) represents the coordinates of the principal point, which is the point where the optical axis of the camera intersects the image plane. du and dv denote the dimensions of a pixel in the image plane, while U and V are the total number of pixels in the horizontal and vertical directions, respectively. This transformation aligns the pixel grid with the spatial coordinate system used in the image, making it possible to apply geometric calculations accurately.

Based on the calculated image coordinates, the Collinearity equation can be utilized to locate the potential position of the corresponding actual points. The Collinearity equation is a cornerstone in photogrammetry and remote sensing, describing the geometric relationship between the image point, the projection center (camera), and the real-world point. It is formulated as follows:

$$x - x_0 = -f \frac{R_{11}(X-X_0)+R_{21}(Y-Y_0)+R_{31}(Z-Z_0)}{R_{13}(X-X_0)+R_{23}(Y-Y_0)+R_{33}(Z-Z_0)} \tag{4-21}$$

$$y - y_0 = -f \frac{R_{12}(X-X_0)+R_{22}(Y-Y_0)+R_{32}(Z-Z_0)}{R_{13}(X-X_0)+R_{23}(Y-Y_0)+R_{33}(Z-Z_0)} \tag{4-22}$$

where x, y are the image coordinates of the image point which converted from the rows and columns of pixels, x_0, y_0, f are the interior orientation parameters of the image, X, Y, Z are the geographical coordinates of the corresponding ground point, X_0, Y_0, Z_0 are the geographical coordinates of the projection center. The elements R_{ii}(i=1,2,3) are components of a 3×3-matrix R composed of three exterior orientation parameters, which describe the rotation of the camera relative to the ground coordinate system.

This formula describes a straight line formed by the image point, the projection center, and the real-world point. Specifically, by connecting these three points (image point, projection center, and real-world point), the formula establishes 3D rays that extend from the viewpoint to infinity. This ray serves as a crucial element in providing a precise means of tracing back

from the image to the physical scene and indicating the potential location of the real-world points.

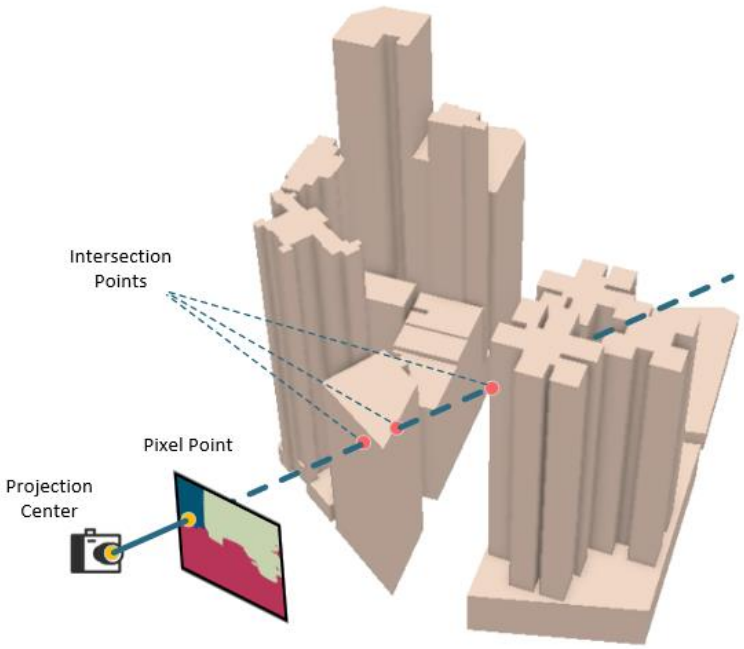

Figure 4.9 Using the collinearity equation to determine the geographic relationship between the camera, street views, and buildings.

As illustrated in Figure 4.9, this geometric concept can be visualized as a ray that extends from the camera's lens, passes through the image point on the plane, and projects outward into the real world, eventually intersecting with an object in the scene. This ray effectively represents the line of sight from the camera to the object, indicating the path that light has traveled. Given the coordinates of the image point and the known position of the projection center, it is possible to calculate this ray with high precision, enabling the determination of the exact spatial relationship between the captured image and the real-world environment.

This means that, given the image point and the projection center, we can obtain a ray pointing to the corresponding geographical point. By knowing the orientation and position of the camera (the projection center) and the position of a point in the image (the image point), we

can use geometric principles to trace this line into the three-dimensional space of the real world. The resulting ray serves as a precise indicator of the direction from the camera to the object, facilitating tasks such as 3D reconstruction and spatial analysis.

Then, based on the occlusion relationship between the camera and the object, the first intersection point obtained by using the ray and buildings is the corresponding real-world point. This intersection point is where the ray, extending from the camera through the image point, first makes contact with a physical object in the environment, thus pinpointing the geographical location of the real-world point with accuracy. By understanding the occlusion relationships in the scene, we ensure that the calculated real-world point corresponds to the visible surface in the image rather than a hidden or obstructed part of the scene. This approach enhances the reliability and accuracy of the spatial data derived from images under the lack of sufficient consecutive photos for each filmed object to identify its location accurately, which is crucial for various applications such as urban planning, navigation, and augmented reality.

4.4.1.5 Accuracy assessment of the projection

Since it is challenging to validate the mapping between all pixels, 3D model points, and real-world points, this study selected feature points like building vertexes to evaluate the accuracy of the projection pipeline. The decision to focus on feature points is driven by the need for precise reference markers that can be easily identified and measured. Building vertexes, being distinct and prominent, serve as reliable reference points for assessing the accuracy of the spatial projection. These points are evenly distributed throughout the study area to avoid spatial bias and systematically evaluate projection accuracy across diverse building geometries. However, due to the obstruction of buildings, the position of the mapping system, and limitations on camera perspectives, in most circumstances, only the middle of the buildings are captured by the camera. This inherent limitation arises from the fact that the camera’s field of view is often restricted by adjacent structures and the specific

angles at which the images are taken. Consequently, the number of captured building vertexes is limited, reducing the sample size available for evaluation.

Table 4.3 Projection accuracy evaluation.

Metrics	Value
Total points	35
Points projected onto the correct building	28
Average distance between the camera and objects	89.19m
RMSE	3.01m
Mean error	2.11m
Error per meter	0.07m

As shown in Table 4.3, 35 feature points are selected to assess the accuracy. Among these, 28 points are projected onto the correct buildings. This initial finding indicates a relatively high level of accuracy in the projection process, with the majority of the feature points correctly aligning with their corresponding real-world counterparts. The root mean square error (RMSE) of these 28 points is 3.01 meters, and the mean error is 2.11 meters. These metrics provide a quantitative measure of the projection accuracy, with the RMSE representing the square root of the average squared errors and the mean error representing the average absolute error.

Considering that most feature points are captured on distant buildings with an average distance of 89.19 meters, which is further than most regular points on street views, the actual error of overall projection points should be considered much lower than this figure. The greater distance of the feature points introduces additional challenges in maintaining projection accuracy, as slight angular deviations can result in larger positional errors.

Therefore, the observed errors are likely inflated due to the increased distances, and the actual error for closer points would be substantially lower. Furthermore, since the segmentation results are building-level and the albedo assignments are based on the distribution of the entire building, a few mislocated pixels on the boundary of buildings have a limited impact on the building material identification. The segmentation process, which groups pixels based on the overall appearance and material properties of the building, ensures that minor inaccuracies at the edges do not significantly affect the overall material classification. The albedo assignments, which determine the reflective properties of the building surfaces, are similarly robust to minor boundary errors.

Given the overall precision of the projection pipeline, the projection accuracy is deemed adequate for subsequent analyses. The root mean square error and mean error values, while indicative of some degree of deviation, are within acceptable limits for the purpose of albedo determination. In summary, the methodology employed in this study provides a reliable means of projecting feature points from images onto 3D models with an accuracy that effectively supports the subsequent solar potential estimation.

4.4.1.6 Albedo determination

This study employs three distinct façade albedo assignment methods to evaluate their effects on solar potential distribution in urban environments. The first method employs a constant albedo value to represent all types of architectural materials within the city, encompassing both façades and rooftops.

According to research conducted by Salleh et al. (2014), as well as Yaghoobian et al. (2012), the albedos of common materials used in urban areas are listed in Table 4.4. For instance, light roofs typically have an albedo range of 0.35 to 0.5, while dark roofs exhibit much lower albedo values, ranging from 0.08 to 0.18. Asphalt ground surfaces show a wide range of albedos from 0.05 to 0.3, highlighting the variability even within a single material type.

Concrete façades, commonly used in urban construction, have albedos ranging from 0.17 to 0.27. Brick façades, which are also prevalent in urban environments, have albedo values between 0.2 and 0.4, while Gypsum façades have relatively high albedo, which is assigned a value of 0.35.

The table indicates that the average albedo for these materials generally falls between 0.2 and 0.4. Zhu et al. (2020) noted that simulations yield optimal performance when the albedo is set to 0.4. Consequently, this study adopts 0.4 as an empirical parameter for the constant albedo scheme, which serves as the control group in the experiments.

Table 4.4 Albedos of common materials in urban areas.

No.	Material	Albedo
1	Light roof	0.35-0.5
2	Dark roof	0.08-0.18
3	Asphalt ground	0.05-0.3
4	Concrete façade	0.17-0.27
5	Brick façade	0.2-0.4
6	Gypsum façade	0.35

The second scheme uses the results of segmentation to assign albedos. After combining the mosaic tile and ceramic, façades are divided into five categories: ceramic tile, glass, hybrid, metal, and paint. There are plenty of studies on the albedos of the five materials. As shown in Table 4.5, (Ilehag et al., 2019) presents an urban spectral library consisting of collected in situ material spectra with imaging spectroscopy techniques in the visible and near-infrared (VNIR) and short-wave infrared (SWIR) spectral range, with 181 façades materials. Similarly, LUMA (Kotthaus et al., 2014) and LBNL (Levinson et al., 2005) present 74 and 87 material

spectra, respectively. Specifically, for glass, the International Glazing Database (IGDB) (Versluis et al., 2012) provides more than 5000 optical data for different glasses, including color, reflectance, emissivity, thickness, and so on. Based on the prevalent façade styles in Hong Kong, this study selects several materials from aforementioned libraries to represent the five categories. The albedos of the five categories in this scheme (ceramic, metal, paint, hybrid, glass) are set as 0.14, 0.31, 0.17, 0.28, and 0.13, respectively. The segmentation-based scheme is the experimental group for evaluating the effect of façade albedos in different districts.

Table 4.5 Material Albedo library.

Library	Size	Including Materials
KLUM	181	Asphalt, Brick (clay), Mortar, Ceramic, Concrete, Granite, Limestone, Metal, Plaster, Sandstone, Conglomerate, Wood
LUMA	74	Quartzite, Stone, Granite, Asphalt, Cement/Concrete, Brick, Roofing shingle, Roofing tile, Metal, PVC
IGDB	5000	Specular glazing
LBNL	87	Conventional and cool pigmented coatings

The third albedo assignment scheme also relies on segmentation results but further classifies façade materials to simulate the complex albedo environment present in urban areas. This detailed classification is essential to represent the diverse material properties found in cityscapes. For this purpose, the study selects 29 materials commonly seen in Hong Kong as subclasses. Each building material is matched to a subclass that exhibits the minimum Jensen–Shannon divergence within the initial segmentation category. The Jensen–Shannon divergence is a statistical method used to measure the similarity between probability distributions, ensuring that each material is classified as accurately as possible based on its spectral properties. Table 4.6 lists the albedos assigned to each subclass within the primary categories of ceramic, paint, metal, hybrid, and glass.

Table 4.6 Assigned albedo of each façade materials.

Class	Albedo	Subclass	Albedo	Class	Albedo	Subclass	Albedo
Ceramic	0.14	C1	0.15	Paint	0.17	P1	0.26
		C2	0.14			P2	0.43
		C3	0.14			P3	0.51
		C4	0.35			P4	0.52
		C5	0.09			P5	0.37
		C6	0.34			P6	0.04
		C7	0.23			P7	0.18
		C8	0.14			P8	0.21
Metal	0.31	M1	0.31			P9	0.33
		M2	0.2			P10	0.17
		M3	0.25			P11	0.12
Hybrid	0.28	H1	0.16	Hybrid	0.28	H5	0.24
		H2	0.28			H6	0.37
		H3	0.23			H7	0.3
		H4	0.18	Glass	0.13		

For instance, within the ceramic category, eight subclasses (C1 to C8) are identified with albedo values ranging from 0.09 to 0.35. The paint category includes eleven subclasses (P1 to

P11) with albedos varying from 0.04 to 0.52. The metal category has three subclasses (M1 to M3) with albedos from 0.20 to 0.31. The hybrid category, consisting of seven subclasses (H1 to H7), features albedos between 0.16 and 0.37. The glass category is represented with a single albedo value of 0.13. This detailed classification allows for a more nuanced simulation of the urban albedo environment, reflecting the real-world complexity of building materials. It provides a deeper understanding of how different materials and their reflective properties interact with sunlight in a densely built environment.

In summary, these three façade albedo assignment methods exhibit different levels of complexity and accuracy in representing the urban environment. The constant albedo method offers a simplified, uniform approach that serves as a baseline for comparison. The segmentation-based scheme provides more accurate material albedo assignments by leveraging advanced image processing techniques to obtain detailed façade information, thus potentially enhancing the precision of solar potential estimation. Finally, the third scheme introduces an additional layer of granularity by further classifying façade materials into specific subclasses and using statistical methods to simulate the intricate albedo environments found in urban areas. Collectively, these three methods contribute to a comprehensive evaluation of the effects of façade albedo on solar potential distribution, providing deeper insights into the interaction between urban materials and solar energy.

4.4.1.7 Solar irradiation estimation

To accurately determine solar radiation at specific times and locations, this study utilized the Point Solar Radiation toolbox in ArcGIS Pro to calculate solar radiation on horizontal surfaces (ArcGIS, 2019; Fu et al., 1999). The Point Solar Radiation toolbox in ArcGIS Pro is a sophisticated tool designed to provide precise calculations of solar radiation by taking into account various factors such as the angle of the sun, the time of year, and the geographical location. This tool is particularly useful for studies that require high accuracy in solar radiation data, such as those related to solar energy potential, climate studies, and

environmental monitoring. Cloud cover is one of the most significant factors affecting radiation in this function. The inclusion of cloud cover data from the Hong Kong Observatory adds another layer of accuracy to the calculations, as cloud cover can significantly affect the amount of solar radiation that reaches the Earth's surface. The 2023 monthly cloud cover data collected by the Hong Kong Observatory (Observatory, 2024) is applied to calculate the diffuse proportion and transmissivity through the following formula (Huang et al., 2008):

$$Diffuse = 0.20P_{clear} + 0.45P_{partlycloud} + 0.70P_{cloudy} \quad (4\text{-}23)$$

$$Transmissivity = 0.70P_{clear} + 0.50P_{partlycloud} + 0.30P_{cloudy} \quad (4\text{-}24)$$

where P_{clear}, $P_{partlycloud}$, and P_{cloudy} represent the proportions of days in clear, partly cloudy, and cloudy conditions, respectively. In this study, P_{clear}, $P_{partlycloud}$, and P_{cloudy} represent the cloud conditions over entire Hong Kong. By accumulating the hourly solar radiation calculated by the Point Solar Radiation toolbox, the annual solar radiation can be determined. This experiment consistently set the spatial resolution to 1 meter, including building façades, rooftops, and the ground. This high spatial resolution required significant computational resources. Therefore, PostgreSQL was used for spatial data intersection, occlusion calculation, shadow computation, and multi-reflection simulation. PostgreSQL is a powerful database management system that is well-suited for handling large datasets and complex spatial queries, making it an ideal choice for this study. Specifically, assuming an albedo of 0.4 for multi-reflection, the remaining radiation after three reflections is less than 7%. When the albedo was set to 0.2, the proportion dropped to 0.8%, and most radiation was emitted toward the sky after reflecting several times. Consequently, three iterations of reflection were simulated for each study area in the experiments.

4.4.2 Experiments setups

This study selects North Point as the research area due to its diverse range of building types and façade styles, including fully glazed office buildings, old residential buildings with

mosaic tile façades, and factories with painted exteriors. The selection of North Point for solar potential analysis balances architectural diversity and computational feasibility. At the same time, North Point contains representative building typologies (e.g., residential high-rises, outlying low-rise structures, and mixed-use commercial façades) and material variations comparable to other urban districts. Computational constraints further necessitated this localized focus: the analysis required about 3 months of processing on 16-core CPU clusters, with iterative ray-tracing simulations accounting for hourly solar angles, façade geometry, and shading effects. Scaling to Hong Kong's entire urban area (~1,100 km²) would likely extend computation time to 12–18 months. This trade-off balances the granularity against infrastructural limitations.

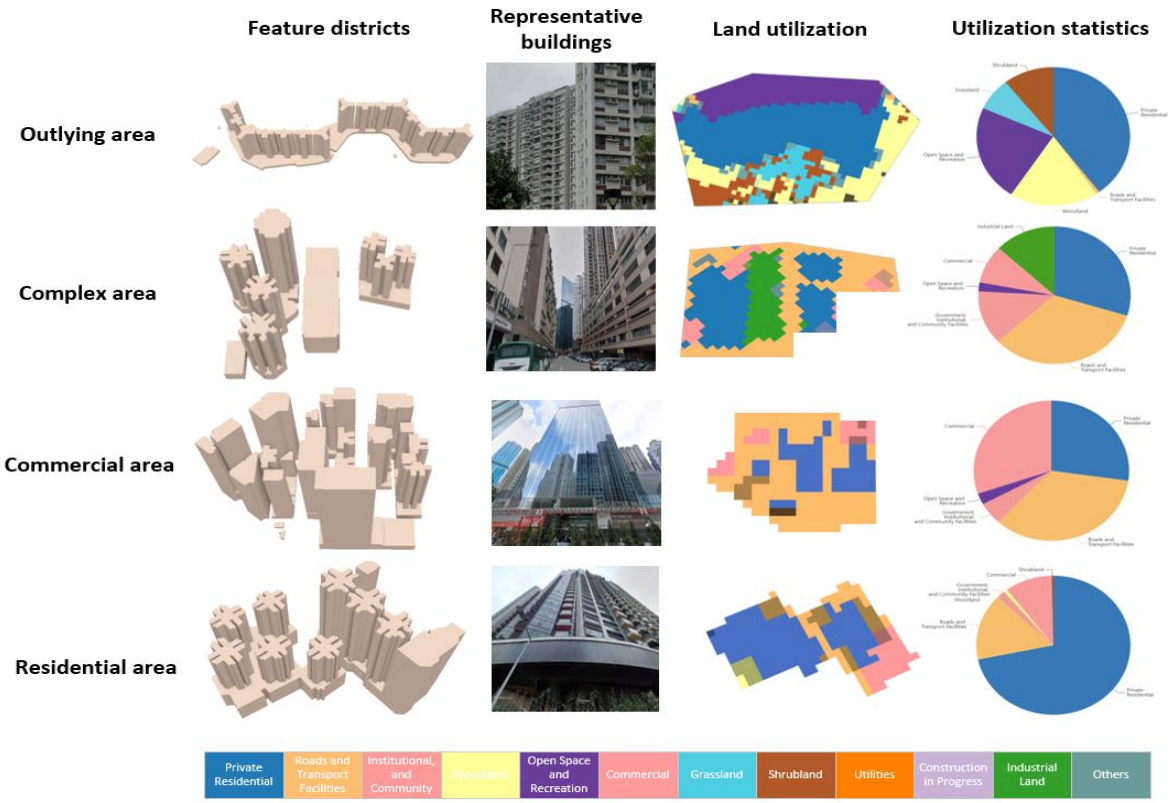

Figure 4.10 The selected districts and corresponding Land utilizations in the North Point. The first column on the left is the 3D models of selected districts. The second column is

sample street views. The third and fourth columns are the corresponding Land utilization statistics.

Based on these building styles, the thesis specifically chose four distinct neighborhoods for experimental comparison, as illustrated in Figure 4.10. These neighborhoods were selected to provide a representative sample of different types of buildings and façade materials found in North Point, thereby ensuring that the study's findings would be broadly applicable.

Area 1, depicted in Figure 4.10, which we refer to as 'outlying area' in this thesis, consists of isolated residential areas with buildings aligned in rows and no adjacent structures nearby. This region serves as a control group to investigate the impact of façade albedos on solar potential with minimal multi-reflection between buildings. The isolation of the buildings in Area 1 minimizes the effects of shadowing and reflection from neighboring structures, allowing for a clearer analysis of how different façade materials affect solar radiation absorption and reflection. The pie chart on the right of the first row in Figure 4.10 shows that, apart from the selected buildings, the majority of this area comprises woodland, open spaces, and grasslands. This composition of the surrounding environment further reduces the potential for complex interactions between buildings and their surroundings. Correspondingly, as shown in the 'outlying area' of Figure 4.11, the façades in this region are relatively homogeneous, predominantly consisting of ceramic tiles. This homogeneity simplifies the analysis and provides a baseline for comparing the effects of different façade materials in other areas.

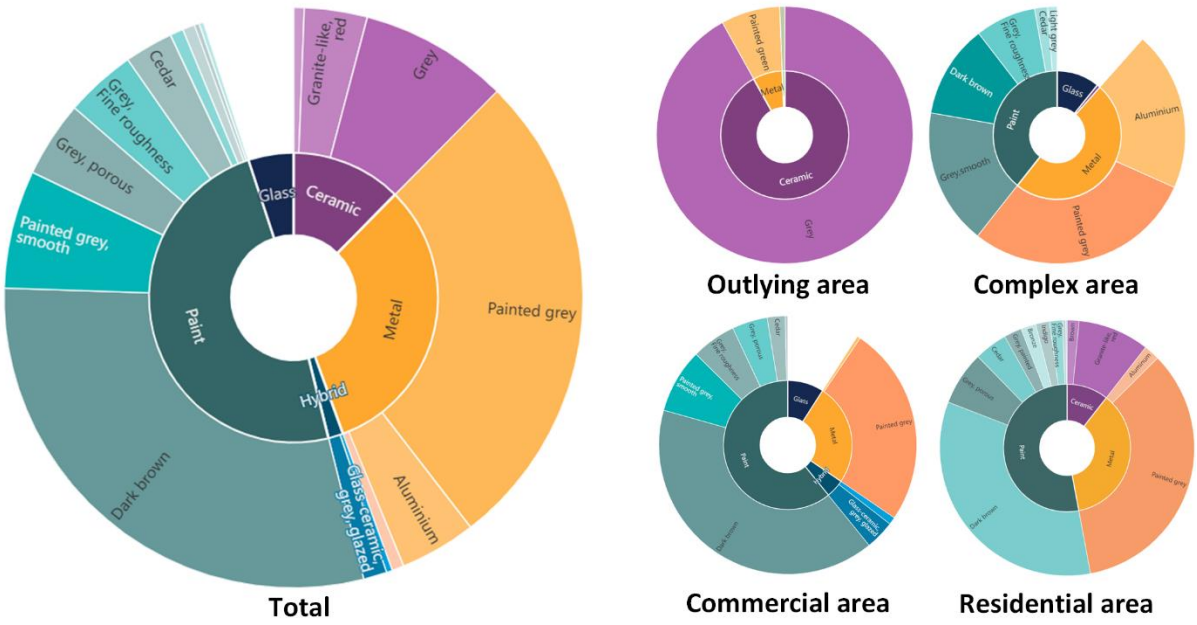

Figure 4.11 The percentage of different building façade materials in each district. The inner circle represents the material percentages obtained by the MSCA network. The outer circle represents the material percentages after further classification.

In contrast, Area 2 is more complex, which we refer to as 'complex area' in this thesis. Located in the northwest corner of North Point, this region includes various types of buildings with different functions and styles, such as private residences, government facilities, factories, and commercial buildings. The diversity of building types in Area 2 introduces a range of variables that can affect solar radiation, including differences in building height, orientation, and façade materials. As shown in Figure 4.11, the materials constituting the façades in this region are also more diverse, including a significant amount of metal (e.g., aluminum), coatings of various colors, and a substantial proportion of glass façades. This diversity in façade materials allows for a more comprehensive analysis of how different surfaces interact with solar radiation, including the effects of reflection, absorption, and transmission.

Similarly, Area 3, which we refer to as ‘commercial area' in this thesis, has a material composition similar to Area 2, with several additional hybrid buildings (e.g., a combination

of glass curtain walls and ceramic tiles). The presence of hybrid buildings in Area 3 introduces additional complexity to the analysis, as these structures often have varying albedo values and reflective properties. Compared to older residential areas, the buildings in Area 3 appear more aesthetically pleasing and modern. Modern architecture in commercial areas often incorporates advanced materials and design features that can significantly influence solar radiation dynamics.

Area 4, which we refer to as ‘residential area' in this thesis, predominantly consists of residential buildings, differing from Area 3, which is a mix of commercial and residential structures. The residential focus of Area 4 provides a contrast to the commercial emphasis of Area 3, allowing for a comparison of how different building uses affect solar radiation. Furthermore, due to the typical structure of buildings in Hong Kong, the lower floors of these residential buildings in Area 4 are often used for commercial purposes, such as shopping malls, which provide extra diversity in façade albedos. Overall, the selected regions encompass commercial, industrial, and residential buildings with façades made from more than twenty different reflective materials, providing a comprehensive basis for research and analysis. The wide range of materials and building types in the selected areas ensures that the study's findings will be applicable to a variety of urban environments.

Additionally, due to the high computational cost associated with a spatial sampling interval of one meter, this experiment conducted temporal sampling at intervals of every 28 days from January 1st 2023 to December 31st 2023.

Chapter 5 Results and discussion

5.1 Segmentation results of Urban Building Surface Materials

5.1.1 Quantitative experimental results

First, this study conducts a series of experiments using the proposed dataset to evaluate the performance of the developed model. As illustrated in Table 5.1, the results reveal that the overall performance of the proposed model is superior to other existing models, achieving a mean Intersection over Union (mIOU) of 72.58%. This notable performance underscores the efficacy of the proposed model in accurately segmenting architectural elements in diverse and complex visual datasets.

A critical aspect of this study is the analysis of the Theoretical Receptive Field (TRF) of various models. The TRF represents the area of the input image, which is a given feature in the output layer that can theoretically receive information. Although a larger receptive field should inherently enhance a model's performance by capturing more contextual information, our findings indicate that there is no significant linear correlation between the size of the receptive field and the model's performance in this task. This observation suggests that merely increasing the receptive field size does not necessarily translate to better performance in façade material identification. The dataset used in this study encompasses a wide range of building sizes, varying from several hundred pixels to approximately two thousand pixels. This variability presents a substantial challenge, requiring the model to effectively adapt to different scales of architectural elements. Given the discrepancy between TRF and ERF, which is the actual area that significantly influences the model's predictions, the study found that Hierarchical MSA and MSCA, which have the TRF closer to the building size, obtained better results. That makes aligning the TRF with the actual sizes of buildings crucial. Thereby, the multi-scale structure could enhance the model's ability to understand different level details.

Table 5.1 Performance of MSCA versus Baselines based on the constructed dataset. Best results in each class are represented in bold.

Method	Backbone	TRF	Ceramic	Glass	Hybrid	Metal	Mosaic	Paint	Tree	mIOU
DeepLabV3	ResNet-101	3459 * 3459	54.5	71.13	54.59	68.96	69.41	85.52	**80.58**	64.25
DeepLabV3+	ResNet-101	3583 * 3583	46.71	63.08	39.9	67.22	65.65	84.2	79.57	60.91
OCR	HRNet-W48	1087 * 1087	**59.48**	73.53	53.43	**74.12**	68.67	84.17	77.95	65.28
Hierarchical MSA	HRNet-W48	2302 * 2302	53.47	66.59	46.43	67.91	68.51	84.52	75.76	69.31
MSCA (ours)	HRNet-W48	2558 * 2558	55.4	**76.46**	**58.44**	64.48	**70.09**	**86.88**	75.95	**72.58**

In evaluating the model's performance across different material classes, it is evident that the MSCA outperforms baseline models in most categories, with notable exceptions being ceramic tile and metal. Specifically, for the metal class, the model achieves a mIOU of only 64.48%, which is the lowest among all material classes. This subpar performance is primarily due to the limited amount of training data available for metal façades. The scarcity of metal façade samples in the dataset hampers the model's ability to learn and generalize effectively for this category. However, it is worth noting that, as shown in Figure 5.1, most of the

misclassified metal pixels are labeled as background rather than being confused with other material classes. This suggests that, while the model struggles to identify metal façades specifically, it can still differentiate these regions from other material types, maintaining a reasonable level of accuracy at a broader classification level.

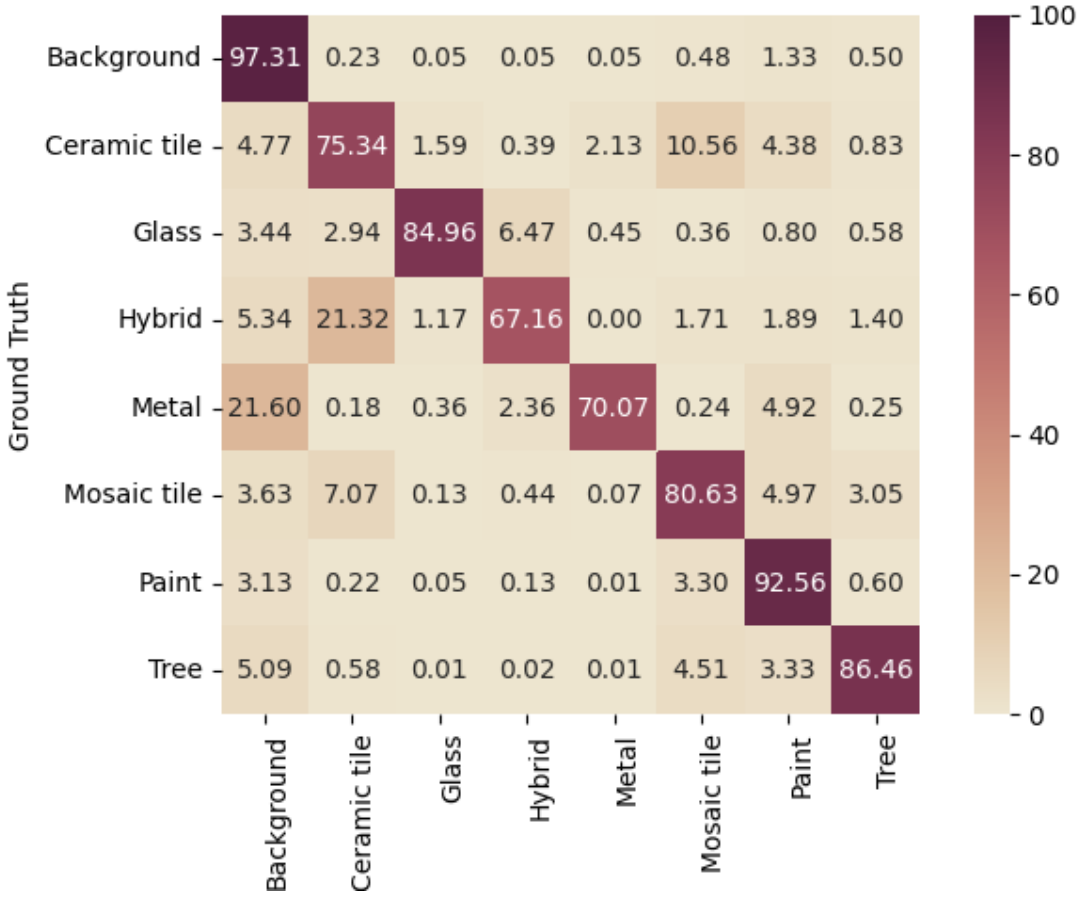

Figure 5.1 The percentage of pixels that are classified into different classes. Rows represent the total pixels of this material (Ground truth). Columns represent all pixels classified into this material (Predicted class).

Moreover, all models tested, including the proposed one, achieved their worst performance metrics in the hybrid and ceramic tile classes. For instance, the DeepLabV3+ model attains an IOU of 39.90% for hybrid and 46.71% for ceramic tile. In comparison, the proposed model performs better, achieving 58.44% and 50.40% for these classes, respectively. Despite these improvements, the performance remains suboptimal. However, the samples of hybrid and

ceramic tile are sufficient compared with metal. The poor performance can be attributed to the inherent ambiguity in the labeling of hybrid materials. The hybrid class, which typically includes a mixture of glass and ceramic materials, poses significant classification challenges due to its vague definition. The models seem challenging to classify a ceramic-like and glass-like object as the class hybrid. As depicted in Figure 5.1, the hybrid category is frequently misclassified as ceramic tile, further illustrating the difficulty in categorizing these mixed-material façades. This issue is compounded by the subjectivity involved in determining the proportion of materials in a hybrid façade. For instance, when the glass component constitutes 30% or 40% of the total material, rather than a clear-cut 50%, it becomes challenging to classify the façade accurately. This subjectivity leads to a significant proportion of glass façades being mislabeled as hybrids. 6.47% of glass façades fall into this misclassification, accounting for 43.02% of all misclassified glass pixels. Similarly, 21.32% of hybrid façades are incorrectly identified as ceramic tiles, representing 64.92% of the total misclassifications within the hybrid category. Consequently, the hybrid class exhibits the lowest precision, with a score of 0.67 as shown in Table 5.2.

Table 5.2 Metrics of the proposed method on the Hong Kong street views dataset.

Metrics	Ceramic	Glass	Hybrid	Metal	Mosaic	Paint	Tree	Mean
Precision	0.75	0.85	0.67	0.7	0.81	0.934	0.86	0.8
Recall	0.68	0.88	0.82	0.89	0.84	0.93	0.86	0.84
F1-score	0.71	0.87	0.74	0.78	0.82	0.93	0.86	0.82

The model's performance in the segmentation of mosaic tiles also highlights several challenges. As illustrated in Figure 5.2, factors such as erosion, fading, and the distance from which images are captured can obscure the distinctive grid patterns of mosaic tiles, making

them difficult to distinguish from metal, paint, and some ceramic tiles. The unique colors and patterns of mosaic tiles, which often serve as critical visual cues, can become less discernible due to these degradations. As a result, the network occasionally misclassifies mosaic tiles as painted façades. This misclassification is quantified in Figure 5.1, where 12.04% of mosaic tile pixels are incorrectly labeled as paint or ceramic, accounting for 62.16% of all misclassifications in the mosaic tile category. Similarly, 3.30% of painted façade pixels are

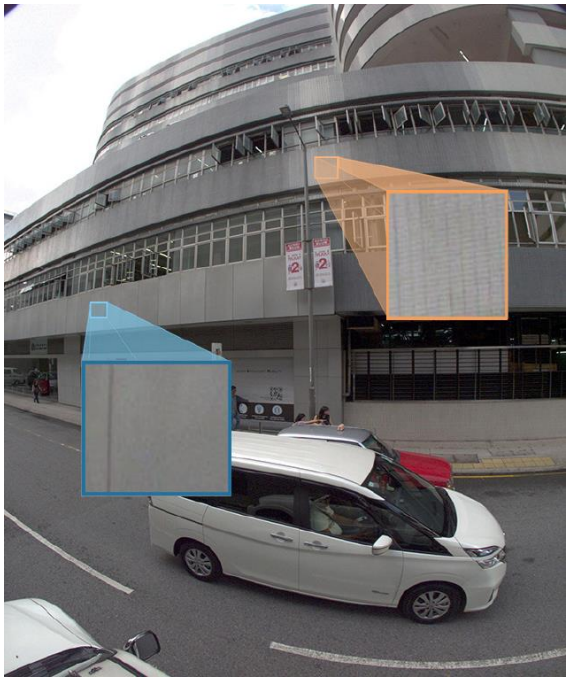

Figure 5.2 Two different materials have almost the same color and luster. The lower left material is metal, and the upper right is mosaic tile. The difference between the two materials in the picture is only reflected in the pixel-level details, i.e., mosaic tiles have grids.

confused with mosaic tiles, representing 44.35% of the misclassifications in the paint category. Despite these challenges, the model performs reasonably well in these categories, achieving F1-scores of 0.82 for mosaic tiles and 0.93 for painted façades, as detailed in Table 5.2.

The extensive training data available for painted façades plays a crucial role in the model's high performance in this category. Painted façades constitute the largest sample size within

the dataset and feature more colourful and varied appearances compared to other material classes. This abundance of diverse training examples enables the model to learn and generalize effectively, resulting in an mIOU of 86.88% for painted façades. The high recognition accuracy in this category underscores the importance of ample and diverse training data in developing robust segmentation models.

The study also presents a detailed percentage matrix in Figure 5.1, showing how ground-truth pixels are predicted across different classes. The matrix reveals a tendency of the proposed model to misclassify pixels as background, especially in categories with significant occlusions. For example, aside from the metal class, 3% to 5% of pixels in other categories are incorrectly labeled as background. This issue primarily arises from the occlusions present in street view images, such as advertisements and other obstacles, which are included as part of the façades due to the building-level annotation principle. These occlusions can confuse the network, leading to errors in identifying and excluding non-relevant objects on the façades.

The annotation principle itself also impacts the segmentation results, particularly for hybrid façades. As previously mentioned, hybrid façades are often composed of a mix of glass and other materials, typically ceramic, with an approximate 50-50 distribution. However, the actual proportion of materials can vary, and this variability can lead to subjective interpretations during the annotation process. For instance, a façade with 30% or 40% glass may not fit neatly into the hybrid category, leading to inconsistencies in the training data and subsequent misclassifications by the model. This subjectivity in labeling is reflected in the misclassification rates, where a significant portion of glass façades is incorrectly labeled as hybrid, and vice versa.

Table 5.3 Performance of MSCA versus Baselines based on FaçadeWHU. Best results in each class are represented in bold.

Method	Window	Door	Wall	Balcony	Roof	Shop	mIOU
DeepLabV3	42.78	19.08	**61.82**	29.15	43.93	19.53	44.27
DeepLabV3+	**45.4**	17.39	59.04	29.39	41.42	16.98	43.24
OCR	43.66	8.23	61.32	25.24	36.94	11.46	40.07
Hierarchical MSA	43.22	20.17	60.68	33.84	42.5	19.67	44.82
MSCA(ours)	44.68	**21.7**	61.26	**36**	**45.41**	**24.34**	**46.69**

To verify the effectiveness of the proposed model, this study also conducts experiments on FaçadeWHU. As shown in Table 5.3, the proposed model achieves the highest overall performance, with a mIOU of 46.69%, and outperforms the baselines in different classes, except for Window and Wall. Even in Window and Wall, MSCA is only 0.72% and 0.56% lower than the best model. Furthermore, compared with Wall, Roof, and Window, all methods have poor performances in Balcony, Shop, and Door. The best IOUs are only 36.00%, 24.34%, and 21.70%, respectively. As shown in Table 5.3, since Wall and Roof have the most expansive area, which makes them the most unlikely to be blocked by obstacles, the metrics of these categories are significantly higher than others. Window also has a relatively satisfactory performance due to its regular shape. The model performs worst in Shop and Door, with the lowest precision of 0.28. The potential reasons leading to the poor results could be the insufficient data volume and the indefinable object boundaries. The latter

requires a strong semantic comprehension ability of models. Nonetheless, the experimental results on two datasets show that MSCA can handle the façade segmentation in street-level images robustly and efficiently.

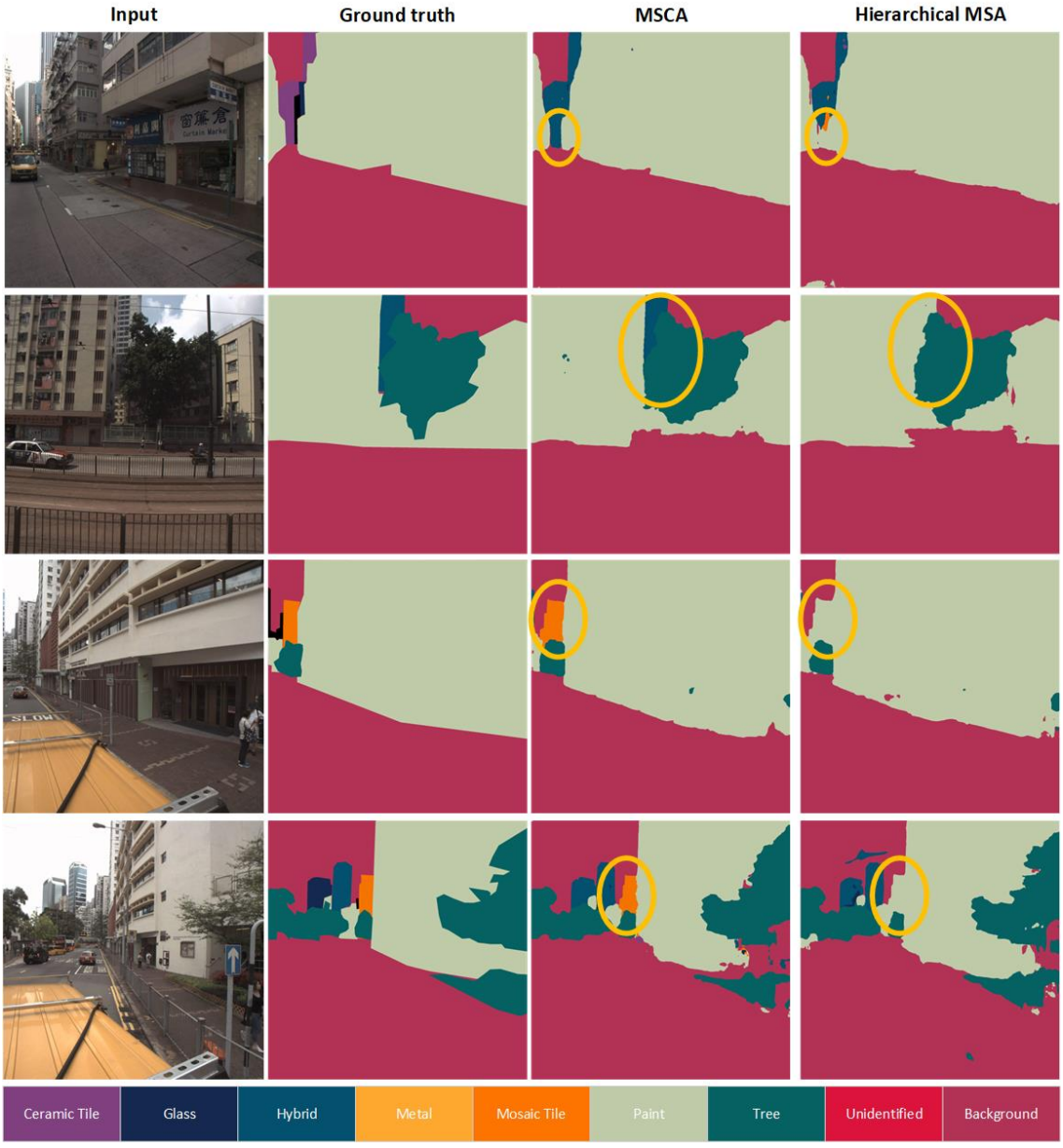

Figure 5.3 Qualitative comparison between MSCA and strong baseline (Hierarchical MSA). From left to right: input, ground truth, our method, and baseline.

5.1.2 Qualitative experimental results

According to Table 5.1, the Hierarchical MSA method outperforms other state-of-the-art approaches in façade segmentation. This demonstrates the efficacy of our proposed model. Furthermore, Figure 5.3 provides a visual comparison of the performance between our proposed method and the Hierarchical MSA. A significant distinction between the proposed model and the Hierarchical MSA is the adaptation of the attention module within the Object Contextual Representation (OCR) module. This adaptation facilitates the fusion of features across multiple scales, leading to improved segmentation performance. The qualitative results illustrate this advantage clearly. As shown in the first row of Figure 5.3, ignoring whether the pixels are correctly inferred as its categories, the proposed model successfully detects the commercial building with the ceramic façades (left in the figure, coloured in light blue) as a separated building. On the contrary, the baseline model only recognizes the upper part of the building and regards it as hybrid (colour in light blue), the same as MSCA. The lower part is reckoned as part of other residential buildings as the environment near the ground is more complex. The baseline's failure to accurately segment the lower part of the building, likely due to the complexity of the ground-level environment, underscores the superior contextual understanding of our model.

Similarly, in the second row, the hybrid building behind the residential one (middle in the figure, coloured in light blue) only shows a limited part of the picture. Hierarchical MSA thus fails to distinguish the two buildings, while MSCA correctly classifies the material and maintains its integrity.

Further, in the third row, the architectural style of buildings on the left side (coloured in orange) is similar to that on the right, presenting a challenging scenario for segmentation models. Despite the similarity, our proposed model successfully separates these buildings, whereas the baseline model fails to do so. This demonstrates the enhanced feature discrimination power of our model, particularly in complex urban scenes. Similarly, in the

fourth row, the proposed model accurately identifies a mosaic façade in the middle of the image (coloured in orange) as an independent structure. The baseline model, however, struggles with this task due to the intricate details and similar visual patterns shared by adjacent structures.

These qualitative results collectively indicate that the hierarchical nature of the MSA allows the model to capture both global contextual information and fine-grained details. This multi-scale approach ensures that the model can effectively segment large, homogenous areas such as walls and roofs, as well as smaller, intricate details like windows and decorative elements. The ability to integrate information across different scales is particularly beneficial in urban environments, where buildings exhibit a wide range of architectural features.

5.1.3 Ablation Study

This section conducted a series of ablation studies to demonstrate the effectiveness of various modules integrated into our network architecture. Specifically, the study focused on four significant modifications to the basic HRNet+OCRNet structure: the incorporation of a multi-scale approach, the addition of a Multi-Head Attention module post-HRNet, the integration of attention within the OCRNet, and the implementation of a residual block at the network's end. These modifications were designed to enhance the model's performance in the specific task of façade segmentation. Among them, the effectiveness of the attention within OCRNet is proved by comparing it with Hierarchical MSA, which adopts the attention module after OCRNet.

Table 5.4 presents the results of the ablation studies, beginning with the adoption of a single-scale pipeline without the MHA module. This initial configuration achieved a mean Intersection over Union (mIOU) of 70.43%, which is 2.15% lower than the proposed multi-scale structure. The subsequent introduction of a multi-scale network architecture provided a performance boost, raising the mIOU by 0.87% over the single-scale pipeline.

This increase underscores the importance of utilizing features across different scales to enhance the comprehension of contextual information. The introduction of a multi-scale network architecture resulted in a significant performance boost. Specifically, it provided an increase of 0.87% in mIOU over the single-scale baseline. This improvement highlights the value of incorporating features from multiple scales, which enriches the contextual understanding necessary for accurate segmentation.

Table 5.4 Quantitative results of the ablation studies.

Ablation	Multi-Scale	MHA	Residual block	mIOU
I			√	70.43
II	√		√	71.3
III		√	√	70.27
IV	√	√		71.61
MSCA	√	√	√	72.58

Furthermore, the impact of MHA, when applied in isolation, was found to be less beneficial. In setting 3, the model incorporating MHA yielded an mIOU of 70.27%, which represents a slight decrease of 0.16% compared to the single-scale baseline. Although this reduction is minor, it indicates that merely adding MHA does not significantly enhance performance. This finding suggests that the integration of MHA needs to be orchestrated with other network components to be effective. Further analysis of the ablation study IV revealed the impact of adding a residual block at the network's end. This modification resulted in a 0.97% increase in mIOU, suggesting that the residual block could be helpful in fine-tuning the preliminary output of the network. The residual block likely aids in refining the features, thereby enhancing the overall segmentation accuracy.

The cumulative results from these ablation studies indicate that each of the proposed modifications contributes to the network's improved performance. The multi-scale approach, attention within OCRNet, MHA, and the residual block each provide unique enhancements that, when combined, result in a robust and effective model for façade segmentation. Specifically, incorporating these modules into the network led to gains of 2.15%, 1.28%, 2.31%, and 0.97% mIOU over the baseline settings, respectively.

5.1.4 Discussion

This study introduces a multi-scale contextual attention network designed to address the dual challenges faced in urban façade analysis: the need for detailed material classification based on spectral characteristics, and the necessity for comprehensive contextual understanding to maintain the integrity of larger structures, such as entire buildings. Hong Kong, with its dense urban environment and diverse architectural styles, was selected as the research site for this study. To effectively evaluate the performance of the proposed model, we developed a detailed street-level dataset tailored to the unique characteristics of Hong Kong's urban landscape.

The experimental results demonstrate that our model excels in accurately classifying building materials, outperforming other existing models. This achievement is particularly significant given the demand for precision in material classification within urban studies. A detailed understanding of materials' spectral characteristics can provide crucial insights into various urban phenomena, such as energy efficiency, thermal regulation, and aesthetic qualities. Our model's ability to balance intricate detail with broader contextual comprehension is facilitated by the innovative use of multi-scale contextual attention mechanisms.

The implications of this study extend beyond the task of imagery analysis from street views. One of the most substantial contributions of this research is its potential to bridge domain gaps in façade information collection. Traditional methods often struggle with inconsistencies

and insufficiencies in data collection, leading to gaps that can hinder comprehensive urban analysis. By providing a reliable and extensive data source, our method could enhance the accuracy and comprehensiveness of urban albedo studies, which are critical for understanding and mitigating urban heat islands and optimizing energy consumption.

The spatial information derived from our model can be seamlessly integrated into three-dimensional Geographic Information Systems (3D GIS). This integration significantly improves the accuracy of solar potential simulations, a critical component in the planning and deployment of solar energy systems. One of the persistent challenges in solar potential estimation is the accurate simulation of reflected light. Existing models often either ignore reflected solar radiation or use a uniform albedo to represent entire urban areas, which can lead to substantial inaccuracies. Our study offers a pathway to more precise and context-specific simulations by providing detailed albedo data for various urban surfaces, including rooftops, façades, and ground areas.

This detailed albedo information can refine strategies for photovoltaic (PV) deployment. For example, using the model proposed by Zhu et al. (2022), our work can enhance understanding of the relationship between urban morphology and solar capacity by incorporating accurate albedo measurements into the simulation framework. This approach allows for more effective and targeted PV deployment strategies, optimizing solar energy capture and reducing energy consumption in urban areas.

However, despite these advancements, several limitations must be acknowledged. The high cost of annotation necessitated certain compromises in this study. This study assumed that each building is composed of no more than two primary materials, simplifying the annotation process but introducing inconsistencies when dealing with buildings featuring complex or novel designs, such as theatres or museums. Moreover, the building-level annotation approach results in vague classifications, particularly for hybrid façades, which can lead to confusion during the classification process.

Another significant limitation relates to the complex reflection characteristics of materials. Our study was unable to classify façade materials strictly based on their reflectivity due to these complexities and the limitations of visual observation. This shortcoming could potentially limit the application of our work in solar potential estimation, as accurate albedo differentiation for each building is crucial.

Future research should focus on achieving more fine-grained classifications to enhance the precision and applicability of the model for solar potential simulations. Additionally, integrating the proposed model with other urban data sources, such as thermal imaging and LiDAR, could provide a more comprehensive understanding of urban environments. This multidisciplinary approach would further enhance the precision and utility of urban simulations, supporting more effective and sustainable urban planning initiatives. Moreover, the integration of our model into broader urban analytics frameworks can facilitate more nuanced and dynamic urban planning processes. For instance, by combining the detailed façade material classifications with environmental and socio-economic data, city planners can develop more informed strategies for energy distribution, building retrofits, and sustainable urban development. This holistic approach can contribute to the creation of smarter, more resilient cities capable of adapting to the evolving challenges of urbanization and climate change.

In conclusion, the proposed multi-scale contextual attention network effectively identifies material categories from street-level images in metropolitan settings like Hong Kong. This work not only provides a viable solution for the precise simulation of reflective radiation accumulation processes but also explores the potential for conducting detailed urban analyses through street-level imagery. The ability to accurately classify materials and understand their spatial distribution opens up new avenues for urban planning and energy efficiency studies, particularly in the context of solar energy deployment and the mitigation of urban heat island effects.

5.2 Effect of Façade Albedo on Solar Potential Distribution in Different Urban Districts

5.2.1 Estimation based on different albedo schemes

As shown in Figure 5.4, the distribution of annual solar potential across the four study areas under the segmentation-based albedo assignment strategy is illustrated. The figure reveals that, compared to the vertical surfaces of buildings, the points on horizontal surfaces exhibit higher solar potential across all study areas. This is attributed to the fact that horizontal surfaces can evenly receive sunlight from east to west throughout the day and experience fewer obstructions, thereby maximizing solar exposure. In contrast to the solar potential distribution at 3 PM on August 13, depicted in Figure 5.5 for Area 4, where noticeable solar potential concentration occurs due to reflected sunlight, such phenomena are less apparent in the annual solar potential distribution. This discrepancy arises because the distribution of reflected light varies and is uneven at any specific moment. No particular area is consistently illuminated by reflected solar radiation throughout the day. Even at the same time on different days, variations in the sun's azimuth angle and altitude cause the positions of reflected light from the buildings, which have fixed spatial relationships, to differ. Consequently, the indirect components are unevenly distributed across the study areas on an annual scale. Furthermore, since reflected components do not dominate the overall solar potential distribution, their impact becomes less noticeable in annual-scale visualizations. Figure 5.6 further demonstrates that, regardless of the study area or albedo allocation strategy, the proportion of solar potential contributed by direct sunlight consistently ranges between 77% and 90%, constituting majority of the total solar potential.

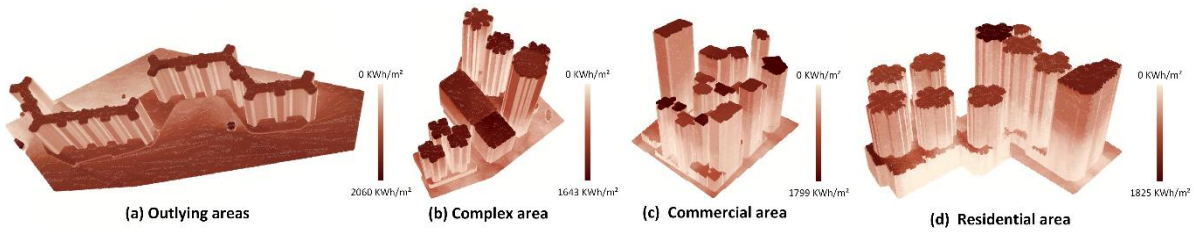

Figure 5.4 The distribution of annual solar potential across the four study areas under the segmentation-based albedo assignment strategy.

However, it does not diminish the role of façade albedo in the total solar potential distribution. Even without considering multiple reflections between buildings, the retained irradiation is still determined by the building's albedo. Then, the redistribution of the reflected irradiation is significantly influenced by the interaction between buildings, which is dominated by their spatial relationship and albedos. As shown in Figure 5.6, in Area 1, the total solar potential under the simulation albedo scheme is 9.1% higher than that under the constant albedo scheme. In Area 2, the segmentation-based scheme results in an 8.0% higher total solar potential compared to the constant scheme. For Area 3, the simulation albedo scheme surpasses the constant scheme by 8.3%, and in Area 4, the segmentation-based scheme exceeds the constant scheme by 8.9%. If multiple reflections are not considered, the differences between various albedo assignment strategies would be even more pronounced. The maximum differences between the different strategies across the regions are 11.9%, 15.6%, 17.8%, and 14.0%, respectively. This is because the albedo strategy, in conjunction with the spatial relationship between buildings and the sun, determines the initial distribution and subsequent redistribution of solar potential.

Figure 5.5 The distribution of solar energy potential in Area 4 at 3 p.m. on August 13th. The purple circles represent the concentration of solar potential caused by the reflection of sunlight.

Specifically, the impact of different façade albedo on solar potential distribution is illustrated in Figure 5.7. This figure highlights the differences in annual solar potential distribution resulting from varying façade albedo strategies. Figure 5.7(a) compares the annual solar potential under the segmentation-based strategy to that under the constant strategy. The primary differences are observed in the façades, while the rooftops and ground surfaces maintain consistent albedo values. Apart from insignificant visual differences due to multiple reflections, the horizontal surfaces exhibit nearly zero discrepancy. This can be seen from the predominantly purple points on the horizontal surfaces in the image, which indicates negligible differences. Points closer to blue represent negative differences, while points approaching yellow indicate positive differences, with a shift toward yellow signifying more significant differences.

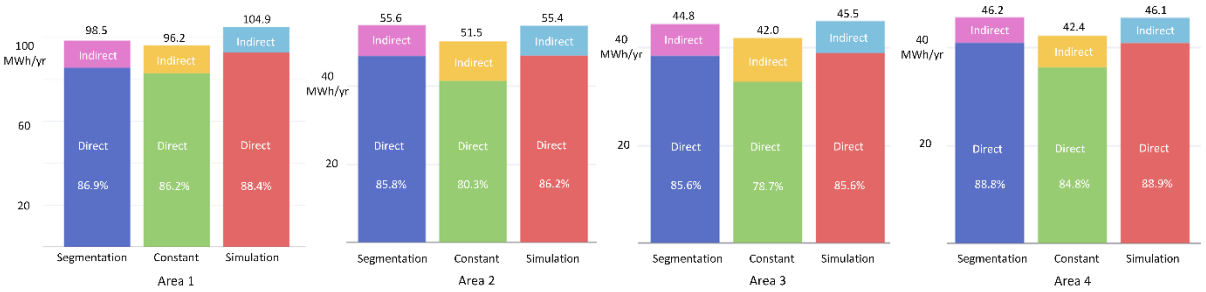

Figure 5.6 Comparison of the total annual solar potential of each study area under different albedo assignment strategies. 'Direct' refers to the part of solar irradiation that comes from direct sunlight. 'Indirect' represents the indirect components.

Figure 5.7 shows that in Area 1, the 'outlying area', the differences in solar potential distribution across the three strategies are relatively evenly distributed. This is because Area 1 consists of isolated buildings with minimal obstructions in the north-south direction, providing similar lighting conditions and consistent building materials for façades. As a result, the east-west distribution of buildings has little impact on solar potential. The favorable lighting conditions result in significant and uniform differences in solar potential from the upper to lower levels on the sun-facing sides of buildings. Compared with the distribution under the constant scheme, both the segmentation-based and the simulation schemes believe that based on the façade material of Area 1, in the actual distribution, the façade of this area should exhibit higher solar energy potential. This difference becomes more pronounced with more abundant sunlight.

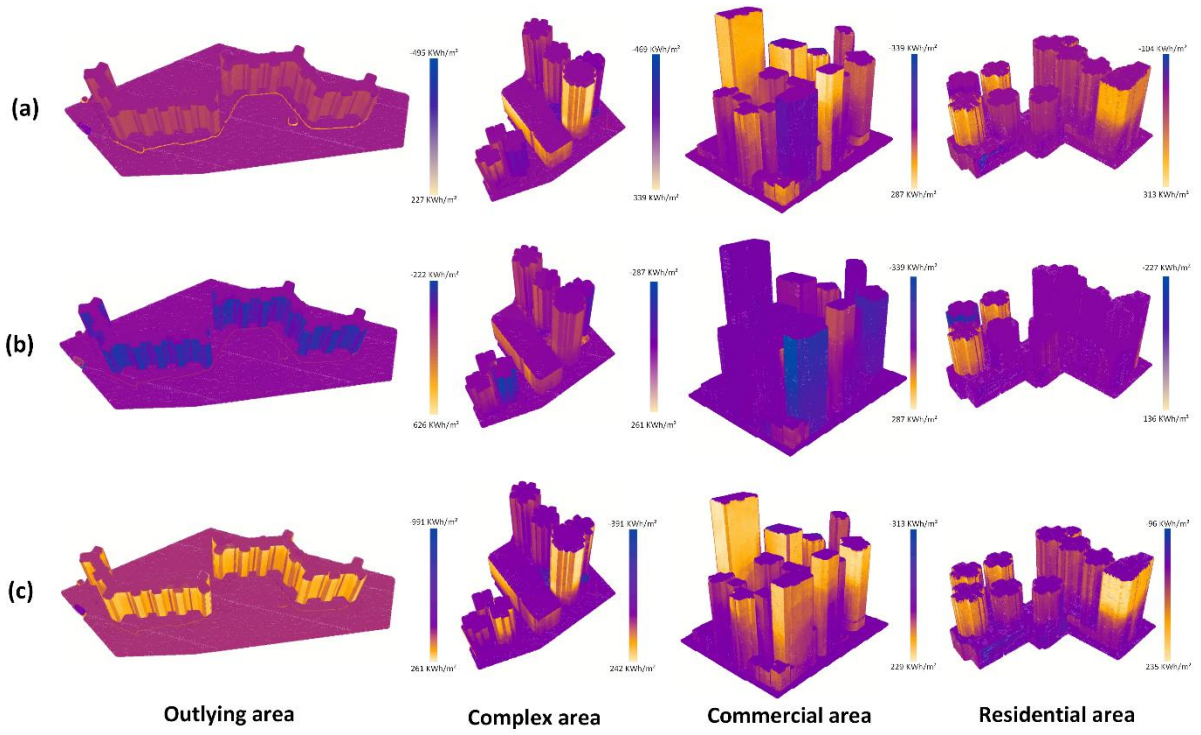

Figure 5.7 Differences in façade albedo impact the distribution of annual solar potential across four study areas. (a) shows the annual solar potential under the segmentation-based strategy minus that under the constant strategy. (b) represents the solar potential under the segmentation-based strategy minus that under the simulation strategy. (c) illustrates the difference between the simulation and constant strategies.

In addition to the patterns above, other factors also influence the distribution of solar potential in Area 2. As shown in the second, third, and fourth columns of Figure 5.7, the differences in solar potential tend to increase with building height. This is because, in the metropolitan environment, the complex spatial layout results in dynamic shading relationships among buildings at different moments. Higher floors are less likely to be shaded, allowing façade areas at these levels to receive better sunlight. As previously noted, better lighting conditions amplify the differences caused by varying albedo values.

However, a different pattern emerges in the second column of Figure 5.7(b). In the residential building located at the lower right corner of Area 2, near an industrial-like structure, the solar

potential difference initially decreases and then increases from top to bottom on the side facing the factory. This trend is also evident in rows (a) and (c), where the façade shows a similar pattern of weakening and then strengthening differences. This anomaly is difficult to explain by considering only the direct component of irradiation. This variation in solar potential differences can be attributed to reflections from the factory roof and the different absorptivity of irradiation by the façade materials themselves. Due to these indirect components, the lower floors exhibit greater differences in solar potential.

Similarly, the southwest façade of the industrial-like building displays a complex and uneven pattern of differences, with inconsistent trends observed in rows (a) and (b). This suggests that direct irradiation differences are not the dominant factor in this case. Instead, the complexity and heterogeneity of the sources of indirect components play a significant role. Additionally, discrepancies in recognition of the source buildings' materials, which the reflected solar radiations come from, among the three albedo strategies further contribute to the complexity and unevenness of the patterns. Consequently, the heterogeneous nature of the indirect components, combined with the weak direct components, results in the observed complexity and unevenness in the patterns.

5.2.2 Solar potential distribution influenced by the district morphology

The morphology and function of the study areas significantly influence the distribution of solar potential. The four study areas are categorized based on their primary use: outlying residential area, complex area, commercial area, and downtown residential area. Relevant indicators for these areas are listed in Table 5.5. Undoubtedly, building height, density, and façade orientation significantly affect the amount of accessible radiation. However, in this section, we mainly focus on the impact of albedo differences caused by morphology and function on the distribution of solar potential.

Table 5.5 Detail indicators of each area.

	Area 1	**Area 2**	**Area 3**	**Area 4**
Open area (m^2)	198131	16007	12522	9990
Building footprint (m^2)	20705	13974	10242	17126
Open space ratio (%)	90.5	53.3	55.0	36.8
Average height (m)	65.7	72.9	74.1	91.6

Firstly, Areas 1 and 4 are primarily residential. These residences are in consistent building heights and styles within the selected regions. In high-density urban environments, similar heights mean rooftops are less likely to be shaded, while façades are more likely to be shaded. In such areas, the solar potential of the roof accounts for a more significant proportion of the total distribution. Consequently, as shown in Table 5.6, compared to Areas 2 and 3, which are also in the downtown area, Area 4 has the lowest R values (1.82, 1.58, and 1.81 under the three albedo allocation strategies) despite having the highest average building height of 91.6 meters. In this study we defined the R as the solar potential ratio of façade to roof in different areas, which can be formula as follow:

$$R = \frac{Potential_{\text{façade}}}{Potential_{\text{roof}}} \tag{5-1}$$

In contrast, the situation differs in suburban areas. Due to the lack of obstructions (after excluding a large podium area from Area 1 rooftop statistics), the extensive façade area results in a considerable R value. Under the segmentation-based, constant, and simulation albedo strategies, the R values reach 5.42, 5.08, and 6.39, respectively. Simultaneously, the impact of façade material albedo on solar potential distribution is greatest in Area 1, with the difference between R_{sim} and R_c reaching 1.31, compared to a maximum difference of 0.38 in other areas. As described in Table 5.6, in this context, R_{max}-R_{min} can serve as an indicator for observing the changes in solar distribution caused by changes in albedo.

Table 5.6 The solar potential ratio of façade to roof in different areas. R_{seg}, R_c, and R_{sim} represent the ratio under segmentation-based, constant, and simulation strategies, respectively. R_{max} is the maximum value of R_{seg}, R_c, and R_{sim} in the study area. R_{max}-R_{min} serves as an indicator representing the changes in solar distribution caused by changes in albedo.

	R_{seg}	R_c	R_{sim}	R_{max}-R_{min}
Area 1	1.91	1.79	2.25	0.46
Area 1 w/o podium	5.42	5.08	6.39	1.30
Area 2	2.31	1.93	2.31	0.38
Area 3	2.71	2.39	2.76	0.37
Area 4	1.82	1.58	1.82	0.25

Area 2, in contrast, is more inclined to be multifunctional, with more architectural styles and larger differences in building heights. The situation is similar in the commercial area. Compared with the homogeneous residential building, it is difficult to ensure that the office buildings in a commercial area are built in the same period and maintain the same height, which results in significant portions of rooftop areas frequently being shaded. Thus, although the average heights of Areas 2 and 3 (72.9m and 74.1m, respectively) are lower than Area 4 (91.6m), their R values are higher.

Moreover, the diversity in building types and uses leads to heterogeneous façade albedos. This diversity allows us to observe more significant changes in the distribution of solar potential when the façade albedo shifts from the constant to the simulation strategy in Areas 2 and 3, with changes of 0.38 and 0.37, respectively, both higher than the 0.25 observed in the urban residential area.

5.2.3 Albedo-caused effects under different temporal scales

5.2.3.1 Intra-annual scale

Throughout different seasons, even at the same time, the solar elevation angle and incident angles vary. The unequal daylight duration across seasons also results in a non-uniform annual distribution of solar potential. For example, in Area 4, as illustrated in Figure 5.8, solar potential in the first (January, February, and March) and fourth (October, November, and December) quarters is significantly lower than in the second and third quarters.

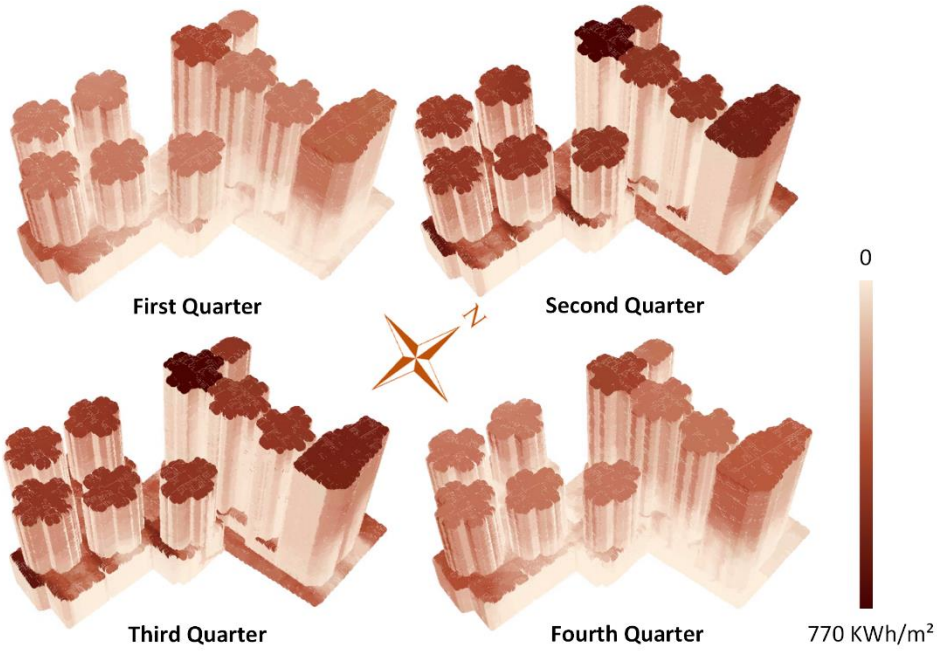

Figure 5.8 Distribution of solar potential in different quarters of the year for Area 4 under segmentation-based scheme.

Due to the study area standing near 22.28°N, the sun reaches its highest elevation angle during the summer. A higher solar elevation angle means that building façades are less likely to be shaded, exposing a larger surface area to sunlight. This is evident in Figure 5.8, where, during the second and third quarters, the red areas on the rightmost building, indicative of strong solar irradiation, extend to lower floors. However, factors beyond the elevation angle

also influence the illuminated area. In the first and fourth quarters, sunlight predominantly impacts the south-facing sides of buildings, while in summer and autumn, the north-facing sides receive more sunlight. This means the projection area of buildings in different azimuths also determines the illuminated surface area. Unlike the commercial area's rectangular buildings with high aspect ratios, the residential buildings have nearly circular cross-sections, ensuring a relatively consistent illuminated area throughout the year.

Table 5.7 The solar potential ratio of façade to roof in Area 4 under different quarters.

	R_{seg}	R_c	R_{sim}	R_{max}-R_{min}
Quarter 1	2.66	2.33	2.66	0.34
Quarter 2	1.31	1.11	1.30	0.20
Quarter 3	1.32	1.11	1.30	0.20
Quarter 4	2.83	2.47	2.83	0.36

Despite minimal changes due to azimuth angles, the increased elevation angle still enhances the illuminated façade area. However, in Figure 5.8, the façade colors remain relatively consistent across all four quarters, with only the illuminated surface varying. This consistency occurs because the higher solar elevation angle reduces the angle of incidence on the façades, thereby decreasing the component of irradiation projected perpendicularly onto the façade. Consequently, the overall irradiation intensity on the façades does not significantly vary across the seasons.

In contrast, the solar potential on horizontal surfaces shows significant seasonal variations. Both rooftops and ground surfaces receive irradiation primarily influenced by the solar elevation angle and daylight duration. In the second (April, May, and June) and third (July, August, and September) quarters, when the elevation angle is higher, horizontal surfaces exhibit markedly greater solar potential. Considering both the façade and horizontal surfaces,

the results in Table 5.7 can be derived. During fourth and first quarter, the R values in the study area are relatively high, reaching up to 2.66 and 2.83, respectively. In these seasons, differences in façade albedo significantly impact the overall solar potential distribution, with R_{max}-R_{min} values reaching 0.34 and 0.36.

5.2.3.2 Hourly scale

Compared to the distribution of solar potential over longer time scales, the hourly distribution within a single day is influenced by more factors and thus exhibits greater variability. The most significant influence is the weather. In Hong Kong, typhoons and heavy rains are common, especially in summer, leading to discrepancies when calculating diffuse proportions, transmissivity, and comparing sampled results. To mitigate the impact of these factors on the hourly solar distribution, data from morning, noon, and afternoon throughout the year were collected and analyzed.

Table 5.8 The solar potential ratio of façade to roof in Area 3 under different hours.

	R_{seg}	R_c	R_{sim}	R_{max}-R_{min}
Morning	3.95	3.54	4.06	0.51
Noon	1.93	1.68	1.96	0.27
Afternoon	5.28	4.63	5.37	0.74

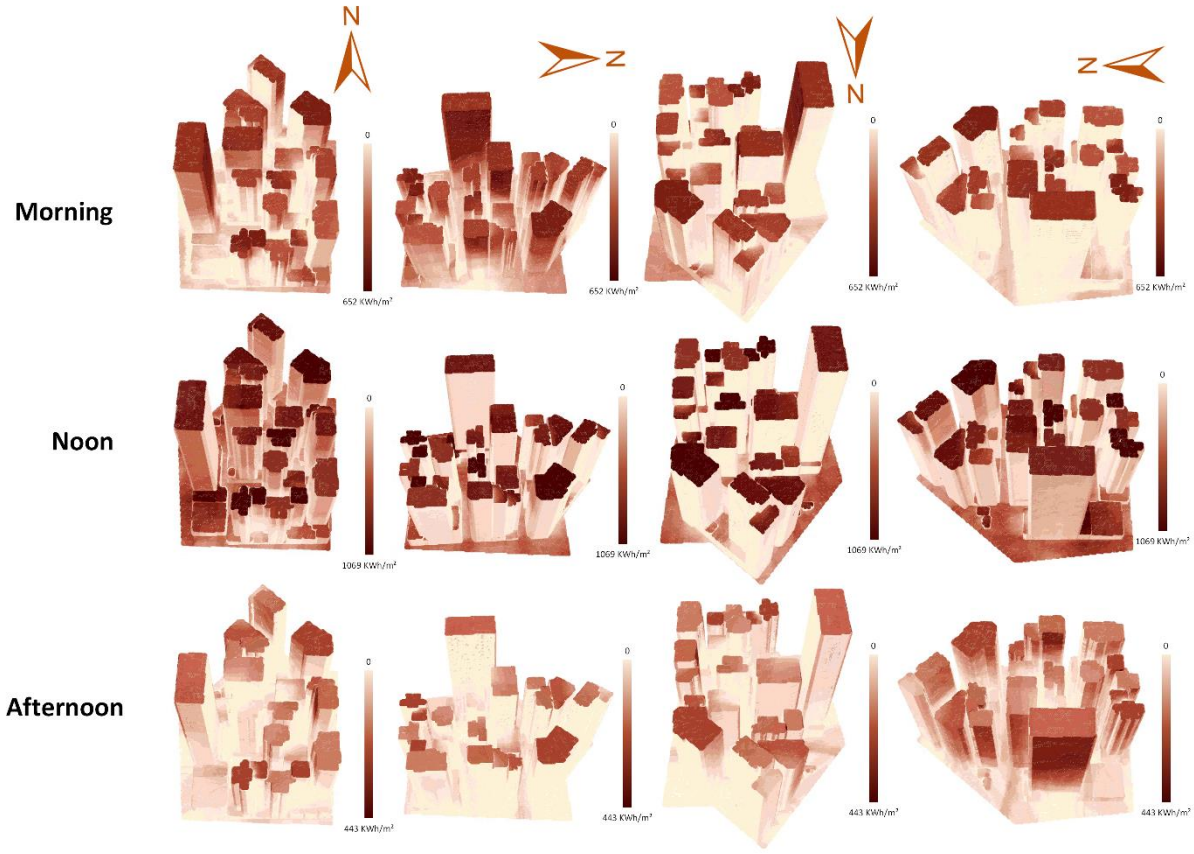

Figure 5.9 The distribution of solar potential in Area 3 at different time periods throughout the day. The four columns of images represent views of the study area in different orientations.

The experiment divided the daytime sunshine into three time periods: morning (before 11 a.m.), noon (between 11 a.m. and 2 p.m.), and afternoon (after 2 p.m.). After minimizing the impact of weather, the solar potential distribution within the study area is primarily influenced by geometric factors such as building layout and spatial relationship with the sun. Similar to the results on the seasonal scale, the potential distribution in the study area shows a strong correlation with the solar elevation angle. As illustrated in Figure 5.9, taking Area 3 as an example, the elevation angle is the largest at noon, and the sunshine also reaches its peak. Consequently, the overall solar potential of the study area is increased, especially in the horizontal roof and ground areas. Figure5.9 reveals that in the morning and afternoon, the maximum solar potential on horizontal surfaces is comparable to that on façades. However, at

noon, there is a noticeable color difference between these surfaces (rooftops appear dark red, while façades are merely orange-red). This phenomenon is also observed in Table 5.8, where the R_{sim} value decreases from 4.06 in the morning to 1.96 at noon and rises back to 5.37 in the afternoon. This indicates that horizontal surfaces benefit more from the increased solar elevation. On the contrary, as the elevation angle decreases, the irradiance received by horizontal surfaces drops sharply. This means that the solar potential is more concentrated on the façades during morning and afternoon periods. This is further evidenced by the morning and afternoon R_{max}-R_{min} values, where 0.74 and 0.51 reflect that selecting different façade albedo strategies during these times will significantly impact the solar potential distribution.

5.2.4 Discussion

The experimental results provide insights into how albedo distribution strategies impact the solar potential within the selected study area of North Point, Hong Kong. The study revealed that variations in albedo across different urban surfaces could lead to significant changes in the overall solar potential, with observed effects ranging from 8.0% to 9.1%. Another observation of the study is that when multiple reflections effect within buildings are disregarded, the impact of albedo on solar potential increases markedly, ranging from 11.9% to 17.8%. This suggests that the internal reflections within an urban environment can mitigate some of the potential gains or losses caused by varying albedo levels. The amplification of the effect, when these reflections are ignored, emphasizes the importance of considering both direct and indirect solar radiation in urban energy models. This finding also highlights the need for more detailed and accurate modelling of solar potential estimation, which should account for complex interactions between surfaces, materials, and urban morphology.

The study further delves into the differences observed across various morphological study areas, comparing regions with different building densities and uses. In areas with a high open space ratio, such as suburban isolated residential areas, the impact of façade materials on solar potential distribution is more pronounced. This can be attributed to the larger surface

areas of façades exposed to direct sunlight and the reduced shading effects from neighboring structures. In these environments, the choice of façade material and its albedo can significantly influence the overall solar energy that can be harnessed.

In contrast, metropolitan downtown areas, characterized by mixed-use regions and commercial districts with diverse building styles and functions, exhibit a different pattern. In these areas, the influence of façade albedo on solar potential distribution is more complex due to the intricate interplay of shading, reflections, and varying building heights. The study found that in such densely built environments, the differences in façade albedo are more impactful compared to residential areas with more consistent architectural styles. This suggests that in mixed-use and commercial districts, careful consideration must be given to the selection of façade materials, as they can have a substantial effect on the distribution of solar potential across the area.

The temporal analysis of solar potential distribution further extends the findings of the study. It was observed that during the first and fourth quarters of the year, horizontal surfaces such as rooftops and the ground receive weaker irradiation. This seasonal variation makes the differences in façade reflectance more influential on overall distribution changes. During these periods, when the elevation angle of the sun is lower in the sky, the angle of incidence of sunlight on façades becomes larger, thereby increasing the direct component of the incident irradiation and the impact of albedo variations. The study's focus on temporal dynamics could be valuable for the design of solar energy systems, as it suggests that albedo strategies may need to be adjusted seasonally to maximize solar potential.

On shorter time scales, such as daily variations, the study noted that after mitigating incidental factors like weather conditions, the solar potential distribution in the morning and afternoon is more dependent on façade albedo values. This daily fluctuation can be linked to the changing angle of sunlight throughout the day, which alters the amount of solar radiation absorbed or reflected by different surfaces. In the morning and afternoon, when the sun's rays

strike façades at lower angles, the reflectivity of these surfaces becomes a more critical factor in determining solar potential. This finding underscores the importance of considering the diurnal cycle in the planning and implementation of urban solar energy strategies.

Overall, the study highlights the significant role of albedo in shaping solar potential in urban environments. The findings underscore the importance of considering both spatial and temporal variations in albedo when designing and optimizing solar energy systems. By understanding the complex interactions between urban morphology, façade materials, and solar potential, we can develop more effective strategies to harness solar energy, contributing to more sustainable and energy-efficient cities.

Chapter 6 Conclusion and future work

6.1 Conclusion

This study proposes a comprehensive evaluation framework to quantitatively assess the impact of urban façade albedo on solar potential distribution. The framework incorporates several key components, including a deep learning network designed to efficiently acquire large-scale urban building façade information, a projection method for converting single, discontinuous 2D images into cohesive 3D models, three distinct albedo distribution strategies, and a methodology for quantitatively evaluating the influence of façade albedo on solar potential distribution.

Acquiring detailed information on urban façade albedos has long been a significant challenge in the field of urban energy research. The process has traditionally been hindered by the vast workload involved, making manual data collection both time-consuming and costly. Complex urban environments, a wide variety of facade materials, and occlusions caused by billboards or trees, place high demands while trying to use algorithms to automatically identify façade materials. Thus, the simulation of reflected light continues to be one of the most challenging aspects of indirect solar radiation estimation. Conventional approaches, which often involve ignoring reflected solar radiation or applying a constant albedo value to represent an entire urban area, can lead to significant inaccuracies in solar potential simulations. This research seeks to address this gap by proposing a novel multi-scale contextual attention network (MSCA) that is specifically designed to efficiently identify façade materials at the city scale. The MSCA network is crafted to balance the need for high levels of detail, such as capturing the spectral characteristics of materials, with the requirement for contextual comprehension of larger objects, such as preserving the structural integrity of buildings within complex urban environments.

The research was conducted in the densely built metropolitan area of Hong Kong, which serves as an ideal testing ground for the proposed model due to its diverse architectural styles and challenging urban conditions. A street-level dataset was developed to evaluate the effectiveness of the proposed model. The experimental results demonstrate that the MSCA network is capable of accurately classifying façade materials and outperforming existing models in this domain. This finding is significant, as it addresses a critical gap in the collection of façade information, which has long been a bottleneck in urban albedo research. The ability to accurately classify façade materials enables the projection of this data onto 3D geographic information system (GIS) platforms, which can greatly enhance the precision of solar potential simulations by incorporating detailed and location-specific data.

The significance of this study extends beyond the immediate task of image analysis and into the broader field of urban energy management. By providing a comprehensive evaluation framework that quantitatively analyzes and discusses how façade materials influence solar potential distribution, the research offers valuable insights into the relationship between urban morphology and solar capacity. Specifically, the incorporation of precise albedo values for urban envelopes, comprising rooftops, façades, and ground surfaces, into simulations allows for a more accurate assessment of solar potential. This nuanced approach not only improves the reliability of solar energy forecasts but also supports the development of more effective photovoltaic (PV) deployment strategies tailored to the unique characteristics of urban environments.

One of the key contributions of this work lies in its ability to bridge the domain gaps that have plagued façade information collection. The reliable data obtained through the MSCA network provides a more nuanced understanding of urban albedos, which is particularly important in the context of solar energy simulations. Traditional

methods, which often neglect the complexities of reflected solar radiation or rely on constant albedo values, introduce significant errors in solar potential estimations. By contrast, the approach presented in this study, incorporating precise and context-specific albedo measurements, results in more accurate and reliable simulations of solar potential distribution. This advancement holds particular relevance for urban planners and energy policymakers, who require precise data to make informed decisions regarding the integration of renewable energy sources into the urban fabric.

Moreover, this study is the first to quantitatively assess the impact of different façade albedos on solar potential distribution. The inclusion of precise albedo data in these simulations facilitates a more detailed analysis of how various urban forms and materials affect solar energy potential. This enhanced understanding of the complex interactions between surfaces, materials, and urban morphology is crucial for advancing the field of urban energy management. By elucidating the relationship between urban materials and solar capacity, the study provides a foundation for future research that seeks to optimize the design and placement of solar energy systems within the built environment.

6.2 Limitations and recommendations for future research

Despite the contributions of this work, the study acknowledges several limitations that must be addressed in future research. One such limitation is related to the high cost of data annotation, which led to the assumption that each building is composed of no more than two primary materials. While this assumption simplifies the data collection process, it may not accurately reflect the complexity of modern architectural designs, such as theatres, museums, and other buildings that feature a diverse array of materials and design elements. This simplification could lead to inaccuracies in the

classification process, particularly for buildings with hybrid façades that do not conform to the binary material assumption.

Furthermore, the simplified classification of rooftop materials, based solely on satellite imagery color intensity, introduces uncertainties in albedo estimation compared to the detailed façade-level analysis. This approach overlooks material heterogeneity (e.g., variations in roofing tiles, solar panels, or weathering effects). Additionally, roof geometries (e.g., slopes, obstructions) not captured by street-view imagery may further bias reflectance calculations. Future work could integrate aerial LiDAR or multispectral data to refine rooftop material characterization and reduce reliance on assumptions.

Additionally, the reliance on building-level annotations introduces a degree of ambiguity in the classification of hybrid façades. Hybrid façades, which incorporate multiple materials and design features, pose a challenge for the MSCA network, as the model may struggle to accurately categorize these complex surfaces. This ambiguity can lead to classification errors that reduce the overall accuracy of the model and limit its applicability in certain urban contexts. Addressing this limitation will require the development of more sophisticated annotation techniques and the incorporation of finer-grained classification methods that can accurately capture the diversity of materials present in contemporary urban architecture.

Another limitation of the study is related to the complex reflective properties of materials. The MSCA network's reliance on visual methods for classification means that it does not fully account for the reflectivity of different façade materials. Reflectivity plays a crucial role in determining the solar potential of a surface, as highly reflective materials can reduce the amount of solar energy absorbed by a building. The study's failure to categorize materials strictly by their reflectivity could diminish the potential applicability of the model in solar potential estimation. Future

research will need to focus on achieving a more fine-grained classification of materials based on their reflective properties to enhance the accuracy and usefulness of the model for solar energy simulations.

Furthermore, due to computational limitations, the study only analysed four areas within Hong Kong's North Point district, which may limit the generalizability of the findings. The study concludes that while the proposed methods significantly advance the field, more extensive and precise research is needed to fully understand the complex relationship between urban morphology and solar capacity.

To address these limitations, future research should aim to expand the scope of analysis to include a wider variety of urban areas, both within Hong Kong and in other cities with different architectural and environmental conditions. This expanded scope would provide a more comprehensive understanding of how urban morphology and material characteristics influence solar potential across diverse contexts. Additionally, further development of the MSCA network and related algorithms will be necessary to improve the accuracy and reliability of façade material classification.

In conclusion, this study presents a robust and innovative framework for accurately simulating urban solar potential distribution by leveraging street view imagery to acquire detailed building façade information. Through a combination of deep learning techniques and projection methods, the research quantitatively assesses the impact of façade reflectance on solar potential, thereby enhancing our understanding of how urban materials influence solar capacity. Despite its limitations, the study makes significant contributions to the field of urban energy research, providing a strong foundation for future work aimed at optimizing photovoltaic deployment strategies and improving solar energy simulations.

References

Ağbulut, Ü., Gürel, A.E., Biçen, Y., 2021. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. Renewable and Sustainable Energy Reviews 135, 110114.

An, Y., Chen, T., Shi, L., Heng, C.K., Fan, J., 2023. Solar energy potential using gis-based urban residential environmental data: A case study of shenzhen, china. Sustainable Cities and Society 93, 104547.

ArcGIS, 2019. Points Solar Radiation. http://desktop.arcgis.com/en/ arcmap/10.3/tools/spatial-analyst-toolbox/points-solar-radiation. htm.

Assouline, D., Mohajeri, N., Scartezzini, J.L., 2015. A machine learning methodology for estimating roof-top photovoltaic solar energy potential in switzerland, in: Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale, LESO-PB, EPFL. pp. 555–560.

Assouline, D., Mohajeri, N., Scartezzini, J.L., 2017. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. Solar Energy 141, 278–296.

Bell, S., Upchurch, P., Snavely, N., Bala, K., 2013. Opensurfaces: A richly annotated catalog of surface appearance. ACM Transactions on graphics (TOG) 32, 1–17.

Bell, S., Upchurch, P., Snavely, N., Bala, K., 2015. Material recognition in the wild with the materials in context database, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3479–3487.

Besharat, F., Dehghan, A.A., Faghih, A.R., 2013. Empirical models for estimating global solar radiation: A review and case study. Renewable and sustainable energy reviews 21, 798–821.

Bill, A., Mohajeri, N., Scartezzini, J.L., 2016. 3d model for solar energy potential on buildings from urban lidar data., in: UDMV, pp. 51–56.

Boccalatte, A., Fossa, M., Ménézo, C., 2020. Best arrangement of bipv surfaces for future nzeb districts while considering urban heat island effects and the reduction of reflected radiation from solar façades. Renewable Energy 160, 686–697.

Calcabrini, A., Ziar, H., Isabella, O., Zeman, M., 2019. A simplified skyline-based method for estimating the annual solar energy potential in urban environments. Nature Energy 4, 206–215.

Chen, J.L., He, L., Yang, H., Ma, M., Chen, Q., Wu, S.J., Xiao, Z.l., 2019. Empirical models for estimating monthly global solar radiation: A most comprehensive review and comparative case study in china. Renewable and Sustainable Energy Reviews 108, 91–111.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4), 834–848.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 .

Chen, L.C., Yang, Y., Wang, J., Xu, W., Yuille, A.L., 2016. Attention to scale: Scale-aware semantic image segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3640– 3649.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmen- tation, in: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

China Light and Power Company, 2021. CLP Sustainability Report 2021. https://sustainability.clpgroup.com/en/2021/standard-esg-disclosures/customers.

Choi, Y., Suh, J., Kim, S.M., 2019. Gis-based solar radiation mapping, site evaluation, and potential assessment: A review. Applied Sciences 9, 1960.

Dai, M., Meyers, G., Tingley, D.D., Mayfield, M., 2019. Initial investi- gations into using an ensemble of deep neural networks for building façade image semantic segmentation, in: Remote Sensing Technologies and Applications in Urban Environments IV, International Society for Optics and Photonics. p. 1115708.

Dana, K.J., Van Ginneken, B., Nayar, S.K., Koenderink, J.J., 1999. Re- flectance and texture of real-world surfaces. ACM Transactions On Graphics (TOG) 18, 1–34.

Dehwah, A.H., Asif, M., Rahman, M.T., 2018. Prospects of pv application in unregulated building rooftops in developing countries: A perspective from saudi arabia. Energy and Buildings 171, 76–87.

Electrical and Mechanical Services Department, 2021. Hong Kong Energy End-use Data. https://www.emsd.gov.hk/en/energy_efficiency/energy_end_use_data_and_consumption_indicators/hong_kong_energy_end_use_data/data/ index.html.

Erdélyi, R., Wang, Y., Guo, W., Hanna, E., Colantuono, G., 2014. Three-dimensional solar radiation model (soram) and its application to 3-d urban planning. Solar Energy 101, 63–73.

Fouad, M., Shihata, L.A., Morgan, E.I., 2017. An integrated review of factors influencing the perfomance of photovoltaic panels. Renewable and Sustainable Energy Reviews 80, 1499–1511.

Freitas, S., Catita, C., Redweik, P., Brito, M.C., 2015. Modelling solar potential in the urban environment: State-of-the-art review. Renewable and Sustainable Energy Reviews 41, 915–931.

Fritz, M., Hayman, E., Caputo, B., Eklundh, J.O., 2004. The kth-tips database. Gassar, A.A.A., Cha, S.H., 2021. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. Applied Energy 291, 116817.

Fu, P., Rich, P.M., 1999. Design and implementation of the Solar Analyst: An ArcView extension for modeling solar radiation at landscape scales. Proceedings of the 19th Annual ESRI User Conference, San Diego, USA, 1–31.

Gadde, R., Marlet, R., Paragios, N., 2016. Learning grammars for architecture-specific facade parsing. International Journal of Computer Vision 117, 290–316.

Gassar, A.A.A., Cha, S.H., 2021. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. Applied Energy 291, 116817.

Gu, J., Dong, C., 2021. Interpreting super-resolution networks with local attribution maps, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9199–9208.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

HO, D.C., Lo, S., Yiu, C., Yau, L., 2004. A survey of materials used in external wall finishes in hong kong. ol. 15 Issue 2 December 2004.

Hong Kong Government, 2021. Hong Kong's Climate Action Plan 2050. https://www.gov.hk/en/residents/environment/global/climate.htm.

Hong Kong Government, 2021. 2021 Population Census. https://www.census2021.gov.hk/sc/district_profiles.html.

Huang, S., Rich, P.M., Crabtree, R.L., Potter, C.S., Fu, P., 2008. Modeling monthly near-surface air temperature from solar radiation and lapse rate: Application over complex terrain in yellow stone national park. Physical Geography 29, 158–178.

Ibrahim, I.A., Khatib, T., 2017. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversion and Management 138, 413–425.

IEA, 2023. Renewable Energy Market Update. https://www.iea.org/reports/renewable-energy-market-update-june-2023.

Ilehag, R., Schenk, A., Huang, Y., Hinz, S., 2019. Klum: An urban vnir and swir spectral library consisting of building materials. Remote Sensing 11, 2149.

Izquierdo, S., Montañés, C., Dopazo, C., Fueyo, N., 2011. Roof-top solar energy potential under performance-based building energy codes: The case of spain. Solar Energy 85, 208–213.

Jakubiec, J.A., Reinhart, C.F., 2013. A method for predicting city-wide electricity gains from photovoltaic panels based on lidar and gis data combined with hourly daysim simulations. Solar Energy 93, 127–143.

Jiang, H., Dong, Y., Xiao, L., 2017. A multi-stage intelligent approach based on an ensemble of two-way interaction model for forecasting the global horizontal radiation of india. Energy Conversion and Management 137, 142–154.

Kong, G., Fan, H., 2020. Enhanced facade parsing for street-level images using convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing 59, 10519–10531.

Korc, F., Förstner, W., 2009. etrims image database for interpreting images of man-made scenes. Dept. of Photogrammetry, University of Bonn, Tech. Rep. TR-IGG-P-2009-01.

Kotak, Y., Gul, M., Muneer, T., Ivanova, S., 2015. Investigating the impact of ground albedo on the performance of pv systems.

Kotthaus, S., Smith, T.E., Wooster, M.J., Grimmond, C., 2014. Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. ISPRS Journal of Photogrammetry and Remote Sensing 94, 194–212.

Leica Geosystems, 2024. Leica Pegasus:Two Mobile Sensor Platform. https://leica-geosystems.com/products/mobile-mapping-systems/capture-platforms/leica-pegasus_two.

Levinson, R., Berdahl, P., Akbari, H., 2005. Solar spectral optical properties of pigments—part ii: Survey of common colorants. Solar Energy Materials and Solar Cells 89, 351–389.

Li, Y., Ding, D., Liu, C., Wang, C., 2016. A pixel-based approach to estimation of solar energy potential on building roofs. Energy and Buildings 129, 563–573.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117– 2125.

Liu, H., Zhang, J., Zhu, J., Hoi, S.C., 2017. Deepfacade: A deep learning approach to facade parsing, IJCAI.

Liu, B.Y., Jordan, R.C., 1963. The long-term average performance of flat-plate solar-energy collectors: with design data for the us, its outlying possessions and canada. Solar energy 7, 53–74.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.

Ma, W., Ma, W., Xu, S., Zha, H., 2020. Pyramid alknet for semantic parsing of building facade image. IEEE Geoscience and Remote Sensing Letters 18, 1009–1013.

Makade, R.G., Chakrabarti, S., Jamil, B., 2019. Prediction of global solar radiation using a single empirical model for diversified locations across india. Urban Climate 29, 100492.

Mallikarjuna, P., Targhi, A.T., Fritz, M., Hayman, E., Caputo, B., Eklundh, J.O., 2006. The kth-tips2 database. Computational Vision and Active Perception Laboratory, Stockholm, Sweden 11.

Meenal, R., Selvakumar, A.I., 2016. Estimation of global solar radiation using sunshine duration and temperature in chennai, in: 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), IEEE. pp. 1–6.

Meenal, R., Selvakumar, A.I., 2018. Assessment of svm, empirical and ann based solar radiation prediction models with most influencing input parameters. Renewable Energy 121, 324–343.

Nguyen, H.T., Pearce, J.M., 2013. Automated quantification of solar photovoltaic potential in cities overview: A new method to determine a city's solar electric potential by analysis of a distribution feeder given the solar exposure and orientation of rooftops. International Review for Spatial Planning and Sustainable Development 1, 49–60.

Nwokolo, S.C., Obiwulu, A.U., Ogbulezie, J.C., 2023. Machine learning and analytical model hybridization to assess the impact of climate change on solar pv energy production. Physics and Chemistry of the Earth, Parts A/B/C 130, 103389.

Observatory, H.K., 2024. Daily mean amount of cloud. https://www.hko.gov.hk/sc/abouthko/opendata_intro.htm.

Park, S., Kim, Y., Ferrier, N.J., Collis, S.M., Sankaran, R., Beckman, P.H., 2021. Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods. Atmosphere 12, 395.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32.

Planning Department, 2023. Land Utilization in Hong Kong. https://www.pland.gov.hk/pland_en/info_serv/open_data/landu/.

Quej, V.H., Almorox, J., Ibrakhimov, M., Saito, L., 2016. Empirical models for estimating daily global solar radiation in yucatán peninsula, mexico. Energy conversion and management 110, 448–456.

Redweik, P., Catita, C., Brito, M., 2013. Solar energy potential on roofs and facades in an urban landscape. Solar energy 97, 332–341.

Richter, M.L., Byttner, W., Krumnack, U., Wiedenroth, A., Schallner, L., Shenk, J., 2021. (input) size matters for cnn classifiers, in: International Conference on Artificial Neural Networks, Springer. pp. 133–144.

Riemenschneider, H., Krispel, U., Thaller, W., Donoser, M., Havemann, S., Fellner, D., Bischof, H., 2012. Irregular lattices for complex shape grammar facade parsing, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE. pp. 1640–1647.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. Labelme: a database and web-based tool for image annotation. International journal of computer vision 77, 157–173.

Salleh, S., Latif, Z.A., Pradhan, B., Wan Mohd, W., Chan, A., 2014. Functional relation of land surface albedo with climatological variables: a review on remote sensing techniques and recent research developments. Geocarto International 29, 147–163.

Sánchez, E., Izard, J., 2015. Performance of photovoltaics in non-optimal orientations: An experimental study. Energy and Buildings 87, 211–219. Sharan, L., Rosenholtz, R., Adelson, E.H., 2014. Accuracy and speed of material categorization in real-world images. Journal of vision 14, 12–12.

Schwartz, G., Nishino, K., 2016. Material recognition from local appearance in global context. arXiv preprint arXiv:1611.09394 .

Sharan, L., Rosenholtz, R., Adelson, E., 2009. Material perception: What can you see in a brief glance? Journal of Vision 9, 784–784.

Sharan, L., Rosenholtz, R., Adelson, E.H., 2014. Accuracy and speed of material categorization in real-world images. Journal of vision 14, 12– 12.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information processing & management 45, 427–437.

Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., Wang, J., 2019. High-resolution representations for labeling pixels and regions. arXiv preprint arXiv:1904.04514.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114.

Tao, A., Sapra, K., Catanzaro, B., 2020. Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv:2005.10821.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P., Paragios, N., 2011. Shape grammar parsing via reinforcement learning, in: CVPR 2011, IEEE. pp. 2273–2280.

Teboul, O., Simon, L., Koutsourakis, P., Paragios, N., 2010. Segmentation of building facades using procedural shape priors, in: 2010 IEEE computer society conference on computer vision and pattern recognition, IEEE. pp. 3105–3112.

Versluis, R., Powles, R., Yazdanian, M., Rubin, M., Jonsson, J., 2012. International glazing database: Data file format.

Walch, A., Castello, R., Mohajeri, N., Scartezzini, J.L., 2020. Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. Applied Energy 262, 114404.

Wei, Z., Sun, Y., Wang, J., Lai, H., Liu, S., 2017. Learning adaptive receptive fields for deep image parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2434– 2442.

Willenborg, B., Pültz, M., Kolbe, T.H., 2018a. Integration of semantic 3d city models and 3d mesh models for accuracy improvements of solar potential analyses. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42, 223–230.

Willenborg, B., Sindram, M., Kolbe, T.H., 2018b. Applications of 3d city models for a better understanding of the built environment. Trends in Spatial Analysis and Modelling: Decision-Support and Planning Strategies, 167–191.

Xu, F., Wong, M.S., Zhu, R., Heo, J., Shi, G., 2023. Semantic segmentation of urban building surface materials using multi-scale contextual attention network. ISPRS Journal of Photogrammetry and Remote Sensing 202, 158–168.

Xu, S., Huang, Z., Wang, J., Mendis, T., Huang, J., 2019. Evaluation of photovoltaic potential by urban block typology: A case study of wuhan, china. Renewable Energy Focus 29, 141–147.

Yaghoobian, N., Kleissl, J., 2012. Effect of reflective pavements on building energy use. Urban Climate 2, 25–42.

Yuan, Y., Chen, X., Chen, X., Wang, J., 2019. Segmentation transformer: Object-contextual representations for semantic segmentation. arXiv preprint arXiv:1909. 11065.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890.

Zhu, R., Anselin, L., Batty, M., Kwan, M.P., Chen, M., Luo, W., Cheng, T., Lim, C.K., Santi, P., Cheng, C., et al., 2022a. The effects of different travel modes and travel destinations on covid-19 transmission in global cities. Science bulletin 67, 588.

Zhu, R., Cheng, C., Santi, P., Chen, M., Zhang, X., Mazzarello, M., Wong, M.S., Ratti, C., 2022b. Optimization of photovoltaic provision in a three- dimensional city using real-time electricity demand. Applied Energy 316, 119042.

Zhu, R., Kondor, D., Cheng, C., Zhang, X., Santi, P., Wong, M.S., Ratti, C., 2022. Solar photovoltaic generation for charging shared electric scooters. Applied Energy 313, 118728.

Zhu, R., Kwan, M.P., Perera, A., Fan, H., Yang, B., Chen, B., Chen, M., Qian, Z., Zhang, H., Zhang, X., et al., 2023. Giscience can facilitate the development of solar cities for energy transition. Advances in Applied Energy, 100129.

Zhu, R., Wong, M.S., You, L., Santi, P., Nichol, J., Ho, H.C., Lu, L., Ratti, C., 2020. The effect of urban morphology on the solar capacity of three- dimensional cities. Renewable Energy 153, 1111–1126.

Zhu, R., You, L., Santi, P., Wong, M.S., Ratti, C., 2019. Solar accessibility in developing cities: A case study in kowloon east, hong kong. Sustain- able Cities and Society 51, 101738.