https://doi.org/10.1038/s40494-025-01888-4

Semantic segmentation of building façade materials and colors for urban conservation

Check for updates

Jinfeng Xie[1], Minhua Li[2], Jiaqi Wu[3], Xiaohu Zhang[1] & Jie Zhang[3] ✉

Building façade materials are fundamental to understanding cultural heritage and preserving the historic urban environment. While traditional material analysis requires extensive manual effort, this study introduces an automated methodology of material identification and feature extraction. We construct the Building Façade Material Segmentation (BFMS) database, which surpasses previous datasets in its alignment with architectural standards and the diversity of material categories included. Based on the database, we develop a transformer-based semantic segmentation model that achieves an overall accuracy of 76.5%. Furthermore, the algorithm to extract detailed textures and colors for each material is formulated with K-means clustering and grid statistics. Applied in the Taiping Alley, a historic area in Jingdezhen, the methodology demonstrates robustness and efficiency, yielding valuable guidance for architectural landscape assessment and urban renewal planning. The study promotes the application of cutting-edge deep learning algorithms in urban conservation and contributes to the broader understanding of urban material characteristics.

Building materials stand as fundamental elements in architectural heritage, serving as tangible witnesses to construction techniques, cultural preferences, and historical developments of urban areas[1,2]. As William Morris[3] asserted, "The subject of material is clearly the foundation of architecture." From the earliest use of rocks, clays, and woods to the contemporary employment of bricks, concrete, and glass, building materials have not only shaped our built environment but also carried invaluable information about our architectural heritage. Experts treat the selection and application of building materials as important proofs of economic prosperity, urban transformation, and even cultural identities[4–6], providing a crucial lens through which to study the historic evolution.

Materials are also critical components of unique cityscapes and irreplaceable sources of historic sense, and they are often emphasized in the assessment of architectural landscape and the development of conservation planning and policies. Since 1909 in the UK, town planning schemes have implemented reviews and controls of building materials in heritage areas, with the 1925 Bath Corporation Act formally establishing material approval processes through consulting committees[7]. The Venice Charter in 1964 states that the historic urban setting is also part of the historic monument. The Washington Charter in 1987 further emphasizes building materials, as part of the physical composition of the character of historic towns and cities, are important objects of protection. Today, most cities renowned for their historic significance, such as Istanbul, Rome, and Xi'an, maintain strict regulations governing material use in new construction and renovation within and around heritage zones[8,9], demonstrating the continued importance of material control in preservation efforts.

However, material surveys, which are the basis of material documentation and preservation, primarily relying on traditional methods like field observation[4], prove increasingly inadequate for comprehensive urban conservation. These methods, while detailed, are labor-intensive and cost-prohibitive for large-scale historic urban areas, limiting our ability to quantitatively analyze and digitally document material characteristics. This limitation necessitates a shift toward automated methods for analyzing building materials in heritage contexts.

Recent advances in computer vision, particularly convolutional neural networks (CNN) and self-attention mechanisms, have revolutionized the ability to analyze buildings or cities from images[10,11]. While these technologies have been successfully applied to various aspects of heritage protection, including façade reconstruction[12], crack or deterioration detection[13,14], mural segmentation[15], their application to material recognition remains limited. Basically, there are two predominant technological approaches to recognizing and segmenting materials from images. The first approach is "patch slicing & image recognition". The images are first divided into uniform patches, and then each patch is analyzed by a trained neural network to determine the material types. For example, Bell et al. trained a sliding CNN to predict the materials in each patch, followed by a dense fully connected

[1]Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, Hong Kong. [2]Tsinghua Tongheng Urban Planning and Design Institute, Beijing, China. [3]School of Architecture, Tsinghua University, Beijing, China. ✉e-mail: zjzhangjie@mail.tsinghua.edu.cn

Table 1 | Typical Material Database

Dataset name	Annotation type	Outdoor building images	Material categories
CUReT[22]	Texture capture	No	61
OpenSurfaces[21]	Localized label assignment	No	37
FMD[23]	Label assignment	No	10
MINC[16]	Point Annotation & Localized label assignment	19.9%	23
GTOS[46]	Texture slicing	No	40
Multi-illumination dataset[25]	Semantic segmentation	No	14
LMD[26]	Semantic segmentation	56.5%	16
CMD[18]	Texture slicing	100%	5
DMS[19]	Semantic segmentation	7.0%	52
MSD[27]	Localized label assignment	88.0%	1
KITTI-Materials[20]	Semantic segmentation	100%	20
BFMS (Ours)	Semantic segmentation	100%	42

conditional random field module to integrate the outputs of the CNN module[16]. Chen et al. adopted ResNet-50 to recognize materials in street view images across six cities[17]. Sun &Gu assessed the accuracy of five typical models including VGG, GoogleNet, Inception-V3, ResNet50, and DenseNet, in recognizing materials from image patches[18]. The second approach is semantic segmentation, which typically employs an encoder-decoder structure to classify each pixel in original images. A notable implementation of this method is the dense material segmentation (DMS) Model, introduced by Upchurch & Niu[19], which uses a ResNet-50 architecture as the encoder and a Pyramid Pooling Module as the Decoder. Additionally, the RMSNet proposed by Cai et al.[20] features a Mix-Transformer-based encoder and a SAMixer-based decoder, which achieves a better performance.

In the field of the heritage and built environment, most current research approaches involve the "patch slicing & image recognition" method. This preference might stem from the relatively lower accuracy and greater complexity associated with semantic segmentation techniques. Furthermore, typical material recognition tasks require only a basic identification of the presence or absence of certain materials or an understanding of the main façade materials of buildings, rather than a detailed analysis of the complex material compositions found in urban context. However, in historic areas, where buildings may have undergone years of repairs and modifications, a more sophisticated and detailed segmentation approach is often necessary to accurately distinguish between the mixed materials present. Therefore, our study necessitates the development and application of a more refined segmentation model to meet these specific requirements.

A significant barrier to the development of material semantic segmentation models is the lack of suitable datasets for building materials. Systematic literature review has been done to better reveal the current gap in the dataset (Table 1). Typical large-scale material databases include OpenSurfaces, the Materials in Context (MINC) database, and DMS database. OpenSurfaces, as one of the pioneer crowdsourced material databases, comprises 25,357 residential indoor images, with over 58,000 surfaces annotated with material names[21]. The MINC database encompasses over 430,000 images annotated through more than 2.3 million interactions and includes 72,000 segments refined and expanded from OpenSurfaces, in addition to 3 million patches derived from clicks and segments[16]. These two datasets are mainly designed for material recognition and are mostly comprised of indoor photos. The DMS database which focuses on material segmentation features a comprehensive set of polygon annotations for 52 material categories across more than 44,000 images[19]. In addition to these big databases, several medium-sized collections have been constructed to analyze material characteristics through material slices. Dana et al.[22] captured images across 61 material classes under varying observation and illumination angles, establishing the Columbia-Utrecht Reflectance and Texture (CURET) Database to support the study of three-dimensional textures. The Flickr Material Database (FMD) was developed to examine the challenges of material recognition from images, compiling ten materials with 100 images each sourced from the Flickr website[23]. Xue et al. documented 40 types of outdoor terrain materials using a robotic system, resulting in the Ground Terrain in Outdoor Scenes (GTOS) database, which includes 34,243 images[24]. Several smaller databases have been designed to recognize specific materials in distinct settings. For example, the multi-illumination dataset captures 1016 indoor images under 25 lighting conditions to explore the relationship between appearance and illumination[25]. The Local Materials Database (LMD) annotates 16 types of materials in 5151 images drawn from multiple existing image datasets, like MS COCO and ImageNet[26]. However, the dataset is not densely annotated, which means only a small percentage of frames in each photo are labeled. The Construction Material Database (CMD) comprised of five categories of material slices, provides trustworthy data for building materials recognition[18]. The KITTI-Materials Database[20], which segments materials in 1000 outdoor images, aligns most closely with the focus of our study. Notably, images in this dataset were predominantly taken by the in-vehicle camera, with the road being the subject of the view and the building presenting only a side view. Another relevant dataset is the Mirror Segmentation Dataset (MSD) by Gonzales et al.[27], which is a specialized collection that includes 454 images of mirrors.

Though many material datasets have been constructed, they face three major challenges when applied to identifying materials on the urban facades. Firstly, most of these datasets target indoor scenes, and based on the current literature review, no dataset has been specifically focuses on the building facades. Consequently, there is a significant dearth of annotations for common facade materials like rammed earth and tiles, and representations of materials typically found outdoors are scarcely included. For instance, rubber in indoor datasets frequently appears on shoe soles and fitness equipment, whereas outdoors, it is predominantly found on vehicle tires and electrical wire sheathing. Although some datasets incorporate outdoor images, these constitute only a minor portion of the total collection. Training models exclusively on these datasets results in a limited variety of outdoor materials, compromising the models' applicability and accuracy in recognizing outdoor materials.

Secondly, the material classification in previous datasets deviates from standard architectural nomenclature. Professional scrutiny has identified misclassifications and omissions in these databases. For instance, the DMS dataset expands the definition of "paint" to include other building materials like metal and stone, because they are typically coated with a protective layer of paint. Classifying all these materials simply as "paint" overlooks critical information about the built environment. This issue of oversimplification is prevalent in other datasets as well.

Thirdly, for material segmentation tasks that necessitate fine textural information, some existing datasets exhibit limitations concerning image resolution. While our survey of 12 relevant datasets reveals a diversity in

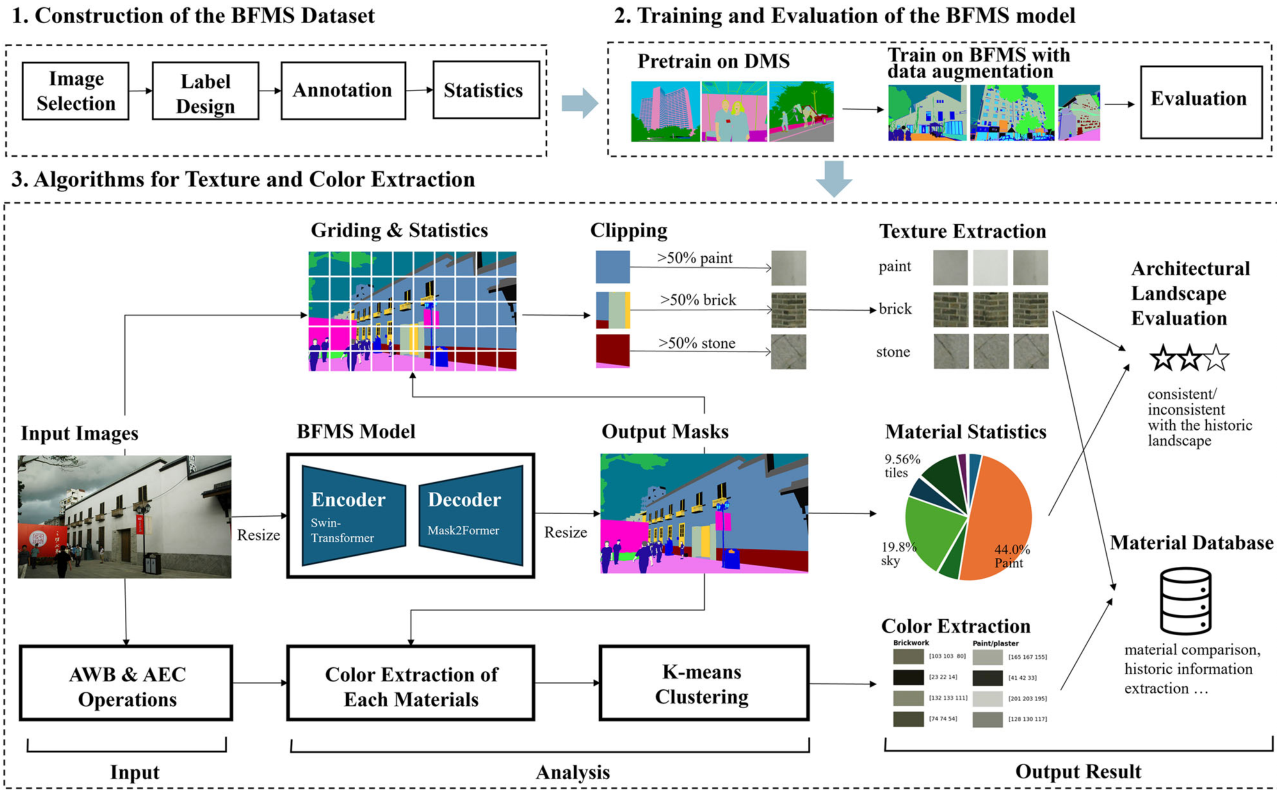


Fig. 1 | Framework of methodology. The methodology consists of three steps: construction of the BFMS dataset, training and evaluation of the BFMS model, and algorithms for texture and color extraction. In the final part, the prediction result from BFMS model will be used to calculate the material statistics and facilitate the texture and color extraction. These results will constitute the material database and provide objective metrics to support architectural landscape assessment.

resolutions, datasets specifically tailored for applications such as building facade analysis and consistently offering resolutions significantly exceeding 2K appear to be relatively uncommon. For instance, images in the DMS dataset typically have their longest dimension not exceeding 1024 pixels, the multi-illumination dataset offers 1500×1000 pixel images, and the KITTI-material dataset provides 320×1216 pixel images. However, the accurate identification of building facade materials heavily relies on capturing their fine-grained textural features. Research has also demonstrated that high-resolution datasets are beneficial for effectively training segmentation models[28].

Consequently, there is a pressing need for high-resolution semantic segmentation datasets specifically tailored to building façades. Such datasets should predominantly feature images of building exteriors and include comprehensive information about the surrounding environment. Moreover, architectural knowledge should inform the design of the labels to ensure the datasets are both accurate and applicable to studies and practices in built environment.

Therefore, our study aims to bridge this gap by constructing a dedicated database for material segmentation of the building facades, along with developing deep-learning models and feature extraction algorithms to enhance material analysis in heritage contexts. This approach is aimed at automating the traditional documentation and analysis process and providing fundamental information for architectural landscape assessment and regeneration guidelines development. The methodology will be tested in the Taiping Alley area of Jingdezhen, China, a historically significant ceramic production center, to examine the model's robustness and discuss its practical implications for urban conservation.

Methods

Framework

Our methodology encompasses a comprehensive framework for building facade material segmentation. The methodology commences with the development of the BFMS dataset through systematic image selection, hierarchical label design, and rigorous annotation protocols. Subsequently, the BFMS model is trained and evaluated utilizing specialized data augmentation techniques. The framework further incorporates a sophisticated pipeline for texture and color extraction, comprising three integrated components: material segmentation, texture extraction via grid statistical analysis, and color extraction through K-means clustering. This pipeline significantly enhances the efficiency of traditional material documentation and architectural landscape evaluation processes.

During implementation, input images undergo preprocessing where they are resized to dimensions compatible with the neural network architecture. The trained BFMS model then performs pixel-wise material classification. Post-processing includes rescaling segmentation results to original dimensions and calculating material proportions. Concurrently, the segmentation output facilitates texture pattern extraction and color palette generation, providing comprehensive material characterization for architectural applications (Fig. 1). The resulting color templates and texture image samples are systematically archived in a material database, enabling comparative analysis with other historic areas and extraction of historical information. Additionally, the material statistics and color clustering results provide objective metrics to support architectural landscape evaluation procedures.

Image selection and label design of the BFMS dataset

To enhance the model's proficiency in recognizing the common scales and material appearances in historic urban areas, this study strategically selects a total of 530 photographs that embody a human-centric viewpoint of building facades, although some exhibit distortion due to the constrained spaces of narrow streets. This distortion has been deliberately included as it represents a common challenge that the model will need to address in real-world scenarios. Additionally, the dataset is enriched by a few photos that offer detailed close-ups of architectural features and unique rooftop views. Selected photos are taken by dozens of professional urban planners during

Table 2 | The spatial distribution of the photos in the BFMS dataset

Province	Picture numbers	Province	Picture numbers
Anhui	46	Jiangxi	100
Beijing	107	Liaoning	17
Fujian	65	Shandong	17
Gansu	35	Shanxi	27
Hebei	35	Hong Kong	16
Heilongjiang	62	Jiangsu	3

Table 3 | Building types in the BFMS dataset photos

Building Type		Numbers
Residential	Low-rise (1–2 floors, e.g., hutong, urban village, village residence)	231
	Multi-story (three or more floors, e.g., gated community, high-rise residence)	63
Commercial buildings (e.g., street shops, shopping malls)		103
Public buildings (e.g., temples, galleries, monuments and other facilities)		72
Office buildings		31
Industrial buildings (e.g., warehouses, factories)		30

their field research. The original use of photos is either for designers to get familiar with the sites before the formal design process or to evaluate the design after construction. More than three-quarters of the frame in each photo are buildings, roads, or decorations. Over 60% of these images are captured using professional cameras, while the remainder are taken using smartphones. Over 90% of the images boast resolutions exceeding the standard 2K (2560 × 1400 pixels), offering detailed insights into the textural nuances of urban materials.

The dataset encompasses a diverse range of images representing a balanced distribution of geographical regions and building typologies across China. Geographically, as shown in Table 2, the collection includes photographs from multiple provincial-level administrative regions, ensuring representation from northern, southern, eastern, and western parts of China, including both major urban centers and less developed areas. This geographical diversity allows the dataset to capture regional variations in architectural materials and styles. In terms of building typologies, Table 3 demonstrates our effort to include various architectural forms, balancing traditional structures such as hutongs and village residences with more contemporary buildings including commercial complexes, public facilities, multi-story residential structures, office buildings, and industrial facilities. This balanced distribution ensures the dataset comprehensively represents diverse Chinese architectural contexts while enabling material analysis applicable across varied urban environments.

The photos cover a span of twenty years, from 2003 to 2023. Most of the photographs were taken within the local historic district, but often some new construction in the neighborhood was attached to add variety to the materials. Besides the primary materials like brick, stone, cement concrete, glass, and metal, the dataset also encompasses relatively rare building materials such as rammed earth. This diverse collection of images aids in training a model that is adept at handling varied visual scenarios typical of urban environments.

Although our database mainly targets buildings in historic areas, we have also tried to set up some elements of the natural environment including sky, water, etc., so that the dataset can also be adapted to multiple types of heritage environments and achieve better segmentation results. The process of designing the label has three steps. Firstly, the study initially adopts 31 categories of materials, such as brick, paper, and tree, from the DMS dataset, which forms the foundation of the labeling system. The second step involves comparing these categories with the national standard, specifically the *Standard for Terminology of Building Materials (JGJ/T 191-2009)*[29]. 19 materials not listed in the DMS dataset, like cement and airbrushed fabric, are added to the labeling system. The definitions of the categories from the DMS dataset are also critically evaluated and redefined to better meet the specific requirements of the built environment studies. This involves providing more detailed definitions for materials that were previously defined broadly or vaguely in the DMS dataset. Special attention is given to materials that do not conform to traditional architectural categorization criteria, such as paint and metal, ensuring that the labels are both precise and practical for real-world applications. The final step is test labeling, where materials that are commonly seen yet challenging to distinguish—such as windows with metal fences and window screens—are specifically identified and added to the labeling categories. Ultimately, 42 categories of materials are developed, with Supplementary Table 1 providing detailed definitions for each category and guidelines for handling contentious cases. 26 categories of materials are related to building facades and grounds, and 16 categories are for the surrounding environment.

It's worth emphasizing that the labeling approach prioritizes the material visible to the naked eye as the primary layer. For instance, when annotating glass windows, even if objects behind the glass are visible, only the glass is labeled. Similarly, in cases of brick walls covered by paint, the decision to label the material as brick or paint depends on whether the brick joints are visible through the paint. If they are visible, the area is labeled as brick; otherwise, it is labeled as paint (Fig. 2). An exception is made for metal, which, despite being coated with paint to prevent rust and corrosion, retains a metallic luster and is therefore still labeled as metal. This nuanced approach ensures that the dataset accurately reflects the complex material compositions encountered in urban environments.

Annotation process of the BFMS dataset

11 senior undergraduate or graduate students majoring in architecture or urban planning are recruited as image annotators. These students are considered more knowledgeable about building materials than peers from other disciplines. To ensure high-quality annotation, all annotators have finished a three-hour training on the material definitions and annotation skills and passed a test labeling assessment of three images. All annotators are required to label as many and as accurately as possible the materials in the picture that they can identify. The annotation process utilizes the CVAT platform, and its layer function is used to reduce labeling effort. The Segment Anything Model (SAM) on the platform is also employed to help annotate objects with complex and organic contours, like trees and cars.

As photos of exterior buildings have more fine-grained details, which makes them more difficult to label, we allow annotators to adopt different approaches to certain materials. Elements such as wires and branches, which often have elongated contours and relatively organic shapes, could be grouped and labeled collectively if they are parallel or closely aligned. Conversely, regular-edged elements like window frames and door frames require precise outline in the labeling process.

The annotation process extends over 45 days, with each annotator handling over 50 images. The average labeling time for each photo exceeds 90 minutes. After labeling, a rigorous review process is conducted in two rounds by the director, focusing on correcting any inaccurately labeled edges or misclassified categories.

After the above labeling process, the study achieves a dataset of 530 densely labeled photos containing more than 330 thousand separate closed regions, an average of over 640 regions per photo. Compared with others, BFMS dataset has a significantly high annotation density (Table 4). For a visual representation of the annotation quality and detail, Fig. 3 showcases examples of the annotated results. Additional raw images and their corresponding masks are available in Supplementary Fig. 1 for further examination. Table 5 shows the segment numbers and percentages calculated by pixels of each category. The dataset is strategically split into two parts: a training set consisting of 430 photos and a validation set comprising 100 images, assigned randomly. The distribution of material categories within the dataset exhibits long-tail characteristics, where some materials such as brick, paint, cement, and metal are more commonly represented due to their prevalence in urban settings. Natural elements like the sky, water, snow, and trees are also included, reflecting the dataset's comprehensive approach to representing both built environments and natural landscapes.

Structure of the BFMS model

The state-of-the-art model in image segmentation, Mask2Former is adopted for this task. Introduced by Cheng et al.[30], the Mask2Former model features a transformer-based decoder that unifies segmentation and detection tasks, achieving excellent results in numerous segmentation and detection challenges upon its release. Our adaptation maintains much of the original architecture, which is described briefly below (Fig. 4).

The model is basically comprised of three main components: a four-stage pixel encoder module, a pixel decoder module and a transformer decoder module (Fig. 4). The encoder is based on the Swin Transformer architecture proposed by Liu et al.[31]. Input images of size 768 x 768 x 3 (height x width x channel) which are acquired from the image processing pipeline are first processed through a patch embedding procedure that converts them into 4 × 4 patches. These patches are transformed into 48-dimensional vectors, resulting in an initial feature map of 192 x 192 x 48. This map then undergoes further transformation through four stages of Swin Transformer blocks, each including a down-sampling step. This process extracts features from images and turns them into high-dimensional matrices with different resolutions. The output feature maps from the four stages have dimensions of 192 x 192 x 96, 96 x 96 x 192, 48 x 48 x 384, and 24 x 24 x 768, respectively.

Following the encoding stages, the pixel decoder module adjusts the channel dimension of four feature maps from the pixel encoder to 256. More

Table 4 | Comparison of annotation density across datasets

Dataset	Annotation segments	Images	Annotation segments per image
OpenSurfaces[21]	137k	19,447	7.04
LMD[26]	9.4k	5845	1.61
DMS[19]	3200k	44,560	71.81
BFMS	337k	530	654.37

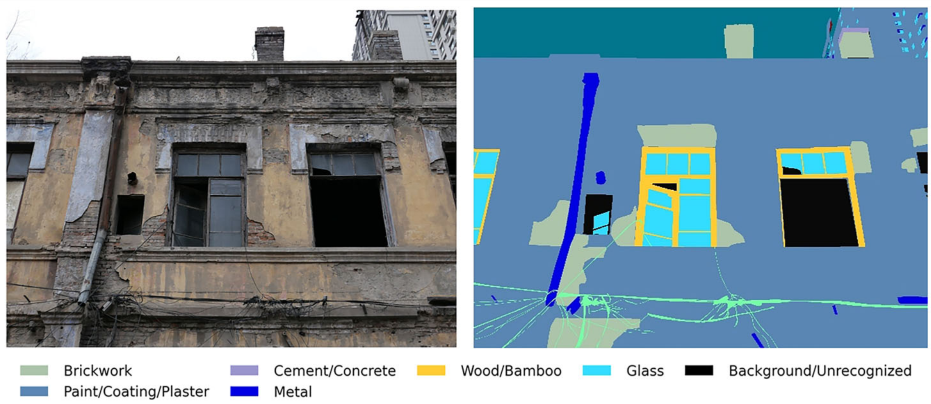


Fig. 2 | Examples of material definitions for glass, paint, and brick

Table 5 | Statistics of the BFMS Dataset

Material	Type	Segment Numbers	Percentages by Pixels	Material	Type	Segment Numbers	Percentages by Pixels
Brickwork	B	30,826	16.270%	Rubber/Latex	E	7216	0.550%
Sky	E	27,851	15.068%	Cardboard/Paper	B	1336	0.374%
Paint/Coating/Plaster	B	28,814	12.791%	Snow	E	797	0.278%
Foliage	E	13,700	6.580%	Photograph/Painting/Airbrushed Fabric	B	401	0.276%
Cement/Concrete	B	10,761	6.487%	Windows with metal fences on the outside	B	1526	0.236%
Metal	B	24,724	5.627%	Plastic, clear	E	870	0.177%
Ground Tile	G	6306	4.126%	Pottery	B	556	0.165%
Wood/Bamboo	B	10,731	4.216%	Sand	E	39	0.106%
Glass	B	37,724	4.038%	Roofing waterproof material	B	104	0.102%
Natural Stone	B	4704	3.218%	Hair	E	669	0.086%
Background/Unrecognized	-	75,677	4.017%	Carved Brick	B	53	0.073%
Tree	E	15,047	2.427%	Skin/Lips	E	1376	0.063%
Asphalt	G	1853	2.039%	Leather	E	366	0.026%
Roof tile	B	3945	1.795%	Enamel	B	117	0.013%
Fabric/Cloth	E	6463	1.734%	Chalkboard/blackboard	B	3	0.006%
Engineered Stone/ Imitation Stone	B	3064	1.460%	Mirror	B	83	0.003%
Composite Decorative Board	B	3816	0.957%	Window screen	B	162	0.002%
Ceramic	B	4967	0.917%	Food	E	17	0.003%
Plastic, non-clear	E	14,506	0.912%	Whiteboard	B	5	0.001%
Rammed earth	B	1330	0.842%	Fur	E	24	0.001%
Soil/mud	G	1684	0.691%	Ice	E	2	0.000%
Water	E	115	0.555%				

*In the column "type", B is short for building, which means that this material is a subdivision of the materials found in buildings. Similarly, E is short for environment, and G is short for ground.

Fig. 3 | Examples of annotations.

cross-scale information is also integrated in this module. Then the adjusted feature maps from the last three encoder stages are sequentially fed into the transformer decoder. This decoder employs masked cross-attention mechanisms to iteratively refine the image features. The transformer module also processes the trainable query features with a dimension of 100×256, which means the 100 queries for objects within the picture. The transformer module outputs a corresponding feature map of the same dimensions. This output is then combined with the feature map from the first encoder stage to produce the final segmentation masks and classes.

Training and evaluation of the BFMS model

In our image processing pipeline, we implemented a series of data augmentation techniques to enhance the model's generalization capabilities. Initially, images are randomly resized to various resolutions such that their shortest side is scaled to one of several target lengths from 512 to 4096 pixels, while maintaining the original aspect ratio. This multi-scale approach exposes the model to material appearances and contextual information at different levels of detail. Subsequently, to improve viewpoint invariance, images undergo random rotation within a ±60-degree range and random horizontal flipping with 50% probability. Following these global transformations, fixed-size patches are randomly cropped from the potentially large, transformed images. This cropping step significantly increases the effective number of training samples derived from each source image and leverages the high resolution of our data by presenting detailed local views containing rich annotation information (Fig. 5). Finally, standard normalization using pre-computed mean and standard deviation values is applied to each patch, ensuring consistent input distribution for optimal network training dynamics. The patches and masks of size 768×768 pixels are finally prepared for model training.

Our model's base parameters were derived from a model pre-trained by Facebook Inc. on the ADE20k dataset. Then, the model is trained on the DMS dataset for 100 epochs, allowing it to learn material-specific information and serve as a baseline for performance comparison. After that, the model is trained on the BFMS dataset for 200 epochs, obtaining the BFMS model. The training process is conducted on a Nvidia 3090 GPU, with Pytorch as the framework. Limited by the memory of devices, the batch size is set to 3, and the learning rate declines from $5*10^{-5}$ to $5*10^{-8}$ gradually.

To quantitatively evaluate the performance of our proposed method for material segmentation result, several widely accepted metrics, namely Pixel Accuracy (PA), Mean Intersection over Union (mIoU), Weighted Precision, and Weighted F1 Score are employed. These metrics collectively provide a comprehensive assessment by considering overall pixel-wise correctness, per-class spatial overlap, and performance metrics weighted by

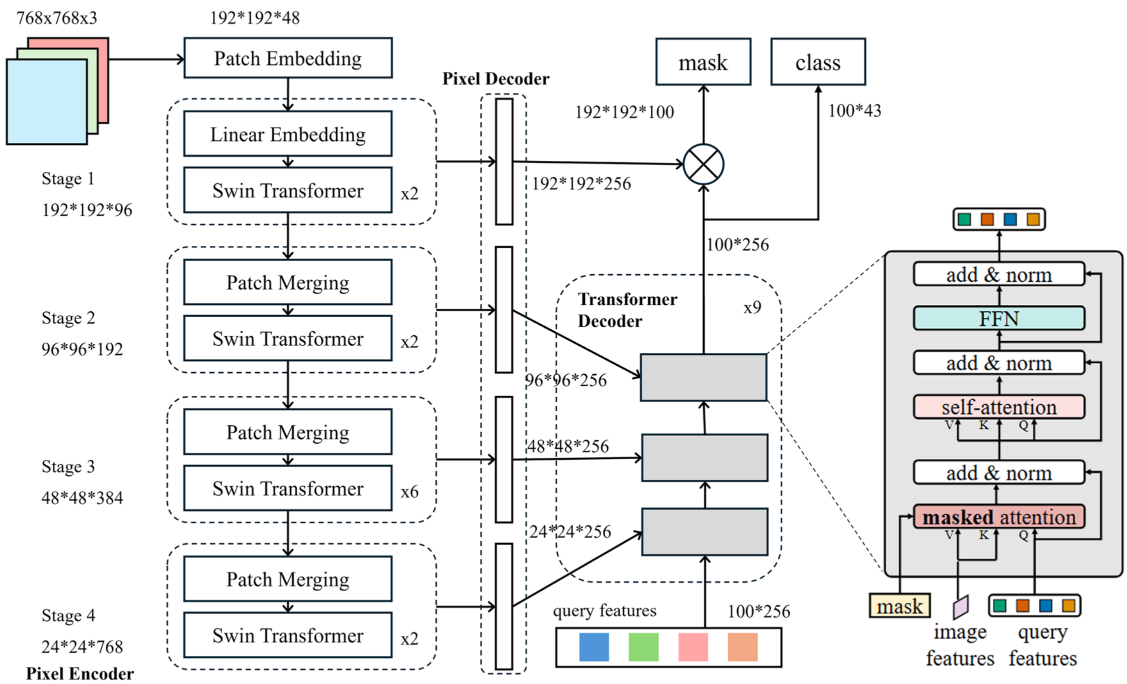
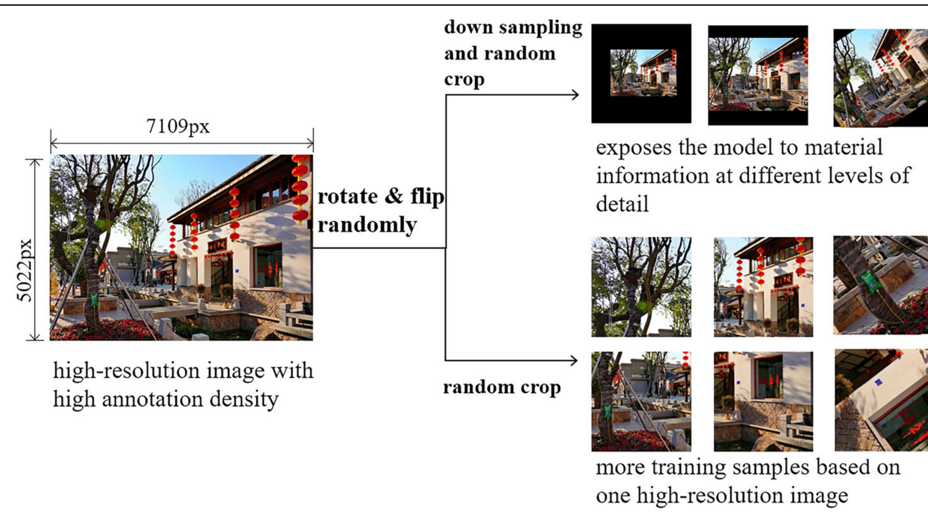


Fig. 4 | The architecture of the BFMS model. The left side of the figure illustrates the complete architecture of the BFMS model, which includes a four-stage pixel decoder, a pixel decoder, and a transformer decoder. The dimensions of each matrix are displayed next to the respective steps, presented in the order of height, width, and channel. The right side features a copy of the structure of the transformer decoder module, originally depicted in Cheng et al. (2022).

Fig. 5 | Examples of the data augmentation pipeline.



class support. Let k denote the number of classes, and for each class $i \in \{1, \ldots, k\}$, let TP_i, FP_i, FN_i, FN_i represent the number of True Positive, False Positive, False Negative and true negative pixels. The metrics are defined as follows:

$$mIoU = \frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{1}$$

$$PA = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k}(TP_i + FP_i)} \tag{2}$$

$$precision_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

$$recall_i = \frac{TP_i}{TP_i + FN_i} \tag{4}$$

$$F1\ Score_i = 2*\frac{Precision_i * Recall_i}{Precision_i + Recall_i} \tag{5}$$

The support for class i is given by $Support_i = TP_i + FN_i$. The weighted precision is the average of the per-class precision scores, weighted by the support of each class. Similarly, the weighted F1 score is the weighted average of per-class F1 scores. Since the weighted recall is equal to the pixel accuracy in numbers, this metric is not adopted in the evaluation. These metrics allow us to evaluate different aspects of the model's performance, providing a robust quantitative analysis, especially in challenging scenarios involving class imbalance.

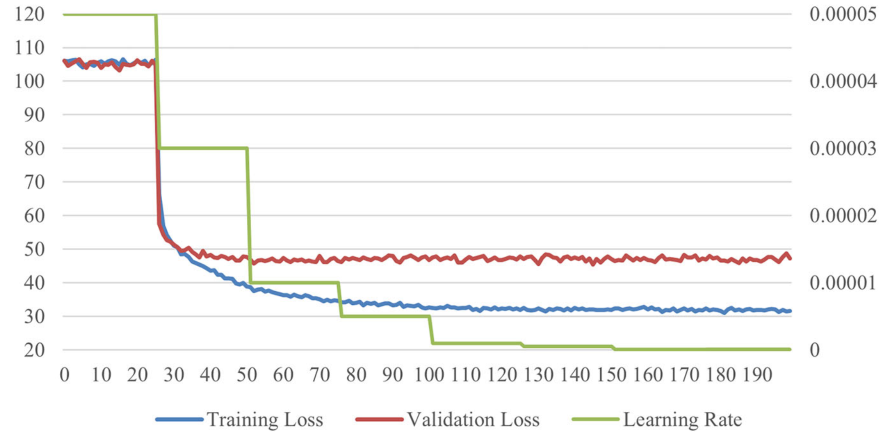


Fig. 6 | The training process of the BFMS model. The loss curves have proven the convergence of the model.

Algorithms for texture and color extraction

In this study, texture extraction from images is performed using a grid statistics method. The process involves segmenting raw images and masks into square patches of 100 pixels on each side, as determined from experimental comparisons and prior research[17]. The composition of each patch is analyzed, with those containing over 50% of a dominant material being categorized accordingly. These patches are then ranked by material prevalence and compiled into candidate lists for potential inclusion in a final texture library.

For color extraction, it has been reflected that white balance and exposure conditions have a large impact on the color extraction result[32]. Therefore, we apply the automatic white balance (AWB) correction on the image developed by Afifi & Brown[33], which employs a trained neural network. Following AWB, exposure is adjusted using a histogram cropping method that enhances image quality by cropping 1% of the extremes from both ends of the grayscale histogram and linearly stretching the remaining range. The experimental results show that this method can improve the quality of the image and restore the original color of the building façades relatively realistically.

In the process of identifying dominant colors in building materials, color variations were observed within the same material type across different buildings and even on different sections of the same building. For instance, bricks might generally appear orange. However, factors such as lighting and viewing angle can cause their color to vary from dark brown to orange-red, creating a gradient effect. Relying solely on counting pixel percentages for specific colors could lead to inaccuracies in classification. To address these variations more effectively, the K-means clustering method from Zhai et al.[34] is adopted. We first isolate pixels associated with a specific material using a masking technique, then transform the colors of these pixels into the HSV color space because the perceptual differences in color within HSV are more aligned with human vision compared to RGB space. We apply K-means clustering within this color space to ascertain the predominant color. To mitigate the influence of shadows, which can skew color perception, each color's value in the clustering algorithm is adjusted by a coefficient that is proportional to the square root of the V-value (brightness component) in the HSV model. This adjustment helps in achieving a more accurate representation of the material colors by reducing the distortion caused by varying light conditions. After that, we gather these representative colors derived from all instances of the same material type across the entire study area and apply K-means clustering again to identify the primary colors of this material in the whole area.

Results

Performance evaluation of the BFMS model

Figure 6 reflects the progression of training loss, validation loss, and learning rate throughout the training process of the BFMS model. Between epochs 25 and 30, a significant reduction in training loss is observed, coinciding with the learning rate adjustment, leading to rapid convergence. Similarly, the validation loss decreases sharply during this period, ultimately stabilizing around 100 epochs.

The mIoU of the final trained model on the validation set is calculated to be 0.415, with an overall accuracy of 76.50%, indicating relatively strong performance. Additionally, the IoU and accuracy for each category are computed, as shown in Table 6. Most materials achieve over 60% accuracy. However, it is important to note that 12 categories, including non-clear plastic, rubber, and pottery, exhibit a lower accuracy and IoU values. This is likely due to their infrequent appearance in the dataset, varied visual characteristics, or inherent difficulty in recognition. Accurately identifying these materials solely from images remains a challenge.

The study also compares the performance of the BFMS model on the validation set with the Mask2Former model trained on the DMS dataset (Table 7). Due to differences in material definitions between the DMS and BFMS datasets, IoU values are calculated only for materials with identical definitions. Table 4 demonstrates that the BFMS model outperforms the DMS model in terms of IoU for most commonly occurring outdoor materials, such as brick, glass, metal, and rubber. This suggests that the BFMS model has superior capability in accurately categorizing materials and identifying material edge contours.

The study also compares the performance of Mask2Former architecture with several classical sematic segmentation model, namely U-Net[35], DeepLab V3+[36], and SegFormer[37], which are the typical representatives of distinct segmentation approaches. The results reveal that the BFMS model has achieved the best results in all metrics including mIoU, pixel accuracy, weighted precision, and weighted F1 score (Table 8).

Application case: Taiping Alley area in Jingdezhen

The study applies the above method to the Taiping Alley historic landscape area to test its reliability and robustness. Taiping Alley is located in the center of Jingdezhen, an ancient cultural city in Jiangxi Province, China, which is renowned as the "porcelain capital" of the world. Taiping Alley area features a typical traditional Chinese urban style, with most buildings not exceeding three floors and relatively narrow streets (Fig. 6). Most buildings in the area use traditional building materials like brick and wood etc., and a few buildings have incorporated new materials during renovations. The project owner and the planning department have requested that the new buildings on the south side of the area be consistent with the surrounding area. For this reason, we apply the methods to the existing buildings on the north side, hoping to refine their landscape characteristics to provide guidance on the design of the new buildings.

The study investigates and photographs every building façade in Taiping Alley, and for some buildings with multiple façades, all shootable façades are photographed. Exclusions are made for a small number of privately owned buildings that cannot be entered after negotiation, as well as collapsed or severely damaged buildings (Fig. 7). The study filters out images

Table 6 | IoU and accuracy of the BFMS model on each category of materials

Material	IoU	Accuracy	Material	IoU	Accuracy
Roofing waterproof material	0.778	95.73%	Snow	0.591	65.64%
Sky	0.924	94.71%	Tree	0.525	64.71%
Foliage	0.691	88.11%	Natural Stone	0.382	64.37%
Hair	0.648	87.54%	Plastic, clear	0.507	64.19%
Water	0.850	86.86%	Food	0.235	62.39%
Brickwork	0.787	86.06%	Soil/ Mud	0.454	58.30%
Glass	0.711	85.29%	Engineered Stone/ Imitation Stone	0.476	57.30%
Skin/Lips	0.637	82.51%	Fabric/ Cloth	0.512	56.50%
Rammed earth	0.714	79.48%	Mirror	0.351	55.46%
Wood/Bamboo	0.532	76.73%	Ceramic	0.260	54.14%
Ground Tile	0.599	76.26%	Composite Decorative Board	0.288	51.79%
Asphalt	0.712	74.13%	Cardboard/ Paper	0.361	50.65%
Paint/ Coating/ Plaster	0.535	68.10%	Plastic, non-clear	0.228	33.12%
Photograph/ Painting/ Airbrushed Fabric	0.264	67.62%	Rubber/Latex	0.270	32.93%
Roof tile	0.585	67.34%	Pottery	0.165	21.35%
Metal	0.501	67.24%	Windows with metal fences on the outside	0.077	9.28%
Cement/ Concrete	0.435	66.22%	Leather	0.004	0.54%

Table 7 | Comparison of IoU of the BFMS and DMS model on the validation dataset

Material	BFMS	DMS	Material	BFMS	DMS
Sky	**0.924**	0.906	Snow	**0.591**	0.204
Foliage	**0.691**	0.519	Plastic, clear	**0.507**	0.006
Hair	0.648	**0.748**	Soil/ Mud	0.454	**0.491**
Brickwork	**0.787**	0.567	Fabric/ Cloth	**0.512**	0.415
Glass	**0.711**	0.606	Cardboard/ Paper	**0.361**	0.014
Skin/ Lips	**0.637**	0.623	Plastic, non-clear	**0.228**	0.058
Leather	0.004	**0.026**	Rubber/ Latex	**0.270**	0.039
Cement/ Concrete	**0.435**	0.320			

bold values indicate superior performance between the two models for each material category.

Table 8 | Comparison of model performance

Model	mIoU	Accuracy	Weighted Precision	Weighted F1 Score
U-Net (ResNet50)	0.205	0.634	0.642	0.517
DeeplabV3+ (ResNet50)	0.194	0.629	0.623	0.620
SegFormer-b3	0.241	0.672	0.676	0.668
BFMS (Mask2Former)	**0.415**	**0.765**	**0.782**	**0.770**

bold values indicate the best performance for each metric. Higher values represent better performance of Mask2Former architecture across all metrics.

with repetitive perspectives, ultimately resulting in 393 photos of 230 buildings, with each image having a resolution of 3072×4096 pixels.

Material segmentation result analysis in Taiping Alley area

The study conducts material segmentation along with color and texture extraction on all the collected photos. The task is done on a workstation equipped with a GTX 3090, and it takes less than 15 min to complete the computation and saving of 393 images. The average processing time is under 2.5 s per image, with computation time alone not exceeding 1 second per image. Figure 8 shows several examples of material segmentation results, and more images are provided in Supplementary Fig. 2. Overall, the model performs well in identifying materials. It not only achieves good recognition results for materials like brick walls, which occupy a large portion of the image but also demonstrates strong recognition accuracy for smaller objects, such as wooden sticks and wires. However, there are still some limitations: rough cement surfaces are occasionally misclassified as stone, and the model struggles to accurately predict ceramics when the camera is positioned too close.

The study also analyzes the material composition of the entire area and each building. The proportions of the materials show that the top building materials are cement/concrete, brick, paint, wood, and metal, followed by stone, ground tiles, and glass. Figure 9 displays the dominant materials for each building alongside examples of typical textures. The data confirm that brick and cement are the most common façade materials, scattered throughout the neighborhood. A smaller number of buildings primarily feature wood or tile veneer. This result is consistent with the researcher's actual observation made during the study, which notes that the main structural components in the area consist of brick and wood, with some facades coated in cement mortar for moisture protection.

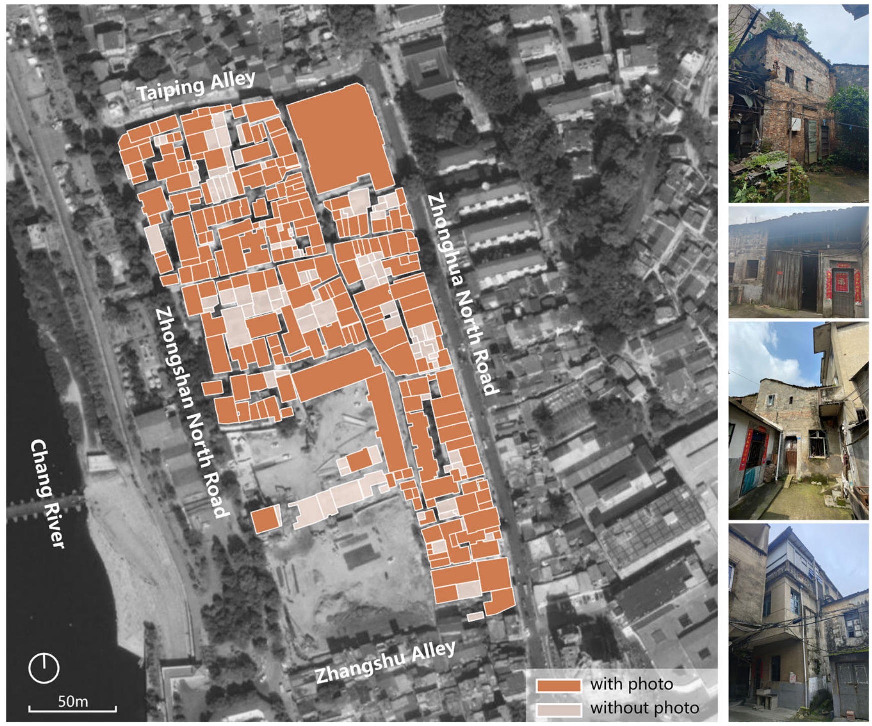
The study extracts the eight main colors of each type of material of each building and then clusters each type of material to ascertain the primary colors of building materials across the entire area (Fig. 10). For instance, gray emerged as the dominant color for brick materials in this region, complemented by some brownish-red bricks, and light blue waterproof paint is also applied to some of the brick surfaces. The colors for paint and plaster are mainly gray, supplemented by blue and dark red. Most of the wood is mainly dark brown, with few logs in bright colors. The tile veneer is mostly gray and white tiles, with some tiles having a blue or purple color bias. Ground tiles predominantly exhibit gray tones with a greenish tint, which is related to the damp ground surfaces in this area that are mostly covered with moss and other vegetation.

The application of material analysis results of Taiping Alley

The analytical results concerning materials and colors provide an objective and quantitative foundation for the classification of the architectural landscape and the formulation of subsequent preservation and renewal strategies in historic area.

The classification and grading of the architectural landscape is one of the most critical underpinnings for the development of planning policies. However, previous classification efforts often relied heavily on the subjective assessments of planners, lacking a systematic framework of indicators. The

Fig. 7 | Distribution of buildings in the study area and examples of taken photos.



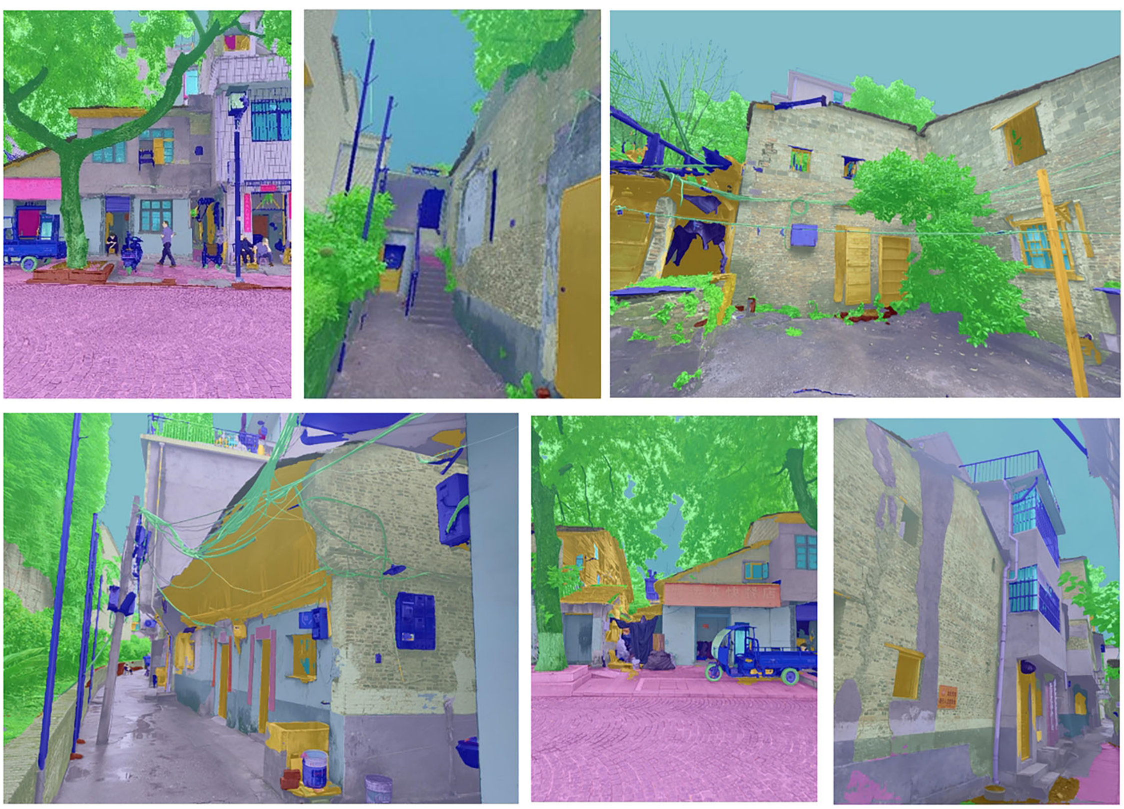


Fig. 8 | Examples of the prediction result of the BFMS model. The colored classification maps obtained from the model prediction are overlaid on the original photos with 50% transparency to better represent the analysis result. Each color represents the same material as shown in Table 5.

proportional composition of façade materials, conversely, offers a significant basis for such assessments. Utilizing the generated classification results for the primary façade materials of individual buildings, we conducted a preliminary data-driven evaluation of urban landscape consistency. Based on preliminary investigations confirming that locally listed historic buildings predominantly feature brick and natural stone as their primary material, this study categorizes buildings with brick, natural stone, rammed earth, or timber as the dominant façade material as "consistent with the prevailing urban landscape" Conversely, buildings primarily featuring imitation stone or ceramic tiles are classified as "inconsistent". Accordingly, 107 buildings

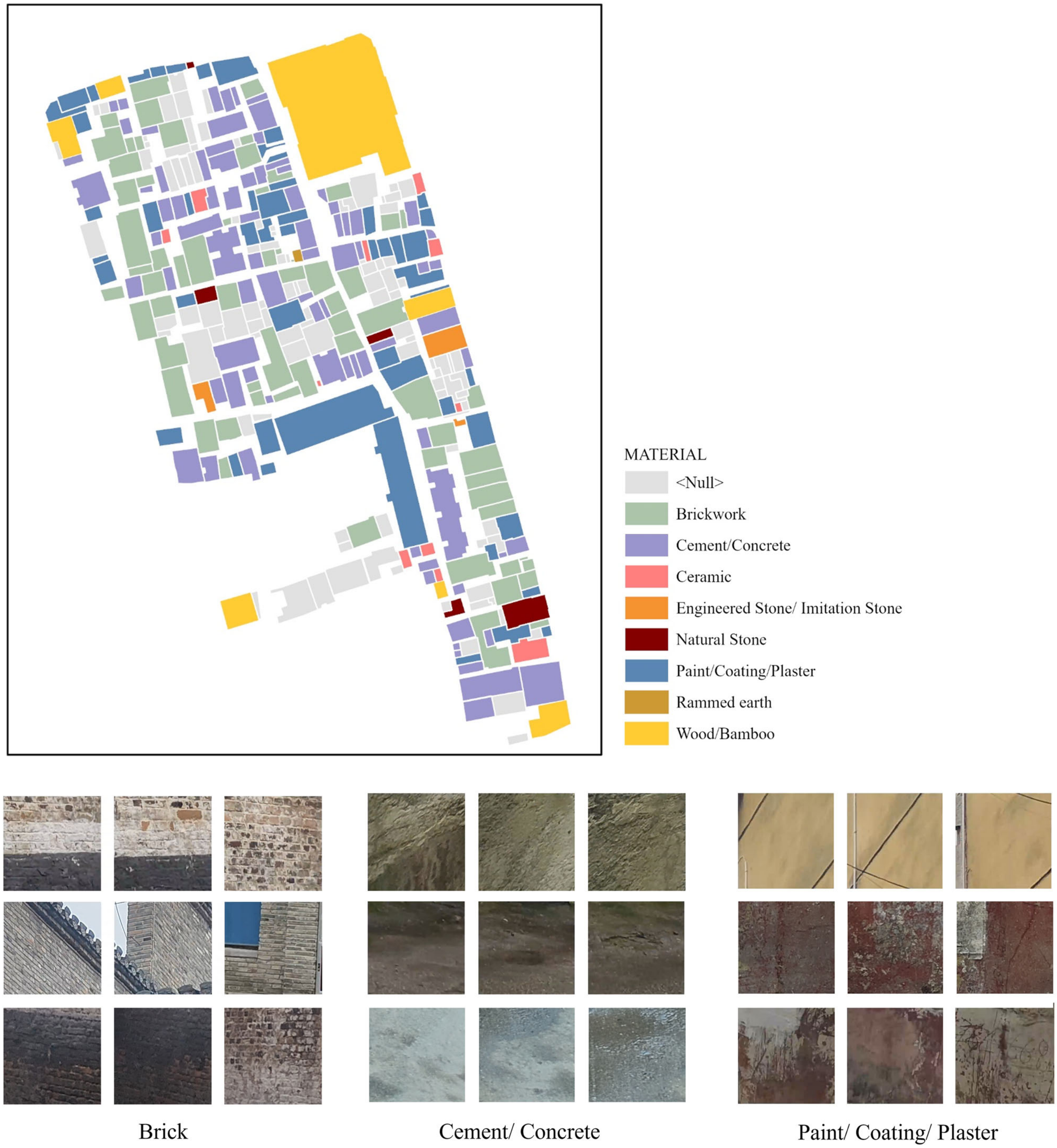


Fig. 9 | Dominant materials in each building in the area and the examples of extracted texture slices.

were assessed for urban landscape consistency. For structures with paint or concrete as the principal exterior wall material, further manual classification was performed, integrating considerations of glass, metal, and overall façade form. To validate the efficacy of this approach, three independent planners were invited to conduct assessments based solely on photographic evidence, without prior knowledge of the material proportions. The results indicated that among the 107 buildings directly classified by the system, discrepancies with the planners' assessments arose in only six instances. All such discrepancies were attributable to system misclassification stemming from the application of modern imitation antique veneers.

Furthermore, the analysis of materials and colors has informed the refinement of urban landscape control measures within the area's conservation and renewal plan. The revised planning regulations stipulate that the renovated or repaired building façades within this area should primarily utilize brick, potentially supplemented with exterior plaster and adorned with timber elements. Regarding color, renovated or newly constructed buildings are required to adopt an achromatic palette (black, white, gray) as the principal color scheme, appropriately augmented with hues of natural wood brown, purple, and dark red. This series of specific requirements, derived from quantitative analysis, significantly enhances the reliability and validity of planning control for the urban landscape. As a practical application guided by these planning principles, the subsequently designed community center in the southern part of the area (Fig. 11) employed red and gray bricks as

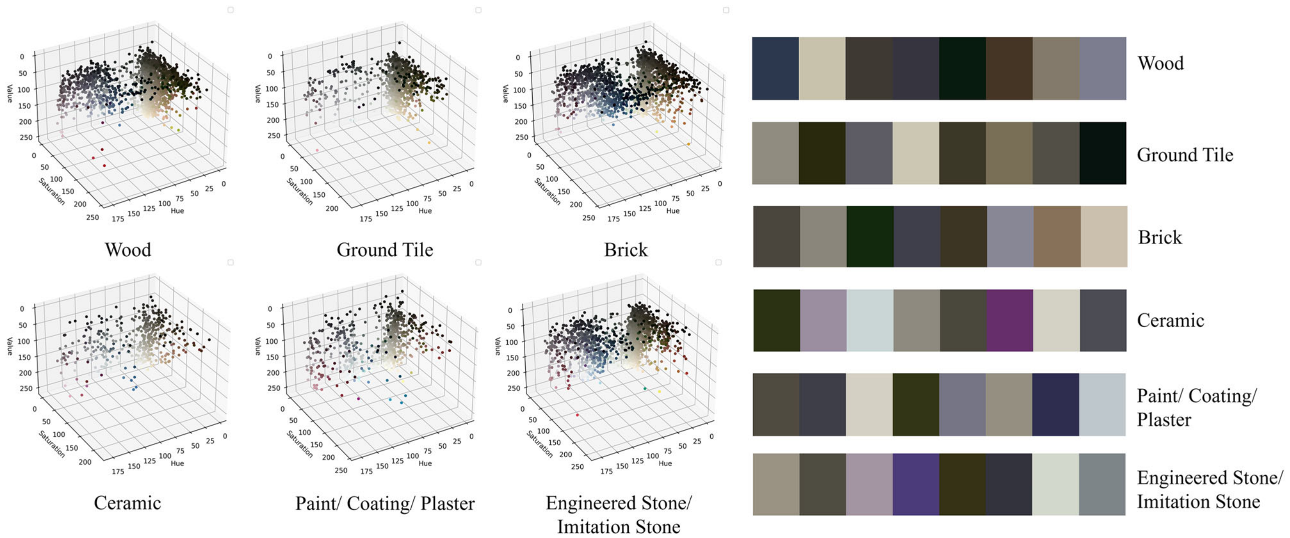


Fig. 10 | Examples of extracted colors of each material. Each color point in the 3D plot on the left represents the clustering result for an individual building, and on the right is the result obtained by clustering all building results again.

Fig. 11 | Architectural design for the southern blocks of the Taiping Alley neighborhood. The designed community center employed red and gray bricks as its primary façade materials, aiming to achieve harmonious integration with the existing urban landscape of the region.

its primary façade materials, aiming to achieve harmonious integration with the existing urban landscape of the region.

Discussion

In this study, we introduced the BFMS database, which includes 42 distinct building materials, to mitigate the prevalent dataset deficiencies in building façade materials segmentation. Compared to prior datasets, this dataset targets the building facades in various urban and rural contexts, which offers a broader range of material categories and aligns more closely with architectural standards. It also features higher-resolution images and higher density of label segments in each image, which preserves detailed material information and enhances opportunities for data augmentation.

Based on the database, a BFMS model is developed which is capable of automatically recognizing various types of building materials in urban imagery. This model surpasses previous approaches by enabling precise, pixel-level material recognition. We have demonstrated that our proprietary dataset significantly enhances the model's accuracy in identifying common exterior building materials. The comparison with other models has shown its superiority. Additionally, we develop an algorithm that utilizes the model's output to extract the texture and color features of each material through K-means clustering and grid statistics. The practical application of our methodology in Jingdezhen's Taiping Alley validates its effectiveness and precision.

The core value of this methodological system lies in three aspects. Firstly, this methodological system facilitates the automation and efficiency of material survey and documentation, enabling the originally labor-intensive task of material analysis to be completed in a few minutes. The material and color selection in the historic area may imply certain historic information including the built years, construction method, and cultural characteristics[38,39]. These databases are also invaluable for analyzing the quantitative characteristics of material usage and color preferences across different historical periods, regions, or architectural styles, thereby allowing for deeper insights into the historic urban landscapes. For instance, our analysis of the Taiping Alley area revealed a predominance of gray bricks, contrasting sharply with the prevalence of red bricks in kilns built in 1980s in Jingdezhen. Such distinct differences in material and color effectively reflect variations in construction practices and preferences tied to different historical periods and building typologies.

Secondly, the system provides a significantly more objective basis for architectural character assessment and management within heritage preservation efforts. By automatically extracting and quantifying the material and color information from building facades, the system can quickly identify buildings that are inconsistent with the prevailing historical character or esthetic of an area. This capability transforms the process of architectural landscape characterization, which traditionally relied heavily on the subjective judgment of urban planners and experts, into a more evidence-based and thus more scientific approach, substantially enhancing the reliability of preservation decisions.

Thirdly, the system offers direct, actionable support for formulating effective conservation and renewal strategies. By providing detailed insights into existing material palettes and historical precedents, it can directly inform and guide the design strategies for new interventions or restoration work on surrounding buildings, ensuring compatibility and contributing to the overall landscape of the historic environment. Application in the Taiping Alley area in Jingdezhen has demonstrated this. The results of the material analysis are rapidly integrated into the design process, providing valuable material selection references for renovation projects in adjacent areas.

We also envision that the application of the model extends beyond urban conservation area to broader research and practice in the built environment. Although the images used for material segmentation are primarily manually captured high-definition building images, the methodology is equally applicable to material analysis of other urban images, such as street view and window view images[40]. Recent studies have highlighted the importance of indicators such as the green view index (GVI) and sky view factor (SVF) in evaluating built environments[41–43]. It is conceivable that with the application of material segmentation algorithms, similar indicators such as glass visual ratio and wood visual ratio can also be proposed. For example, the glass visual ratio may become an indicator for measuring the feeling of modernization in an urban area, and the wood visual ratio may serve as an important measurement for the protection of areas with traditional buildings. The metrics based on visual elements possibly contribute to the future research of the built environment. Moreover, material segmentation may also contribute to the research and practice in predicting the carbon emissions of buildings, identifying the need for renewal and renovation[44], and designing building facades with generative models[45]. Further quantitative research can be carried out in the above areas based on the methodology proposed in this paper.

Despite the promising results, this study faces several constraints that warrant acknowledgment. First, the BFMS dataset remains relatively small, which limits the model's accuracy with unfamiliar materials. Since the current dataset focus on Chinese historic areas, a globally oriented dataset collection may be necessary for subsequent research. Considering the time and cost of dataset construction, future efforts might consider adopting the fuzzy labeling approach used in the Cityscapes dataset. Generative models like Stable Diffusion may also be applied into the data augmentation process to solve the imbalance of the datasets. Second, considering the material segmentation remains a challenging problem in the area of computer vision, our segmentation method remains to be improved. Subsequent research may try to integrate more semantic information, such as object categories, to provide richer contextual cues for the model. Such integration may facilitate improved discrimination between materials which are visually similar but intrinsically different. Thirdly, our approach relies on planar visual analysis rather than comprehensive three-dimensional material assessment. While this methodology effectively captures the perceptual qualities of building façades that define urban character, which aligns with established conservation practice paradigms, integration of multi-source spatial information like building footprints and LiDAR point clouds represents a promising direction for future research that could extend the current framework toward more comprehensive material analyses. Such enhanced methodologies would be particularly valuable for applications beyond landscape assessment, including material stock inventories and embodied energy calculations. These limitations notwithstanding, the present study establishes a foundational approach for objective, data-driven assessment of material composition in historic districts that addresses a critical gap in current heritage conservation practice.

Data availability

The validation dataset used in this study is publicly available through Figshare (https://figshare.com/s/4b171bf1f657d12cbfd5). The implementation code and trained model are also publicly accessible via Figshare (https://figshare.com/s/fd38d547fdb8708381f5). For any additional data or materials not included in these repositories, please contact the corresponding author.

Received: 1 February 2025; Accepted: 15 June 2025;
Published online: 30 July 2025

References

1. Ogunkah, I. & Yang, J. Investigating factors affecting material selection: the impacts on green vernacular building materials in the design-decision making process. *Buildings* **2**, 1–32 (2012).
2. Feng, L. & Hu, Z. Research on the Use of Building Cladding Materials in the Concessions of Modern Tianjin: In *East Asian Architecture in Globalization* (eds. Xu, S., Aoki, N. & Vieira Amaro, B.) 322–333 (Springer International Publishing, Cham, 2021).
3. Morris, W. The influence of building materials upon architecture. *Century Guild Hobby-. horse* **6**, 1–14 (1892).
4. Jalali, S., Parapari, D. M. & Mahdavinejad, M. J. Analysis of Building Facade Materials Usage Pattern in Tehran. *AEF* **31**, 46–62 (2019).
5. Adam, J.-P. *Roman Building: Materials and Techniques*. (Routledge, 2005). https://doi.org/10.4324/9780203984369.
6. Choplin, A. Building concrete futures: materiality and urban lives in the West. *Africa*. **93**, 20–39 (2023).
7. Punter, J. A history of aesthetic control: part 1, 1909-1953: the control of the external appearance of development in England and Wales. *Town Plan. Rev.* **57**, 351 (1986).
8. Hong, C. et al. Is the design guidance of color and material for urban buildings a good choice in terms of thermal performance. *Sustain. Cities Soc.* **83**, 103927 (2022).
9. Rezafar, A. & Turk, S. S. History of aesthetic control management in planning system, the case of Turkey. *Int. Plan. History Soc. Proc.* 165–175 https://doi.org/10.7480/IPHS.2016.6.1327 (2016).
10. Ibrahim, M. R., Haworth, J. & Cheng, T. Understanding cities with machine eyes: a review of deep computer vision in urban analytics. *Cities* **96**, 102481 (2020).
11. Li, M., Yeh, A. G. O. & Xue, F. CIM-WV: A 2D semantic segmentation dataset of rich window view contents in high-rise, high-density Hong Kong based on photorealistic city information models. *Urban Info* **3**, 12 (2024).
12. Hou, J., Zhou, J., He, Y., Hou, B. & Li, J. Automatic reconstruction of semantic façade model of architectural heritage. *Herit. Sci.* **12**, 1–14 (2024).
13. Wu, M., Chai, R., Zhang, Y. & Lu, Z. Dadnet: dual-attention detection network for crack segmentation on tomb murals. *Herit. Sci.* **12**, 1–12 (2024).
14. Zhang, C. et al. Application of deep learning algorithms for identifying deterioration in the ushnisha (Head Bun) of the Leshan Giant Buddha. *Herit. Sci.* **12**, 1–17 (2024).
15. Cao, J. et al. Ancient mural segmentation based on multiscale feature fusion and dual attention enhancement. *Herit. Sci.* **12**, 1–15 (2024).
16. Bell, S., Upchurch, P., Snavely, N. & Bala, K. Material recognition in the wild with the Materials in Context Database. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3479–3487 (IEEE, 2015). https://doi.org/10.1109/CVPR.2015.7298970.
17. Chen, X., Ding, X. & Ye, Y. Mapping sense of place as a measurable urban identity: Using street view images and machine learning to identify building façade materials. *Environ. Plan. B Urban Anal. City Sci.* **0**, 1–20 (2024).
18. Sun, Y. & Gu, Z. Using computer vision to recognize construction material: A Trustworthy Dataset Perspective. *Resour., Conserv. Recycl.* **183**, 106362 (2022).
19. Upchurch, P. & Niu, R. A Dense Material Segmentation Dataset for Indoor and Outdoor Scene Parsing. In *Computer Vision – ECCV 2022* (eds. Avidan, S., Brostow, G., Cissé, M., Farinella, G. M. & Hassner, T.) 450–466 (Springer Nature Switzerland, Cham, 2022).
20. Cai, S., Wakaki, R., Nobuhara, S. & Nishino, K. RGB Road Scene Material Segmentation. in *Computer Vision—ACCV 2022* (eds. Wang, L., Gall, J., Chin, T.-J., Sato, I. & Chellappa, R.) vol. 13842 256–272 (Springer Nature Switzerland, 2023).
21. Bell, S., Upchurch, P., Snavely, N. & Bala, K. OpenSurfaces: a richly annotated catalog of surface appearance. *ACM Trans. Graph.* **32**, 1–17 (2013).
22. Dana, K. J., Van Ginneken, B., Nayar, S. K. & Koenderink, J. J. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.* **18**, 1–34 (1999).

23. Sharan, L., Rosenholtz, R. & Adelson, E. H. Accuracy and speed of material categorization in real-world images. *J. Vis.* **14**, 12 (2014).
24. Xue, J., Zhang, H., Nishino, K. & Dana, K. J. Differential viewpoints for ground terrain material recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 1205–1218 (2022).
25. Murmann, L., Gharbi, M., Aittala, M. & Durand, F. A Dataset of Multi-Illumination Images in the Wild. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 4079–4088 (2019).
26. Schwartz, G. & Nishino, K. Recognizing material properties from images. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 1981–1995 (2020).
27. Gonzales, M. E. M., Uy, L. C. & Ilao, J. P. Designing a lightweight edge-guided convolutional neural network for segmenting mirrors and reflective surfaces. in 107–116 https://doi.org/10.24132/CSRN.3301.14 (2023).
28. Plaksyvyi, A., Skublewska-Paszkowska, M. & Powroznik, P. A comparative analysis of image segmentation using classical and deep learning approach. *Adv. Sci. Technol. Res. J.* **17**, 127–139 (2023).
29. Ministry of Housing and Urban-Rural Development of the People's Republic of China. Standard for Terminology of Building Materials (JGJ/T 191-2009). (2009).
30. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A. & Girdhar, R. Masked-attention Mask Transformer for Universal Image Segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 1280–1289 (2022).
31. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF international conference on computer vision (ICCV)* 9992–10002 (2021).
32. Zhang, J., Fukuda, T. & Yabuki, N. Development of a city-scale approach for façade color measurement with building functional classification using deep learning and street view images. *ISPRS Int. J. Geo-Inf.* **10**, 551 (2021).
33. Afifi, M. & Brown, M. S. Deep White-Balance Editing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 1394–1403 (IEEE, 2020).
34. Zhai, Y., Gong, R., Huo, J. & Fan, B. Building façade color distribution, color harmony and diversity in relation to street functions: using street view images and deep learning. *ISPRS Int. J. Geo-Inf.* **12**, 224 (2023).
35. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, Cham, 2015).
36. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer vision – ECCV 2018* (eds. Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y.) 833–851 (Springer International Publishing, Cham, 2018).
37. Xie, E. et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Advances in Neural Information Processing Systems* vol. 34 12077–12090 (Curran Associates, Inc., 2021).
38. Hou, X., Cheng, B. & Yang, J. A quantitative study on the exterior wall texture of stone-built dwellings in traditional villages in China: a case study of the xisuo village in the Jiarong Tibetan area. *J. Build. Eng.* **42**, 102357 (2021).
39. Xue, X., Tian, Z., Yang, Y., Wang, J. & Cao, S.-J. Sustaining the local color of a global city. *Nat Cities* **2**, 400–412 (2025).
40. Li, M., Xue, F., Wu, Y. & Yeh, A. G. O. A room with a view: automatic assessment of window views for high-rise high-density areas using City Information Models and deep transfer learning. *Landsc. Urban Plan.* **226**, 104505 (2022).
41. Li, X. et al. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* **14**, 675–685 (2015).
42. Lin, T., Tsai, K., Hwang, R. & Matzarakis, A. Quantification of the effect of thermal indices and sky view factor on park attendance. *Landsc. Urban Plan.* **107**, 137–146 (2012).
43. Miao, C. et al. Review of methods used to estimate the sky view factor in urban street canyons. *Build. Environ.* **168**, 106497 (2020).
44. Dai, M., Ward, W. O. C., Meyers, G., Densley Tingley, D. & Mayfield, M. Residential building facade segmentation in the urban environment. *Build. Environ.* **199**, 107921 (2021).
45. Wang, B., Li, M., Peng, Z. & Lu, W. Hierarchical attributed graph-based generative façade parsing for high-rise residential buildings. *Autom. Constr.* **164**, 105471 (2024).
46. Xue, J., Zhang, H., Dana, K. & Nishino, K. Differential angular imaging for material recognition. In *2017 IEEE conference on computervision and pattern recognition (CVPR)* 6940–6949 (2017).

Acknowledgements

This research was partly supported by the National Office for Philosophy and Social Sciences (Grant No. 23VJXT019). I would like to thank the Tsinghua Tongheng Urban Planning and Design Institute for providing valuable image resources to support this study.

Author contributions

X.J.—Conceptualization; Methodology; Validation; Formal analysis; Investigation; Resources; Data Curation; Writing; Visualization. L.M.—Conceptualization; Resources. W.J.—Resources: Writing. Z.X.-Supervision. Z.J.—Conceptualization; Supervision; Project administration; Funding acquisition. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s40494-025-01888-4.

Correspondence and requests for materials should be addressed to Jie Zhang.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025