Response to Reviewer 1

Manuscript: *BMAT: A footprint-level building facade material dataset for 73 major cities worldwide*

We are grateful to the reviewer for the careful reading of our manuscript and for the constructive comments. We have carefully considered each point and revised the manuscript accordingly. Below, we respond to the comments individually. The reviewer's comments are shown in blue, and the revised text from the manuscript is quoted in *gray italics* where appropriate.

General comment.

> **Comment General.** *This manuscript presents BMAT, a large scale facade material dataset covering 22.09 million building footprints across 73 cities, derived from approximately 147 million Google Street View and Baidu Street View images. The dataset addresses a gap in existing building databases. However, the current validation does not sufficiently establish the reliability of the final building level and multi temporal product. The reported F1 score of 0.91 evaluates image classification performance, but the dataset also depends on correct image to building correspondence, adequate facade visibility, consistent aggregation of multiple observations, stable building identities across years, and representative spatial coverage. These components have not been comprehensively validated. In addition, each record appears to describe the dominant material visible from a street facing viewpoint rather than the complete building envelope. The claims concerning whole building characteristics, temporal transitions, embodied carbon, and energy modeling should therefore be moderated. I recommend major revision.*

Response:

Thank you for recognizing the importance of this dataset and for these valuable suggestions. We strengthened the reliability assessment at three levels. The material classifier achieved an F1 score of 0.91 on the held-out test set. Comparison with 50,823 Overture Maps facade-material records yielded an agreement of 0.72 for the latest BMAT records and 0.79 under the historical matching strategy. We conducted an AI-assisted audit of 2,000 cases across the complete pipeline, including 1,000 final records retained for material inference, to assess target-building visibility, image–footprint correspondence, and material labels, with a material-label agreement of 87.3%. We further clarified that BMAT records the dominant visible street-facing facade material linked to each building footprint, rather than the complete building envelope or all facade materials. The temporal records are now presented as temporally stamped observations, and material changes are interpreted as exploratory evidence. We also refined the application claims by describing facade material as an important building-attribute input for studies of energy, embodied carbon, urban microclimates, and hazard risk. Detailed methods and results are provided in the responses below.

Changes in the manuscript:

Abstract:

"Dataset reliability was further assessed through comparison with 50,823 Overture Maps facade-material records, yielding agreements of 0.72 for the latest BMAT records and 0.79 under a historical matching strategy. In cities with sufficient repeated imagery, the temporally stamped records provide exploratory evidence of facade-material transitions. By linking building geometry with facade-material information, BMAT provides facade material as an important building-attribute input for studies of building energy, embodied carbon, urban microclimates, and hazard risk."

Discussion:

"BMAT provides building-level information on dominant visible street-facing facade materials, an important attribute that is commonly absent from large-scale building datasets. These records provide a material input for studies of building stocks, urban microclimates, building energy, embodied carbon, and material-specific hazard vulnerability, reducing reliance on uniform or assumed material assignments. The temporally stamped records can also support exploratory analyses of persistent facade-material changes. These applications should be interpreted within the scope of BMAT, which represents the dominant material visible from street-facing views rather than the complete material composition of the building envelope."

Major comments.

> **Comment 1.** *The meaning of a BMAT material record needs to be defined more precisely. The manuscript describes BMAT as a footprint level building facade material dataset, although each label appears to represent the dominant material visible from one or more street facing images rather than the entire building envelope. This distinction is important for mixed material buildings, where different facades or facade sections may contain glass, stone, concrete, metal, or other materials. The authors should clearly define what each label represents, explain how the dominant material is determined, and describe how mixed facades are handled. The dataset should also include information on facade orientation, number of images and viewpoints, visible facade proportion, and visibility status. If such information is unavailable, the manuscript should consistently describe BMAT as a dataset of dominant visible street facing facade materials.*

Response: Thank you for this important and constructive comment. We agree that BMAT material labels should be defined more precisely. BMAT does not provide a complete per-facade inventory of all materials across the entire building envelope. Instead, each material record represents the dominant material visible from street-view images on the street-facing facade of the target building. Therefore, BMAT should be understood as a footprint-level dataset of dominant visible street-facing facade materials.

For the material categories, we first referred to prior studies on building facade materials and street-view-based material recognition, and defined seven common facade material types: brick, concrete, glass, metal, stone, stucco, and wood. We further included an "other" category for buildings under construction, renovation, or with materials that could not be assigned to the main categories. During manual annotation, we also observed that tile-clad facades are common in many Chinese cities and some Middle Eastern cities, so we added tile as an additional category. As a result, BMAT uses nine material classes: brick, concrete, glass, metal, stone, stucco, tile, wood, and other.

The dominant material is defined as the material that occupies the main visible part of the street-facing facade. In most urban buildings, one facade material is visually dominant, so a primary material label can be assigned. For mixed-material facades, BMAT does not attempt to record the full material composition. Instead, the label is assigned according to the dominant material on the main visible facade body. For example, when the ground-floor base, storefront, or local decorative elements differ from the upper facade, the material of the upper main facade is used as the primary material label. If the upper facade itself contains multiple materials, the material occupying the largest visible proportion is used as the primary material.

Regarding facade orientation, BMAT material records are mainly derived from the side of the building visible from street-view imagery, namely the street-facing facade. Thus, for multi-facade or corner buildings, BMAT is better interpreted as a record of the dominant material on the street-visible side. However, for most urban buildings, the material of the street-facing facade is

generally consistent with the materials on other sides of the building, so the street-visible facade provides a practical and representative observation for large-scale building material mapping.

Regarding visibility, we applied two filtering steps before material inference to reduce the influence of low-quality images and incorrect target views. First, a visibility model removes images in which the target building is not visible or the image quality is insufficient, such as cases affected by severe overexposure, tree occlusion, mosaicking, blurring, or other visual degradation. Second, an obstruction model is used to identify cases where the line of sight between the street-view sampling point and the target building is blocked by intervening buildings; such images are excluded because they may show another building rather than the intended target. After these filtering steps, material inference is performed mainly on street-view images in which the target building is visible and the image quality is relatively suitable for facade material interpretation.

We appreciate the reviewer's suggestion and agree that these metadata would further support data interpretation. As BMAT is primarily designed to provide footprint-level facade-material labels, detailed image-level attributes are not included as building-level fields in the current release; instead, we report city-level statistics on candidate, not-visible, obstructed, and valid images in Table 6 (Supplementary Table S6).

We have revised the manuscript to define the scope and role of BMAT more precisely as a footprint-level record of dominant visible street-facing facade materials. The key revisions are listed below:

- **Dataset definition in the Abstract.** The definition of BMAT was revised to clarify the scope of each material label:

  > *"Here we present BMAT, a footprint-level dataset of dominant visible street-facing facade materials covering 22.09 million buildings across 73 major cities worldwide."*

- **Dataset definition in the Introduction.** We revised the initial description of BMAT in the Introduction to make clear that the dataset represents dominant visible street-facing facade materials rather than the complete material composition of the building envelope:

  > *"To address these gaps, this study presents BMAT, a footprint-level dataset of dominant visible street-facing facade materials derived from over 147 million multi-temporal street-view images covering 73 major cities worldwide."*

- **Dataset description in the Discussion.** We revised the description of the dataset in the Discussion to maintain this distinction when summarizing the main contribution of BMAT:

  > *"To our knowledge, BMAT is the first large-scale dataset that links dominant visible street-facing facade material labels to individual building footprints."*

- **Limitation of street-facing observations.** The limitation related to facade orientation was clarified:

> *"Third, BMAT records the dominant material visible from street-facing views. For a small proportion of buildings whose materials differ across facades, the street-facing facade material may not fully represent the material composition of the entire building."*

- **Dataset description in the Conclusions.** The description of the dataset in the Conclusions was revised to use the same definition consistently:

> *"This study presents the first large-scale, footprint-level dataset of dominant visible street-facing facade materials, effectively bridging the gap between geometric building footprint data and facade material attributes."*

Comment 2. *The correspondence between street view images and target building footprints has not been adequately validated. The material classification model assumes that the central building in each image is the intended target, but the geometric retrieval process does not fully guarantee this correspondence. Multiple buildings may appear in one image, while footprint errors, panorama positioning, road geometry, camera parameters, building height uncertainty, and unmodelled obstructions such as trees, walls, vehicles, and elevated roads may affect target visibility. The assumption that partial facade exposure is sufficient also requires empirical validation. The authors should therefore conduct a stratified manual audit of the final dataset to assess whether the intended building is correctly identified, sufficiently visible, and assigned the correct material label. The audit should cover different cities, platforms, building densities, sampling distances, building heights, and visibility conditions, with separate results for fully and partially visible buildings. Without such validation, image classification performance cannot be directly interpreted as the accuracy of the footprint level dataset.*

Response: Thank you for this valuable suggestion. To verify whether material labels derived from street-view images were correctly assigned to the target building footprints, we evaluated three key aspects: (1) the visibility of the target building in the street-view image; (2) the correspondence between the building shown in the image and the target footprint; and (3) the correctness of the assigned material label.

We conducted a stratified audit of 2,000 unique cases covering different cities, street-view platforms, sampling distances, building densities, building heights, and material categories. The sample covered 18 cities, including 13 Google Street View cities and five Baidu Street View cities. It comprised 500 images predicted as not visible, 400 cases predicted as fully obstructed, 100 cases predicted as partially visible, and 1,000 images in which the building was predicted to be visible and the central line of sight to the target building was unobstructed. The last group was retained for material inference.

We developed an AI-assisted visual audit platform that jointly displayed the street-view image, target footprint, sampling point, viewing direction, and surrounding buildings (Figure 1). We used GPT-5.6-Sol with the high-reasoning setting because the audit required both image understanding and spatial reasoning. The model was blinded to the original BMAT results and assessed each case independently. Rather than relying solely on judgments from the authors or a small group of human annotators, the model served as a consistent auxiliary evaluator. It applied the same predefined criteria to every case, thereby reducing variation in the evaluation process and providing a standardized assessment. According to our audit objectives, the model evaluated building visibility, obstruction status, image–footprint correspondence, and facade material.

1. Visibility of the target building

We evaluated target-building visibility in two steps. First, the visibility model selected street-view images that contained a building view suitable for material recognition. Second, the

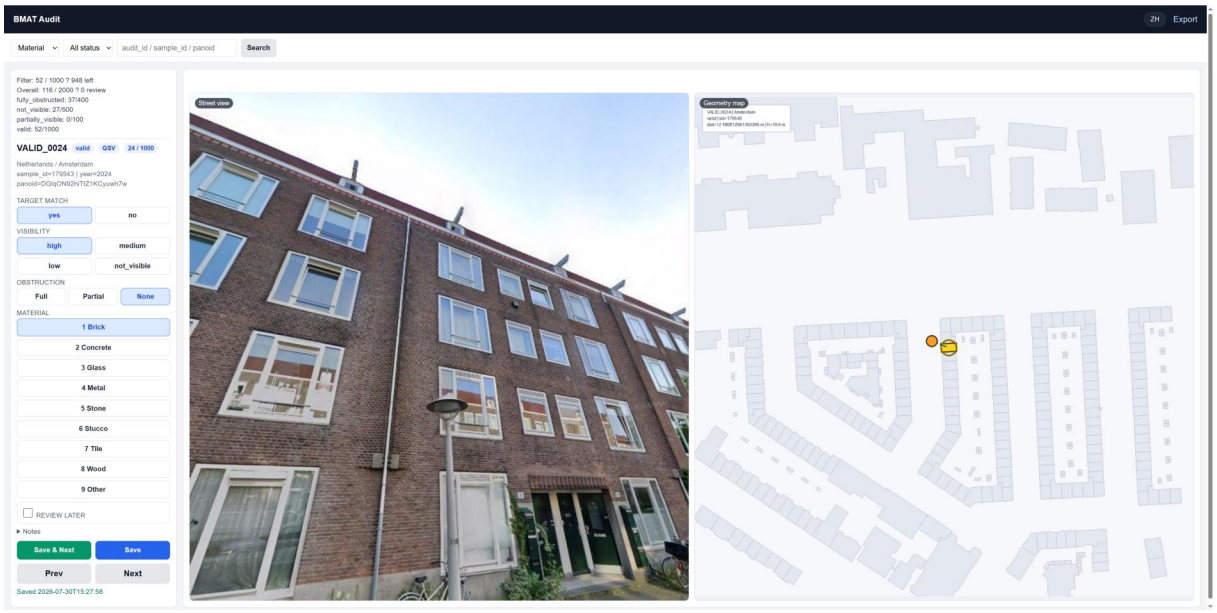

Figure 1: AI-assisted visual audit platform for BMAT validation. The interface jointly displays the street-view image and a geometry map containing the target footprint, surrounding buildings, street-view sampling point, and viewing direction. It supports structured assessment of image-footprint correspondence, target-building visibility, obstruction status, and dominant facade material.

obstruction analysis determined whether the target building itself remained visible from the sampling point or was blocked by an intervening building.

1.1 Image-level visibility

Street-view images may contain samples unsuitable for material recognition, including over-exposed or blurred images, indoor scenes, mosaicking artifacts, and views in which buildings are heavily blocked by trees or vehicles. To identify suitable building images, we trained a binary visibility classifier based on MobileNetV2. The training dataset contained 8,934 labeled images, including 5,903 valid and 3,031 invalid images, and was divided into training and validation sets at a ratio of 80:20. The model was initialized with ImageNet-pretrained weights and trained for 100 epochs using the Adam optimizer with a learning rate of 1×10^{-4}. It achieved a validation accuracy of 88.75%. For each image, the model returned a `building=yes/no` label and the corresponding confidence score. Images classified as `building=yes` were retained for subsequent obstruction analysis and material inference.

To evaluate whether the images retained by the visibility classifier were genuinely suitable for subsequent inference, we conducted an AI-assisted audit of a stratified sample of 1,000 images predicted as `building=yes`. The target-building visibility in each image was classified as *high*, *medium*, *low*, or *not visible*. In our workflow, any image in which the target building remains visible is considered a candidate image; therefore, the high-, medium-, and low-visibility categories were all treated as visible. We first compared the four audited visibility levels with the confidence scores produced by the original classifier. The confidence distributions were generally consistent with the audited visibility levels, indicating that the model confidence

scores were broadly consistent with the audited visibility levels (Figure 2). Among the 1,000 images, 760 were classified as highly visible (mean model confidence: 0.993), 106 as moderately visible (0.888), 101 as having low visibility (0.825), and 33 as not visible. Thus, 967 of the 1,000 images predicted as visible were confirmed as visible, corresponding to a visible-image confirmation rate of 967/1,000 = 96.7%. We additionally audited 500 images predicted as `building=no`. 426 were confirmed as not visible, corresponding to a not-visible confirmation rate of 426/500 = 85.2%.

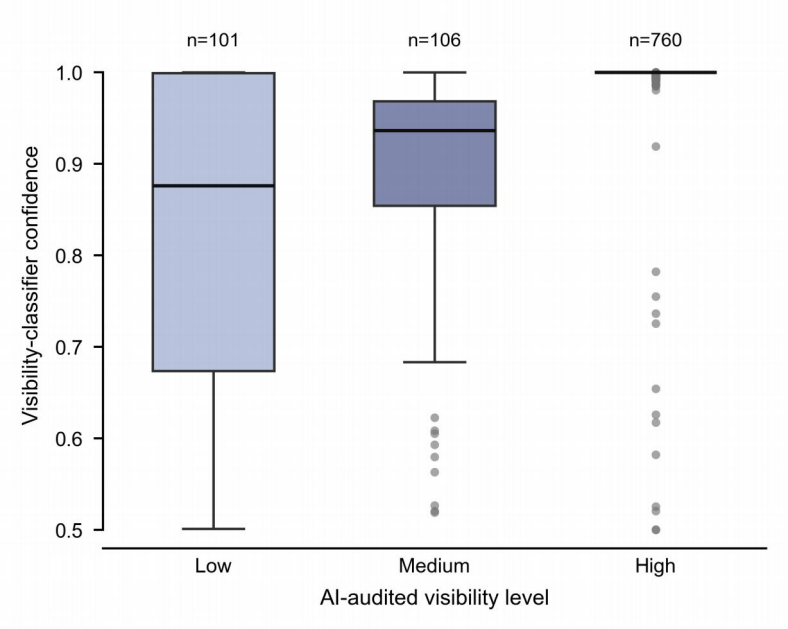

Figure 2: Visibility-classifier confidence across AI-audited visibility levels.

1.2 Building-obstruction analysis

The visibility classifier identifies images that contain a suitable building view, but the visible building may be an intervening building rather than the target building, particularly in dense urban areas. We therefore applied a geometric obstruction analysis to determine whether the target building was visible from the street-view sampling point. A sightline was first constructed from the sampling point to the centroid of the target footprint. The target building was classified as unobstructed if this central sightline did not intersect any surrounding building footprint. If the central sightline was blocked, additional sightlines were constructed from the sampling point to the boundary vertices of the target footprint. The building was classified as partially visible if at least one boundary sightline remained unobstructed, and as fully obstructed if all tested sightlines were blocked.

The AI-assisted audit evaluated 1,000 cases predicted as unobstructed, 400 predicted as fully obstructed, and 100 predicted as partially visible by comparing each street-view image with its footprint–sightline map. Among the predicted-unobstructed cases, 934 were confirmed as unobstructed, corresponding to a confirmation rate of 934/1,000 = 93.4%. Among the predicted-fully-obstructed cases, 374 were confirmed as fully obstructed, corresponding to a confirmation rate of 374/400 = 93.5%. Among the 100 cases predicted as partially visible,

78 were confirmed to retain at least a partial view of the target building, corresponding to a confirmation rate of $78/100 = 78.0\%$. These results show that the obstruction analysis reliably distinguishes unobstructed, fully obstructed, and partially visible target buildings.

2. Image–footprint correspondence

After confirming that the target building was visible, we examined whether the building shown in the image matched the target footprint. This correspondence was first established during data collection through a building-specific sampling strategy. For each target footprint, we identified its nearest location on the road network, retrieved the nearest available street-view point, and calculated the heading, pitch, and field of view from the spatial relationship between the sampling point and the target building. Unlike conventional road-based sampling, which captures images at fixed intervals and then searches nearby buildings, our method assigns a dedicated sampling location and camera view to each target footprint. This reduces ambiguity when multiple buildings appear in the same image and improves image-footprint correspondence. We repeatedly inspected the sampled images during data collection to confirm that the building in the image center was the target building.

As an additional check, the AI-assisted audit evaluated 1,000 sampled images by comparing the building and its surroundings in each street-view image with the target footprint, surrounding footprints, sampling point, and viewing direction. Representative street-view–spatial-view pairs used in this audit are shown in Figure 3. The paired information allowed the AI to determine image–footprint correspondence from both the visual content of the street-view image and the spatial configuration of the target building and its surroundings. The audit confirmed correct image–footprint correspondence in 999 of the 1,000 cases, corresponding to an accuracy of 99.9%. In the single unmatched case, the target building was hidden behind a corrugated metal fence (Figure 3(m)). The visibility classifier incorrectly retained this image as a valid building view, whereas the AI-assisted audit determined that the visible content did not correspond to the target footprint.

3. Material-label accuracy

After confirming the image-footprint correspondence, the material identified from the target facade could be assigned to the corresponding building footprint. The AI-assisted audit classified the dominant visible facade material in the 1,000 sampled images using the same nine categories as BMAT: brick, concrete, glass, metal, stone, stucco, tile, wood, and other. The audited labels were then compared with the BMAT predictions. Of the 1,000 images, 873 received the same material label, corresponding to a material-label accuracy of $873/1{,}000 = 87.3\%$.

As summarized in Table 1, the audit results remained generally consistent across street-view platforms, building heights, building densities, and sampling distances. Visibility confirmation exceeded 95% in all major strata, image–footprint correspondence ranged from 99.5% to 100.0%, and material-label accuracy ranged from 83.8% to 90.9%. The main variation occurred in the obstruction-related metrics. The confirmation rate for predicted-unobstructed cases decreased from 95.4% at near sampling distances to 88.9% at far distances, and from 95.2% in low-density

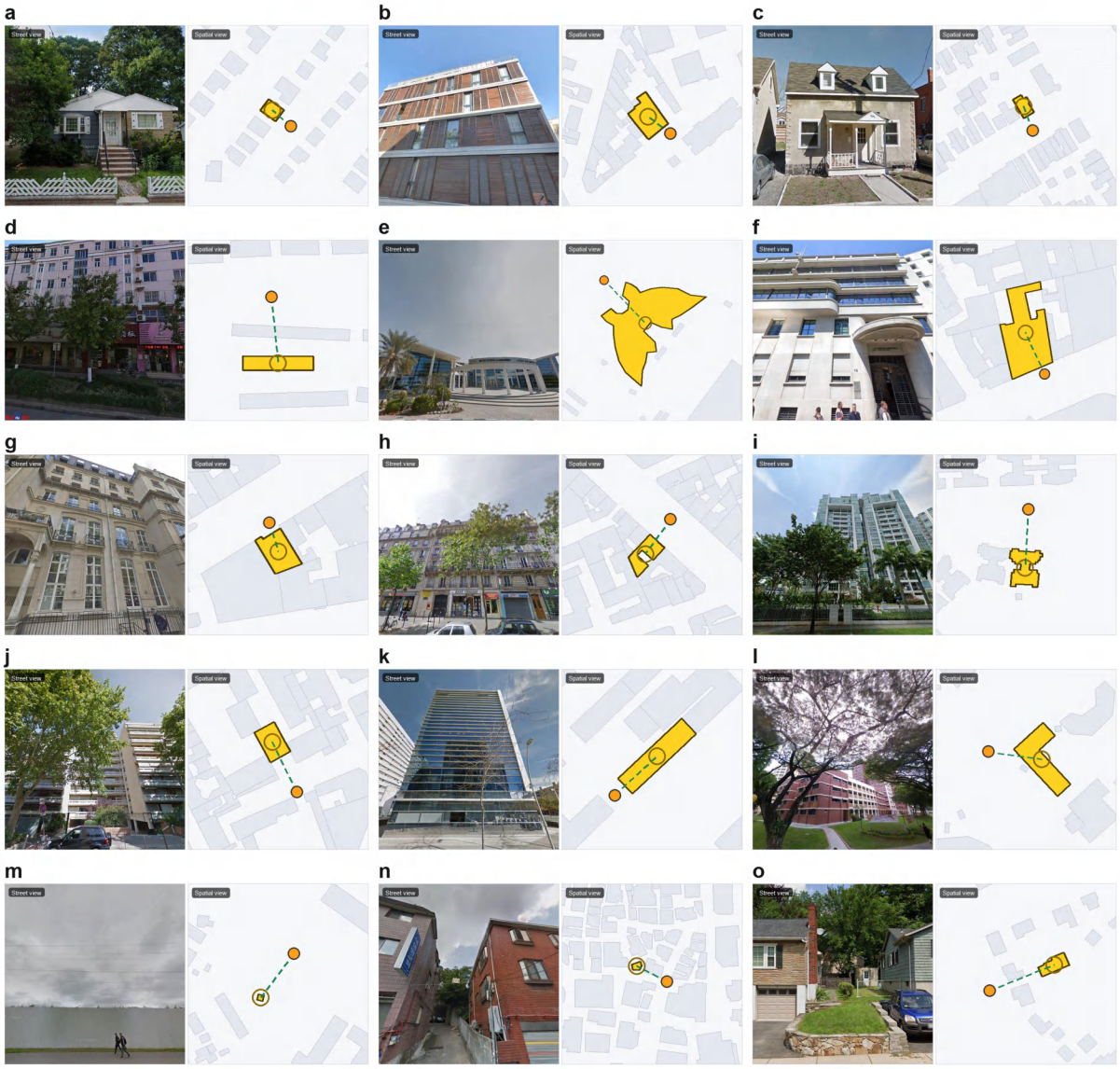

Figure 3: Representative street-view–spatial-view pairs used to audit image-footprint correspondence. Each spatial view shows the target footprint in yellow, the street-view sampling point in orange, and the viewing direction as a green dashed line.

areas to 91.3% in high-density areas. This pattern is consistent with the greater likelihood that longer sightlines or densely distributed buildings intersect other building footprints, making the target building more susceptible to partial or complete obstruction. Overall, the results indicate that the validation performance was robust across different sampling conditions.

Together, these analyses validate the main steps from street-view image acquisition to footprint-level material assignment. The AI-assisted audit confirmed 96.7% of the images retained by the visibility classifier and 85.2% of those excluded as not visible. The confirmation rates for unobstructed, fully obstructed, and partially visible cases were 93.4%, 93.5%, and 78.0%, respectively. Image–footprint correspondence was correct in 99.9% of the sampled cases, and the material-label accuracy was 87.3%.

Table 1: AI-assisted audit performance across platforms and sampling conditions.

Factor	Group	*N*	Rate, % (confirmed/total)						
			Visible	Not visible	No obstruction	Full obstruction	Partial visibility	Image–footprint match	Material label
Overall	**All**	**2,000**	**96.7 (967/1,000)**	**85.2 (426/500)**	**93.4 (934/1,000)**	**93.5 (374/400)**	**78.0 (78/100)**	**99.9 (999/1,000)**	**87.3 (873/1,000)**
Platform	BSV	525	96.0 (336/350)	89.1 (156/175)	100.0 (350/350)	–	–	100.0 (350/350)	86.9 (304/350)
	GSV	1,475	97.1 (631/650)	83.1 (270/325)	89.8 (584/650)	93.5 (374/400)	78.0 (78/100)	99.8 (649/650)	87.5 (569/650)
Density	Low	787	96.1 (440/458)	89.6 (163/182)	95.2 (436/458)	95.0 (114/120)	77.8 (21/27)	100.0 (458/458)	88.6 (406/458)
	Medium	627	97.3 (292/300)	84.6 (126/149)	92.3 (277/300)	92.7 (127/137)	90.2 (37/41)	100.0 (300/300)	86.3 (259/300)
	High	586	97.1 (235/242)	81.1 (137/169)	91.3 (221/242)	93.0 (133/143)	62.5 (20/32)	99.6 (241/242)	86.0 (208/242)
Height	Low	397	95.6 (259/271)	100.0 (6/6)	93.7 (254/271)	98.9 (91/92)	75.0 (21/28)	99.6 (270/271)	83.8 (227/271)
	Medium	413	95.7 (270/282)	100.0 (8/8)	92.6 (261/282)	95.5 (84/88)	74.3 (26/35)	100.0 (282/282)	85.5 (241/282)
	High	818	97.9 (429/438)	82.6 (194/235)	93.8 (411/438)	97.4 (113/116)	86.2 (25/29)	100.0 (438/438)	90.9 (398/438)
	Unknown	372	100.0 (9/9)	86.9 (218/251)	88.9 (8/9)	82.7 (86/104)	75.0 (6/8)	100.0 (9/9)	77.8 (7/9)
Distance	Near	671	97.1 (403/415)	85.4 (105/123)	95.4 (396/415)	90.5 (105/116)	76.5 (13/17)	100.0 (415/415)	87.0 (361/415)
	Medium	704	96.7 (357/369)	85.3 (145/170)	93.8 (346/369)	91.7 (121/132)	84.8 (28/33)	100.0 (369/369)	85.9 (317/369)
	Far	625	95.8 (207/216)	85.0 (176/207)	88.9 (192/216)	97.4 (148/152)	74.0 (37/50)	99.5 (215/216)	90.3 (195/216)

Note. Values are percentages, with confirmed/total shown in parentheses. Visible denotes retained images audited as having low, medium, or high target-building visibility. Not visible denotes excluded images audited as not visible. Partial visibility includes predicted partially visible cases audited as either partially obstructed or unobstructed. Image–footprint matching and material-label accuracy were evaluated using images retained for material inference. *N* is the total number of cases contributing to the reported audit tasks in each stratum. Dashes indicate that no corresponding cases were sampled.

Comment 3. *The necessity of using a vision language model is insufficiently justified. Facade material identification is formulated as a closed set classification task with nine predefined categories, and the model is required to return only one label from a fixed list. Under this setting, the language generation and instruction reasoning capabilities of Qwen2.5-VL appear to be largely unused, making the model function mainly as a computationally expensive image classifier. Although the fine-tuned model outperforms several conventional vision models, the comparison does not demonstrate that a vision language model is necessary because the performance gain may result from differences in model scale, pretraining, optimization, or data partitioning. This concern is especially important because the model is applied to approximately 96 million images and is substantially slower than traditional classifiers. The authors should justify the use of the Vision Language Model through fair comparisons with modern pretrained visual encoders under consistent experimental settings and report model size, inference speed, memory use, total computing cost, and cross city or cross platform generalization. If no clear advantage beyond classification accuracy is demonstrated, a smaller visual model may be more appropriate for this large scale task.*

Response: Thank you for this suggestion. We selected Qwen2.5-VL-7B-Instruct because its pretrained visual-semantic knowledge provides a strong prior for building facade material recognition. As shown in Table 2, the unfine-tuned Qwen2.5-VL-7B-Instruct achieved an accuracy of 0.73, indicating that it already has a certain ability to recognize facade materials without task-specific training. However, its general pretrained knowledge is not fully aligned with the material categories and labeling criteria used in BMAT. For example, some visually similar or locally specific facade materials may be assigned to categories differently from our annotation standard. We therefore fine-tuned the model using manually annotated street-view samples to align its predictions with the BMAT label definitions, which increased the accuracy to 0.91 and outperformed the fine-tuned traditional vision models, whose accuracies ranged from 0.79 to 0.84. Although Qwen2.5-VL-7B-Instruct requires longer inference time, with the full material inference requiring approximately 6,353 GPU-hours on NVIDIA RTX 4090 GPUs, we prioritized inference accuracy to produce a higher-quality dataset.

The primary goal of this study is not to identify the optimal accuracy-efficiency trade-off among material classification models, but to construct a building facade material dataset with high reliability. Therefore, computational cost was not the primary criterion in our model selection. Nevertheless, our model comparison results can serve as a reference for future studies on building material classification or facade attribute extraction. If subsequent studies need to balance inference efficiency and classification accuracy, they can choose an appropriate model based on our results.

We also added the computational cost of full-scale material inference to make the inference process more transparent:

"The full material inference over the filtered street-view images required approximately 6,353 GPU-hours on NVIDIA RTX 4090 GPUs."

Table 2: Comprehensive performance comparison of different models.

Model	**Fine-tuned**	**Inference time (ms)**	**Accuracy**	**Precision**	**Recall**	**F1 score**
Traditional vision backbones						
VGG16	✓	1.26	0.83	0.82	0.80	0.81
ViT-B/16	✓	4.40	0.82	0.81	0.79	0.80
ResNet-50	✓	5.57	0.82	0.80	0.80	0.80
ResNet-101	✓	10.52	0.82	0.81	0.80	0.80
ShuffleNetV2	✓	5.57	0.79	0.78	0.76	0.77
EfficientNet-B0	✓	6.81	0.84	0.82	0.82	0.82
DenseNet-161	✓	18.37	0.84	0.83	0.82	0.82
MobileNetV3-Large	✓	5.54	0.82	0.81	0.80	0.80
Vision-language models (VLMs)						
InternVL2.5-8B	×	185.55	0.69	0.62	0.53	0.54
MiniCPM-V-2.6	×	631.59	0.74	0.55	0.48	0.48
Qwen2-VL-7B-Instruct	×	276.06	0.70	0.51	0.52	0.51
mLLaMA3.2-11B	×	829.64	0.66	0.38	0.36	0.36
Qwen2.5-VL-7B-Instruct	×	221.48	0.73	0.76	0.65	0.66
Fine-tuned InternVL2.5-8B	✓	163.16	0.89	0.88	0.89	0.89
Fine-tuned MiniCPM-V-2.6	✓	636.26	0.90	0.89	0.90	0.89
Fine-tuned Qwen2-VL-7B-Instruct	✓	436.29	0.91	0.90	0.90	0.90
Fine-tuned mLLaMA3.2-11B	✓	545.18	0.91	0.90	0.90	0.90
Fine-tuned Qwen2.5-VL-7B-Instruct (Ours)	✓	237.73	**0.91**	**0.91**	**0.91**	**0.91**

Comment 4. *The current test split may contain building, spatial, or temporal information leakage. The manuscript states that 39,405 annotated images were divided into training and testing sets in a ratio of 7 to 3, but the unit used for data partitioning is unclear. A random image level split may place repeated views of the same building or visually similar images from nearby locations in both subsets, leading to overestimated performance. The authors should therefore use building independent and city independent splits, report separate results for Google Street View and Baidu Street View, and consider a temporal split to assess generalization across years.*

Response: Thank you for this comment. The annotated dataset contains a small number of multi-year images of the same buildings, which were retained to expose the model to variations in viewpoint, image conditions, and acquisition year. We acknowledge that the original random image-level split may therefore include limited building-level overlap and may slightly overestimate performance. Given that only a small number of annotated images are repeated observations of the same buildings, any resulting inflation in the overall performance estimate is expected to be limited. To provide an additional evaluation independent of this split, we conducted a blinded AI-assisted audit of 1,000 final BMAT records sampled across GSV and BSV, 18 cities, acquisition years from 2007 to 2025, and diverse urban conditions. GPT-5.6-Sol independently assigned the dominant visible street-facing facade material, without access to the original BMAT labels, and agreed with BMAT predictions in 87.3% of cases (Table 3).

The agreement rates were similar between GSV (87.5%) and BSV (86.9%) and remained stable across acquisition periods: 88.1% before 2015, 86.6% in 2015–2019, and 87.6% in 2020–2025. These results provide complementary evidence that the trained model maintains stable material-recognition performance across diverse platforms, periods, and urban observation conditions in the final BMAT dataset.

Table 3: Material-label accuracy across street-view platforms and acquisition periods.

	Accuracy, % (correct/total)		
Acquisition period	GSV	BSV	All platforms
Before 2015	89.2 (141/158)	85.5 (59/69)	88.1 (200/227)
2015–2019	85.8 (188/219)	87.5 (168/192)	86.6 (356/411)
2020–2025	87.9 (240/273)	86.5 (77/89)	87.6 (317/362)
Overall	**87.5 (569/650)**	**86.9 (304/350)**	**87.3 (873/1,000)**

Note. Correct cases are those for which the BMAT material label agreed with the independent AI-assisted annotation.

Comment 5. *The Overture Maps comparison is not a sufficient independent validation of the dataset. The comparison with 47,434 Overture Maps records should not be described as independent cross validation because the provenance, observation date, and accuracy of these material attributes are unclear. The semantic meaning of the Overture labels may also differ from the dominant visible facade material inferred from street view images. Moreover, BMAT uses Overture Maps building footprints, so the validation is not fully external at the spatial object level. This analysis should therefore be presented as an agreement assessment with an auxiliary reference source. The authors should clarify the origin and definition of the Overture attributes and report agreement by city and material class. They should also develop an independently annotated validation set drawn directly from BMAT and stratified by city, platform, material class, building density, visibility condition, and acquisition year.*

Response: Thank you for this suggestion. We used the `facade_material` attribute of the Overture Maps building records, which describes the outer surface material of a building facade and therefore corresponds closely to the dominant visible facade material recorded in BMAT. The Overture records contained material labels including brick, cement block, concrete, glass, metal, plaster, steel, stone, timber framing, and wood. Because the category systems differ, we matched semantically similar categories: "cement_block" was mapped to "concrete", "plaster" to "stucco", "steel" to "metal", and "timber_framing" to "wood". Brick, concrete, glass, metal, stone, and wood were retained unchanged, while rare categories without a reliable BMAT correspondence were excluded. After this harmonization, 50,823 Overture material records were included in the comparison. In the original manuscript, the reported total of 47,434 records was based on an incomplete city-level aggregation. We have corrected the aggregation to include all eligible cities, yielding 50,823 records; the city-specific sample sizes are reported in Table 4.

Comparing each Overture material record with the latest BMAT record for the same building yielded an agreement of 0.72. Because Overture does not provide the observation date of each facade-material value, we also considered a record matched if it agreed with any available annual BMAT record for the same building, increasing the agreement to 0.79. We regard 0.72 as the primary conservative estimate and 0.79 as an upper-bound sensitivity estimate, because buildings with more annual records have more opportunities to produce a match (Figure 4). We further summarized the agreement by city and material class (Table 4). The agreement was highest for brick (0.88), followed by glass (0.79), stucco (0.67), stone (0.61), metal (0.49), wood (0.44), and concrete (0.20). The relatively low agreement for concrete mainly reflects the different semantic boundary between concrete and stucco in the two sources. In BMAT, facades with a plastered or cement-rendered surface are classified as stucco, whereas such buildings may still be recorded as concrete in Overture Maps. Therefore, many Overture concrete records correspond to BMAT stucco labels (Figure 4(a)). The city-level results also show high agreement in cities with sufficient Overture material records (Table 4). In addition, some Overture material attributes may be outdated or incorrect. The reported agreement should therefore be interpreted as a conservative consistency estimate with an auxiliary reference source, rather than as the

absolute accuracy of BMAT.

For traceability, we retained the Overture feature ID and its `sources` metadata, including the contributing dataset, upstream source-record identifier, and available update time. The update time refers to the source record and does not necessarily represent the observation date of the facade material.

Figure 4: Comparison of model predictions with Overture Maps records for (a) latest and (b) historical material attributes, including (c) visual examples with green for correct and red for error.

Table 4: Agreement between Overture Maps and the latest BMAT material labels.

City	*N*	Overall	Brick	Concrete	Glass	Metal	Stone	Stucco	Wood
Overall	**50,823**	**0.72 (36542/50823)**	**0.88 (26381/30031)**	**0.20 (1379/6929)**	**0.79 (1271/1612)**	**0.49 (1034/2100)**	**0.61 (700/1147)**	**0.67 (5316/7965)**	**0.44 (461/1039)**
London	16,510	0.90 (14924/16510)	0.93 (13462/14404)	0.23 (30/131)	0.80 (76/95)	0.41 (17/41)	0.78 (335/431)	0.73 (990/1365)	0.33 (14/43)
Moscow	12,675	0.61 (7775/12675)	0.75 (3423/4557)	0.33 (537/1641)	0.63 (112/179)	0.51 (892/1753)	0.24 (10/41)	0.62 (2679/4329)	0.70 (122/175)
Toronto	7,165	0.83 (5958/7165)	0.88 (5316/6071)	0.27 (41/150)	0.81 (127/156)	0.48 (13/27)	0.75 (203/269)	0.54 (242/451)	0.39 (16/41)
Tokyo	2,379	0.08 (186/2379)	0.09 (4/45)	0.05 (107/2215)	0.69 (43/62)	0.62 (23/37)	0.22 (2/9)	0.67 (2/3)	0.62 (5/8)
Amsterdam	1,979	0.66 (1302/1979)	0.98 (1028/1050)	0.34 (169/497)	0.55 (6/11)	0.07 (1/15)	–	–	0.24 (98/406)
Berlin	1,894	0.80 (1517/1894)	0.87 (274/314)	0.84 (256/305)	0.59 (27/46)	0.56 (10/18)	0.68 (15/22)	0.79 (929/1172)	0.35 (6/17)
New York	1,296	0.84 (1087/1296)	0.86 (937/1090)	0.32 (6/19)	0.90 (113/126)	0.56 (9/16)	0.83 (15/18)	0.24 (6/25)	0.50 (1/2)
Sydney	1,147	0.67 (774/1147)	0.75 (550/730)	0.18 (12/68)	0.58 (14/24)	0.22 (2/9)	0.80 (12/15)	0.66 (170/257)	0.32 (14/44)
Singapore	820	0.11 (91/820)	0.04 (4/94)	0.02 (13/582)	0.67 (33/49)	0.44 (24/55)	0.20 (5/25)	0.92 (12/13)	0.00 (0/2)
Istanbul	760	0.15 (115/760)	0.00 (0/2)	0.13 (90/711)	0.85 (11/13)	0.67 (2/3)	0.44 (4/9)	0.44 (4/9)	0.31 (4/13)
Boston	577	0.86 (495/577)	0.93 (444/479)	0.53 (17/32)	0.53 (10/19)	0.50 (3/6)	0.51 (19/37)	0.00 (0/1)	0.67 (2/3)
Bogota	504	0.84 (422/504)	0.92 (334/362)	0.25 (6/24)	0.81 (72/89)	0.24 (4/17)	1.00 (1/1)	0.57 (4/7)	0.25 (1/4)
Paris	346	0.70 (243/346)	0.88 (168/192)	0.42 (8/19)	0.50 (31/62)	0.50 (5/10)	0.48 (26/54)	–	0.56 (5/9)
Shenzhen	314	0.81 (253/314)	0.00 (0/1)	0.04 (1/25)	0.93 (249/267)	0.17 (3/18)	0.00 (0/1)	0.00 (0/2)	–
Bangalore	307	0.75 (229/307)	0.50 (9/18)	0.08 (2/26)	0.76 (26/34)	0.00 (0/3)	0.30 (7/23)	0.93 (185/200)	0.00 (0/3)
Ottawa	296	0.86 (254/296)	0.89 (173/194)	0.38 (5/13)	0.86 (6/7)	0.80 (4/5)	0.91 (21/23)	0.29 (2/7)	0.91 (43/47)
San Francisco	227	0.63 (142/227)	0.75 (12/16)	0.25 (1/4)	0.75 (6/8)	0.00 (0/1)	–	0.50 (1/2)	0.62 (122/196)
Mumbai	211	0.16 (34/211)	–	0.15 (28/193)	–	0.17 (2/12)	–	0.80 (4/5)	0.00 (0/1)
Buenos Aires	188	0.74 (140/188)	0.84 (59/70)	0.33 (7/21)	0.84 (73/87)	0.10 (1/10)	–	–	–
Rome	152	0.27 (41/152)	0.56 (5/9)	0.11 (2/19)	0.67 (4/6)	0.61 (11/18)	0.13 (11/87)	0.67 (8/12)	0.00 (0/1)
Chicago	150	0.70 (105/150)	0.86 (85/99)	0.11 (2/19)	0.73 (11/15)	0.33 (1/3)	0.20 (1/5)	0.00 (0/1)	0.62 (5/8)
Santiago	124	0.57 (71/124)	0.46 (33/71)	0.33 (6/18)	0.94 (30/32)	–	1.00 (1/1)	1.00 (1/1)	0.00 (0/1)
Hong Kong	95	0.77 (73/95)	1.00 (1/1)	0.19 (4/21)	0.94 (67/71)	1.00 (1/1)	0.00 (0/1)	–	–
Delhi	93	0.25 (23/93)	0.24 (12/50)	0.11 (3/27)	–	0.33 (2/6)	1.00 (4/4)	0.40 (2/5)	0.00 (0/1)
Guangzhou	80	0.23 (18/80)	0.25 (1/4)	0.10 (6/59)	0.85 (11/13)	0.00 (0/4)	–	–	–
Shanghai	69	0.84 (58/69)	0.00 (0/1)	0.00 (0/1)	0.75 (9/12)	0.00 (0/1)	1.00 (1/1)	0.91 (48/53)	–
Beijing	68	0.46 (31/68)	0.70 (19/27)	0.00 (0/3)	–	0.00 (0/1)	0.20 (3/15)	0.50 (9/18)	0.00 (0/4)
Cape Town	58	0.09 (5/58)	0.30 (3/10)	0.00 (0/8)	0.00 (0/2)	–	0.00 (0/35)	1.00 (2/2)	0.00 (0/1)
Seoul	55	0.56 (31/55)	0.67 (8/12)	0.18 (3/17)	0.86 (19/22)	–	0.00 (0/3)	–	1.00 (1/1)
Mexico City	55	0.55 (30/55)	0.25 (1/4)	0.48 (10/21)	0.79 (15/19)	0.40 (2/5)	0.40 (2/5)	–	0.00 (0/1)
Jakarta	53	0.38 (20/53)	0.00 (0/28)	0.00 (0/4)	0.95 (20/21)	–	–	–	–
Los Angeles	50	0.58 (29/50)	0.62 (5/8)	0.20 (1/5)	0.64 (9/14)	0.50 (1/2)	0.00 (0/1)	0.72 (13/18)	0.00 (0/2)
Barcelona	49	0.59 (29/49)	0.69 (11/16)	0.00 (0/3)	0.93 (13/14)	1.00 (1/1)	0.14 (1/7)	0.50 (3/6)	0.00 (0/2)
Ningbo	24	0.58 (14/24)	–	0.10 (1/10)	0.93 (13/14)	–	–	–	–
Nairobi	10	0.40 (4/10)	–	0.00 (0/3)	1.00 (2/2)	–	0.00 (0/3)	–	1.00 (2/2)
Taipei	9	0.33 (3/9)	–	–	0.38 (3/8)	–	–	–	0.00 (0/1)
Dubai	7	0.71 (5/7)	–	0.00 (0/1)	1.00 (5/5)	0.00 (0/1)	–	–	–
Dakar	6	0.50 (3/6)	–	0.50 (3/6)	–	–	–	–	–
Chengdu	3	1.00 (3/3)	–	–	1.00 (3/3)	–	–	–	–
Hohhot	3	0.00 (0/3)	0.00 (0/1)	–	0.00 (0/2)	–	–	–	–
Lanzhou	3	0.67 (2/3)	–	–	0.67 (2/3)	–	–	–	–
Lagos	2	0.00 (0/2)	–	0.00 (0/2)	–	–	–	–	–
Rio de Janeiro	2	0.50 (1/2)	–	0.50 (1/2)	–	–	–	–	–
Bangkok	2	0.00 (0/2)	0.00 (0/1)	0.00 (0/1)	–	–	–	–	–
Nanjing	2	0.50 (1/2)	–	1.00 (1/1)	–	0.00 (0/1)	–	–	–
Harbin	1	1.00 (1/1)	–	–	–	–	1.00 (1/1)	–	–
Taiyuan	1	0.00 (0/1)	–	0.00 (0/1)	–	–	–	–	–
Tianjin	1	0.00 (0/1)	–	0.00 (0/1)	–	–	–	–	–
Changsha	1	0.00 (0/1)	–	–	–	–	–	0.00 (0/1)	–

Note. Entries report agreement as ratio (matched/comparable). Dashes indicate no comparable records. Tile is omitted because no comparable Overture reference records were available.

To address the request for an independently annotated validation set directly sampled from BMAT, we constructed an AI-assisted audit set of 1,000 images from the final BMAT records. The samples were selected to cover different cities, street-view platforms, material classes, building densities, building heights, sampling distances, visibility conditions, and acquisition years. GPT-5.6-Sol with the high-reasoning setting was used to independently annotate the dominant visible facade material in each image using the same nine material categories as BMAT. The audited labels were then compared with the corresponding BMAT records. Overall, 873 of the 1,000 labels were confirmed, corresponding to a material-label accuracy of 87.3%. We report the audit results by city in Table 5, by street-view platform and acquisition period in Table 3, and by sampling condition, including building density, building height, sampling distance, and visibility-related strata, in Table 1. These stratified results show that the validation accuracy remained broadly consistent across the requested subsets.

We have also released the audit results for 2,000 BMAT images sampled across cities, platforms, material classes, building-density levels, acquisition years and visibility conditions at https://github.com/yhyJoy/BMAT/blob/main/data/label/image_label.csv

Table 5: City-level material-label accuracy in the AI-assisted audit.

Country	City	*N*	Accuracy, % (correct/total)
Overall	**All**	**1,000**	**87.3 (873/1,000)**
China	Beijing	70	87.1 (61/70)
China	Chengdu	70	84.3 (59/70)
China	Guangzhou	70	88.6 (62/70)
China	Shanghai	70	85.7 (60/70)
China	Shenzhen	70	88.6 (62/70)
Netherlands	Amsterdam	50	84.0 (42/50)
Spain	Barcelona	50	94.0 (47/50)
United States	Boston	50	88.0 (44/50)
Senegal	Dakar	50	94.0 (47/50)
United Arab Emirates	Dubai	50	86.0 (43/50)
Russia	Moscow	50	88.0 (44/50)
India	Mumbai	50	78.0 (39/50)
Kenya	Nairobi	50	82.0 (41/50)
Canada	Ottawa	50	82.0 (41/50)
France	Paris	50	88.0 (44/50)
Italy	Rome	50	90.0 (45/50)
South Korea	Seoul	50	94.0 (47/50)
Singapore	Singapore	50	90.0 (45/50)

Note. Accuracy was calculated by comparing the BMAT material label with the GPT-5.6-Sol high-reasoning audit label for each sampled image.

Comment 6. *The historical matching strategy systematically inflates the reported agreement. The manuscript reports an accuracy of 0.72 when the latest BMAT prediction is compared with the Overture Maps record. The value increases to 0.80 when a prediction is considered correct if any available historical BMAT label matches the undated Overture record. This approach introduces an upward bias because the probability of at least one match increases with the number of available observations. It also creates unequal validation conditions across cities because Google Street View and Baidu Street View provide different temporal coverage. The 0.80 value should therefore not be interpreted as a conventional accuracy estimate. The authors should retain the latest observation comparison as the more transparent result and, where possible, compare records from similar dates. If reference dates are unavailable, agreement should be reported according to the number of historical observations, with the inflation caused by chance matching explicitly quantified and discussed.*

Response:

Thank you for this suggestion. The latest-record comparison yielded an agreement of 0.72 $(36,542/50,823)$, whereas allowing any historical BMAT record to match the undated Overture material label increased the agreement to 0.79 $(40,296/50,823)$. Because buildings with more historical observations have more opportunities to produce a match, the historical comparison may overestimate agreement.

As the acquisition time of the Overture material records is unavailable, we further examined the cases responsible for the 0.07 increase, namely buildings whose latest BMAT material did not match Overture but whose historical BMAT record did. We summarized the number of annual BMAT observations for these buildings (Figure 5) and the corresponding latest-historical material patterns (Figure 6). These historical-match gain cases involved 3,754 buildings. Most gain cases had repeated observations, and 25.1% had at least 10 annual records. The material patterns were concentrated mainly among brick, concrete, and stucco, indicating that part of the difference may be related to temporal facade changes or semantic ambiguity among common exterior wall materials.

These discrepancies may arise from errors in the Overture records, errors in BMAT inference, or genuine facade changes over time. For example, Overture may record an earlier facade material, while the latest BMAT observation captures a subsequent renovation. We therefore regard 0.72 as the conservative agreement estimate and 0.79 as an upper-bound estimate when historical records are considered.

We also revised the validation description in the manuscript to clarify that the two Overture comparisons should not be interpreted as a single conventional accuracy estimate. The latest-record comparison is now described as a conservative agreement estimate, whereas the historical-match comparison is described as an upper-bound sensitivity estimate:

"Because Overture Maps material records do not contain acquisition dates, two comparison strategies were used. First, matching the latest BMAT material record with the Overture Maps record yielded an agreement of 0.72. This comparison is transparent and conservative, but it may underestimate agreement if the undated Overture record corresponds to an earlier facade state rather than the latest BMAT observation. Second, we used a historical matching strategy in which a building was considered matched if any available annual BMAT material record agreed with the Overture Maps record, yielding an agreement of 0.79. This value should be interpreted as an upper-bound sensitivity estimate rather than a conventional accuracy estimate, because buildings with more annual observations have more opportunities to produce a match by chance."

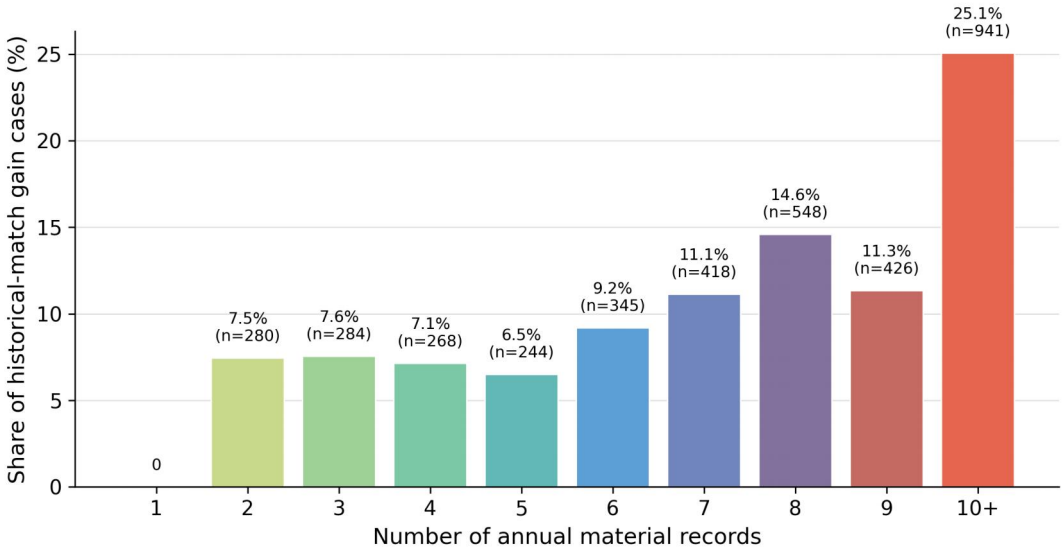

Figure 5: Number of annual material records in historical-match gain cases. Gain cases refer to buildings unmatched by the latest BMAT record but matched by at least one historical BMAT record.

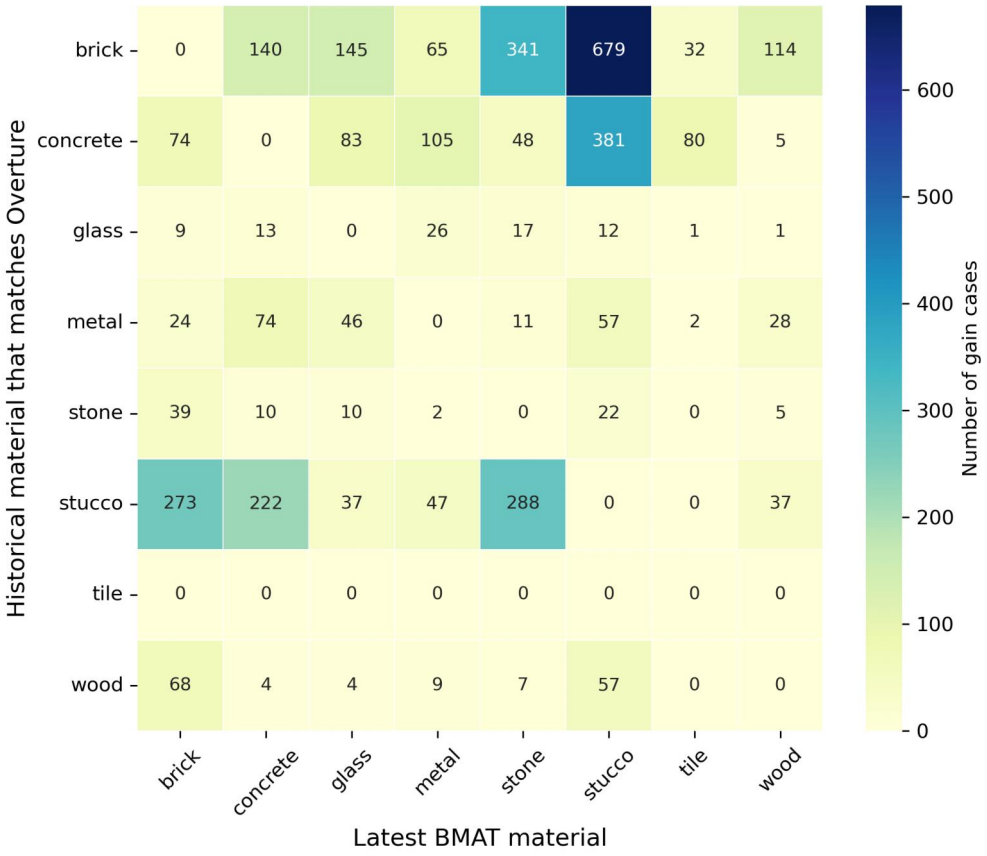

Figure 6: Material correspondence in historical-match gain cases. Rows indicate the historical BMAT material that matches Overture, and columns indicate the latest BMAT material.

Comment 7. *The manuscript validates the classifier but does not adequately validate the complete dataset production pipeline. The final BMAT records are generated through multiple stages, including street view sampling, camera orientation, obstruction detection, image quality filtering, material classification, and footprint association, so errors may accumulate throughout the pipeline. Although the MobileNetV2 filter achieves 88.75 percent validation accuracy, class specific metrics and external validation are not reported, and even a modest false positive rate could introduce many unsuitable images at this scale. The manuscript also does not explain how approximately 96 million valid images are aggregated into records for 22.09 million buildings, including the number of images used per building and year, the treatment of repeated panoramas and conflicting predictions, and the selection of the latest material label. The authors should therefore provide end to end validation of the final dataset and fully document the aggregation procedure, including the handling of low confidence predictions, inconsistent views, and ties.*

Response: Thank you for this helpful suggestion. We have added a clearer description of the complete data-generation workflow, which consists of two main stages: (1) building-specific image acquisition and candidate-image filtering and (2) material inference and assignment of image-level predictions to building footprints. We also added city-level processing statistics to make each stage more transparent.

Candidate images were acquired using building-specific viewing parameters. For each target building, we used its footprint, height, and street-view sampling point to determine the heading, pitch, and field of view for image acquisition. Because not all candidate images provided a suitable view for material recognition, they were filtered in two steps. First, a trained MobileNetV2 visibility model removed images without a sufficiently visible building facade or with unsuitable image quality, including severe overexposure, vegetation occlusion, blurring, and other visual degradation. Second, a geometric line-of-sight analysis identified whether the target building was unobstructed, partially visible, or fully obstructed by intervening buildings. Fully obstructed cases were excluded, while unobstructed and partially visible cases that passed the image-level visibility filter were retained for material inference.

The city-level numbers of candidate, not-visible, fully obstructed, and valid images are reported in Table 6. Overall, approximately 147 million candidate building-view images were processed, of which approximately 96 million were retained for material inference.

Image-level material labels were then assigned back to building footprints. Qwen2.5-VL-7B-Instruct predicted one material label from the nine predefined categories for each valid image. Because each image was acquired specifically for a target building from its nearest sampling point, the inferred label could be directly assigned to the corresponding building and acquisition year. For buildings with valid images from multiple years, the prediction from each year was retained as a separate annual material record. Predictions from different years were not aggregated because they represent time-specific observations. BMAT records these annual material labels as a temporal sequence in the `materials` field, for example, `2007:brick;`

`2011:brick; 2013:brick`. The `material` field records the material label from the latest available valid observation.

We also added building-level statistics to describe the number of annual material records available for each building. Specifically, we calculated the proportions of buildings with 0, 1, 2, 3, 4, and 5 or more annual material records in each city (Figure 7). The results show that the number of annual records varies substantially across cities. Many GSV cities contain a large share of buildings with multiple records, whereas BSV cities and cities with less complete street-view coverage are more frequently dominated by single-year records. We also calculated the distribution of the latest material-record year for each city (Figure 8), showing which observation years provide the latest material labels. For most buildings, the latest available material records were acquired after 2020.

We also revised the manuscript to document the dataset generation and aggregation process more clearly. The key revisions are listed below:

- **Street-view image filtering statistics.** We added a reference to the new supplementary table that summarizes the image filtering process:

  > *"After filtering out images in which the target buildings are not visible or are obstructed by intervening buildings, approximately 96 million (65.3%) images remained as valid samples for facade material inference. The city-level numbers of total, not-visible, obstructed, and valid images are reported in Supplementary Table S6."*

- **Aggregation from image-level predictions to building records.** We clarified how valid image-level material predictions were assigned to building footprints and organized into annual records:

  > *"Each valid image was linked to its target building footprint and acquisition year. For buildings captured at multiple timestamps, the year-specific material labels were retained as a temporal sequence in the "materials" field, while the latest available label was recorded in the "material" field."*

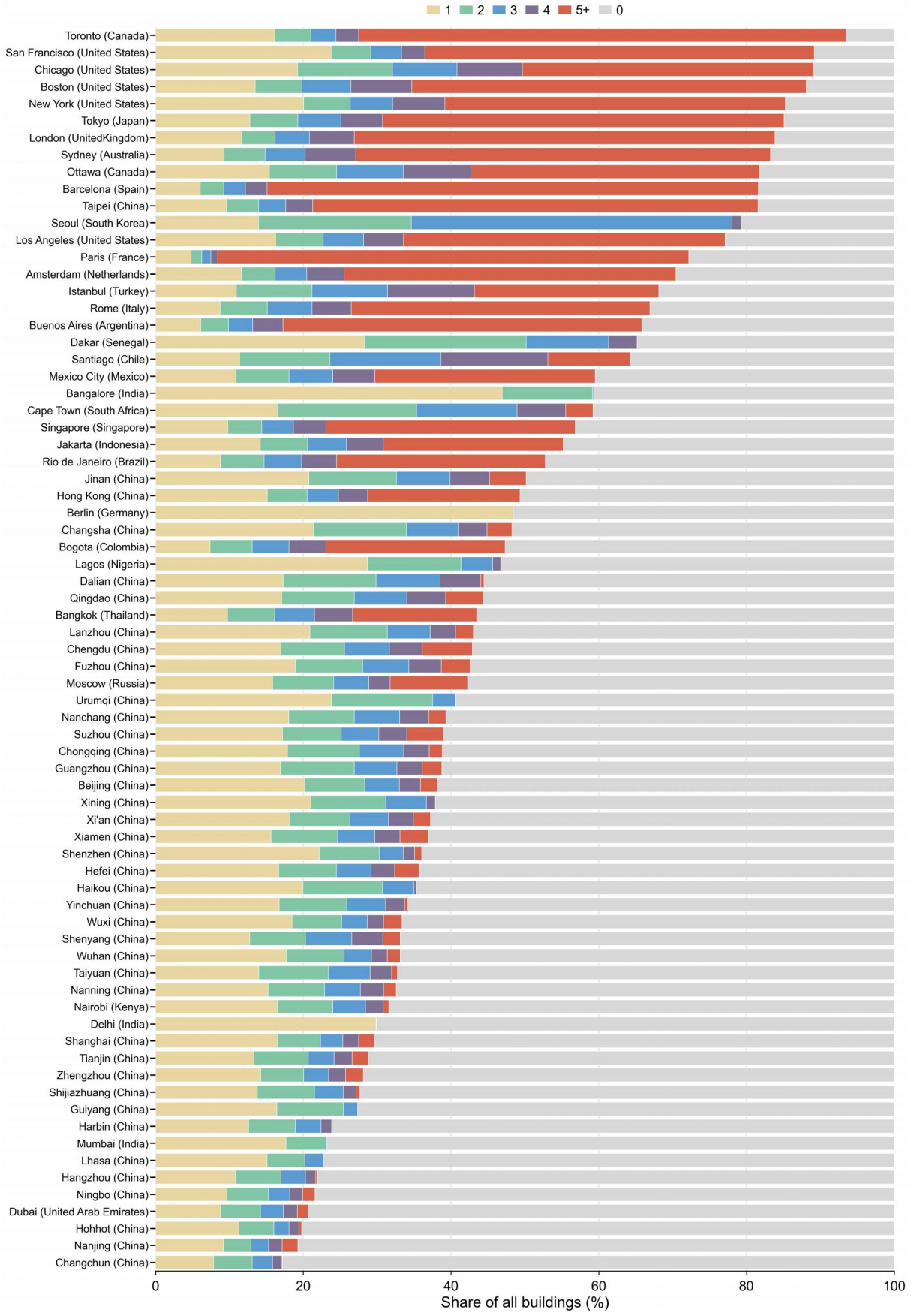

Figure 7: Number of annual material records per building by city.

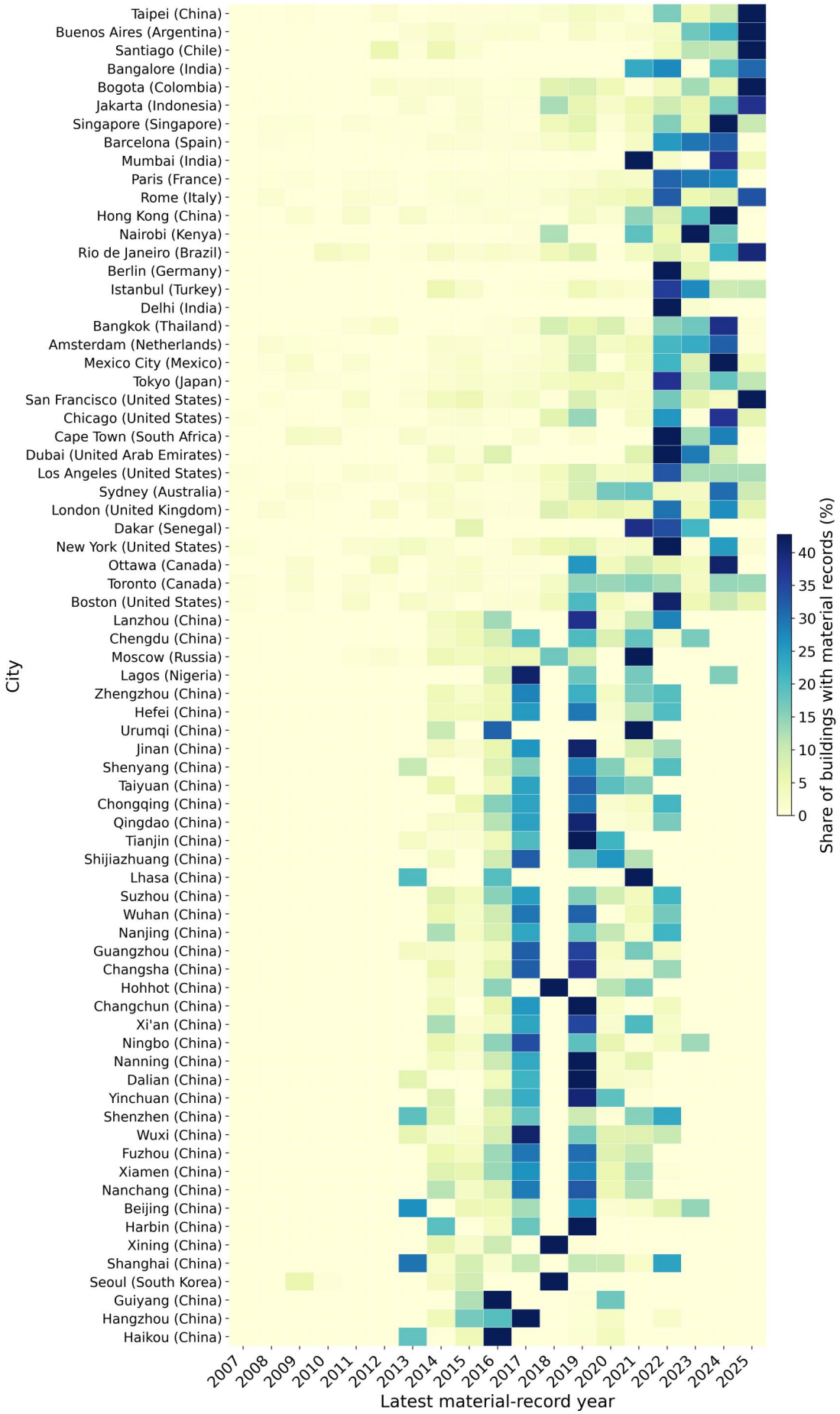

Figure 8: City-level distribution of the latest material-record year among buildings with material information.

Table 6: Statistics of street-view image filtering across different cities in the dataset.

Country	City	Total	Not vis.	Obstr.	Valid	Country	City	Total	Not vis.	Obstr.	Valid
Argentina	Buenos Aires	3,691,075	241,544	1,353,511	2,169,425	China	Beijing	1,308,198	491,254	320,469	606,231
Australia	Sydney	7,988,218	1,388,955	557,780	6,144,243	China	Changchun	225,686	68,079	27,596	138,558
Brazil	Rio de Janeiro	7,932,407	1,346,897	2,479,444	4,473,121	China	Changsha	277,121	96,306	34,821	157,195
Canada	Ottawa	1,665,343	214,816	100,362	1,369,580	China	Chengdu	564,062	266,229	103,548	243,300
Canada	Toronto	3,032,183	287,873	63,800	2,686,640	China	Chongqing	194,703	100,829	37,482	75,427
Chile	Santiago	3,589,647	862,442	856,301	2,056,960	China	Dalian	257,435	64,646	26,420	174,254
China	Hong Kong	730,858	219,522	116,559	405,753	China	Fuzhou	196,855	78,325	44,951	90,529
China	Taipei	790,411	77,164	68,214	651,649	China	Guangzhou	1,108,667	319,159	378,249	498,665
Colombia	Bogota	11,939,130	537,527	6,146,494	5,278,437	China	Guiyang	83,574	27,174	24,666	38,911
France	Paris	1,063,186	38,133	227,981	803,720	China	Harbin	166,606	46,889	23,772	102,379
Germany	Berlin	422,810	68,319	87,030	279,883	China	Haikou	98,083	29,879	29,465	45,941
India	Bangalore	1,282,825	412,898	155,589	769,212	China	Hangzhou	502,079	171,420	89,783	268,052
India	Delhi	779,055	271,699	207,651	364,075	China	Hefei	299,201	113,059	42,090	146,432
India	Mumbai	304,717	111,246	116,888	117,814	China	Hohhot	133,325	42,624	21,560	75,599
Indonesia	Jakarta	10,338,013	2,133,039	3,439,052	5,424,938	China	Jinan	197,384	57,423	33,870	115,492
Italy	Rome	1,498,340	303,008	155,798	1,073,982	China	Lanzhou	55,897	17,921	10,099	31,103
Japan	Tokyo	17,167,258	1,099,183	2,549,686	13,570,895	China	Lhasa	68,678	21,235	21,717	31,844
Kenya	Nairobi	1,122,070	407,887	381,690	450,027	China	Nanchang	278,395	87,963	57,590	148,349
Mexico	Mexico City	9,148,932	2,200,996	2,531,610	4,959,146	China	Nanjing	596,523	250,503	128,845	265,089
Netherlands	Amsterdam	1,295,660	93,175	286,657	930,393	China	Nanning	261,521	96,306	62,774	123,528
Nigeria	Lagos	1,931,150	329,575	548,899	1,138,120	China	Ningbo	461,510	167,651	90,346	231,774
Russia	Moscow	698,796	282,458	114,763	341,973	China	Qingdao	337,520	118,710	60,385	178,388
Senegal	Dakar	271,230	54,855	53,068	174,333	China	Shanghai	1,983,485	756,890	482,351	902,507
Singapore	Singapore	1,093,070	220,712	141,958	758,626	China	Shenyang	284,973	82,926	35,993	177,268
South Africa	Cape Town	3,453,124	676,859	811,437	2,131,480	China	Shenzhen	744,952	252,262	225,363	328,356
South Korea	Seoul	853,709	47,774	134,929	676,952	China	Shijiazhuang	227,703	77,200	52,490	114,494
Spain	Barcelona	722,801	61,747	76,949	592,409	China	Suzhou	212,466	88,051	44,904	96,241
Thailand	Bangkok	5,819,874	1,534,129	1,226,771	3,361,497	China	Taiyuan	191,173	55,212	30,091	114,262
Turkey	Istanbul	3,679,331	517,384	259,261	2,954,732	China	Tianjin	450,318	120,770	83,770	265,967
UAE	Dubai	267,049	50,622	44,001	179,782	China	Urumqi	35,860	12,803	4,748	19,996
US	Boston	647,335	76,195	40,789	533,801	China	Wuhan	530,456	207,249	100,037	256,893
US	Chicago	4,062,497	276,462	390,594	3,419,038	China	Wuxi	366,617	154,035	66,910	173,012
US	Los Angeles	7,105,527	1,095,311	1,251,386	4,941,347	China	Xiamen	254,584	84,420	62,332	124,815
US	New York	5,789,156	422,844	633,284	4,778,535	China	Xi'an	392,046	117,801	107,027	193,547
US	San Francisco	1,285,824	69,814	66,464	1,155,622	China	Xining	47,974	15,105	7,528	27,404
UK	London	9,964,843	600,769	1,159,575	8,267,664	China	Yinchuan	66,201	22,782	8,342	37,817
						China	Zhengzhou	450,554	199,253	102,443	193,351

Note: Not vis. denotes images classified as not containing visible buildings. Obstr. denotes images whose target building is obstructed. Valid denotes images with visible and unobstructed buildings. Region abbreviations: UAE = United Arab Emirates; UK = United Kingdom; US = United States.

Comment 8. *The temporal dimension does not yet provide reliable evidence of facade material transitions. The temporal dimension is presented as a major contribution, but differences between annual labels cannot be directly interpreted as physical facade changes. Historical street view images are linked to contemporary building footprints, although buildings may have been constructed, demolished, expanded, merged, subdivided, or replaced between 2007 and 2025. Label changes may also result from different viewpoints, mixed facades, illumination, image quality, vegetation, occlusion, or incorrect building alignment. The manuscript is therefore conceptually inconsistent in treating label changes as evidence of renovation while also acknowledging that such changes may reflect mixed materials or varying visible surfaces. The authors should establish a stable building cohort, exclude or flag buildings with uncertain historical identities, require repeated or persistent evidence for material transitions, and manually validate a representative sample to distinguish genuine renovation from building replacement, viewpoint effects, mixed materials, association errors, and classification errors.*

Response: We thank the reviewer for this important and constructive comment. We agree that differences between annual material labels should not be interpreted unconditionally as confirmed physical facade changes for every individual building. Such differences may reflect real material replacement, but may also arise from mixed-material facades or uncertainty in model inference. We therefore added analyses to characterize the temporal records more explicitly and to distinguish stable material sequences from more ambiguous fluctuations.

Regarding image quality and visibility, we filtered the street-view images before material inference. A visibility classifier first removes images in which the target building is not visible or the image is unsuitable for material recognition, for example because of severe overexposure, vegetation occlusion, or other visual degradation. A geometric line-of-sight obstruction model then excludes cases where the target building is blocked by intervening buildings. Thus, the subsequent material inference is performed on visible, relatively high-quality street-view images of the target buildings.

After this filtering, buildings retain different numbers of valid annual observations. Some buildings have no material record, some have only a single-year record, and others have repeated observations across multiple years. We therefore added city-level summaries of the number of annual material records per building (Figure 7) and the temporal span between the first and last available material records (Figure 9). These results show that BMAT contains multi-year material records for many buildings, although the density and temporal span of the records vary across cities.

For buildings with material records, we further classified their material sequences into four groups: single-year records, stable repeated records, stable changes, and complex fluctuations (Figure 10). Single-year records contain only one year of material information. Stable repeated records have the same material label across multiple years. Stable changes are defined as persistent shifts from one dominant material to another, for example sequences in which one

material dominates the earlier years and a different material consistently appears in later years. Complex fluctuations contain frequent or irregular label changes and are treated cautiously, because they may reflect mixed-material facades, changing visible facade portions, or model uncertainty. This classification helps distinguish stable and interpretable temporal patterns from more ambiguous sequences, rather than treating all label changes as confirmed material transitions.

We also added representative visual examples for these sequence types (Figure 11). The examples show that stable repeated records generally correspond to buildings with consistent facade materials across years, while stable changes often capture visually apparent changes in the dominant facade material. Complex fluctuations account for a smaller subset of cases and often occur in buildings with mixed facade materials or visually ambiguous surfaces; these fluctuations may therefore arise from either genuine mixed-material appearances or model uncertainty. Overall, the additional analyses show that most multi-year records are stable or show persistent candidate changes, while fluctuating sequences are explicitly separated and interpreted with caution.

Finally, we added these supplementary analyses of the temporal dimension of BMAT to the "Image counts and temporal record statistics of BMAT" section in the Supplementary Materials.

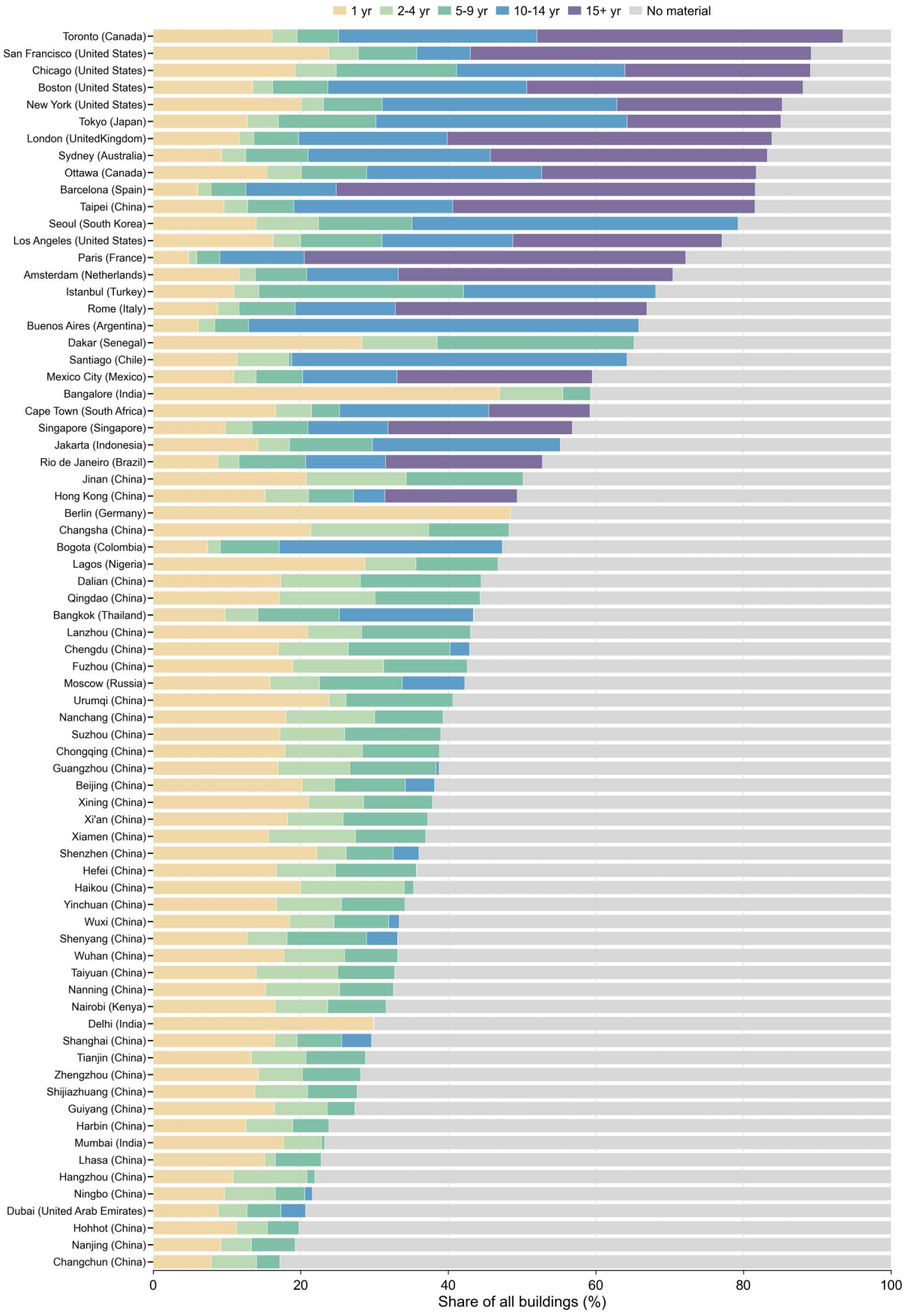

Figure 9: Temporal span of building material records by city. The span is defined as the number of years between the first and last observed years.

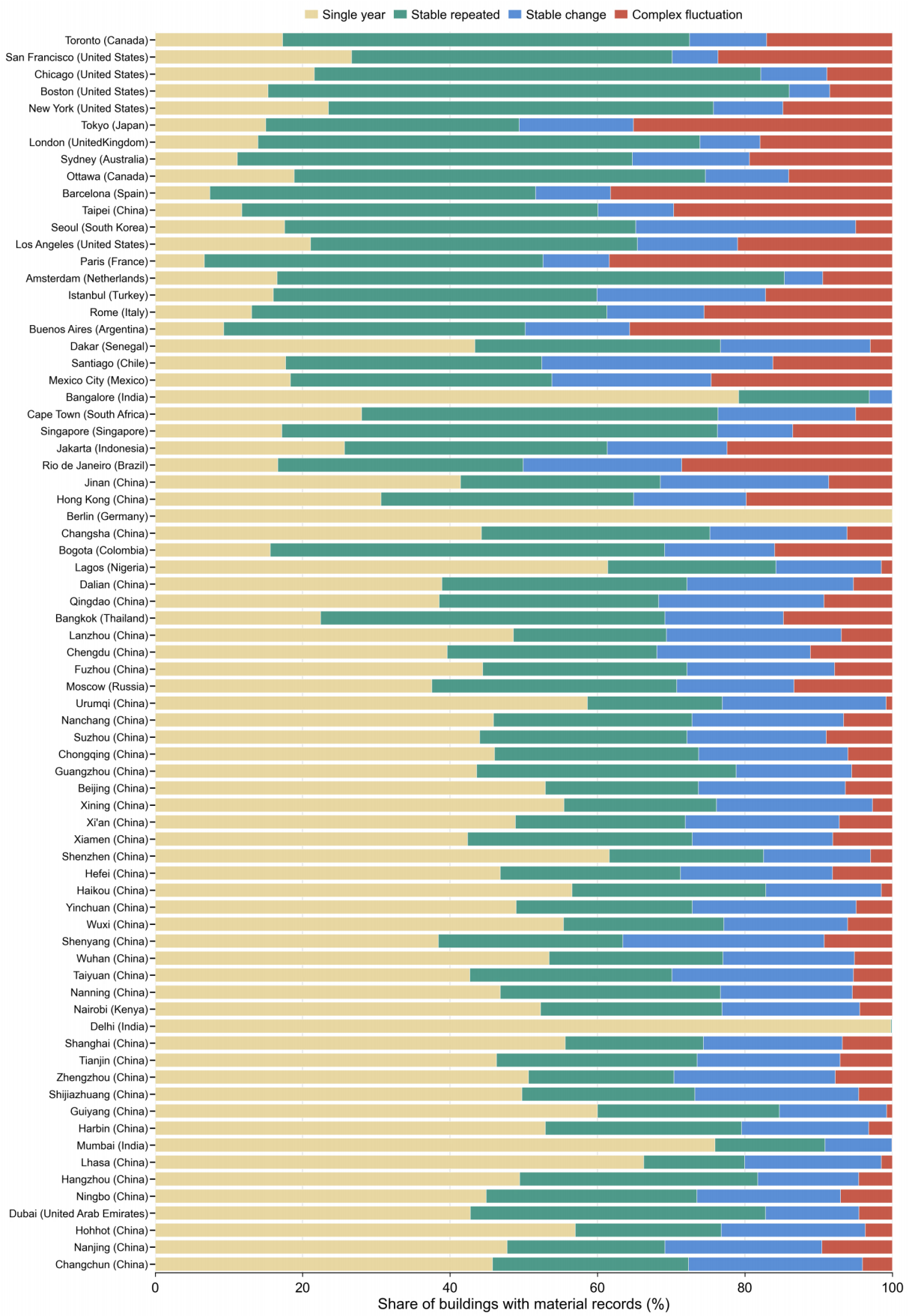

Figure 10: Temporal sequence types of building material records by city. Single-year records indicate buildings with material information from only one year; stable repeated records indicate multi-year records with the same material label; stable changes indicate a transition from one material type to another; complex fluctuations indicate changes among multiple material types.

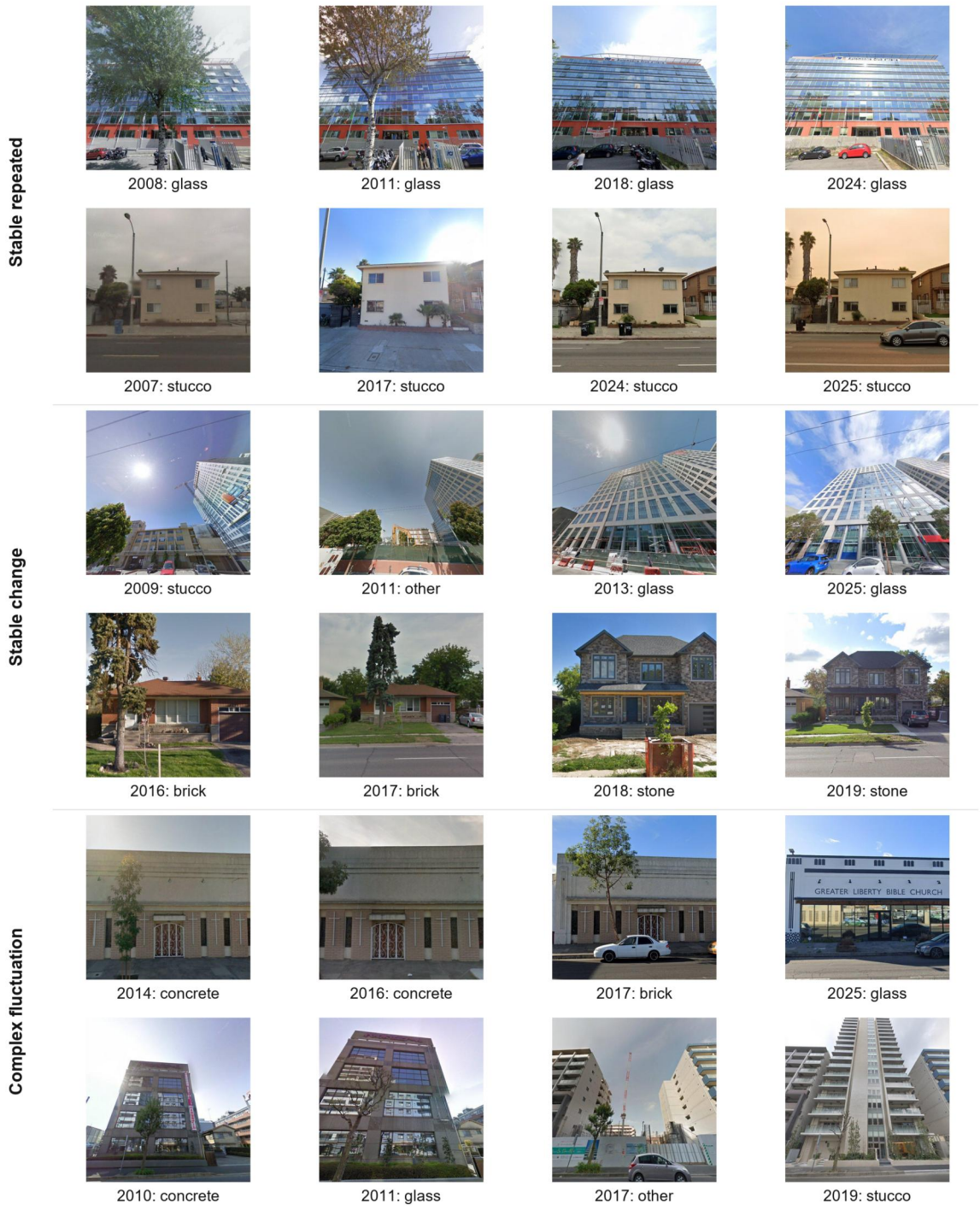

Figure 11: Representative temporal sequences of BMAT material records, including stable repeated records, stable changes, and complex fluctuations.

Comment 9. *Spatial coverage is incomplete and nonrandom, which may bias cross city comparisons. Material information is available for approximately 54 percent of the buildings, with substantial variation across cities. Because missing records are associated with street accessibility, building density, gated development, platform coverage, and distance from urban centers, the observed buildings are unlikely to represent a random sample. Consequently, city level material proportions and geographic patterns may partly reflect differences in imagery availability, study boundaries, and platform characteristics rather than actual facade distributions. The authors should assess coverage bias using building size, height, density, road distance, urban location, and neighborhood morphology, report coverage rates and uncertainty, and restrict or adjust cross city comparisons accordingly. Furthermore, the statement on page 20, lines 281 to 284 that missing material records can indirectly identify underdeveloped areas, informal housing, and marginalized neighborhoods is insufficiently supported. Missing street view imagery may also result from platform policies, legal or access restrictions, private roads, incomplete updates, and technical limitations. This interpretation should therefore be removed or validated using independent accessibility, settlement, and socioeconomic data.*

Response: Thank you for this important comment. We agree that BMAT does not provide material records for every building in each city. To examine the sensitivity of city-level material composition to incomplete material coverage, we performed two additional checks. First, we examined whether material proportions estimated from different fractions of buildings remain close to the material composition estimated from the full BMAT records for each city (Figure 12). Second, following your suggestion, we compared material proportions across different building volumes, local building densities, and street-view sampling distances (Figures 13 and 14).

For the first check, we randomly sampled buildings within each city at sampling fractions of 10%, 20%, ..., 90%. For each sampling fraction, the random sampling was repeated 100 times. We then compared the material composition estimated from each random subset with the material composition estimated from all BMAT records in the same city. The difference was measured using the total variation distance (TVD):

$$TVD(p, q) = \frac{1}{2} \sum_{k=1}^{K} |p_k - q_k|,$$

where p_k is the proportion of material class k in the random subset, q_k is the proportion of material class k in the full BMAT records for that city, and K is the number of categories, including both material classes and the additional "no material record" category. A smaller TVD value indicates that the subset-based material composition is closer to the full-city BMAT composition. The results show that material compositions estimated from random subsets of the observed records remain close to the full observed BMAT composition across sampling fractions (Figure 12). This suggests that, within the observed BMAT records, the dominant material structure is relatively stable under random reductions in sample size. This analysis

does not assume that missing records are random; it only evaluates the sampling stability of the available records.

For the second check, we assessed how material proportions (Figure 13) and missing-record shares (Figure 14) vary across coverage-related urban and observational conditions. Within each city, buildings were divided into low, medium, and high groups according to building volume, 1 km grid-level building density, and street-view sampling distance to the target building. For each group, we calculated the proportions of all material classes, with "no material record" included as an additional category. To show the most sensitive cases, we selected the GSV and BSV cities showing the largest differences in material composition across the low, medium, and high groups, and compared their material proportions across these strata. The results show that coverage-related factors mainly affect the completeness of material records. Smaller buildings and buildings in low-density areas have higher missing-material shares, especially in BSV cities. This pattern likely reflects the weaker street-view coverage of many small houses in peripheral areas. Buildings with longer sampling distances also have more missing records, probably because distant facades are more easily blocked by intervening buildings or vegetation. Nevertheless, in most selected cities, the leading material classes and their relative ordering remain generally similar across strata. This suggests that coverage-related factors influence record completeness, but they do not generally lead to a complete reordering of the main city-level material composition.

We added these coverage-related analyses to the Supplementary Materials. We also revised the discussion in the manuscript to provide a more careful interpretation of areas without material records.

> *"Spatially, missing material records may also contain useful spatial information. In some areas, large clusters of missing records may result from narrow roads, private or restricted access that make it difficult for street-view vehicles to enter or capture usable facade images (Kamalipour and Dovey, 2019). In other cases, missing records may reflect platform-specific coverage gaps (Fan et al., 2025)."*

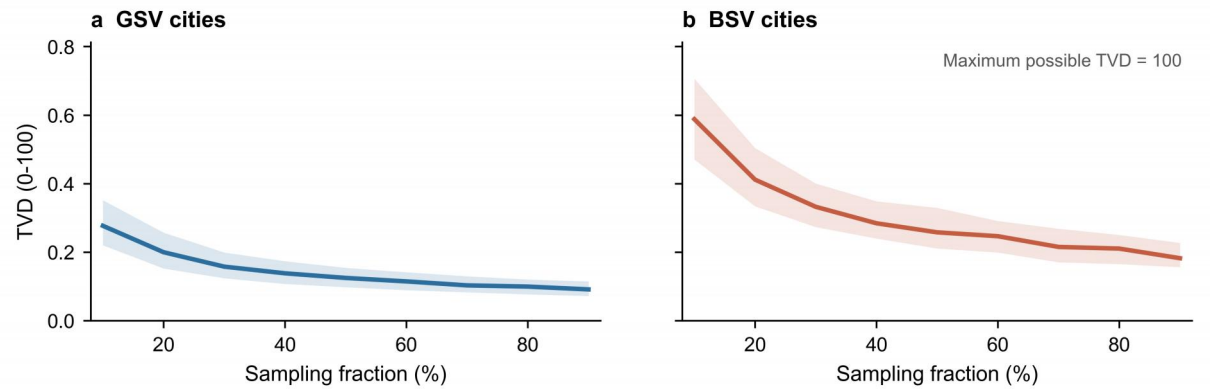

Figure 12: Random-subsampling stability of city-level material composition. TVD measures the difference between material proportions estimated from random building subsets and those estimated from all BMAT records in each city. Sampling was repeated 100 times at each fraction.

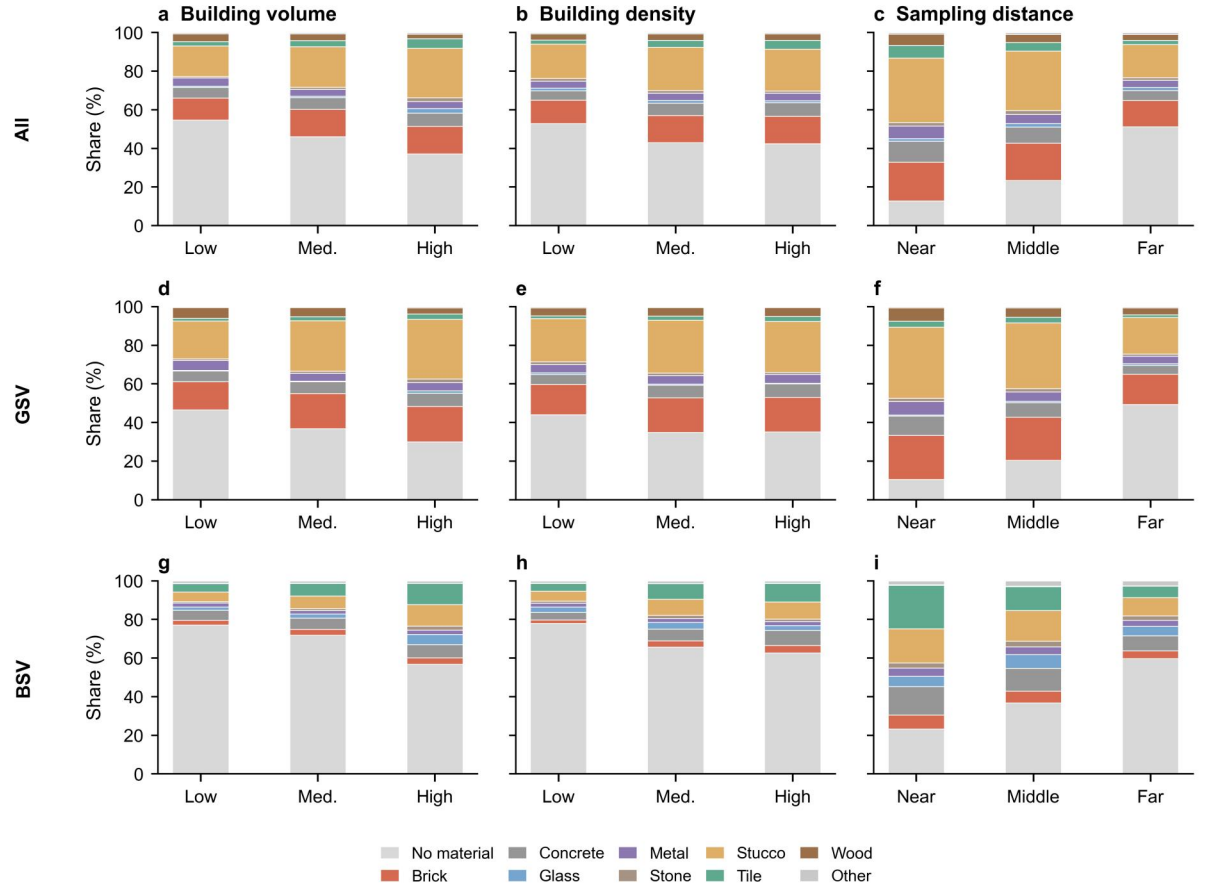

Figure 13: Material composition across coverage-related strata. Buildings are grouped by tertiles of building volume, local building density, and street-view sampling distance. "No material record" is included as an additional category.

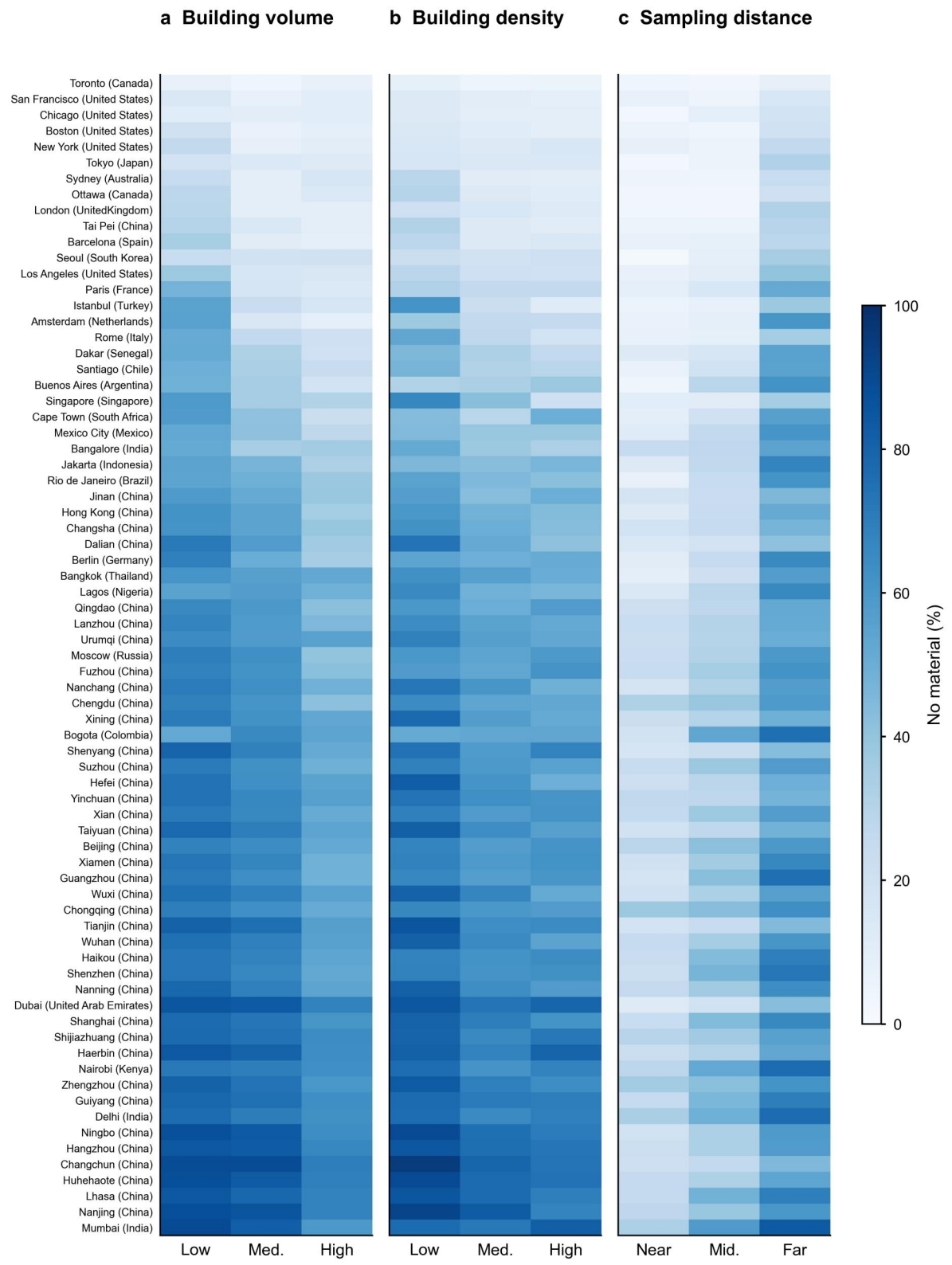

Figure 14: Missing-material share across coverage-related strata by city. Colors indicate the percentage of buildings without material records within each tertile of building volume, local building density, and street-view sampling distance.

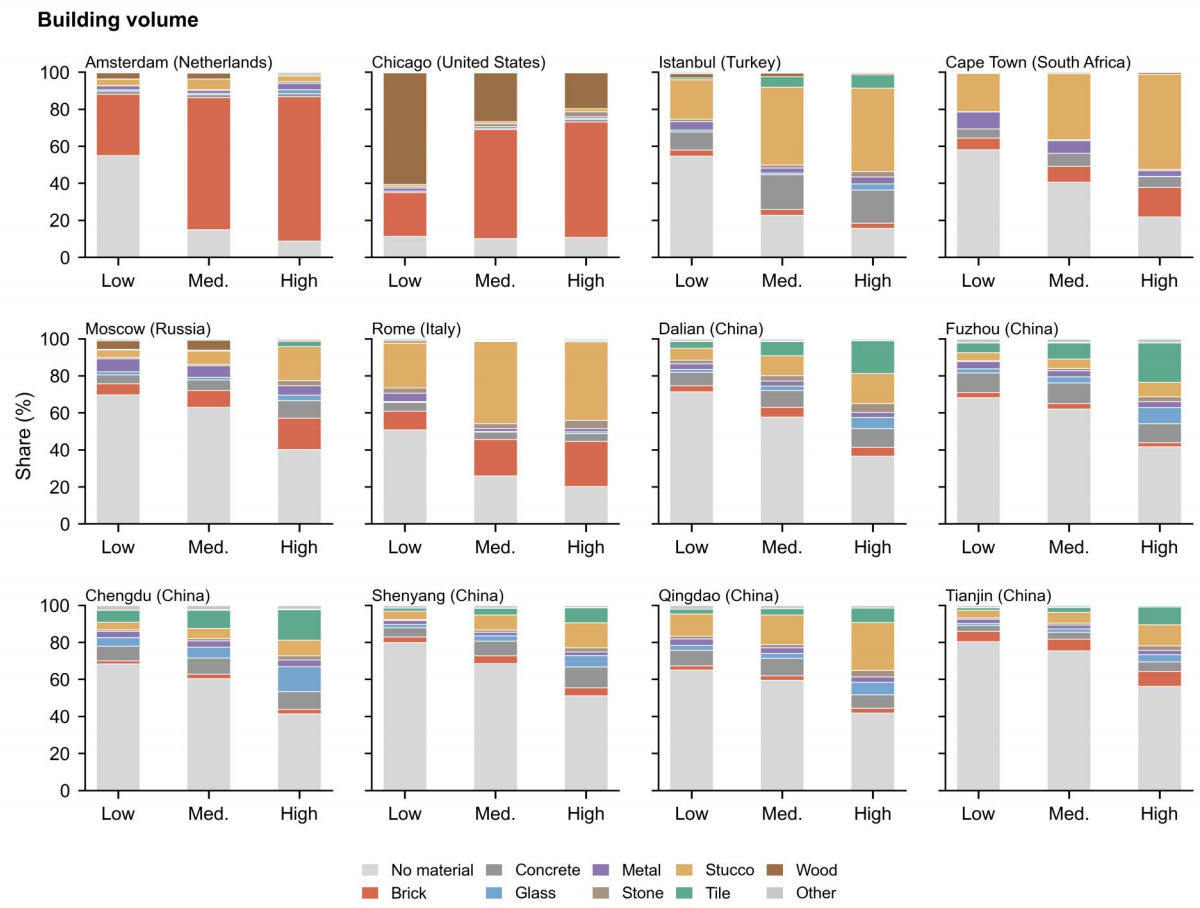

Figure 15: Cities with the largest material-composition differences across building-volume strata. For each platform, the six cities with the largest TVD across tertiles are shown. "No material record" is included as an additional category.

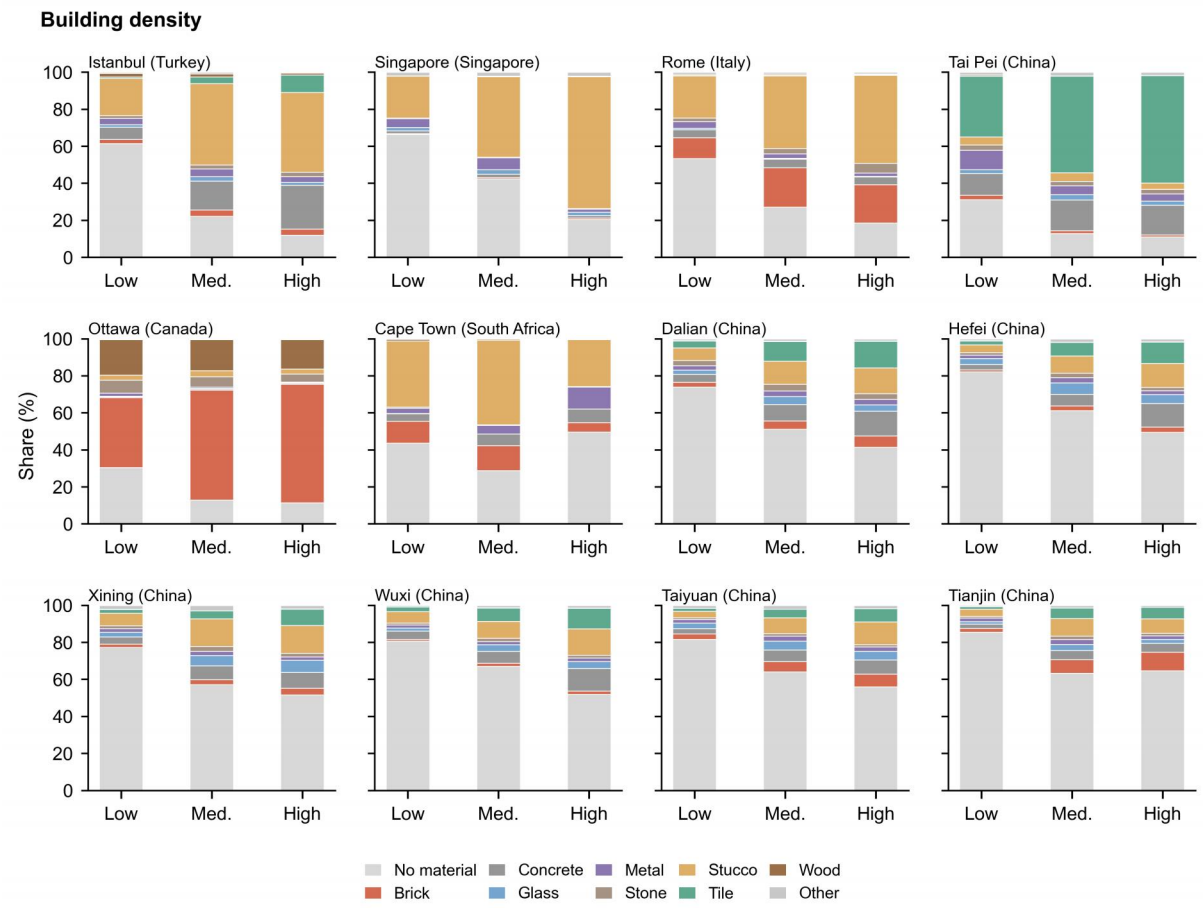

Figure 16: Cities with the largest material-composition differences across local building-density strata. For each platform, the six cities with the largest TVD across tertiles are shown. "No material record" is included as an additional category.

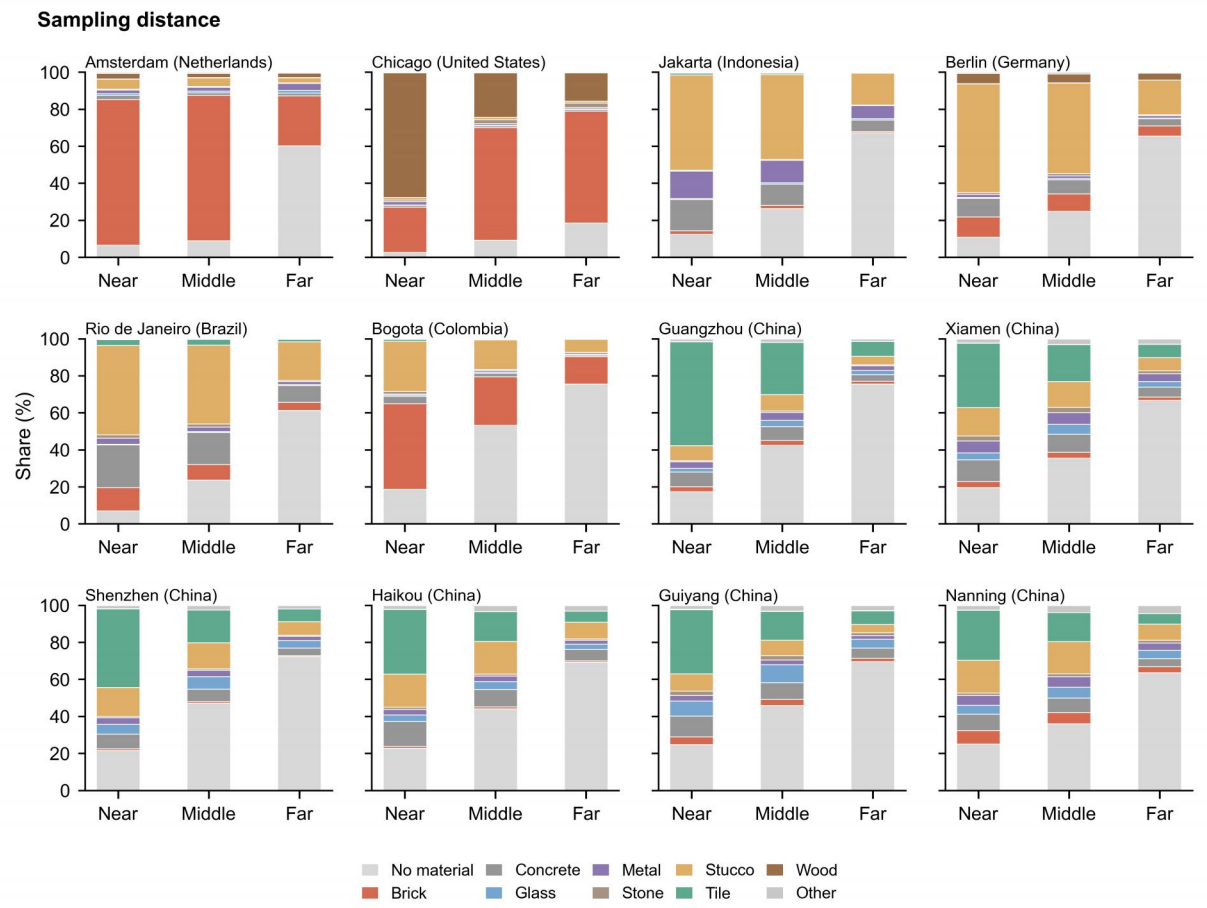

Figure 17: Cities with the largest material-composition differences across street-view sampling-distance strata. For each platform, the six cities with the largest TVD across tertiles are shown. "No material record" is included as an additional category.

Comment 10. *Reproducibility and data documentation need substantial improvement. A data paper should allow users to understand, reproduce, and responsibly reuse the product. The manuscript should report the exact versions, release dates, access dates, or image acquisition periods for Overture Maps, GlobalBuildingAtlas, GADM, OpenStreetMap, Google Street View, and Baidu Street View, as applicable.*

Response: Thank you for this suggestion. We used Overture Maps v1.13.0, GADM v4.1, and OpenStreetMap data downloaded using OSMnx v2.0.4. Building height data were obtained from GlobalBuildingAtlas (Zhu et al., 2025). Google Street View images cover 2007-2025, and Baidu Street View images cover 2013-2023. These versions and time periods are now clearly reported in the revised manuscript. The relevant data versions and image acquisition periods reported in the manuscript are listed below:

- **City administrative boundaries.** The GADM version was added:

  > *"This study uses the GADM database v4.1 to obtain administrative boundaries of major cities worldwide."*

- **Building footprint data.** The Overture Maps schema version was added:

  > *"Building footprint data were acquired from the Overture Maps Foundation global dataset (schema version: v1.13.0)."*

- **Road network data.** The OSMnx version was added:

  > *"Road network data were retrieved from OpenStreetMap (OSM) using the OSMnx Python library (version: v2.0.4)."*

- **Google Street View images.** The image acquisition period was clarified:

  > *"GSV images were collected for 36 cities outside mainland China, covering available observations between 2007 and 2025."*

- **Baidu Street View images.** The image acquisition period was clarified:

  > *"BSV images were collected for 37 cities, covering the period 2013–2023."*

Minor comments.

> **Comment Minor 1.** *Figure 13(a) presents material transitions from 2015 to 2025 for Beijing and Shenzhen, although the manuscript states that Baidu Street View imagery is available only from 2013 to 2023. The authors should clarify the source of the 2025 records or correct the figure and the associated text. More generally, all temporal figures should use years that are consistent with the actual observation periods of the corresponding street view platforms.*

Response: Thank you for pointing this out. Most of the latest Baidu Street View images for Beijing and Shenzhen were acquired in 2023. We have therefore changed the end year in this figure from 2025 to 2023.

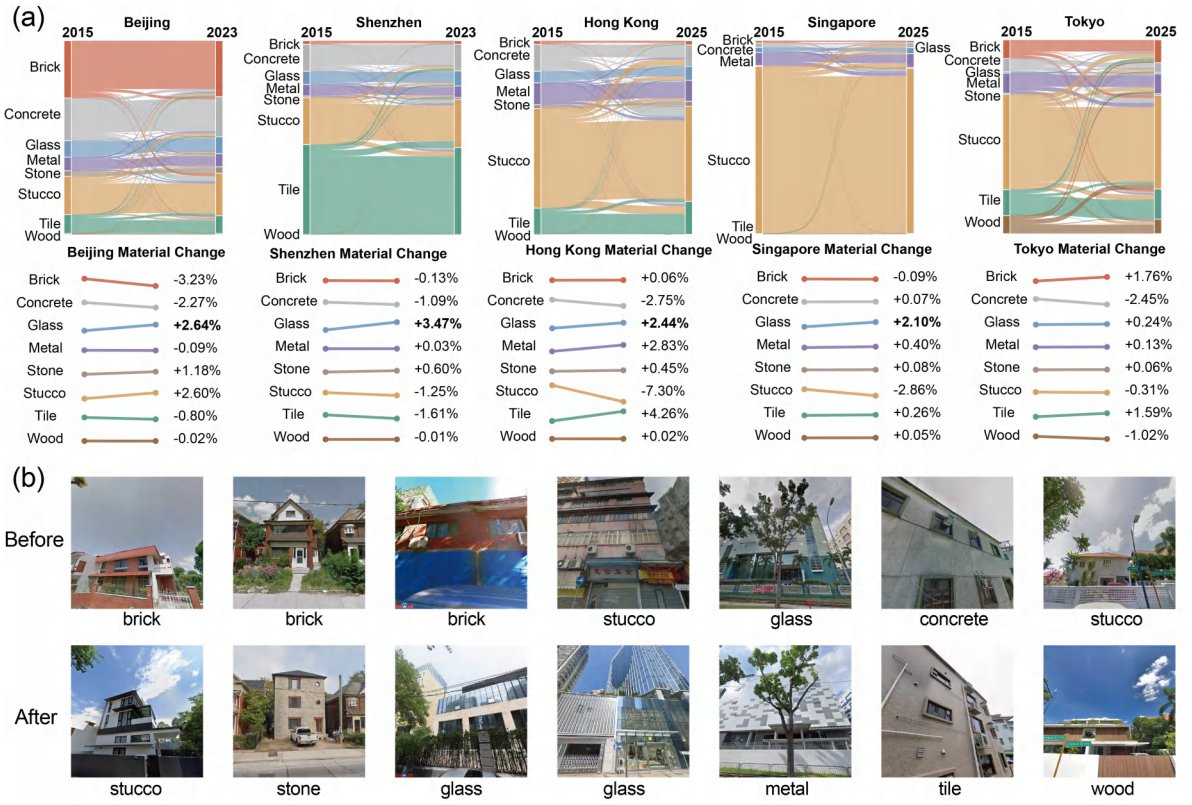

Figure 18: Building material transitions over time: (a) Flows and statistical shifts of material composition in case study cities; (b) Visual comparison of building facade before and after material changes.

Comment Minor 2. *The visualization in Figure 9(b) requires clarification. If the values represent proportions calculated within cumulative 5 km, 10 km, and 15 km buffers, these measures are not additive and should not be displayed as stacked bars. Grouped bars or line plots would provide a more appropriate comparison. If the segments instead represent nonoverlapping annular zones, this should be clearly stated in the caption and Methods section.*

Response: Thank you for this helpful suggestion. The values in Figure 9(b) (Figure 19 in this document) represent cumulative material-record coverage within increasing distances from the city center, such as 5 km, 10 km, and 15 km buffers. We revised the visualization to use grouped bars, with bars for different buffer distances placed side by side to facilitate comparison while saving space. We also clarified in the manuscript and figure caption that these values are cumulative coverage proportions within each buffer distance.

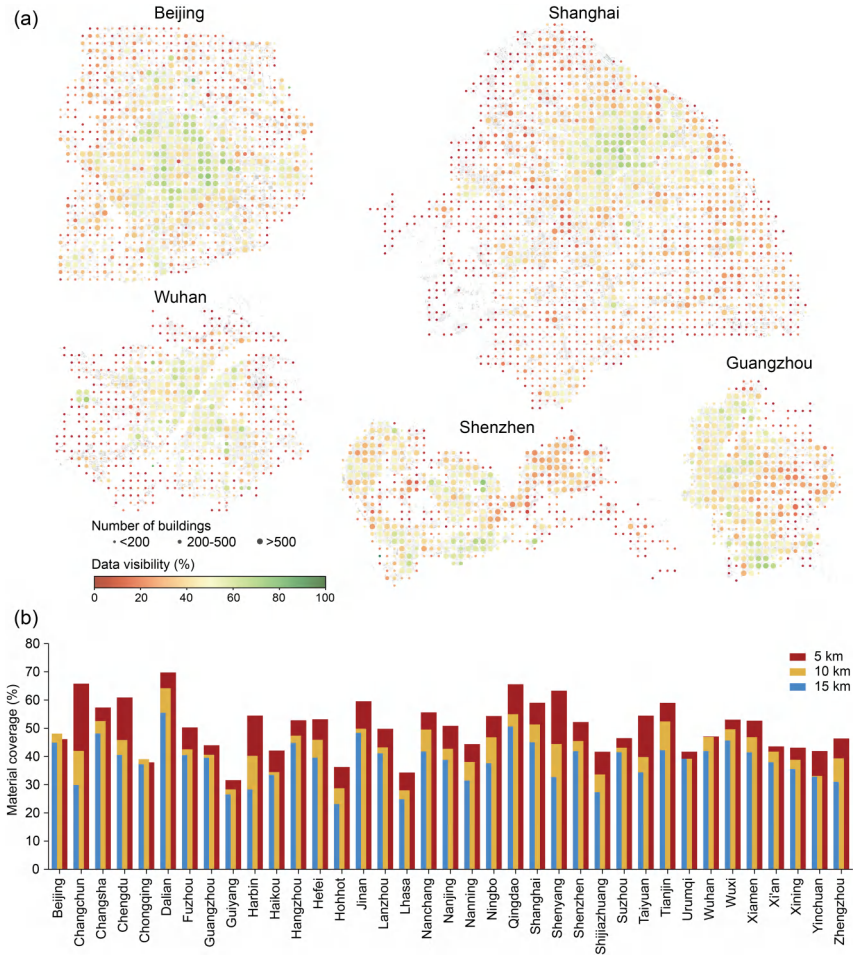

Figure 19: Spatial coverage of building material data inferred from BSV images: (a) Building counts and material coverage proportions at 2 km grid resolution in representative cities; (b) Proportions of buildings with material records at varying distances from city centers.

Comment Minor 3. *The manuscript reports an F1 score of 0.91 in the Results section but refers to the same value as overall accuracy in the Discussion section. The terminology should be corrected throughout.*

Response: Thank you for pointing out this inconsistency. Both the F1 score and overall accuracy of the model were 0.91, as reported in Supplementary Table S2 (Table 7 in this document). To ensure consistency throughout the manuscript, we have revised the Discussion to refer to this value as the F1 score.

Table 7: Class-wise performance comparison with traditional models

Category	Metric	VGG16	ViT	Res50	Shuff	Res101	EffB0	Dense	MbNet	Ours
Overall	Accuracy	0.83	0.82	0.82	0.79	0.82	0.84	0.84	0.82	**0.91**
	Precision	0.82	0.81	0.80	0.78	0.81	0.82	0.83	0.81	**0.91**
	Recall	0.80	0.79	0.80	0.76	0.80	0.82	0.82	0.80	**0.91**
	F1 score	0.81	0.80	0.80	0.77	0.80	0.82	0.82	0.80	**0.91**
Brick	Precision	0.87	0.82	0.84	0.80	0.84	0.86	0.85	0.85	**0.92**
	Recall	0.84	0.82	0.84	0.79	0.82	0.83	0.85	0.83	**0.94**
	F1 score	0.86	0.82	0.84	0.79	0.83	0.85	0.85	0.84	**0.93**
Concrete	Precision	0.77	0.76	0.78	0.71	0.75	0.77	0.81	0.77	**0.86**
	Recall	0.75	0.73	0.71	0.70	0.74	0.78	0.74	0.72	**0.85**
	F1 score	0.76	0.75	0.75	0.70	0.75	0.77	0.78	0.74	**0.86**
Glass	Precision	0.85	0.88	0.86	0.87	0.88	0.88	0.86	0.86	**0.95**
	Recall	0.93	0.95	0.94	0.92	0.93	0.95	0.95	0.94	**0.96**
	F1 score	0.89	0.91	0.90	0.89	0.90	0.91	0.91	0.90	**0.95**
Metal	Precision	0.78	0.77	0.74	0.73	0.79	0.78	0.79	0.76	**0.88**
	Recall	0.72	0.73	0.72	0.69	0.72	0.73	0.74	0.72	**0.88**
	F1 score	0.75	0.75	0.73	0.71	0.76	0.76	0.77	0.74	**0.88**
Other	Precision	0.81	0.83	0.77	0.76	0.76	0.74	0.78	0.79	**0.89**
	Recall	0.67	0.66	0.72	0.60	0.66	0.74	0.69	0.66	**0.91**
	F1 score	0.73	0.73	0.75	0.67	0.70	0.74	0.74	0.72	**0.90**
Stone	Precision	0.89	0.88	0.87	0.87	0.89	0.89	0.92	0.89	**0.95**
	Recall	0.87	0.87	0.88	0.84	0.87	0.89	0.88	0.88	**0.94**
	F1 score	0.88	0.87	0.88	0.86	0.88	0.89	0.90	0.88	**0.95**
Stucco	Precision	0.82	0.81	0.80	0.77	0.82	0.85	0.83	0.81	**0.89**
	Recall	0.84	0.82	0.81	0.81	0.83	0.83	0.85	0.83	**0.89**
	F1 score	0.83	0.81	0.80	0.79	0.82	0.84	0.84	0.82	**0.89**
Tile	Precision	0.73	0.73	0.68	0.66	0.69	0.76	0.74	0.73	**0.88**
	Recall	0.71	0.68	0.71	0.67	0.71	0.73	0.72	0.72	**0.83**
	F1 score	0.72	0.70	0.70	0.66	0.70	0.74	0.73	0.72	**0.85**
Wood	Precision	0.87	0.85	0.86	0.84	0.85	0.87	0.86	0.85	**0.96**
	Recall	0.90	0.88	0.86	0.83	0.89	0.91	0.91	0.90	**0.97**
	F1 score	0.88	0.87	0.86	0.84	0.87	0.89	0.88	0.88	**0.97**

Note: Abbreviations: VGG16: VGG16; ViT: Vision Transformer (ViT-B/16); Res50: ResNet-50; Shuff: ShuffleNetV2; Res101: ResNet-101; EffB0: EfficientNet-B0; Dense: DenseNet-161; MbNet: MobileNetV3-Large; **Ours**: Our fine-tuned Qwen2.5-VL-7B-Instruct model.

Comment Minor 4. *High image classification accuracy should not be interpreted as evidence of high overall dataset reliability without validation of the complete data production pipeline.*

Response: Thanks for your comment. We have added a detailed description of the complete BMAT production pipeline, from building-specific street-view sampling to footprint-level material aggregation. Specifically, the revised "Data sources" and "Methods" now describe how street-view sampling points and camera parameters were determined for each target footprint, how unsuitable images were removed using the visibility classifier, how building-level obstruction was assessed using geometric sightlines, how facade material was inferred from retained images, and how multiple image-level predictions were aggregated into annual building-level material records and the latest BMAT label. We also reported the scale of each processing stage: the pipeline started from approximately 147 million raw street-view images, retained approximately 96 million valid building-facade images after visibility and quality filtering, and finally generated material records for 22.09 million building footprints.

To support reproducibility, we publicly released the source code (https://github.com/yhyJoy/BMAT), trained model (https://huggingface.co/yinjoy30/BMAT), and human-annotated data (https://github.com/yhyJoy/BMAT/tree/main/data/label). The classifier achieved an F1 score of 0.91 on the held-out test set. Agreement with Overture Maps was 0.72 under the latest-record strategy and 0.79 under the historical-match sensitivity analysis. In addition, an AI-assisted audit of a stratified sample of 1,000 final inference records yielded a material-label agreement of 87.3%.

Comment Minor 5. *The claims that BMAT is the first dataset of its type and that the annotated training set is the largest to date require a transparent comparison with existing datasets or more cautious wording.*

Response: Thank you for this helpful suggestion. We added Supplementary Table S3 (Table 8 in this document) to summarize recent studies that use street-view imagery to infer building facade materials. The table compares these resources in terms of city coverage, number of annotated material labels, material classes, openness, and main output. This comparison shows that BMAT provides the largest dataset in terms of both city coverage and annotated material-label volume among the reviewed street-view-based facade material resources.

Table 8: Comparison of recent street-view-based facade material resources.

Study	Year	Cities	Labels	Classes	Open	Main output
Raghu et al. (2023)	2023	3	985	7	Image data	Annotated images and material detection model
Wang et al. (2024)	2024	2	2,567	6	Image data	Facade image dataset for cladding classification
Chen et al. (2025)	2025	8	17,914	9	No	Training data and city-scale material mapping
Sun et al. (2025)	2025	4	11,787	9	No	Roof and facade material identification
Liang et al. (2025)	2025	7	2,871	7	Toolkit	OpenFACADES attribute-enrichment workflow
BMAT (ours)	2026	73	39,405	9	Dataset	Footprint-level material records for 22.09 million buildings

Note: Labels refer to the number of annotated facade or material samples used for material classification or evaluation. Classes refer to material categories. Open indicates the type of publicly available resource reported by each study.

Comment Minor 6. *The source and definition of building age information used in the supplementary analyses should be described in the Methods section.*

Response: Thank you for this helpful suggestion. We added information on the building-age data used in the supplementary analysis to both the manuscript and the Supplementary Materials. In the manuscript, we now state that building-level construction-year data were collected for four cities to explore the relationship between building age and facade material types. In the Supplementary Materials, we further specify the purpose of this analysis and the data sources used for Boston, New York City, Amsterdam, and Paris.

> *"We also collected building-level construction-year data for four cities to explore the relationship between building age and facade material types (Supplementary Section S10)."*

> *"Building facade materials are closely related to construction periods, because material choices reflect the construction technologies, architectural styles, and urban development practices of different eras. To explore this relationship, we collected building-level construction-year data for four cities with publicly available building/property records: Boston (https://www.placekey.io/datasets/boston-property-assessment-data), New York City (https://data.cityofnewyork.us/City-Government/BUILDING/5zhs-2jue), Amsterdam (https://eubucco.com/), and Paris (https://eubucco.com/)."*

Comment Minor 7. *Figure 7 should display or reference the material coverage rate and the number of buildings with valid material records for each city because proportions derived from substantially different coverage levels are not directly comparable.*

Response: Thank you for this helpful suggestion. We revised both the figure and the corresponding manuscript description to make the material proportions more comparable across cities with different coverage levels. In the revised figure, buildings without valid material records are included as a separate "No material record" category, so the stacked bars are calculated relative to all building footprints in each city.

"Figure 7 illustrates the city-level composition of BMAT material records by building count, with buildings without valid material records included as a separate "No material record" category. Cities outside mainland China generally show higher proportions of valid material records, reflecting the broader spatial coverage and more frequent update cycles of GSV compared with BSV (Supplementary Figure S2). Among buildings with valid material records, distinct material patterns emerge across cities. Coastal cities (e.g., Singapore, Los Angeles, San Francisco, Tokyo, Mumbai) show a high proportion of stucco facades, consistent with its suitability for humid maritime environments due to moisture resistance. Cities in mid-to-high latitude cold climate regions (e.g., Ottawa, Toronto, Amsterdam, Chicago, London) are characterized by a high proportion of brick, a material valued for its thermal mass and heat retention. Chinese cities exhibit considerably higher proportions of tile facades than other regions, likely reflecting local ceramic manufacturing traditions and architectural preferences. Glass facades are also more prevalent in Chinese cities, corresponding to rapid urbanization and the widespread adoption of curtain wall construction in recent decades."

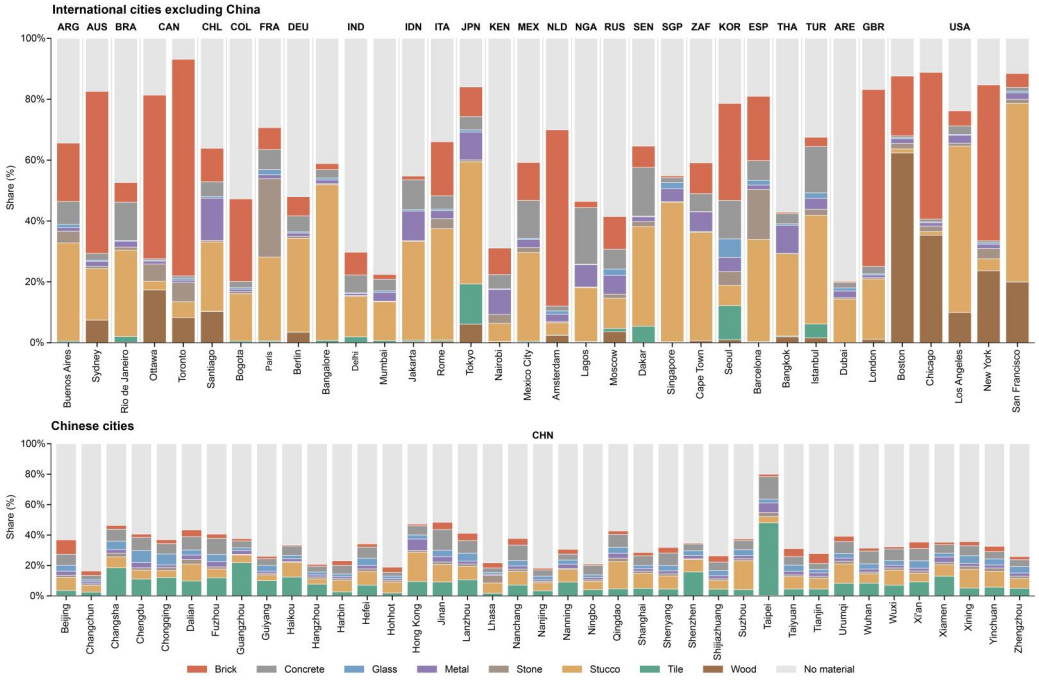

Figure 20: Distribution of building facade materials across the 73 studied cities.

Comment Minor 8. *Several city names in Table 2 use nonstandard English spellings. Haerbin, Huhehaote, Lasa, Taibei, Wulumuqi, and Xian should be revised to Harbin, Hohhot, Lhasa, Taipei, Urumqi, and Xi'an, respectively. All city names should be checked and standardized throughout the manuscript.*

Response: Thank you for pointing this out. We have checked and standardized city names throughout the manuscript, including all tables and figures. These city names have been revised using standard English spellings.

Comment Minor 9. *The manuscript should undergo careful language editing. Several sentences contain grammatical or capitalization problems.*

Response: Thank you for this suggestion. We have carefully edited the manuscript to improve grammar, capitalization, and overall readability.

We sincerely thank the reviewer once again for these thoughtful and constructive comments, which have substantially helped us improve the clarity, reliability, transparency, and documentation of both the dataset and the manuscript.

References

Chen, X., Ding, X., and Ye, Y. (2025). Mapping sense of place as a measurable urban identity: Using street view images and machine learning to identify building façade materials. *Environment and Planning B: Urban Analytics and City Science*, 52(4):965–984.

Fan, Z., Feng, C.-C., and Biljecki, F. (2025). Coverage and bias of street view imagery in mapping the urban environment. *Computers, Environment and Urban Systems*, 117:102253.

Kamalipour, H. and Dovey, K. (2019). Mapping the visibility of informal settlements. *Habitat International*, 85:63–75.

Liang, X., Xie, J., Zhao, T., Stouffs, R., and Biljecki, F. (2025). Openfacades: an open framework for architectural caption and attribute data enrichment via street view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 230:918–942.

Raghu, D., Bucher, M. J. J., and De Wolf, C. (2023). Towards a 'resource cadastre' for a circular economy–urban-scale building material detection using street view imagery and computer vision. *Resources, Conservation and Recycling*, 198:107140.

Sun, K., Li, Q., Liu, Q., Song, J., Dai, M., Qian, X., Gummidi, S. R. B., Yu, B., Creutzig, F., and Liu, G. (2025). Urban fabric decoded: High-precision building material identification via deep learning and remote sensing. *Environmental Science and Ecotechnology*, 24:100538.

Wang, S., Park, S., Park, S., and Kim, J. (2024). Building façade datasets for analyzing building characteristics using deep learning. *Data in Brief*, 57:110885.

Zhu, X. X., Chen, S., Zhang, F., Shi, Y., and Wang, Y. (2025). Globalbuildingatlas: an open global and complete dataset of building polygons, heights and lod1 3d models. *Earth System Science Data*, 17:6647–6668.