Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Heatmap-Guided Fire Appearance Stylization of Building Facade Textures for Urban Digital Twins

HyeonJoong Wi
Korea University of Science and Technology
Insung Jang
Electronics and Telecommunications Research Institute
Sang Gi Hong
sghong@etri.re.kr

Electronics and Telecommunications Research Institute

Research Article

Keywords: 3D building model, geovisualization, urban digital twin, generative artificial intelligence, fire disaster visualization

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11273288/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Heatmap-Guided Fire Appearance Stylization of Building Facade Textures for Urban Digital Twins

HyeonJoong Wi[1,2], Insung Jang[1], Sang Gi Hong[3,2*]

[1]Air Mobility Intelligence Research Section, Electronics and Telecommunications Research Institute (ETRI), Gajeong-ro, Yuseong-gu, Daejeon, 34129, Republic of Korea.
[2]ETRI School, University of Science and Technology (UST), Gajeong-ro, Yuseong-gu, Daejeon, 34113, Republic of Korea.
[3*]Future Defense AX Research Section, Electronics and Telecommunications Research Institute (ETRI), Gajeong-ro, Yuseong-gu, Daejeon, 34129, Republic of Korea.

*Corresponding author(s). E-mail(s): sghong@etri.re.kr;
Contributing authors: wjh4877@etri.re.kr; e4dol2@etri.re.kr;

Abstract

Textured 3D building models are increasingly used as visual and spatial assets in urban digital twins, yet their fragmented texture-atlas layouts make localized and structure-preserving disaster appearance adaptation difficult. This study proposes a heatmap-guided facade fire appearance stylization framework that modifies building textures without changing the original mesh geometry, texture coordinates, or material configuration. The framework first rectifies fragmented atlas regions into facade-aligned images while preserving triangle-wise correspondences with the source atlas. Wall–window masks and edge-based structural guidance are then combined with a normalized fire heatmap that controls the spatial extent of the fire appearance. The modified facade images are subsequently reconstructed into the original texture atlas through piecewise-affine mapping. Experiments were conducted on 1,435 upright facade images extracted from 428 real building models. As the normalized appearance scale increased from 0.25 to 1.00, Target-LPIPS increased from 0.1708 to 0.2925. Full-dataset ablation showed that heatmap guidance substantially reduced unintended changes in non-wall regions and improved structural preservation. Compared with four representative diffusion-based stylization methods, the proposed method achieved the lowest Protected-LPIPS of 0.0816 and the highest Edge-SSIM of 0.6999.

These results demonstrate controllable fire-appearance modification while preserving non-target facade content and architectural structure. Within this scope, the framework is intended as an AI-driven geovisualization approach for visual scenario generation that complements, rather than replaces, physics-based fire simulation.

Keywords: 3D building model, geovisualization, urban digital twin, generative artificial intelligence, fire disaster visualization

1 Introduction

As shown in Figure 1, a city model is a digital representation of urban infrastructure and objects, such as terrain and buildings [1, 2].



Fig. 1 Example of a city model in which topography, building models, etc. are visualized

It is mainly used in simulations to reproduce and analyze urban phenomena [3, 4]. In particular, city models are used as fundamental data for disaster reproduction, response, and management. This indicates that city models are not limited to simply visualizing the locations of disasters in urban areas. Rather, they can also visualize the shapes of buildings, surrounding infrastructure, and urban structures in disaster-affected areas, thereby supporting situational awareness, disaster scenario review, and response activities [5–8]. In particular, building models, which constitute a major component of city models, are used as important fundamental data for evacuation, recognition, and assessment in disaster situations [9, 10]. Accordingly, a considerable number of related simulation methods have also been developed.

Among disaster simulations that use city models, fire simulations can visualize disaster scenes, smoke dispersion, and surrounding conditions, as shown in Figure 2.

Previous fire simulation studies have mainly focused on human evacuation [11, 12], the spread of fire inside and outside buildings [13], and smoke dispersion [5]. These studies have played an important role in fire safety assessment and the establishment



Fig. 2 Example of fire-simulation visualization within a city model

of response plans. Accordingly, recent fire simulations using city models have been specialized in heat, smoke, evacuation, and facility control.

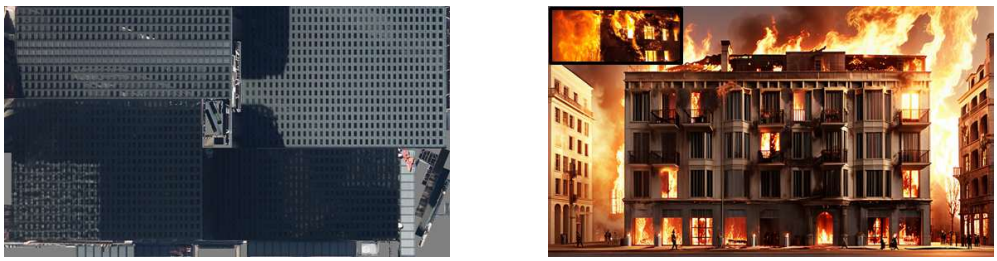
However, existing fire disaster simulations have mainly focused on phenomena occurring inside buildings. Although such simulations concentrate on internal building conditions to minimize human casualties, exterior visualization is also important as a key tool for visualizing and managing risks at disaster sites and supporting decision-making by relevant authorities [11, 14]. Previous studies have reported that representing the exterior appearance of buildings under disaster conditions is useful for the intuitive understanding of such situations and that there is a demand for this type of visualization. They have also highlighted the need to represent the exterior appearance of buildings during urban fire disasters, even in the absence of an actual disaster, for applications such as generating artificial intelligence training data for fire disaster response [15, 16].

Among fire disasters occurring in urban areas, fires originating in buildings exhibit different externally observable flame shapes depending on factors such as flame height, plume rise, air inflow, and ejection through openings such as windows [5, 17]. In fire simulations, the intensity of these factors is adjusted to represent different fire conditions, including the initial, intermediate, and final stages of a fire [12].

To apply style transfer to the exterior walls of buildings, we select representative factors that determine flame appearance and visually represent fire appearance on the exterior-wall textures of building models using diffusion-based style transfer. In particular, the vertical rise and horizontal spread of fire are simplified and represented as a heatmap. The generated heatmap is then used as soft guidance for applying fire appearance to building textures.

General diffusion-based style transfer can apply a fire style to textures using text guidance. However, when style transfer is applied to texture atlas images of building models, the building components and shapes may become distorted or reconstructed into entirely different forms, as shown in Figure 3. This occurs because diffusion-based style transfer tends to modify the global composition of an image [18].

To address this problem, we use a wall–window mask and edge information as spatial control elements to preserve the building structure. At the same time, fire



(a) (b)

Fig. 3 Failure cases of direct atlas stylization: (a) original 3D building model and (b) 3D building model after direct atlas stylization

appearance stylization is applied to the building texture within the heatmap-guided region so that the resulting texture reflects the intended fire appearance.

Previous studies of fire disasters in urban areas have reported relatively high fire occurrence rates in commercial, residential, and mixed-use buildings [19–21]. Accordingly, the building models used in fire disaster simulations have also primarily represented commercial or residential buildings [22]. In particular, commercial and residential buildings have been found to account for the majority of the building models constituting city models [21]. These cases indicate that commercial, residential, and mixed-use buildings are representative building types for fire simulations in urban areas.

Based on the building types identified above, commercial, residential, and mixed-use building models are selected as the targets for fire appearance stylization in this study. Fire-disaster simulations also represent flame conditions across different stages, from fire initiation to development [6, 23]. To provide a graded visual representation without treating the scale as a physical fire stage, the normalized appearance scale is evaluated at four levels, from a localized fire appearance to a more spatially developed appearance.

The building models used in this study were obtained from V-World, a Korean 3D spatial information service that provides building geometry and texture data [24]. Buildings with heights of 100 m or less were used. This sampling criterion was determined based on a previous study reporting that 95% of buildings in urban areas are below 100 m [25].

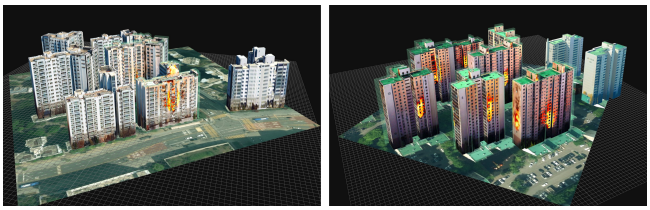


Fig. 4 Example of fire-appearance visualization using the proposed method within a city model

The main contributions of this work are as follows:

Facade-level fire appearance stylization A facade-level fire appearance stylization framework is proposed for existing 3D building models, in which a heatmap is used as soft guidance to spatially control fire-related visual effects while preserving the original building geometry.
Structure-preserving conditioning A conditioning framework combining wall–window masks and edge-based structural guidance is introduced to preserve facade components and structural layouts during heatmap-guided fire appearance stylization.
Evaluation on 3D building model textures The proposed framework is evaluated on real building models obtained from the V-World service through quantitative and qualitative experiments, demonstrating its applicability to fire-disaster scenario visualization in city models, as shown in Figure 4.

2 Related work

Previous studies on fire-disaster simulation using city models have primarily developed toward the simulation of indoor fires and evacuation models [12, 13, 26, 27]. Studies have modeled indoor fire conditions and evacuation routes for commercial, residential, and mixed-use buildings [12, 27]. In these simulations, fire scenarios corresponding to the initial, developing, and decay stages were defined based on heat release rate curves, while systems were also developed to evaluate evacuation time and safety.

Studies have also been conducted to simulate how a building affected by fire appears externally. For example, simulation research has combined building footprints with smoke models based on computational fluid dynamics to estimate how smoke spreads from the location of a building fire [13]. These simulations primarily focused on smoke dispersion and its effects on surrounding areas and demonstrated the three-dimensional visualization of real-time fire events and smoke at the urban scale. However, the 3D building models used in these city models were low level-of-detail models without textures. This limitation suggests that improvements are still possible in terms of risk communication, situational awareness, and training-data generation for disaster response [28].

However, methods for directly representing disaster conditions at the texture level of existing building models for disaster managers and decision-makers have not yet been sufficiently investigated. Previous disaster-response simulation studies have repeatedly reported that presenting realistic disaster conditions to disaster managers and decision-makers can support meaningful decision-making [29–32]. Nevertheless, existing studies and simulations have generally represented fire conditions using simple particles or color maps on building models, as shown in Figure 5, rather than directly applying facade-level fire appearances with controllable spatial progression to the textures of individual building models.

Texture style transfer for building models in city models has primarily focused on generating shaded regions [1, 33], creating textures that are visually consistent with surrounding buildings, or applying aesthetic style transformations [34, 35]. These studies demonstrate a demand for generating and modifying building-model textures and have further developed toward the application of style-transfer techniques to real city models [18, 36, 37]. However, many existing approaches still focus on the generation

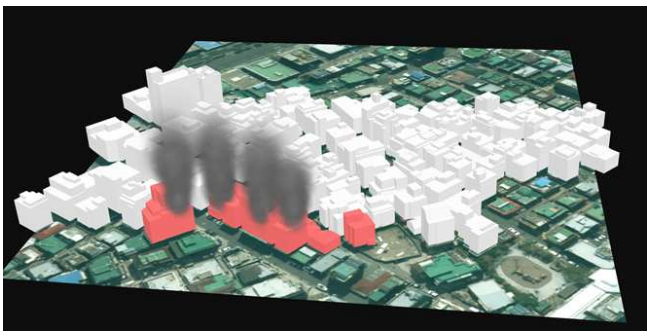


Fig. 5 Example of fire visualization using a low-LOD city model

of virtual buildings and do not directly process textures obtained from real buildings. In disaster simulation, minimizing the gap between the simulated environment and reality is particularly important. Therefore, using textures from real buildings is more effective than relying on virtual textures.

However, the textures used in building models are commonly organized as texture atlases rather than as facade images in which architectural components are clearly represented. When conventional texture style-transfer methods are applied directly to such atlas images, existing building components may disappear or become distorted, as shown in Figure 3. Applying style transfer to facade images can partially alleviate this problem. However, major architectural components or even the overall building shape may still be distorted, as shown in Figure 6.



(a) (b)

Fig. 6 Failure cases of direct facade stylization: (a) original building facade with style image and (b) building facade image after direct stylization

This occurs because the building shapes and architectural components represented in building textures are relatively small, while conventional style-transfer methods often prioritize global style over detailed architectural elements such as windows, walls,

and doors. More broadly, domain mismatch [38] also makes it difficult to perform building-texture style transfer while preserving the original architectural components.

Existing studies have highlighted the need to directly represent fire conditions on building models in city model-based fire-disaster applications. However, this issue has been addressed only to a limited extent. In contrast, studies on building-model texture style transfer have mainly focused on aesthetic transformation or the generation of building textures for shaded regions. These approaches have not sufficiently addressed detailed components, such as windows, walls, and overall building shapes, in atlas-based building textures, which are the texture format commonly used by real building models. Consequently, applying disaster appearances, such as fire effects, to the textures of building models remains challenging. Therefore, this study proposes heatmap-guided facade fire appearance stylization as a complementary approach for visually representing fire disasters on building models currently used in operational city model services and applicable to existing fire simulations.

To apply fire appearance to building-model textures, a heatmap is generated by reflecting the spatial characteristics of horizontal spread and vertical progression. The generated heatmap is then used as soft guidance during diffusion-based stylization to control the spatial distribution and relative strength of the applied fire appearance. In addition, wall–window masks and edge information are jointly used to preserve the major components of the building model. The original atlas texture is first converted into a facade-form representation, stylized, and then mapped back to the original texture structure. Through this process, a stylized atlas texture is generated in a form that can be directly applied to the original 3D building model.

3 Method

The proposed heatmap-guided facade fire appearance stylization method performs style transfer by using a fire heatmap as guidance while preserving the building components represented in the building texture. Figure 7 presents an overview of the proposed method, which applies a fire style to building-model textures through four main stages.

The first stage generates facade images by mapping the geometric information of the building model to the texture atlas through UV correspondence. The second stage generates guidance information for preserving building components and incorporating fire-related characteristics during style transfer. The third stage performs style transfer by applying the generated guidance information to the style-transfer network. Finally, inverse mapping is conducted using the correspondence information retained during facade generation, and the resulting stylized facade images are reconstructed into a texture atlas that can be used as the final texture of the building model.

3.1 Atlas-to-facade rectification and correspondence construction

A texture atlas consolidates multiple surface texture regions into a single image for efficient storage and rendering. In textured 3D building models, however, facade texture patches are often fragmented, rotated, and spatially disjoint within the atlas, so

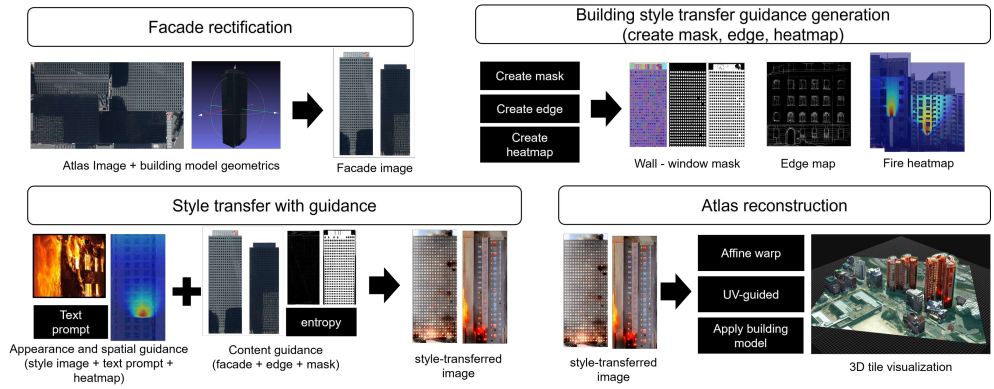


Fig. 7 Functional components and workflow of the proposed method

the coherent arrangement of facade components such as windows, walls, and architectural edges is not directly represented. The source texture atlas is therefore converted into facade-aligned images before guidance generation and stylization.

To convert the texture atlas into facade images, the geometric information of the building model is used to project the atlas regions onto facade-aligned planes. A building model generally consists of vertex coordinates, indexed triangular meshes defining the building geometry, and UV coordinates defining the texture mapping. For each mesh triangle, the triangle indices establish the correspondence between its 3D vertex coordinates and the associated UV texture coordinates. The UV coordinates are then converted into pixel coordinates in the texture atlas using Equation 1.

$$\mathbf{p}_i^A = (u_i W, (1 - v_i)H) \tag{1}$$

where (u_i, v_i) denotes the normalized UV coordinate of vertex i, and W and H denote the width and height of the texture atlas, respectively. Combining the resulting pixel coordinates with the triangle indices locates the texture region of each mesh triangle in the atlas, as shown in Figure 8.

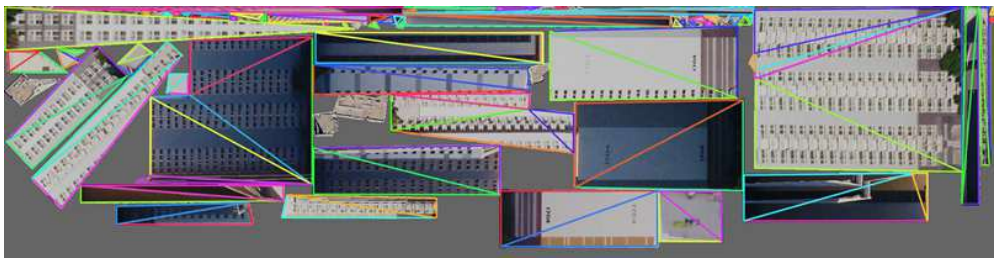


Fig. 8 Visualization of polygon mapping within the atlas image using parsed UV texture coordinates and face indices

As shown in Figure 8, the UV mapping identifies the location of each texture region on the building surface. The building geometry is then aligned using Equation 2 and projected onto the corresponding facade planes.

$$R_{\text{align}} = R_{\text{wall}} R_{\text{z-up}} R_{\text{PCA}} \tag{2}$$

where R_{PCA} is estimated from the area-weighted distribution of face normals so that the dominant wall planes drive the alignment, $R_{\text{z-up}}$ aligns the longest principal axis with the provisional vertical axis, and R_{wall} aligns the wall normals with the horizontal axes. Face normals are then computed from the aligned triangle vertices, and each triangle is assigned to an axis-aligned orientation class according to the dominant component of its normal vector. Only triangles assigned to the four vertical directions $(\pm X, \pm Y)$ are retained, whereas roof, ground, and strongly slanted surfaces are excluded from facade image generation. For each vertical orientation group, the aligned triangle vertices are orthographically projected onto the corresponding facade plane:

$$\mathbf{p}_i^F = \pi_d \left(R_{\text{align}} \mathbf{v}_i \right) \tag{3}$$

where $\mathbf{v}_i$ is a 3D vertex coordinate and π_d denotes the direction-specific projection that removes the depth axis parallel to the viewing direction. The projected coordinates are then normalized by the 2D bounding box of the orientation group and scaled to the facade image resolution, yielding pixel coordinates in the facade domain.

Within each orientation group, the triangles are sorted from far to near according to their mean depth and rasterized sequentially. This ordering allows foreground surfaces to overwrite background surfaces and prevents recessed or protruding elements from being projected in an unnatural arrangement. The resulting positions and rendering order of the atlas-mapped triangles form the projected polygon layout shown in Figure 9.

After rasterization, each facade image is reoriented so that the estimated building-up direction points upward and the facade is viewed from the exterior. This correction is performed using rotations in 90° increments and, when required, mirror flipping, with the same transformation applied to the corresponding coordinates used for inverse mapping. The generated layout information is subsequently used for inverse mapping, through which the stylized facade image is reconstructed in the original texture-atlas layout of the 3D building model.

3.2 Guidance generation

Once the facade image for stylization has been obtained, guidance information is generated for the subsequent stylization process. To preserve the building components and their spatial relationships during stylization, the principal facade components, namely walls and windows, are represented using a wall–window mask, while an edge map is generated to retain the overall building structure. These structural guidance signals are intended to support component-aware stylization rather than applying a uniform style to the entire building. In addition, a heatmap is generated from the facade image and the wall–window mask and is used as soft guidance to control the

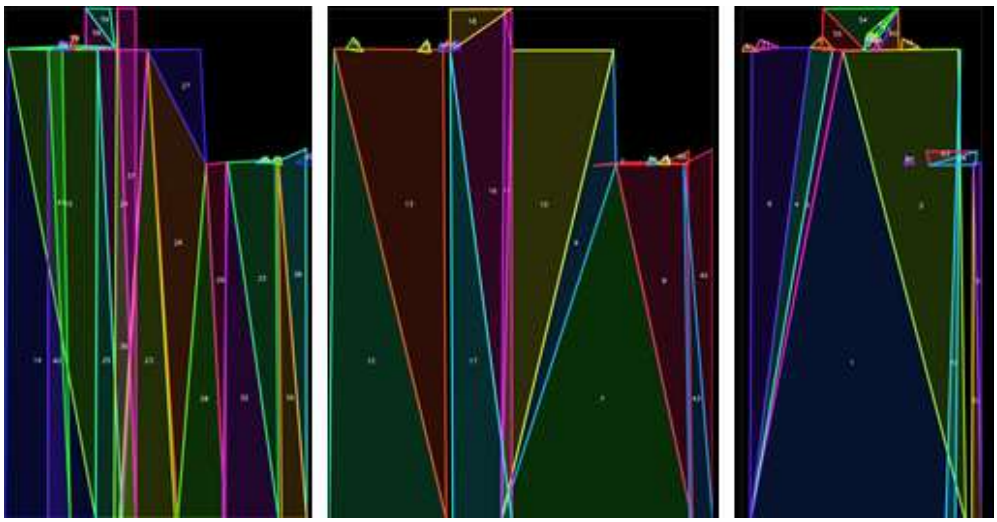


Fig. 9 Projected layout and rendering order of aligned 3D mesh triangles on the 2D facade planes

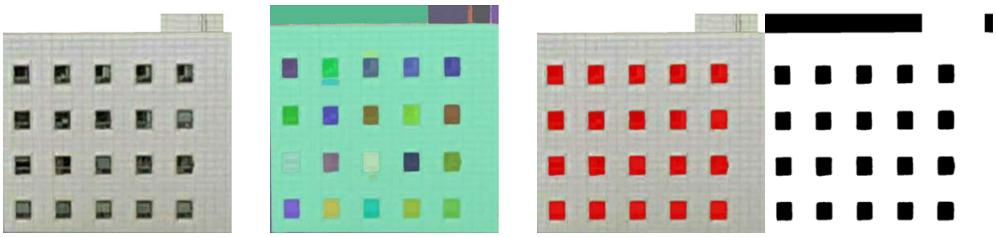
spatial location of the applied style. Accordingly, the guidance information is generated from two perspectives: content preservation and style control.

3.2.1 Structure guidance

A wall–window mask is generated to help the stylization network more clearly distinguish walls from windows. The mask is used to preserve the positions and shapes of these major facade components during stylization. SAM 2.1[39] is employed for mask generation because its point-grid prompting enables dense, class-agnostic mask extraction and is suitable for facades in which windows are repeatedly arranged in grid-like patterns. In the implementation, the SAM 2.1 Hiera-Small automatic mask generator is used with a 64×64 point grid, a predicted-IoU threshold of 0.7, a stability-score threshold of 0.85, and a minimum mask region area of 50 pixels.

Because SAM 2.1 generates class-agnostic masks, its candidate regions are further filtered to identify window regions. The building region is first estimated from grayscale pixels below 250 and refined using 5×5 morphological closing and opening. The resulting building region is initialized as the wall region. Among the SAM 2.1 masks that sufficiently overlap the building region, regions occupying between 0.01% and 15% of the facade image and having a mean grayscale intensity below 180 on a 0–255 scale are classified as windows and removed from the wall mask. Candidate masks that do not satisfy these window criteria do not alter the initialized building-region mask. Pixels outside the building region and detected window regions are therefore encoded as black, whereas the remaining building region is encoded as white. The use of size and intensity cues is motivated by previous facade and window extraction studies [40], whereas the specific thresholds used here are heuristic settings adopted for the present implementation. The same threshold values were applied to all facades without building-specific adjustment. As shown in Figure 10, the wall regions are

encoded in white, whereas the window and background regions are encoded in black. The resulting mask provides structural guidance containing the shapes of the major facade components and their spatial arrangement.



(a) (b) (c) (d)

Fig. 10 Wall–window mask generation using SAM2.1 and filtering: (a) input facade, (b) SAM2.1 class-agnostic masks, (c) filtered window candidates, and (d) final wall–window mask

Although the wall–window mask represents the component layout, it does not sufficiently retain detailed architectural shapes and patterns. Therefore, a line-art edge map is additionally extracted from the facade image to provide detailed structural guidance. In the implementation, the Realistic Lineart annotator provided by the ControlNet auxiliary pre-processing toolkit is used in coarse mode at a processing resolution of 768 pixels. Unlike a threshold-based edge detector, the resulting line-art map emphasizes architectural contours and detailed facade patterns. It is supplied to the T2I-Adapter as the LineArt structural condition, complementing the wall–window mask in preserving window boundaries, floor divisions, local facade patterns, and the overall building outline.

3.2.2 Fire heatmap guidance

To apply fire appearance stylization to building-model textures, we construct heatmap-based guidance by incorporating selected spatial characteristics of flames observed on building exteriors. The heatmap provides a controllable spatial region for diffusion-based stylization rather than allowing the appearance change to affect the facade globally. It is not generated from predictions of physical quantities such as temperature, heat release rate, or fire propagation. Instead, its spatial distribution reflects simplified tendencies of externally visible building fires, in which fire appearance extends predominantly upward from an opening near the ignition location and broadens laterally as the affected region develops.

The heatmap is constructed based on previous studies reporting that flames and smoke emitted through openings tend to rise upward and spread laterally with height [6, 15, 17, 41]. Fire simulations commonly represent changes in fire conditions over multiple progression levels. Inspired by this need for graded visualization, the proposed heatmap is also generated at multiple appearance-control levels rather than as physically defined fire stages. Candidate parameters affecting the extent and shape of the heatmap were selected from the spatial characteristics reported in previous studies

and evaluated through a parameter-isolation experiment. The results are presented in Figure 11.

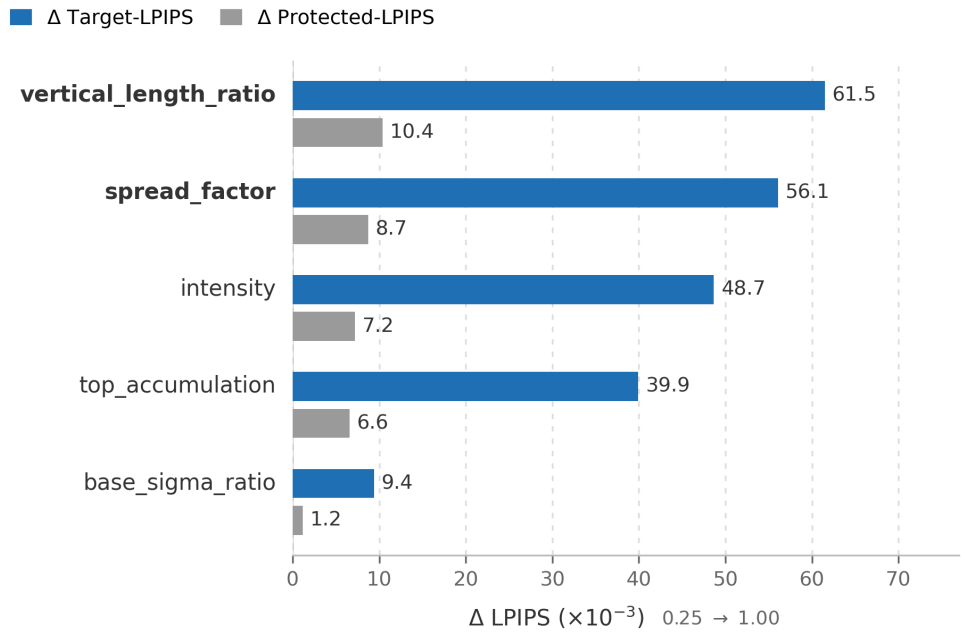


Fig. 11 Effects of individual heatmap parameters on Target-LPIPS and Protected-LPIPS

The parameter-isolation experiment for generating the fire-appearance heatmap was conducted by varying each candidate parameter sequentially across four levels while keeping the remaining parameters fixed at their reference settings. Target-LPIPS was used to quantify the magnitude of appearance change in the wall regions as each parameter was varied. The results showed that all candidate parameters affected the generated appearance to different degrees, and vertical_length_ratio and spread_factor were among the parameters that produced relatively large changes in the building textures. Both selected parameters also showed a clear monotonic increase in Target-LPIPS across the tested levels. Considering these experimental results together with the vertical and lateral spreading characteristics of externally visible fire, the two parameters were selected to control the spatial extent of the proposed heatmap as follows:

$$H(x,y) = \exp\left(-\frac{\Delta y}{r H_{\text{img}}}\right) \exp\left(-\frac{(x-x_0)^2}{2\,(\sigma_0 + r\Delta y)^2}\right), \qquad r \in \{0.25, 0.50, 0.75, 1.00\}. \tag{4}$$

In Equation 4, the reference position (x_0, y_0) can be specified by the user. When it is set automatically, the window closest to the horizontal center of the facade among those located in the lower third is first selected, and the center of its top edge is used as the reference position. Here, $\Delta y = y_0 - y \geq 0$ denotes the upward distance from the reference position, and H_{img} is the facade height. The initial width is defined as $\sigma_0 = 0.40 w_{\text{ref}}$, where w_{ref} is the bounding-box width of the detected window whose centroid is closest to (x_0, y_0). The first exponential term controls the vertical attenuation, whereas the second determines the horizontal distribution at each height. A short attenuated region is retained immediately below the reference position to avoid an abrupt spatial transition, and the resulting field is spatially smoothed before masking. As the normalized appearance scale r increases, both the effective vertical extent and the lateral spread increase. Pixels excluded by the wall–window mask, including detected window and background regions, are assigned zero change-map values so that these regions are preferentially preserved during editing.

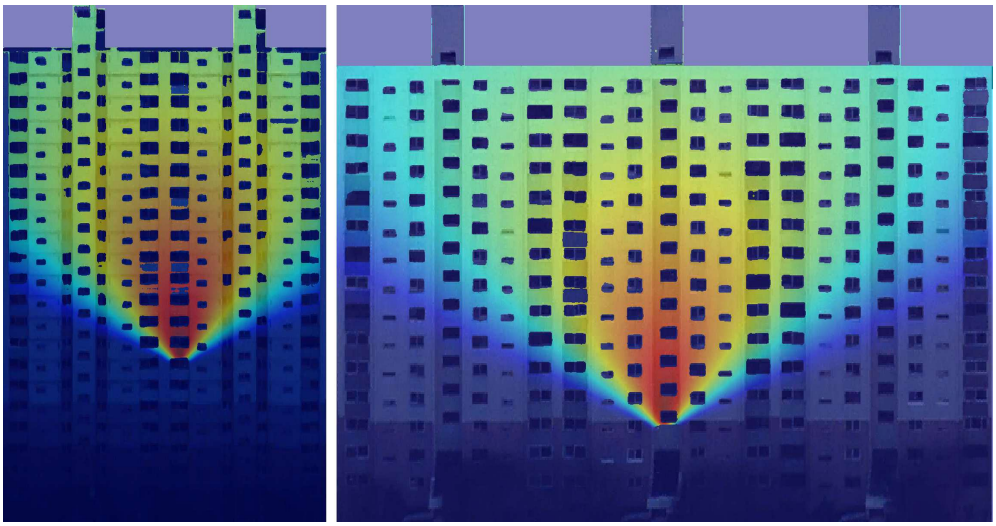


Fig. 12 Examples of generated fire-appearance heatmaps overlaid on building facade images

As shown in Figure 12, the resulting heatmap is spatially anchored to the selected reference window and extends upward and laterally across the facade. Its distribution adapts to the location and width of the reference opening as well as the dimensions of the facade image.

3.3 Stylization and atlas reconstruction

When conventional style-transfer methods are applied, facade components may be distorted or lost, as shown in Figure 3. Such changes disrupt the spatial correspondence between the architectural elements represented in the texture and the underlying geometry of the building model, making the resulting texture difficult to apply to the

original 3D asset [42, 43]. To address this problem, the guidance signals generated in Section 3.2 are provided to the corresponding stylization modules, allowing fire appearance to be applied to the facade while preserving its architectural components.

The stylization pipeline shown in Figure 13 applies fire appearance to the facade-aligned image using the generated structural and heatmap guidance signals.

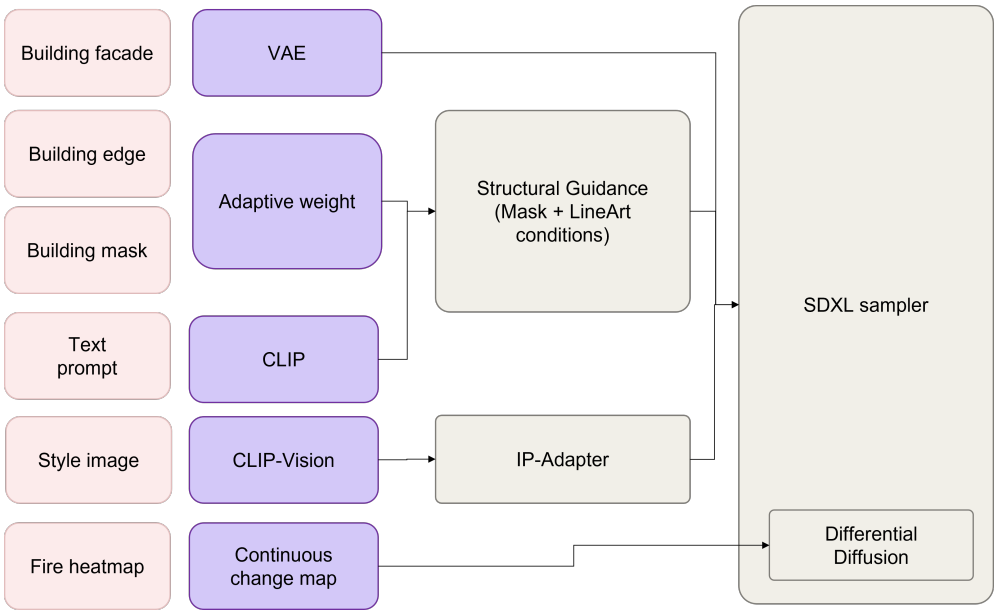


Fig. 13 Stylization flow combining structure-preserving conditions with heatmap-guided fire style injection

The pipeline consists of a diffusion block for image generation and conditioning blocks that provide the fire style and building-structure information. Stable Diffusion XL (SDXL) [44] is employed as the base diffusion model using the fine-tuned checkpoint [45]. SDXL supports high-resolution image generation and is therefore suitable for processing high-resolution building-facade images.

The fire appearance applied to the building-model texture is derived from the reference image provided to IP-Adapter [46]. IP-Adapter is used in its style-transfer weighting configuration so that the reference image provides appearance information without directly determining the spatial region to be edited. In the proposed framework, Differential Diffusion [47] is used to spatially control where and to what extent this appearance modifies the source facade. The heatmap generated in Section 3.2.2 is supplied to Differential Diffusion as a continuous change map whose values range from 0 to 1. Its spatial distribution determines the region in which editing is permitted, while its continuous values regulate the relative degree of modification across the facade. This separation enables the reference image to provide the target fire appearance while the heatmap controls its spatial location and relative editing strength.

The wall–window mask and line-art edge map generated in Section 3.2.1 are supplied as two structural conditions using an SDXL T2I-Adapter [48]. The same pretrained adapter is applied separately to the two conditioning images. The wall–window mask provides coarse component-level spatial information, whereas the line-art condition provides detailed architectural boundaries. Their conditioning strengths are adjusted for each facade according to the visual complexity of the corresponding edge map. To dynamically adjust the effects of the two guidance signals according to facade complexity, Shannon entropy is calculated from the edge map using Equation 5. The resulting value is then used to determine the weights of the two structural conditions according to Equation 6.

$$H_{\text{edge}} = -\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N} p_{k,i}\log_2(p_{k,i}) \tag{5}$$

$$\text{Strength} = \begin{cases} 1 - \dfrac{H_{\text{edge}}}{\log_2 N} & \text{(mask)} \\ \dfrac{H_{\text{edge}}}{\log_2 N} & \text{(edge)} \end{cases} \tag{6}$$

The edge complexity H_{edge} in Equation 5 is calculated by dividing the line-art map into non-overlapping local patches and computing the Shannon entropy of the grayscale-intensity distribution within each patch. The resulting local entropy values are then averaged over the facade. Here, K denotes the number of local patches, N denotes the number of grayscale intensity levels, and $p_{k,i}$ denotes the normalized frequency of intensity level i in the k-th patch. In the implementation, the line-art map is divided into 32×32 pixel patches and $N = 256$ intensity levels are used. The averaged entropy is normalized by its maximum possible value, $\log_2 N$, to obtain a value in the range $[0, 1]$.

Based on this value, complementary strengths are assigned to the two structural conditions according to Equation 6. A higher line-art entropy increases the contribution of the edge condition, whereas a lower entropy increases the contribution of the wall–window mask condition. The two strengths are complementary and sum to one for each facade.

The wall–window mask and line-art map are provided as structural conditions through the T2I-Adapter to support the preservation of the spatial composition and architectural structure of the source facade. In contrast, the reference image provides the visual characteristics of the target fire appearance through IP-Adapter. Differential Diffusion separately uses the generated heatmap as a continuous change map to control the spatial distribution and relative degree of editing. These conditions therefore serve complementary roles in the stylization process, with the structural guidance supporting facade preservation, the reference image providing the target appearance, and the heatmap controlling where and to what extent the source facade is modified.

The facade images produced by the stylization pipeline in Figure 13 must be transferred back to their corresponding regions in the original texture atlas before they can be used by the 3D building model. Using the facade–atlas correspondence established in Section 3.1 through Equation 1 and Equation 3, the stylized facade images are mapped back to the atlas by the per-triangle affine transform given in Equation 7.

$$M_t \begin{bmatrix} \mathbf{p}_{t,i}^F \\ 1 \end{bmatrix} = \mathbf{p}_{t,i}^A, \qquad i = 1, 2, 3 \tag{7}$$

where $\mathbf{p}_{t,i}^F$ and $\mathbf{p}_{t,i}^A$ denote the facade and atlas pixel coordinates of the i-th vertex of triangle t, respectively, and $M_t \in \mathbb{R}^{2\times 3}$ denotes the facade-to-atlas affine transform of that triangle. Because three non-collinear point correspondences uniquely determine a two-dimensional affine transform, M_t is fully determined by the three vertex pairs of each retained triangle.

Before inverse warping, empty pixels surrounding the valid facade regions are filled from their nearest valid neighbors to avoid sampling background values at triangle boundaries. Using M_t, each retained facade triangle is then warped into its corresponding atlas triangle using nearest-neighbor sampling, including the boundary pixels.

Because the reconstructed atlas retains the same resolution and UV layout as the source texture, it can replace the texture referenced by the existing material without modifying the source mesh geometry, topology, or UV coordinates. The stylized fire appearance is therefore rendered by the existing 3D building models and the tile-based visualization pipeline without additional geometric processing. The overall reconstruction process is illustrated in Figure 14.

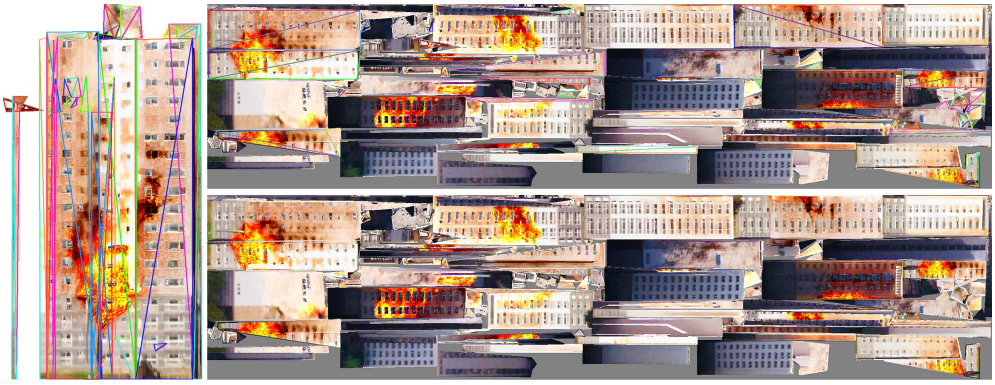


Fig. 14 Atlas reconstruction. The retained triangles of the stylized facade (left) are affinely warped to their corresponding regions in the texture atlas (top right, matching color), and the warped regions are composited to produce the reconstructed atlas (bottom right)

4 Experimental

This section evaluates the proposed fire-appearance stylization pipeline on facade textures extracted from textured 3D building models of a real-world city-model service. The experiments examine whether fire-related appearance can be applied to facade textures while preserving the architectural components and spatial correspondence of the source building assets.

The evaluation addresses the trade-off between the magnitude of appearance modification in the editable wall regions and the preservation of the non-wall facade content, together with the extent to which this trade-off can be explicitly controlled. Building textures are spatially aligned with the geometry and UV mapping of their corresponding 3D models. Therefore, unintended changes to the shapes or positions of windows, walls, and other architectural elements may reduce the applicability of the stylized textures to the original building assets [49, 50]. Accordingly, the experiments measure the perceptual change in the editable wall regions, the preservation of the complementary non-wall regions, and the retention of the overall facade structure.

The experiments are organized to evaluate four aspects of the proposed framework. First, controlled experiments examine the effects of the heatmap design and the structural guidance components. Second, the controllability of the fire appearance is evaluated with respect to appearance scale and spatial location, together with its robustness to changes in generation conditions. Third, the proposed method is compared with representative diffusion-based stylization methods using identical upright facade inputs. Finally, the practical applicability of the complete framework is evaluated through texture reconstruction and reapplication to the original 3D building assets.

4.1 Experimental settings

We used 3D building models provided by the Republic of Korea's V-World platform [24] as a practical source of real building models with atlas textures and geometry. As an operational national geospatial platform, V-World provides a realistic data source for evaluating the proposed method under practical usage conditions. Among the building models provided by V-World, commercial, residential, and mixed-use buildings located in densely built-up urban areas were selected as the experimental dataset. These building-use categories are commonly represented in urban fire-simulation studies. In addition, based on previous reports that 95% of buildings in urban areas are lower than 100 m [25], buildings below 100 m were sampled to reflect the typical requirements of city-scale disaster simulations.

A total of 428 building models were used in the experiments. For each model, up to four upright facade images were extracted depending on the number of valid principal exterior sides, resulting in 1,435 facade images in total. All available valid facade views were used rather than selecting a single representative view for each building. Figure 15 shows examples of facade images generated from the texture atlases of the selected building models.

A single disaster scenario, building fire, was used throughout the experiments. For the main comparison, the target appearance was defined using one reference image of a burning building facade together with a text prompt shared by all methods. The shared text prompt was *Realistic building facade with fire, flames, and smoke. Preserve the original building structure, facade layout, window positions, proportions, and perspective with no deformation.* The prompt combines a scenario-specific description of the appearance to be synthesized with a common constraint that encourages preservation of architectural components such as window positions, proportions, and



Fig. 15 Examples of facade images generated from the texture atlas of V-World building models

perspective. Additional prompt and reference-image variations were used only in the controllability analysis described in Section 4.2.4.

The proposed pipeline is based on SDXL using the fine-tuned checkpoint and uses IP-Adapter for reference-image appearance conditioning. The wall–window mask and

line-art map are provided as structural conditions through an SDXL T2I-Adapter, with their relative strengths adaptively determined for each facade. The generated fire heatmap is supplied to Differential Diffusion as a continuous change map for spatial edit control. Sampling was performed using DPM++ 2M Karras with 35 steps and a classifier-free guidance scale of 7, and the IP-Adapter weight was set to 1.0.

We compare the proposed method with four representative diffusion-based stylization methods. ZeST [51] is a single-image zero-shot material transfer method, and InstantStyle [52] is a style-preserving generation method. SANA [53] is a linear diffusion transformer for high-resolution synthesis, and FLUX is a FLUX.1-based [54] style-transfer workflow combining FLUX.1 Depth [dev] and FLUX.1 Redux [dev]. All methods received the same source facades, style reference image, and positive text prompt. The baseline methods were reproduced using their publicly available implementations or workflows and the pre-processing and hyperparameter settings recommended for each method. All generated results were evaluated at the original upright-facade resolution using the same source-derived masks. The comparison evaluates the end-to-end suitability of each method for the target facade-stylization task, while the contribution of the structural and heatmap guidance used only by the proposed framework is examined separately through the ablation study.

Since our method changes only the facade texture appearance and keeps the building geometry fixed, geometry-based 3D mesh quality metrics are less informative in our setting. Instead, we evaluate how well the method preserves structural content while introducing the target fire appearance. This view is consistent with image-based evaluation practices for 3D reconstruction and rendering, where perceptual appearance is assessed from rendered views using image-based metrics such as LPIPS [55, 56].

We quantitatively evaluate the fire-stylized facades using five metrics, namely LPIPS [57], CLIP-score [58], Edge-SSIM [59], MUSIQ [60], and CLIP-IQA [61]. LPIPS has also been employed to evaluate perceptual similarity in diffusion-generated design images [62]. In this study, LPIPS is reported separately for two facade regions, as described below.

The proposed method is designed to modify the wall regions while preserving the remaining facade content, and a single global perceptual distance cannot distinguish an intended appearance change from an unintended alteration of non-target regions. We therefore use the wall–window mask to define two complementary regions for LPIPS evaluation. Target-LPIPS is evaluated over the wall regions, which are treated as the editable regions for fire-appearance stylization, and quantifies the magnitude of perceptual change relative to the original facade. A larger Target-LPIPS indicates a greater appearance change in the wall regions, but does not by itself indicate better stylization performance. Protected-LPIPS is evaluated over all complementary non-wall regions, including detected windows and background regions, which are not intended to receive the fire appearance. A lower Protected-LPIPS therefore indicates better preservation of the non-target facade content and layout. The same source-derived wall and non-wall regions are used for all compared methods. The two LPIPS measures are interpreted independently because they are computed over different regions.

Edge-SSIM is computed between the Canny edge maps of the original and the fire-stylized facade and evaluates the preservation of the overall architectural structure. CLIP-score measures the semantic consistency between the stylized facade image and the fire-related text prompt, thereby indicating how faithfully the target appearance has been introduced. MUSIQ and CLIP-IQA are no-reference image-quality assessment metrics that require no corresponding ground-truth image [63]. MUSIQ is a multi-scale transformer trained to predict human mean opinion scores, whereas CLIP-IQA estimates quality from the relative similarity between the image and a pair of antonym prompts. Such no-reference measures are appropriate for stylization tasks in which ground-truth fire-stylized facades are difficult to obtain [64], and comparable perceptual-quality measures have been employed in urban and geospatial image-evaluation contexts [65, 66].

These metrics quantify objectives that are not mutually reinforcing, because diffusion-based stylization also introduces changes in regions that are intended to remain unchanged, so the degree of style introduction can conflict with the preservation of the architectural structure. A ranking based on a single metric therefore reflects only one of these objectives and does not provide a balanced assessment. To account for all of them at once, we employ TOPSIS [67, 68], a distance-based multi-criteria decision-making method that ranks each alternative according to its proximity to an ideal combination of criterion values and its distance from the corresponding worst-performing combination. Previous studies have used TOPSIS in the same manner to comprehensively evaluate style-transferred and generated images [69, 70].

For the TOPSIS analysis, we use CLIP-score and MUSIQ to represent prompt alignment and perceptual image quality, respectively, while Edge-SSIM and Protected-LPIPS represent the preservation of the architectural structure and of the protected facade regions. Target-LPIPS is excluded because it measures the magnitude of an intended appearance change rather than an attribute for which a larger value is always preferable. CLIP-IQA is excluded because its CLIP-based representation partially overlaps with CLIP-score, and using two general image-quality criteria would give perceptual quality a disproportionate influence on the overall ranking. All four criteria are assigned equal weight, and the resulting ranking is reported under four normalization schemes to assess its sensitivity to the choice of normalization.

4.2 Experimental results

4.2.1 Ablation study results

The ablation study evaluates the contributions of the spatial and structural guidance components using the full set of 1,435 upright facades.Here, S denotes the SDXL-based stylization backbone with IP-Adapter, H denotes heatmap-guided spatial control using Differential Diffusion, M denotes the wall–window mask condition, E denotes the LineArt structural condition, and A denotes entropy-adaptive weighting of the two structural conditions. For the single structural-condition variants, the corresponding conditioning strength is set to 1.0. For S+H+M+E, fixed mask and LineArt strengths are applied to all facades using the dataset-level mean of the adaptive weights obtained from Equations 5 and 6. In S+H+M+E+A, the two strengths are instead determined

individually for each facade. Configurations containing H use the heatmap at $r = 1.00$, whereas configurations without H use a uniform change map so that editing is permitted over the entire facade. The corresponding ablation results are reported in Table 1.

Table 1 Ablation study of the proposed fire-appearance stylization pipeline

Configuration	CLIP-score ↑	Target-LPIPS	Protected-LPIPS ↓	Edge-SSIM ↑
S (baseline)	0.3331	0.5683	0.4014	0.4403
S + H	0.3000	0.3246	0.0879	0.6549
S + H + M	0.2786	0.3052	0.0848	0.7010
S + H + E	0.2911	0.2891	0.0811	0.6798
S + H + M + E (fixed)	0.2866	0.2931	0.0817	0.6995
S + M + E + A	0.3030	0.5767	0.4263	0.5440
S + H + M + E + A (Ours)	0.2864	0.2925	0.0816	0.6999

Note: All rows are absolute metric values. **S**: style backbone of SDXL with the IP-Adapter, **M**: wall–window mask structural condition, **E**: LineArt structural condition, **H**: fire heatmap used as the Differential Diffusion change map, and **A**: entropy-adaptive weighting of the two structural conditions.

The ablation results show a clear effect of the heatmap guidance. Compared with S, S+H substantially reduced both Target-LPIPS and Protected-LPIPS and increased Edge-SSIM, although its CLIP-score decreased. This indicates that, unlike diffusion-based stylization that can introduce appearance changes across the entire image, the heatmap guidance restricts the spatial extent of the fire appearance and suppresses unintended changes outside the intended region. In particular, the substantial improvements in Protected-LPIPS and Edge-SSIM, despite the decrease in CLIP-score, show that the heatmap contributes to preserving facade components by limiting global style modification. Adding either M or E to S+H further improved Edge-SSIM. This suggests that the mask and edge conditions contribute to preserving the overall facade layout and detailed architectural components, respectively.

For S+H+M+E (fixed), the weights of M and E were fixed to 0.6987 and 0.3013, respectively, using the dataset-level mean, with their sum constrained to one. The quantitative differences between this configuration and Ours were very small. This comparison indicates that the role of A is not primarily to improve the dataset-level mean performance, but to determine the relative strengths of M and E independently for each facade according to its structural complexity, without relying on a single dataset-level fixed ratio. Compared with the configurations using only M or E, S+H+M+E (fixed) produced values that generally remained between those of the two single-condition configurations, including Edge-SSIM, Target-LPIPS, Protected-LPIPS, and CLIP-score. These results suggest that combining M and E provides a balance between the two structural conditions rather than a consistent improvement over either condition alone. As shown in Figure. 16, the qualitative results also illustrate the spatial control provided by the heatmap guidance and the preservation of architectural components supported by M and E within the edited regions.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 16 Qualitative comparison of the ablation configurations: (a) input facade and style reference, (b) S, (c) S+H, (d) S+H+M, (e) S+H+E, (f) S+H+M+E with fixed weights, (g) S+M+E+A, and (h) S+H+M+E+A (Ours)

The spatial-control effect of the heatmap is further illustrated by comparing S+M+E+A with Ours, S+H+M+E+A, in Figure. 16. Without the heatmap, S+M+E+A produced higher CLIP-score and Target-LPIPS values than Ours, indicating more extensive appearance modification across the facade. In contrast, the fire appearance generated by Ours is spatially concentrated according to the distribution and continuous values of the heatmap rather than being applied uniformly across the facade. This comparison shows that H serves as the primary spatial guidance for controlling where and to what extent the fire appearance is introduced. The substantially lower Protected-LPIPS and higher Edge-SSIM of Ours further show that this spatial control also contributes to the preservation of non-wall regions and facade structure.

4.2.2 Appearance-scale results

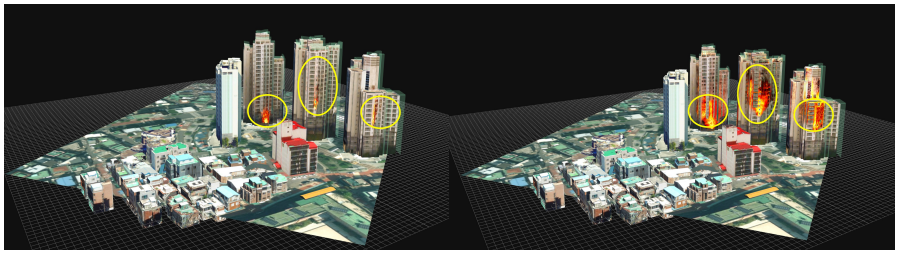
The proposed pipeline applies fire appearance to building facades by using a heatmap as soft guidance. As described in Section 3.2.2, the normalized appearance-scale

parameter r is used as a scalar that controls the extent of the heatmap. In the experiments, four values of r were evaluated, namely $r \in \{0.25, 0.50, 0.75, 1.00\}$, to produce progressively broader fire appearances. The corresponding results are presented in Table 2.

Table 2 Facade-level metrics as the normalized appearance scale r increases

r	CLIP-score↑	Target-LPIPS	Protected-LPIPS↓	Edge-SSIM↑
0.25	0.2532	0.1708	0.0669	0.7786
0.50	0.2723	0.2222	0.0729	0.7537
0.75	0.2811	0.2604	0.0776	0.7271
1.00	0.2864	0.2925	0.0816	0.6999

Table 2 shows that Target-LPIPS increased from 0.1708 to 0.2925 as r increased from 0.25 to 1.00. CLIP-score also increased with r, whereas Protected-LPIPS increased and Edge-SSIM decreased. These results indicate that larger values of r produce progressively greater wall-region appearance modification, accompanied by a trade-off in non-wall and structural preservation.



(a) (b)

Fig. 17 Appearance-scale control on textured 3D building models at (a) $r = 0.25$ and (b) $r = 1.00$

Because the stylized facade textures are reconstructed into the original atlas and reapplied through the existing UV mapping (Section 3.3), the controllable appearance is observed directly on the 3D building assets rather than only on isolated facade images. As highlighted by the yellow circles in Figure 17, the fire appearance at $r = 0.25$ is confined to localized regions of the target facades, whereas at $r = 1.00$ it extends across larger portions of the buildings. Despite this difference in spatial extent, the surrounding buildings and the structural layouts of the target facades, including the window rows, floor divisions, and building silhouettes, remain visually consistent.

This explicit control over the spatial extent of the fire appearance is not provided by the compared stylization methods under the evaluated settings, as these methods generate a single fixed result for each input condition. Having characterized this controllable behavior, we next compare the proposed method with representative

diffusion-based methods. For this comparison, the proposed method is evaluated at its maximum scale level ($r = 1.00$).

4.2.3 Comparison with representative methods

Table 3 reports the quantitative comparison on the common facade set. For the proposed method, we report the maximum scale level ($r = 1.00$), at which the fire appearance covers the largest facade area. The remaining scale levels are analyzed separately in Section 4.2.2.

Table 3 Quantitative comparison with representative diffusion-based stylization methods on the common facade set. The proposed method is reported at the maximum scale level ($r = 1.00$)

Method	CLIP-score↑	CLIP-IQA↑	MUSIQ↑	Target-LPIPS	Protected-LPIPS↓	Edge-SSIM↑
ZeST	**0.3159**	**0.4856**	**56.05**	0.5429	0.3835	0.4632
InstantStyle	0.2834	0.2983	50.36	0.3762	<u>0.2003</u>	0.6180
SANA	<u>0.3076</u>	0.4260	38.61	0.4686	0.3026	<u>0.6945</u>
FLUX	0.2969	<u>0.4751</u>	45.37	0.6114	0.4231	0.5846
Ours	0.2864	0.2970	<u>53.79</u>	0.2925	**0.0816**	**0.6999**

Table 3 shows different trade-offs among the compared methods. The proposed method achieved the lowest Protected-LPIPS of 0.0816 and the highest Edge-SSIM of 0.6999. ZeST achieved the highest CLIP-score, MUSIQ, and CLIP-IQA, whereas SANA obtained the second-highest Edge-SSIM but the lowest MUSIQ. These results show that appearance-related and preservation-related criteria do not necessarily improve simultaneously. Target-LPIPS is interpreted separately because it measures wall-region modification rather than stylization quality. The proposed method obtained the lowest Target-LPIPS of 0.2925, indicating less extensive modification over the entire wall region than the compared methods. This is consistent with the heatmap-guided design, which spatially restricts the appearance modification rather than applying the style globally across the facade.

As shown in Table 4, the proposed method ranked first among the five compared methods under all four TOPSIS normalization procedures, with C^* ranging from 0.644 to 0.924. The vector, MAX, and SUM procedures produced the same ordering, whereas MAX-MIN changed the ordering of the intermediate methods. The top-ranked position of the proposed method remained unchanged across the tested normalization procedures.

Taken together, the quantitative results indicate that the proposed method provides a distinct trade-off between appearance modification and preservation of the source facade. Although it does not maximize the appearance-oriented metrics, it achieves the lowest Protected-LPIPS, the highest Edge-SSIM, the second-highest MUSIQ, and the highest TOPSIS closeness coefficient among the five compared methods. This balance is important for textured 3D building models because the stylized appearance is practically useful only when the spatial arrangement and visual identity of the architectural components remain sufficiently consistent with the source asset.

Table 4 TOPSIS closeness coefficient under four normalization procedures

	Vector[a]		MAX[b]		SUM[c]		MAX-MIN[d]	
Method	C^*	Rank	C^*	Rank	C^*	Rank	C^*	Rank
ZeST	0.270	4	0.301	4	0.257	4	0.515	2
InstantStyle	0.647	2	0.644	2	0.648	2	0.496	4
SANA	0.405	3	0.421	3	0.400	3	0.512	3
FLUX	0.174	5	0.203	5	0.163	5	0.354	5
Ours	**0.918**	**1**	**0.900**	**1**	**0.924**	**1**	**0.644**	**1**

Note: $C^* = D^-/(D^+ + D^-)$, so a larger value indicates a shorter distance to the positive-ideal solution and a longer distance to the negative-ideal one. The criteria are CLIP-score, Protected-LPIPS, Edge-SSIM, and MUSIQ with equal weights, and Protected-LPIPS is treated as a cost criterion. In every procedure x_{ij} denotes the value of method i on criterion j, and $x_j^{\max}$ and $x_j^{\min}$ denote the largest and smallest values observed on that criterion.

[a] Vector normalization, $r_{ij} = x_{ij}/\sqrt{\sum_i x_{ij}^2}$, is the classic TOPSIS formulation and is used as the primary analysis.

[b] MAX normalization, $r_{ij} = x_{ij}/x_j^{\max}$.

[c] SUM normalization, $r_{ij} = x_{ij}/\sum_i x_{ij}$.

[d] MAX-MIN normalization, $r_{ij} = (x_{ij} - x_j^{\min})/(x_j^{\max} - x_j^{\min})$.

Figure 18 presents a qualitative comparison after the stylized facade textures are reconstructed into the original atlases and reapplied to the 3D building models. The yellow dashed boxes indicate regions of interest, with adjacent magnified views provided for closer inspection.

SANA, FLUX, and ZeST applied the fire appearance broadly across the building facade, while visible changes in architectural components were also observed. ZeST showed the lowest Edge-SSIM, and the enlarged regions reveal newly generated windows and patterns on surfaces that originally contained no such architectural components. FLUX showed the largest change in the protected regions, whereas SANA achieved the second-highest Edge-SSIM but the lowest MUSIQ. This indicates that structural similarity alone does not fully describe the visual quality of the stylized facade. InstantStyle retained the overall facade appearance relatively well and achieved the second-lowest Protected-LPIPS. In the qualitative examples, however, its fire appearance is visually limited, while newly generated patterns can still be observed on some otherwise flat surfaces.

Failure modes that the quantitative metrics do not capture were also observed. On flat surfaces without distinct architectural components, the compared methods created openings, windows, and distorted patterns that do not exist in the source. Edge-SSIM evaluates the structural similarity between edge maps but may not fully distinguish source-consistent architectural boundaries from newly generated patterns when their overall edge distributions remain similar. Consequently, some facade hallucinations may not be sufficiently reflected in this metric. For the objective of this study, in which the textures have to be reapplied to the original building models, this type of error directly undermines the practical usability of the result.

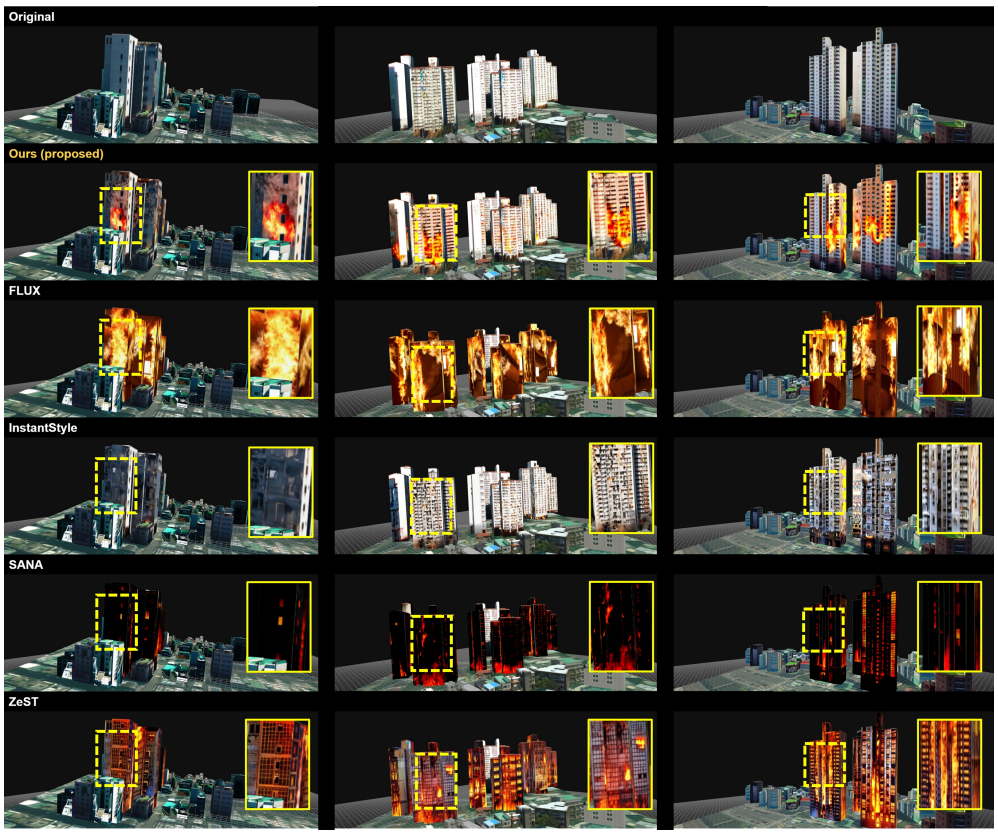


Fig. 18 Qualitative comparison of reconstructed 3D building models. Rows from top to bottom show the original models, Ours, FLUX, InstantStyle, SANA, and ZeST. Yellow dashed boxes indicate regions of interest, with adjacent magnified views

The proposed method, in contrast, specifies the spatial distribution of the fire through heatmap-based soft guidance and constrains the architectural structure through the mask and edge information. Inside the highlighted regions, the fire appearance follows the wall surfaces while the window arrangement and facade patterns remain visually identifiable. This observation is consistent with the quantitative preservation results, in which the proposed method achieved the lowest Protected-LPIPS and the highest Edge-SSIM. A relatively uniform color shift with respect to the source facade is also observed in some results. Nevertheless, the major architectural components and their spatial arrangement remain visually distinguishable.

4.2.4 Controllability and practical applicability

The atlas textures generated by the proposed framework should be applicable to the original building models without modifying their geometry. We therefore evaluated whether the spatial correspondence between geometry and texture was preserved

after converting the 3D building textures into 2D facade images for stylization and reconstructing them back into atlas textures for reapplication to the 3D models. We also examined whether the generated atlas textures could be successfully reapplied and rendered on the original 3D models after fire-appearance stylization. In addition, processing time was measured to assess the practical usability of the proposed framework for building-model texture transformation. Similar evaluation criteria have been reported in previous studies on building models and 3D texture editing [71–73], and our evaluation was conducted with reference to these studies.

We further evaluated the controllability of the proposed method and its practical applicability to 3D building models. To examine controllability at the individual-facade level, controlled experiments were conducted on 60 buildings using two random seeds. When the appearance scale was varied across four levels, Target-LPIPS increased monotonically with the scale for all facade–seed sequences under the primary prompt and reference condition. The consistency of scale control was further evaluated using three prompt conditions with the primary reference image, together with an additional test using a second reference image under the default prompt. Although the amount of generated fire appearance varied depending on the reference image, Target-LPIPS consistently increased from the lower to the higher scale under all evaluated conditions. Spatial controllability was evaluated by changing the heatmap reference position among the left, center, and right regions of the facade. The centroid of detectable flame regions was assigned to the nearest target position, with cases without detectable flames counted as failures. Under this criterion, the proposed method achieved a location hit rate of 94.4%.

The atlas textures generated by the proposed framework should also be reusable on the original 3D building models without modifying the building geometry. We therefore evaluated whether the spatial correspondence between the original geometry and texture was preserved after 2D facade conversion, stylization, and atlas reconstruction. For 20 randomly selected building models, approximately 10.43 million texture pixels were evaluated. The median coordinate error was 0 pixels and the mean error was 0.0769 pixels, while 99.99% of the evaluated texels were reconstructed within 1 pixel of their corresponding positions in the source atlas. These results indicate that the spatial correspondence required to reapply the reconstructed textures to the original 3D building models was preserved after facade rectification and atlas reconstruction. The reconstructed textures were also successfully reapplied and rendered for all 20 selected building models.

Processing time was measured on an Ubuntu 22.04 system equipped with a single NVIDIA GeForce RTX 4090 GPU and an AMD EPYC 9124 CPU. The initial model-loading warm-up required 34.8 s. After initialization, wall–window mask generation required an average of 4.16 s per facade, while diffusion-based fire-appearance generation required an average of 17.24 s per facade. Based on these stage-wise measurements, the complete pipeline required approximately 81.9 s per building model on average, excluding the one-time model initialization. Compared with the execution times of tens of seconds to several minutes per model reported for existing diffusion-based 3D texture-generation methods, this processing time falls within a practical range [74–76]. The proposed framework is therefore more suitable for offline texture

transformation and disaster-scenario generation of existing 3D building assets than for real-time transformation.

5 Discussion and conclusions

To visualize fire-related appearance on a city model, the fire appearance has to be applied to the building model textures while the architectural components are preserved. In practice, however, style transfer involves a trade-off between preserving the content of the image and the strength with which the style is introduced, and because building model textures require the spatial layout to remain aligned, balanced performance matters more than a strong result on either side alone. If this is not taken into account, the architectural components may become substantially distorted, causing a mismatch between the texture content and the underlying building geometry that severely reduces practical usability and makes the texture difficult to reuse in the original building model, as shown in Figure 3 of Section 1. As shown in the comparison with the other methods in Section 4.2.3, style transfer methods tend to emphasize one side, either by applying the style strongly or by preserving the components of the image to a large degree. In that case it becomes difficult to apply not only the fire appearance but also any other appearance to a building model.

In Section 4.2.3 no method obtained good values on all of the quantitative metrics such as Edge-SSIM, LPIPS, and CLIP-score, and Strong performance on an individual metric did not always correspond to practical reuse, as the qualitative evaluation revealed results whose practical reuse in existing building models was severely limited because the qualitative evaluation revealed distortions in which patterns that do not exist in the source were created, cases in which the fire appearance was hardly introduced, and cases in which the architectural components themselves disappeared. These observations indicate that no single quantitative metric fully captures the structural and perceptual failure modes observed in the qualitative comparison. In contrast, the proposed method combines component-aware structural guidance with heatmap-based spatial control, which helps balance appearance adaptation and preservation of the source facade. The full-dataset ablation further shows that heatmap guidance provides the largest improvement in protected-region and structural preservation, while the mask and edge conditions provide additional structural refinement. The comparison results show that the proposed method achieves strong preservation performance without maximizing the appearance-oriented metrics. This trade-off is also reflected in its highest TOPSIS ranking across all four normalization procedures.

This shows that the fire appearance can be applied to building model textures while the architectural components are preserved, so that the result can be reapplied to the original model. The proposed pipeline could potentially be extended to other appearance-transformation scenarios by replacing the style reference image and the corresponding spatial guidance, although the validation in this study was limited to fire appearance. Although the characteristics of fire were taken into account, they were used only as the soft guidance of the heatmap. The proposed method therefore does not reproduce fire on building textures under rigorous physical considerations. Fire disaster simulations present results driven by physical factors such as the internal

or the external spread of smoke for evacuation, whereas applying fire directly to the exterior of building models in a large-scale city model would require far more complex factors, including the weather, the material of the building, the cause of the fire, and the contents stored inside the building.

A further limitation concerns the range of facade appearances and datasets evaluated in this study. The current wall–window separation procedure may be less reliable for curtain-wall or fully glazed buildings with atypical window patterns. In addition, the quantitative evaluation was conducted using building models from a single city-model data source. However, the proposed framework is not specific to that source and can be applied to textured building models with compatible mesh geometry, UV coordinates, and texture information. Evaluation on additional facade types and datasets remains a direction for future work.

The demand for transforming the textures of building models in disaster simulations is expected to continue, and the proposed method needs to be developed further for the collection and reproduction of training data in simulation settings that are difficult to reproduce or for which little data is available. We therefore intend to extend this work beyond disasters to a range of urban phenomena, to style transfer that reflects the material and properties of buildings rather than heatmap guidance alone, and to the direct modification of the appearance of building models in an urban digital twin according to real-time conditions at the application level.

Declarations

Funding. This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (26ZR1110, Research on Intelligent Industrial Convergence).

Clinical trial number. Not applicable.

Ethics, Consent to Participate, and Consent to Publish declarations. Not applicable.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (26ZR1110, Research on Intelligent Industrial Convergence).

Data availability. The 3D building models and texture atlases analyzed in this study were obtained from the V-World platform operated by the Ministry of Land, Infrastructure and Transport of the Republic of Korea, and are available at https://www.data.go.kr/en/data/3073144/openapi.do under the terms of that service. Access to the original data is subject to the availability and terms of the V-World service.

The implementation of the proposed framework is available at https://github.com/IngIeoAndSpare/3D-Building-Heatmap-Guide-for-Specifying-Fire-Styles. To facilitate testing of the released implementation, the repository also provides public example

building models prepared from StyleCity3D scenes. These example models are provided only as demonstration inputs and were not used in the quantitative evaluation reported in this study.

References

[1] Büyükdemircioğlu, M., Oude Elberink, S.: Generating photorealistic urban digital twins with occlusion-aware texture mapping. International Journal of Applied Earth Observation and Geoinformation **145**, 105002 (2025) https://doi.org/10.1016/j.jag.2025.105002

[2] Jeong, T., Jeong, D., Kim, M.: Development of an urban digital twin based on geospatial data: A case study of Busan, South Korea. ISPRS International Journal of Geo-Information **15**(6), 247 (2026) https://doi.org/10.3390/ijgi15060247

[3] Luo, J., Liu, P., Kong, X., Shen, J., Wu, Q., Xu, D.: Urban digital twins for citizen-centric planning: A systematic review of built environment perception and public participation. International Journal of Applied Earth Observation and Geoinformation **143**, 104746 (2025) https://doi.org/10.1016/j.jag.2025.104746

[4] Campagna, L.M., Carlucci, F., Fiorito, F., Marinelli, E.R., Ottomanelli, M., Marinelli, M.: Mapping the integration of urban air mobility into the built environment: A bibliometric analysis and a scoping review. Drones **9**(10), 692 (2025) https://doi.org/10.3390/drones9100692

[5] Węgrzyński, W., Krajewski, G., Kimbar, G., Lipecki, T.: Fire smoke dispersion inside and outside of a warehouse building in moderate and strong wind conditions. Fire Safety Journal **136**, 103760 (2023) https://doi.org/10.1016/j.firesaf.2023.103760

[6] Sun, X., Tang, F., Lu, K., Ren, F., Shi, C., Merci, B., Hu, L.: Fundamentals of window-ejected fire plumes from under-ventilated compartment fires: Recent progresses and perspectives. Progress in Energy and Combustion Science **94**, 101039 (2023) https://doi.org/10.1016/j.pecs.2022.101039

[7] Zhongda, C.: Ai based data-driven point cloud optimization strategy in cloud computing framework for improving rendering performance of virtual city scenes. Discover Computing **28**(1), 146 (2025)

[8] Bhoi, A., Bhuyan, H.B., Nayak, R.P., Balabantaray, R.C., Pattanaik, A., Chinmay, A.: Smart crisis response leveraging social media content for effective disaster management. Discover Computing **28**(1), 263 (2025)

[9] Lee, M., Nam, H., Jun, C.: Multiple exits evacuation algorithm for real-time evacuation guidance. Spatial Information Research **25**(2), 261–270 (2017) https://doi.org/10.1007/s41324-017-0090-x

[10] Sulistyah, U.D., Hong, J.-H.: The use of 3D building data for disaster management: A 3D SDI perspective. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **42**, 395–402 (2019) https://doi.org/10.5194/isprs-archives-xlii-3-w8-395-2019

[11] Kanangkaew, S., Jokkaw, N., Tongthong, T.: A real-time fire evacuation system based on the integration of building information modeling and augmented reality. Journal of Building Engineering **67**, 105883 (2023) https://doi.org/10.1016/j.jobe.2023.105883

[12] Ding, Z., Xu, S., Xie, X., Zheng, K., Wang, D., Fan, J., Li, H., Liao, L.: A building information modeling-based fire emergency evacuation simulation system for large infrastructures. Reliability Engineering & System Safety **244**, 109917 (2024) https://doi.org/10.1016/j.ress.2023.109917

[13] Lewis, R.H., Jiao, J., Seong, K., Farahi, A., Navratil, P., Casebeer, N., Niyogi, D.: Fire and smoke digital twin – a computational framework for modeling fire incident outcomes. Computers, Environment and Urban Systems **110**, 102093 (2024) https://doi.org/10.1016/j.compenvurbsys.2024.102093

[14] Panya, D.S., Kim, T., Heo, M., Choo, S.: A BIM-based virtual reality evacuation simulation for fire safety management. In: Proceedings of the 42nd Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2024), vol. 2, pp. 47–56 (2024). https://doi.org/10.52842/conf.ecaade.2024.2.047

[15] Haeri Boroujeni, S.P., Mehrabi, N., Afghah, F., McGrath, C.P., Bhatkar, D., Biradar, M.A., Razi, A.: Fire and Smoke Datasets in 20 Years: An In-depth Review (2025). https://doi.org/10.48550/arXiv.2503.14552

[16] Arlovic, M., Hrzic, F., Patel, M., Bednarz, T., Balen, J.: Evaluation of synthetic data impact on fire segmentation models performance. Scientific Reports **15**(1), 16759 (2025) https://doi.org/10.1038/s41598-025-01571-5

[17] Heskestad, G.: Fire plumes, flame height, and air entrainment. In: SFPE Handbook of Fire Protection Engineering, pp. 396–428. Springer, New York, NY (2016)

[18] Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3836–3847 (2023). https://doi.org/10.1109/iccv51070.2023.00355

[19] Seo, M.S., Yoo, H.H.: Spatial econometrics analysis of fire occurrence according to type of facilities. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography **37**(3), 129–141 (2019)

[20] Seo, M.S., Castillo-Osorio, E.E., Yoo, H.H.: Fire risk prediction analysis using machine learning techniques. Sensors & Materials **35** (2023) https://doi.org/10.18494/sam4252

[21] Kim, Y., Cho, Y., Heo, H.K., Lim, L.: Estimating casualties from urban fires: A focus on building and urban environment information. Sustainable Cities and Society **115**, 105839 (2024) https://doi.org/10.1016/j.scs.2024.105839

[22] Zhang, J., Jiang, S., Qi, X.: Enhancing fire safety with improved risk index and BIM in building evacuation. Engineering, Construction and Architectural Management **32**(5), 2793–2813 (2025) https://doi.org/10.1108/ecam-09-2023-0923

[23] Hu, H., Qi, Z., Shi, J., Li, H., Ji, J.: Experimental and numerical investigations on fire development in a timber-based compartment with identification of characteristic events. Journal of Building Engineering **87**, 109033 (2024) https://doi.org/10.1016/j.jobe.2024.109033

[24] Ministry of Land, Infrastructure and Transport: V-World (2026). https://www.vworld.kr/v4po_main.do Accessed 15 July 2026

[25] Wi, H.J., Jang, I.S., Lee, A.: Driving altitude generation method with pseudo-3D building model for unmanned aerial vehicles. ETRI Journal **45**(2), 240–253 (2023) https://doi.org/10.4218/etrij.2021-0403

[26] Shkedova, O., Bäßmann, F.N., Feuerhake, U., Sester, M.: Introduction to a voxel-based urban digital twin for emergency response information systems (ERISs). Abstracts of the ICA **9**, 32 (2025) https://doi.org/10.5194/ica-abs-9-32-2025

[27] Wehbe, R., Shahrour, I.: A BIM-based smart system for fire evacuation. Future Internet **13**(9), 221 (2021) https://doi.org/10.3390/fi13090221

[28] Cha, M., Lee, J., Park, S., Choi, B.: A development of fire training simulator based on computational fluid dynamics simulation. Transactions of the Society of CAD/CAM Engineers **14**(4), 271–280 (2009)

[29] Kim, D., Cha, H.-s., Jiang, S.: The prediction of fire disaster using BIM-based visualization for expediting the management process. Sustainability **15**(4), 3719 (2023) https://doi.org/10.3390/su15043719

[30] Yan, F., Hu, Y., Jia, J., Ai, Z., Tang, K., Shi, Z., Liu, X.: Interactive WebVR visualization for online fire evacuation training. Multimedia Tools and Applications **79**(41), 31541–31565 (2020) https://doi.org/10.1007/s11042-020-08863-0

[31] Qazi, M.H., Khan, F., Kim, J., Rojas-Munoz, E.J.: Developing a VR-based training platform for emergency fire handling services using Unity 3D. In: 2023 International Conference on Frontiers of Information Technology (FIT), pp. 102–107 (2023). https://doi.org/10.1109/fit60620.2023.00028 . IEEE

[32] Zhu, J., Zhang, J., Zhu, Q., Li, W., Wu, J., Guo, Y.: A knowledge-guided visualization framework of disaster scenes for helping the public cognize risk information. International Journal of Geographical Information Science **38**(4), 626–653 (2024) https://doi.org/10.1080/13658816.2023.2298299

[33] Shang, Q., Hu, H., Yu, H., Xu, B., Wang, L., Zhu, Q.: Semantic image translation for repairing the texture defects of building models. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–20 (2023) https://doi.org/10.1109/tgrs.2023.3338962

[34] Zhang, J., Huang, Y., Li, Z., Li, Y., Yu, Z., Li, M.: Development of a method for commercial style transfer of historical architectural façades based on stable diffusion models. Journal of Imaging **10**(7), 165 (2024) https://doi.org/10.3390/jimaging10070165

[35] Chen, Y., Huang, H., Vu, T.-A., Shum, K.C., Yeung, S.-K.: StyleCity: Large-scale 3D urban scenes stylization. In: European Conference on Computer Vision (ECCV), pp. 395–413 (2024). https://doi.org/10.1007/978-3-031-73202-7_23 . Springer

[36] Zhou, J., Jin, J., Zhuang, P., Jia, D., Wu, C.: Semantic understanding and controllable generation methods of aigc in virtual reality scene construction. Discover Computing **29**(1), 94 (2026)

[37] Chung, J., Hyun, S., Heo, J.-P.: Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8795–8805 (2024). https://doi.org/10.1109/cvpr52733.2024.00840

[38] Kim, S., Kim, D.-h., Kim, H.: Texture learning domain randomization for domain generalized segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 677–687 (2023). https://doi.org/10.1109/iccv51070.2023.00069

[39] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., *et al.*: SAM 2: Segment anything in images and videos. In: International Conference on Learning Representations, vol. 2025, pp. 28085–28128 (2025)

[40] Hoegner, L., Gleixner, G.: Automatic extraction of facades and windows from MLS point clouds using voxelspace and visibility analysis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **43**, 387–394 (2022) https://doi.org/10.5194/isprs-archives-xliii-b2-2022-387-2022

[41] Han, Y., Sun, X., Yang, Y., Chen, X., Lv, J., Wang, X., Zhang, X., Hu, L.: An experimental study of wall smoke pattern characteristics adjacent to fire source.

International Journal of Thermal Sciences **214**, 109865 (2025) https://doi.org/10.1016/j.ijthermalsci.2025.109865

[42] He, H., Yu, J., Cheng, P., Wang, Y., Zhu, Y., Lin, T., Dai, G.: Automatic, multi-view, coplanar extraction for CityGML building model texture mapping. Remote Sensing **14**(1), 50 (2021) https://doi.org/10.3390/rs14010050

[43] Xiong, W., Zhang, H., Peng, B., Hu, Z., Wu, Y., Guo, J., Huang, H.: TwinTex: Geometry-aware texture generation for abstracted 3D architectural models. ACM Transactions on Graphics (TOG) **42**(6), 1–14 (2023) https://doi.org/10.1145/3618328

[44] Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: SDXL: Improving latent diffusion models for high-resolution image synthesis. In: International Conference on Learning Representations, vol. 2024, pp. 1862–1874 (2024)

[45] RunDiffusion: Juggernaut XI v11. Accessed: 8 September 2026. https://huggingface.co/RunDiffusion/Juggernaut-XI-v11

[46] Ye, H., Zhang, J., Liu, S., Han, X., Yang, W.: IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models (2023). https://doi.org/10.48550/arXiv.2308.06721

[47] Levin, E., Fried, O.: Differential diffusion: Giving each pixel its strength. In: Computer Graphics Forum, vol. 44, p. 70040 (2025). Wiley Online Library

[48] Mou, C., Wang, X., Xie, L., Wu, Y., Zhang, J., Qi, Z., Shan, Y.: T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 4296–4304 (2024)

[49] Lee, A., Jang, I.: Implementation of an open platform for 3d spatial information based on webgl. ETRI Journal **41**(3), 277–288 (2019)

[50] Hensel, S., Goebbels, S., Kada, M.: Facade reconstruction for textured LoD2 CityGML models based on deep learning and mixed integer linear programming. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **4**, 37–44 (2019) https://doi.org/10.5194/isprs-annals-iv-2-w5-37-2019

[51] Cheng, T.-Y., Sharma, P., Markham, A., Trigoni, N., Jampani, V.: Zest: Zero-shot material transfer from a single image. In: European Conference on Computer Vision, pp. 370–386 (2024). Springer

[52] Wang, H., Wang, Q., Bai, X., Qin, Z., Chen, A.: InstantStyle: Free lunch towards style-preserving in text-to-image generation (2024). https://doi.org/10.48550/arXiv.2404.02733

[53] Xie, E., Chen, J., Chen, J., Cai, H., Tang, H., Lin, Y., Zhang, Z., Li, M., Zhu, L., Lu, Y., et al.: SANA: Efficient high-resolution image synthesis with linear diffusion transformers (2024). https://doi.org/10.48550/arXiv.2410.10629

[54] Black Forest Labs, Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., et al.: FLUX.1 Kontext: Flow matching for in-context image generation and editing in latent space (2025). https://doi.org/10.48550/arXiv.2506.15742

[55] Bagdasarian, M.T., Knoll, P., Li, Y., Barthel, F., Hilsmann, A., Eisert, P., Morgenstern, W.: 3DGS.zip: A survey on 3D Gaussian splatting compression methods. In: Computer Graphics Forum, vol. 44, p. 70078 (2025). https://doi.org/10.1111/cgf.70078 . Wiley Online Library

[56] Gong, R., Zhang, J., Pang, Y., Wang, Y., Zhang, X., Meng, W., Zhou, P.: Implicit vs. explicit a comparative survey on nerf and 3dgs for large-scale aerial scene editing. Discover Computing **29**(1), 142 (2026)

[57] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). https://doi.org/10.1109/cvpr.2018.00068

[58] Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., Choi, Y.: CLIPScore: A reference-free evaluation metric for image captioning. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7514–7528 (2021). https://doi.org/10.18653/v1/2021.emnlp-main.595

[59] Chen, G.-H., Yang, C.-L., Po, L.-M., Xie, S.-L.: Edge-based structural similarity for image quality assessment. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 2, p. (2006). https://doi.org/10.1109/ICASSP.2006.1660497

[60] Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: Multi-scale image quality transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5148–5157 (2021). https://doi.org/10.1109/iccv48922.2021.00510

[61] Wang, J., Chan, K.C., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 2555–2563 (2023). https://doi.org/10.1609/aaai.v37i2.25353

[62] Chung, H., Lee, S.: Prompt-guided diffusion approach for deep generative vehicle designs under user-defined specifications. Journal of Computational Design and Engineering **13**(7), 144–159 (2026)

[63] Vishwakarma, A.K., Ranjan, R., Priyadarshi, R., Kollie Jr, D.S.: Recent progression of artificial intelligence driven no-reference video quality assessment techniques in multimedia services. Discover Computing **29**(1), 380 (2026)

[64] Qi, Y., Lou, M., Liu, Y., Li, L., Yang, Z., Nie, W.: Advancing image super-resolution techniques in remote sensing: A comprehensive survey. ISPRS Journal of Photogrammetry and Remote Sensing **231**, 68–100 (2026) https://doi.org/10.1016/j.isprsjprs.2025.10.024

[65] Sung, M., Gong, M.-G., Ham, S.-J., Kim, I.-M., Yun, S., Kang, J.-M.: A novel VLM-guided diffusion model for remote sensing image super-resolution. IEEE Geoscience and Remote Sensing Letters (2025) https://doi.org/10.1109/lgrs.2025.3608178

[66] Zuo, Y., Zheng, Q., Wu, M., Jiang, X., Li, R., Wang, J., Zhang, Y., Mai, G., Wang, L.V., Zou, J., et al.: 4KAgent: Agentic any image to 4K super-resolution (2025). https://doi.org/10.48550/arXiv.2507.07105

[67] Hwang, C.-L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications a State-of-the-art Survey. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-48318-9

[68] Chengna, W., Jiahao, Z., Sen, Z., Xingze, W.: Trust evaluation mechanism for data collection in smart cities. Discover Computing **28**(1), 67 (2025)

[69] Selvam, C., Sundaram, D.: An intelligent low-light image enhancement model using intuitionistic fuzzy generator and genetic algorithm validated through MCDM. Computers and Electrical Engineering **128**, 110730 (2025) https://doi.org/10.1016/j.compeleceng.2025.110730

[70] Hou, M.-j., Dong, H., Ji, X.-y., Zou, W.-b., Xia, X.-s., Li, M., Yin, Y.-j., Li, B.-h., Chen, Q., Zhou, J.-x.: I-DCGAN and TOPSIS-IFP: A simulation generation model for radiographic flaw detection images in light alloy castings and an algorithm for quality evaluation of generated images. China Foundry **21**(3), 239–247 (2024) https://doi.org/10.1007/s41230-024-3094-x

[71] Matsuoka, R., Ishikawa, M., Inazawa, T., Nakanishi, Y., Kawamata, F., Takada, M., Danjo, T.: Software development for producing texture images mapped on a building surface of a 3d city model using aerial images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **11**, 45–52 (2026)

[72] Buyukdemircioglu, M., Oude Elberink, S.: Automated texture mapping cityjson 3d city models from oblique and nadir aerial imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **10**, 87–93 (2024)

[73] Xiang, F., Xu, Z., Hasan, M., Hold-Geoffroy, Y., Sunkavalli, K., Su, H.: Neutex: Neural texture mapping for volumetric neural rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7119–7128 (2021)

[74] Richardson, E., Metzer, G., Alaluf, Y., Giryes, R., Cohen-Or, D.: Texture: Text-guided texturing of 3d shapes. In: ACM SIGGRAPH 2023 Conference Proceedings, pp. 1–11 (2023)

[75] Cao, T., Kreis, K., Fidler, S., Sharp, N., Yin, K.: Texfusion: Synthesizing 3d textures with text-guided image diffusion models. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4146–4158 (2023). IEEE

[76] Cheng, W., Mu, J., Zeng, X., Chen, X., Pang, A., Zhang, C., Wang, Z., Fu, B., Yu, G., Liu, Z., *et al.*: Mvpaint: Synchronized multi-view diffusion for painting anything 3d. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 585–594 (2025). IEEE