Text-based Tactile Graphics Generation for the Visually Impaired

Ruihan Gao[1⋆], Joonghyuk Shin[1,2⋆], Ava Pun[1], Jaesik Park[2], Wenzhen Yuan[3], and Jun-Yan Zhu[1†]

[1] Carnegie Mellon University, Pittsburgh, PA, USA
[2] Seoul National University, Seoul, Republic of Korea
[3] University of Illinois Urbana-Champaign, Urbana, IL, USA

Abstract. Tactile graphics are a primary medium for blind and low-vision (BLV) individuals to access non-textual information. However, they are difficult to scale or personalize. While recent generative models have revolutionized visual content creation, they are optimized for screen-based visual realism and fail to satisfy the haptic perceptual and physical fabrication constraints required for touch. We present the first integrated generative system that produces fabrication-ready 2.5D tactile graphics directly from natural language prompts, jointly generating global base geometry, fine-grained tactile surface textures, and standard-compliant braille within a unified 3D-printable representation. Our approach introduces fabrication-aware techniques, including template-guided relief generation, a fast diffusion-based text-to-texture module for high-resolution tileable normal maps, and strict base flattening to ensure tactile readability and printability, while supporting both automatic generation and interactive texture control. Extensive evaluations, together with in-person user studies with BLV participants and blindfolded sighted participants using physically 3D-printed outputs, show that participants consistently prefer our results over baselines. By extending generative graphics beyond screens to touchable reliefs, our work broadens access to generative AI for the BLV community and beyond. The code and project webpage are available at `https://ruihangao.github.io/Text2TactileGraphics/`.

Keywords: Tactile Graphics · Fabrication-Aware Generation for BLV

1 Introduction

Recent advances in generative models have dramatically expanded the scope of computer graphics, enabling text-driven synthesis of images [9,18,44], videos [11, 36], and 3D assets [27,29,41,64] with unprecedented realism and diversity. These systems have reshaped how visual content is created, edited, and consumed, lowering barriers for artists, designers, and everyday users.

⋆Equal contribution. † Corresponding Author.

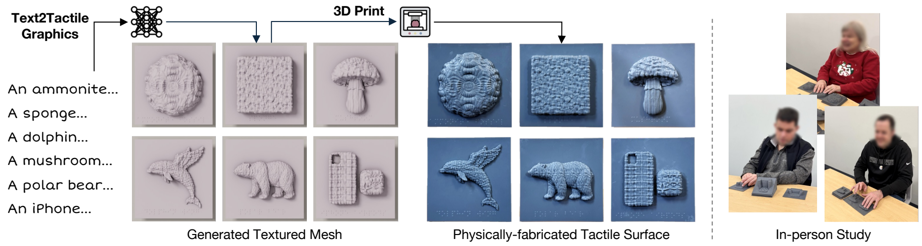

Fig. 1: Given a text prompt describing an object and its surface textures, our method generates *3D-printable 2.5D tactile reliefs* that integrate global base geometry, fine-grained surface textures, and readable braille annotations. By jointly modeling geometry, texture, and braille within a unified fabrication-ready representation, our method produces perceptually salient tactile details and supports expressive, accessible exploration for blind and low-vision users. Please refer to our webpage for a video of an interactive demo and 3D visualizations.

Despite this progress, modern generative pipelines remain almost exclusively *visual*, designed for screen-based viewing. This overlooks a complementary and fundamentally physical modality: *touch*. For blind and low-vision (BLV) individuals, who constitute over 290 million people worldwide, including 43 million who are fully blind [40], touch is a primary means of understanding shapes, structures, and materials. Tactile graphics, which encode information through raised geometry, surface textures, and braille annotations, are widely used in education, accessibility, and cultural contexts such as diagrams, maps, and museum exhibits [16, 45, 52, 57].

Despite their importance, tactile graphics are still largely produced through manual, expert-driven workflows such as swell paper, embossing, collage, and sculptural relief fabrication. Designers must interpret visual or textual content and translate it into touch-friendly representations, making the process time-consuming, labor-intensive, and difficult to scale. As a result, existing tactile graphic collections remain limited in scope and primarily focus on standardized educational materials [1, 39]. Compared to visual graphics, where AI enables rapid creation and personalization, tactile graphics offer little opportunity for user-driven or creative generation.

Tactile graphics also impose constraints not addressed by existing generative pipelines. Unlike images or meshes rendered for viewing, they must satisfy *haptic perceptual constraints* such as tactile contrast, resolution, and style consistency [5], while remaining compatible with *physical fabrication* processes like 3D printing. Simply generating a depth map or visually plausible mesh is insufficient. Current text-to-3D models often produce non-manifold surfaces, fragile structures, and disconnected geometry, while lacking fine tactile textures because they are optimized for global object shape rather than touch-perceptible detail. Consequently, they require extensive post-processing or fail fabrication entirely.

To address these challenges, we introduce a *generative graphics system for tactile content creation*. Given text descriptions of the desired object and textures, our method generates a *2.5D tactile graphic* integrating geometry, tactile texture, and braille annotations, directly ready for standard 3D printing. At its core is a text-to-tactile-texture module that produces fine-resolution surface geometry from text, targeting texture-centric generation rather than global shape synthesis; the outputs are converted into tileable normal and displacement maps for interactive application at arbitrary scales and regions. To ensure physical realizability, we design fabrication-aware processing that unifies template-guided base generation, flat-base enforcement, and standard-compliant braille placement, and provide an interactive interface supporting workflows from fully automatic generation to fine-grained control over texture type, scale, and spatial assignment. Experiments demonstrate that our system produces tactile graphics that are perceptually effective, fabrication-compatible, and faithful to input prompts, significantly lowering the barrier to personalized tactile content for the BLV community and more broadly opening new research directions at the intersection of generative modeling, accessibility, and physical fabrication.

2 Related Work

Tactile Graphics and Physical Accessibility Representations. Tactile graphics are raised-line drawings or textured representations that convey visual and spatial information through touch, enabling blind and low-vision (BLV) users to access diagrams, maps, and illustrations. Traditional tactile graphics are typically produced using swell paper, embossing, or sculptural bas-relief fabrication, but these methods require expert manual design and are time-consuming, limiting scalability and availability [2,5]. A substantial body of work has studied how tactile graphics should be designed to be interpretable through touch. Design guidelines emphasize abstraction, clear contour hierarchies, sufficient spacing, and the use of discriminable textures to avoid tactile clutter. Because touch is local and sequential, overly complex graphics can hinder comprehension. Prior studies highlight the importance of simplifying visual content rather than attempting a direct visual-to-tactile translation [16]. Research has also explored how tactile graphics can convey 3D information using 2D or shallow-relief representations [38, 55]. Together, this literature establishes tactile graphics as a distinct medium with unique perceptual and fabrication constraints.

Visual Content Generation. Generative models [17, 22, 24] have achieved great success in synthesizing 2D and 3D content. While denoising diffusion models have become dominant in text-to-image generation [19,48], generative modeling has also expanded to 3D, evolving from early volumetric approaches [63] to diverse representations including points, meshes, and neural fields [29,31,41,50,51]. Large-scale 3D generative models have since improved visual quality and geometric detail, producing meshes that appear realistic when rendered [27, 64, 65, 68, 75]. Generative models have also been applied to text-to-texture generation [33, 56, 66, 74]. Despite their success, these models are primarily optimized for visual realism and screen-based consumption. They do not enforce fabrication constraints such as watertightness and minimum feature thickness, nor do they account for haptic perception. As a result, outputs may still be unsuitable for physical fabrication or tactile interpretation, motivating the need for fabrication- and perception-aware generative pipelines.

Generative and Computational Approaches for Tactile Content. A growing body of work explores computational methods for generating tactile content to support BLV users. Early approaches focused on algorithmic image-to-tactile translation using edge detection and thresholding techniques [10, 25]. While simple to implement, these methods often generate cluttered or ambiguous tactile outputs that violate tactile design principles. To address these limitations, later systems incorporated higher-level image understanding. For example, Pic2Tac [37] uses object detection to replace foreground elements with readable tactile icons and fills the background with uniform textures. More recent work applies image generation models directly to tactile graphics creation [7, 23, 53], showing the promise of generative AI. However, these methods still focus on two-dimensional embossed or swell graphics, which lack the expressivity and textural richness of 2.5D sculpture. Several works explore the automated production of 2.5D graphics from image, normal map, or 3D inputs through traditional [60, 61, 73] or learning-based [13, 21, 72] methods. Nonetheless, these works focus on fine geometric detail that may not be readable by touch or resolvable by fabrication, leaving the resulting sculptures unsuitable for tactile graphics. Recent works [14, 15] leverage generative priors to produce 2D and 3D objects with aligned color and tactile textures, but they are limited to per-material optimization and fail to generate new material textures. In contrast to prior work, we create 2.5D graphics suitable for touch-based interpretation and physical fabrication, generalizing to any new materials given a text prompt input.

3 Method

Our goal is to generate touch-friendly, fabrication-ready 2.5D tactile graphics from text prompts. This problem is challenging for three reasons: (i) data scarcity, because no large-scale dataset models touch-suitable 2.5D graphics; (ii) representation mismatch, since tactile graphics lie outside the distribution of object-centric 3D generation benchmarks, requiring both global shape abstraction and fine-scale surface relief suitable for touch; and (iii) fabrication constraints, as existing methods do not explicitly enforce watertightness, flat bases, or geometric regularity required for haptic readability and physical printing.

To address these challenges, we propose a text-driven pipeline that generates a 3D-printable tactile graphic integrating three complementary components: *global geometry* capturing object shape, *tactile surface textures* encoding fine-grained material cues, and *braille annotations* providing semantic grounding. As shown in Fig. 2, our pipeline begins by generating a 2.5D base geometry from text (Sec. 3.1), synthesizes and applies tactile surface textures with flexible spatial control (Sec. 3.2, Sec. 3.3), enforces fabrication-aware geometric constraints

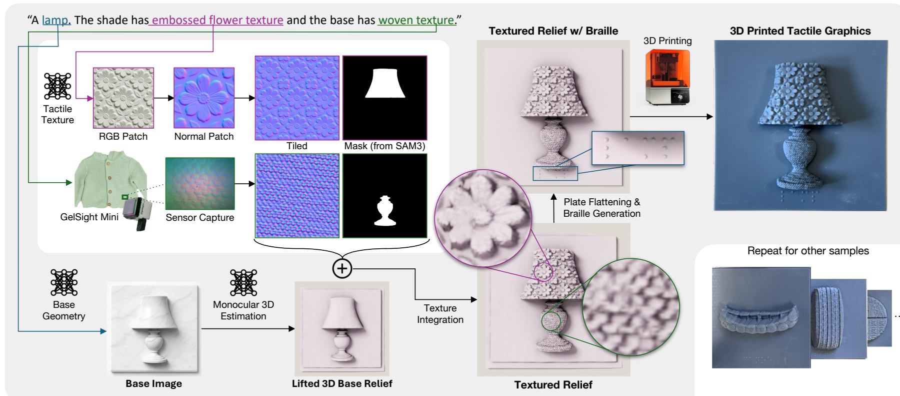

Fig. 2: Method overview. Given a text prompt, we generate 3D-printable 2.5D tactile graphics that jointly encode geometry, surface texture, and braille. We first synthesize a base image constrained by a plate template and lift it to 3D base relief via monocular geometry estimation. In parallel, a tactile texture module produces high-resolution, tileable normal maps from either text prompts or tactile sensor captures, which can be assigned to object parts through text or user interaction. The textures deform the base relief, followed by fabrication-aware base flattening and braille generation. The resulting watertight mesh is fabricated via SLA resin 3D printing.

and integrates braille annotations (Sec. 3.4), and finally outputs a watertight mesh suitable for interactive use and 3D printing (Sec. 3.5).

3.1 Base Geometry Generation

To generate the global shape of the tactile graphic, we first synthesize a text-conditioned image I of a 2.5D relief sculpture using a fixed base-plate template I_{bp}. By reformulating text-to-image generation as an in-context editing task conditioned on this template, we avoid the ambiguous scale and depth often produced by general-purpose models in object- or scene-centric compositions. As shown in Fig. 4 (Qwen-Image-Edit, 4-step), the template encourages the object to rest on a planar support, yielding a consistent spatial layout and more stable depth estimation across diverse prompts. Similar trends are observed across other base models, including Qwen-Image-Edit (40-step) and Nano Banana Pro, both shown in Appendix D.1. Given the generated image $I \in \mathbb{R}^{H\times W\times 3}$, we estimate a monocular depth map $D \in \mathbb{R}^{H\times W}$ using an off-the-shelf depth prediction model f_{depth}, then convert it into a 2.5D relief by mapping depth values to height displacements over the base plate. This representation preserves salient object contours while remaining shallow enough for tactile exploration and efficient fabrication, provideing a stable geometric foundation for subsequent tactile texture refinement.

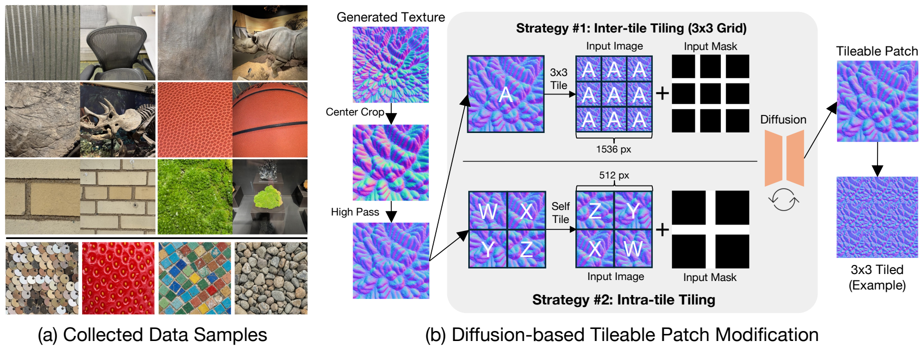

Fig. 3: Texture data and tileable patch generation. (a) We collect paired real texture images at close-up and zoomed-out scales (top three rows) and augment them with synthetic textures to train our text-to-tactile-texture model (bottom row). (b) Generated normal maps are center-cropped and high-pass filtered, then refined through diffusion-based inpainting using either an inter-tile 3×3 arrangement or an intra-tile quadrant rearrangement. The resulting patch tiles seamlessly over larger surfaces.

3.2 Tactile Texture Synthesis

Texture generation from text prompt. A straightforward approach to incorporating surface textures is to generate an image from a single text prompt describing both geometry and material appearance, and then lift the image to 2.5D. In practice, this strategy leads to three major issues: (i) text-to-image models prioritize global coherence over fine-grained detail, and text prompts lack the spatial specificity needed for precise texture placement; (ii) when the desired texture is semantically inconsistent with the object (*e.g.,* avocado skin on a dolphin), models often suppress the specified texture; (iii) monocular depth estimation tends to smooth high-frequency surface details, as these models are optimized for overall geometry rather than subtle relief textures.

To address these limitations, we introduce a dedicated *text-to-tactile-texture* module that synthesizes high-resolution, texture-centric surface representations from text prompts. Unlike object-centric image generation, our goal is to model local, repeatable surface patterns suitable for tactile perception and geometric displacement. Since most text-to-image models are trained on object- and scene-level imagery, they are poorly suited for texture-focused synthesis. We therefore adapt an open-source image generation model [62] by fine-tuning it on a curated texture dataset emphasizing local surface detail. The dataset consists of real captured textures and synthetically generated samples aligned with our target distribution (Fig. 3). Additional details on dataset construction and training are provided in Appendix A.2.

Texture generation from sensor data. While the default texture input is specified via a text prompt, our framework also supports tactile sensor data as an alternative texture source. Specifically, we use the TouchTexture dataset introduced in TactileDreamFusion [14] as a tactile texture library, which contains surface measurements captured by vision-based tactile sensors. We also support processing newly collected tactile data by following the same preprocessing pipeline in [14], enabling users to incorporate custom tactile measurements captured with tactile sensors (*e.g.*, GelSight [59, 70]).

3.3 Tiling

To apply a texture consistently over large surfaces, the generated normal patch must be tileable. Classical texture synthesis methods [6, 8] rely on minimum-cut or graph-cut algorithms to stitch patches along optimal seam boundaries. While effective for RGB images, these approaches can be computationally expensive as the output size grows, and more fundamentally, cannot modify the patch content. They can only select which portions to retain and where to place seams. For normal maps, this can be problematic, as the minimum cut may discard significant portions of the original texture when ideally only narrow boundary regions should be modified.

Recent generative approaches [12, 20, 32, 46, 47, 49] address these limitations by leveraging generative models to synthesize or modify input patches. However, we found that recent training-free diffusion-based methods perform poorly on normal maps due to distribution mismatch. For example, one state-of-the-art approach, Tiled Diffusion [32] often fails to preserve normal map characteristics (see Fig. 5). Instead, it tends to generate arbitrary objects to achieve tileability, as its primary goal is seamless (and mostly non-repeating) tiling rather than content preservation. It is also optimized for direct text-to-(tileable) image setups and can incur noticeable computational overhead. To this end, we propose a simple combination of inpainting and SDEdit that enables fast synthesis with minimal modification to the original normal map.

Tileable patch generation. Motivated by Tiled Diffusion [32] and open-source community efforts [3, 28, 54], we combine SDEdit [34], which enables minor adjustments while preserving low-level features, with inpainting to enforce continuity in specific boundary regions. We explore two spatial arrangement strategies for this formulation, as illustrated in Fig. 3. The first strategy, *inter-tile arrangement*, places the input patch in a 3×3 grid and masks the boundary regions between adjacent tiles. The second, *intra-tile arrangement*, partitions the input patch into a 2×2 grid of quadrants and swaps them both horizontally and vertically (i.e., point-symmetric rearrangement). Both strategies effectively simulate boundary adjacency by bringing seam regions into close proximity within a single image. Given the rearranged image I_{swap} and binary inpainting mask M (where $M = 1$ indicates regions to inpaint), we apply SDEdit [34] with Qwen-Image's flow matching scheduler. Specifically, we first add noise to the input by performing $z_{t_0} = (1 - t_0) \cdot \mathcal{E}(I_{\text{swap}}) + t_0 \cdot \epsilon$, where $\mathcal{E}$ denotes the VAE encoder and t_0 corresponds to the denoising strength (or timestep) and $\epsilon \sim \mathcal{N}(0, I)$.

We then iteratively denoise from t_0 to $t = 0$ using an Euler sampler. At each denoising step $t \to s$, we blend the denoised latent with the original in the unmasked region by performing $z_s = M \odot z_s^{\text{pred}} + (1 - M) \odot z_s^{\text{orig}}$ where z_s^{pred} is the model prediction and $z_s^{\text{orig}} = (1 - s) \cdot \mathcal{E}(I_{\text{swap}}) + s \cdot \epsilon$ is the forward-diffused original latent at timestep s. We use $t_0 = 0.9$ (denoising strength) with 10 inference steps. We find that the intra-tile arrangement is both faster (since it operates in a smaller single image rather than a 3×3 grid) and produces better results. As shown in Fig. 5 and Table 1, compared to Tiled Diffusion, which requires latent wrapping and ∼100 sampling steps, our approach generates fewer artifacts while being substantially faster.

In practice, many generated normal-map patches contain undesired low-frequency geometric variations (*e.g.*, global tilts or smooth warping) as well as outlier normals, which complicate seamless tiling. To mitigate this issue, we first center-crop each generated patch from 1024^2 to 512^2 resolution and apply a per-channel high-pass filter to remove low-frequency components. Specifically, given a normal map $N = (N_x, N_y, N_z) \in [-1, 1]^{H \times W \times 3}$, we filter each channel independently in the Fourier domain:

$$\mathrm{HP}(X) \;=\; \mathcal{F}^{-1}\big(\mathcal{F}(X) \odot G_{\text{hp}}\big), \quad G_{\text{hp}}(u, v) = 1 - \exp\left(-\frac{d(u, v)^2}{2\sigma_f^2}\right), \tag{1}$$

where $\mathcal{F}$ denotes the 2D FFT, $d(u, v)$ is the distance to the frequency origin, and σ_f controls the cutoff frequency. Since high-pass filtering removes the DC component, we restore the flat-normal baseline by adding a constant offset to the z-channel and renormalize the normals per pixel:

$$N^{\text{hp}}(p) = \frac{(\mathrm{HP}(N_x), \mathrm{HP}(N_y), \mathrm{HP}(N_z) + 1)}{\|(\mathrm{HP}(N_x), \mathrm{HP}(N_y), \mathrm{HP}(N_z) + 1)\|_2}. \tag{2}$$

This preprocessing yields a cleaner, more uniform patch that reduces global bias and distortion, making subsequent tiling and seamless boundary synthesis easier. We report an ablation in Fig. 5.

Texture displacement map. To integrate the tiled texture into the base geometry, we integrate the tileable normal texture map into a displacement map. Let $D_{\text{disp}} \in \mathbb{R}^{H_t \times W_t}$ denote the resulting displacement field. For each vertex $\mathbf{v}$ on the mesh within the designated textured region, we displace it along its surface normal $\mathbf{n}$ as $\mathbf{v}' = \mathbf{v} + \lambda\, D_{\text{disp}}(\mathbf{u})\, \mathbf{n}$, where $\mathbf{u}$ denotes the vertex's UV texture coordinate and λ controls the displacement magnitude. The designated region is restricted to a user-defined segmentation mask on the mesh, which is by default inferred from the object part specified in the text prompt through the offline segmentation model SAM3 [4]. To support finer control, we allow users to interactively refine the segmentation mask by adding or subtracting regions through direct clicking. This design enables independent control of texture type, scale, and spatial extent, without altering the underlying global geometry.

3.4 Fabrication-Aware Base Flattening and Braille Placement

To provide semantic grounding for the generated geometry, we integrate braille labels onto the base plate. Because braille readability is highly sensitive to sub-millimeter geometric variation, the support surface must be strictly planar and horizontal. However, monocular depth estimation does not recover geometry at this scale, and lifting geometry from predicted depth may introduce unevenness or global tilt in the base plate. We therefore treat base flattening as a prerequisite for braille generation and enforce a fabrication-aware planar support surface.

Specifically, given the textured mesh, we identify the set of vertices belonging to the base plate using foreground/background segmentation and compute their average height. All base plate vertices are then projected onto the corresponding horizontal plane, producing a base surface with zero height variation and no global tilt. In addition, we remove intermediate boundary offsets introduced during geometry lifting, ensuring that both object relief and braille annotations rest directly on a single outer base. Let $\mathcal{V}_{\text{base}}$ denote the set of base plate vertices. We compute the average base height $h_{\text{base}} = \frac{1}{|\mathcal{V}_{\text{base}}|} \sum_{\mathbf{v} \in \mathcal{V}_{\text{base}}} v_z$. Each base vertex $\mathbf{v} = (x, y, z)$ is then projected as $\mathbf{v}' = (x, y, h_{\text{base}})$. This operation modifies only z-coordinates and preserves mesh connectivity by construction. Additional details on topology preservation, mesh closure for printing, the texture displacement parameter λ, and texture transitions at boundaries are provided in Appendix A.5.

With a flat and stable base plate, we integrate braille annotations into the tactile graphic. Input text is converted into braille dots using a standard lookup table and placed on designated flat regions of the base plate. We follow established U.S. and international standards for braille dot size, height, and spacing to ensure tactile readability. Braille placement is explicitly constrained to regions free of geometric relief and texture displacement, preventing interference during tactile exploration. The enforced flat base guarantees consistent braille height, uniform spacing, and stable tabletop placement, enabling comfortable two-handed exploration.

3.5 Interactive Control and Fabrication

Our system supports multiple levels of user control. In the simplest mode, users provide a single text prompt and obtain a fully generated tactile graphic. Advanced users may optionally specify texture prompts, select object regions for texture application, and adjust texture scale. These interactions operate exclusively at the texture synthesis stage, leaving the global geometry and braille layout unchanged. Appendix A.4 includes a screenshot of our online demo illustrating these interactive controls.

The final output is a watertight, manifold mesh with a flat base and enforced minimum feature thickness, suitable for direct 3D printing. All physical outputs in this work are fabricated using *high-resolution SLA resin 3D printing*, which provides the sub-millimeter geometric accuracy required to faithfully reproduce fine-grained tactile textures and standard-compliant braille features. Specifically, we use a Formlabs Form 4 SLA printer for all physical examples. This fabrication choice enables rapid production of tactile graphics that are portable, durable, and optimized for reliable touch-based exploration.

4 Experiments

We conduct evaluations to validate each module and the overall pipeline. Since a core motivation of this work is to aid blind and low-vision (BLV) individuals, we include a dedicated user study with target users in Sec. 4.3.

Implementation Details. Our modular pipeline combines several off-the-shelf components with a fine-tuned texture model. For reproducibility, we mainly use the open-source Qwen-Image series [62]. Base geometry is generated using Qwen-Image-Edit conditioned on both the text prompt and a marble plate template. The texture module is obtained by fine-tuning a text-to-image variant of Qwen-Image on a mixture of captured and synthetic texture images using the AdamW optimizer [30] (Sec. 3.2). For tiling we use the base diffusion model without additional training. We compare against Nano Banana Pro [43] as a proprietary baseline. Both Qwen-Image-Edit and the texture model optionally use a 4-step distilled variant via LoRA trained with Distribution Matching Distillation [69]. Monocular geometry and normals are estimated using MoGE v2 [58], and object regions are obtained with SAM3 [4]. Additional implementation and training details are provided in Appendix A.

4.1 Quantitative Evaluation

Evaluating text-to-tactile graphics is challenging because no standard metrics capture both geometric relief and tactile surface texture, and no prior method covers the full geometry, texture, and braille pipeline. We therefore evaluate each module independently, and compare end-to-end outputs with recent text-to-3D approaches in qualitative evaluation (Sec. 4.2) and the user study (Sec. 4.3).

Base Geometry Generation. We evaluate the base geometry generation module on 50 diverse prompts, comparing against Qwen-Image-Edit [62] (4-step and 40-step) and Nano Banana Pro, with and without plate template conditioning. We detect plate regions using SAM3 and estimate plate heights with MoGE v2. A stable base plate is critical for fabrication: large height variation leads to inconsistent plate thickness across samples and uneven surfaces that may occlude relief details. To quantify plate stability, for each image i we compute the mean and standard deviation of plate heights within the detected plate region, denoted by μ_i and σ_i. We report the average plate height $\bar{\mu}$ and total plate-height variation on the standard-deviation scale: $\sigma_{\text{total}} = \sqrt{\frac{1}{N}\sum_i \sigma_i^2 + \mathrm{Var}(\mu_i)}$. We also report CLIP score [42] to measure the semantic alignment between the generated base geometry and the input text prompt. As shown in Table 1, plate conditioning consistently improves CLIP scores while stabilizing plate geometry: $\bar{\mu}$ remains tightly around 0.34–0.36 and σ_{total} is significantly reduced across models. Without plate conditioning, Nano Banana Pro fails to detect the plate in 6/50 samples, whereas all plate-conditioned variants achieve 100% detection.

Table 1: Quantitative evaluation. From left to right: **geometry analysis*** shows plate conditioning improves semantic alignment and plate stability; **texture quality**† shows fine-tuning improves consistency and high-frequency detail; and **tileability**‡ shows our intra-tile method achieves the best seam quality and runtime.

Method	Plate Cond.	CLIP↑	Plate Height Avg.	Plate Height SD ↓
QIE-4	✗	34.82	0.181	0.131
QIE-4	✓	36.64	0.337	**0.122**
QIE-40	✗	36.11	0.305	0.205
QIE-40	✓	<u>36.91</u>	0.350	<u>0.123</u>
NBP	✗	36.83	0.467	0.188
NBP	✓	**37.66**	0.365	0.137

Method	CLIP↑	Patch Self-Sim.↓	HF Ratio↑
QI-4	27.14	0.496	0.305
QI-40	27.68	0.461	0.314
Ours-4	**28.98**	0.412	0.396
Ours-40	28.43	<u>0.400</u>	**0.408**
NBP	<u>28.62</u>	**0.395**	<u>0.402</u>

Method	Border Conti.↓	Grad. Conti.↓	Time (s)
Resize	29.80	59.99	–
Crop	26.10	52.82	–
+ HPF	22.84	46.53	–
++ TD	<u>4.43</u>	<u>9.76</u>	<u>10.9</u>
++ Ours-I	7.19	14.42	24.88
++ Ours-A	**1.61**	**3.19**	**3.18**

Text-to-Texture. We evaluate texture generation on 50 diverse prompts, comparing our fine-tuned Qwen-Image model against the base model and Nano Banana Pro. Standard text-to-image models are object-centric: prompting "avocado skin texture" often yields an entire round-shaped avocado body rather than a uniform surface pattern. To measure spatial consistency, we generate all textures at 1024×1024 resolution, randomly sample 32 patches of 128×128 pixels, and compute the average pairwise LPIPS (VGG) distance [71]. Lower values indicate more uniform patterns suitable for tiling. We also compute an FFT-based high-frequency (HF) ratio from estimated normal maps, measuring the fraction of non-DC normal-map variation that remains after applying our Gaussian high-pass normal-map filter, and report CLIP Score to measure alignment with the texture prompt. As shown in Table 1, our fine-tuned model generally achieves lower LPIPS and higher HF ratio than the base model, indicating improved pattern consistency and a greater concentration of high-frequency normal variation. Performance is comparable to Nano Banana Pro while remaining fully reproducible. Fig. 4 shows qualitative comparisons and we provide additional results in Appendix D.1.

Tiling. We ablate each preprocessing stage on 50 normal maps from our fine-tuned model (4-step), targeting 512×512 tileable patches. We compare: (1) plain resize from 1024 to 512, (2) center crop, (3) center crop with high-pass filtering, and (4) diffusion-based tiling. Seam quality is measured using two metrics (lower is better): border continuity, which measures pixel discontinuity at tile edges, and gradient continuity, which measures slope smoothness across seams. As shown in Table 1, each preprocessing stage progressively improves seam quality, and our intra-tile tiling approach achieves the best results. We also compare against the most recent training-free tiling work, Tiled Diffusion [32]. While it

* QIE = Qwen-Image-Edit; NBP = Nano Banana Pro.

† QI-4/40 = Qwen-Image (4/40-step); NBP = Nano Banana Pro. Patch Self-Sim. denotes average patchwise LPIPS, and HF Ratio denotes the proportion of high-frequency normal-map variation.

‡ TD = Tiled Diffusion; Ours-I/A = inter-/intra-tile. Border Conti. measures edge discontinuity, and Grad. Conti. measures slope discontinuity.

Table 2: User preference study. We evaluate the perceptual quality of the generated textures, reporting Likert score ranges 1–5 averaged across each user population group.

Population	Method		
	Hunyuan3D	Naive	**Ours**
Blind and low-vision (BLV)	3.62 ± 1.28	3.62 ± 1.20	$\mathbf{3.93 \pm 1.32}$
Blindfolded sighted (BSP)	2.97 ± 1.31	3.02 ± 1.24	$\mathbf{3.35 \pm 1.39}$

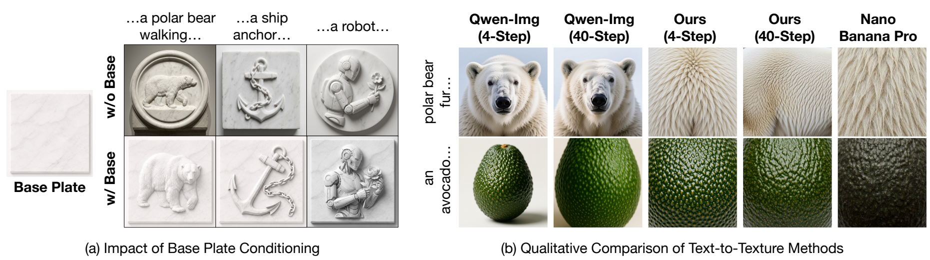

Fig. 4: Base-plate conditioning and text-to-texture generation. (a) Without base-plate conditioning, generated reliefs exhibit inconsistent, uneven, or missing plates. A fixed base-plate template stabilizes plate geometry and relief height/shape across prompts. (b) We compare base Qwen-Image, our fine-tuned variants, and Nano Banana Pro. Our models generate homogeneous, fine-grained texture patterns instead of object-centric images, making them suitable for tiling and tactile synthesis.

produces visually plausible textures, it often alters the underlying normal map structure, resulting in semantically incorrect tactile patterns. Qualitative examples are shown in Fig. 5. Full metric definitions are provided in Appendix B.

4.2 Qualitative Evaluation

Baselines for pipeline evaluation. Unlike prior approaches that produce tactile graphics on swell paper [23], our method outputs *fabrication-ready 3D meshes* that jointly encode geometry, surface texture, and braille. As this constitutes a new problem setting, we compare our approach against recent state-of-the-art 3D generation methods, including Hunyuan3D 2.1 [67] and Trellis 2 [64]. Since they are designed for object-centric 3D generation, they often fail to produce the stable base plate required for tactile graphics (see Appendix B.1). Hunyuan3D may recognize the plate but place it at incorrect depths, while Trellis 2 fails when the plate is present in the input image. To enable comparison, we adapt Trellis 2 by removing the plate background and placing the generated relief onto a fixed base plate. We also include a naive baseline that shares the same base geometry pipeline but omits the explicit text-to-texture and braille modules. Instead, it directly maps the full text prompt to a single relief image, which is then lifted to 3D. Qualitative comparisons across all methods are shown in Fig. 6. Our method

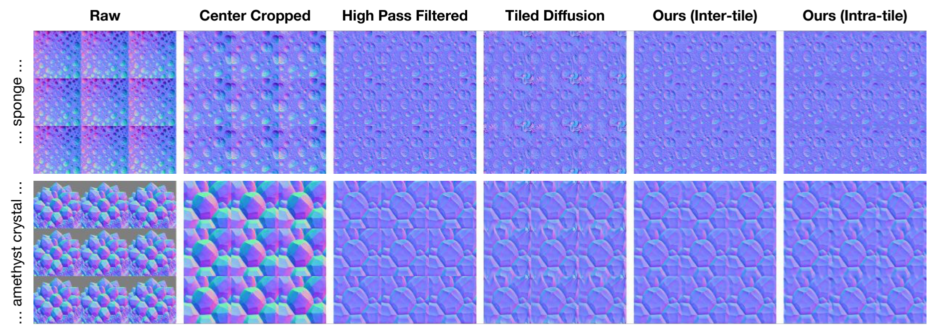

Fig. 5: Tiling comparisons. Our center cropping and high-pass filtering steps make textures more consistent and suitable for application to 3D geometry. Compared to Tiled Diffusion [32], our method produces more seamless textures, with the intra-tile configuration having the best results. Best viewed zoomed in.

achieves the most consistent plate geometry while producing richer and more semantically aligned surface textures. Additional qualitative results for ablations of the base plate template, comparisons against traditional 2.5D representations, and material- and texture- generation methods are provided in Appendix D.

4.3 In-Person Perceptual Study

Study Design. We conduct an in-person perceptual study to evaluate the usability and tactile quality of generated 2.5D tactile graphics, in collaboration with a state-run regional library for accessible media. We fabricate tactile graphics for 16 diverse prompts using three methods: *Hunyuan3D 2.1*, *Ours w/o text-to-texture (Naive)*, and *Ours (full)*. Outputs from Trellis 2 frequently fail fabrication checks due to non-manifold geometry or thin structures and are therefore excluded from physical evaluation (see failure cases in Appendix B.1). In total, 48 tactile samples are fabricated.

Two groups, *(1) eight blind and low-vision (BLV) participants* with varying vision conditions and braille literacy levels, and *(2) seven blindfolded sighted (BSP) participants* who do not read braille, totaling 15 users, participated. The study protocol is approved by the appropriate institutional review process. Each session lasts 45–60 minutes and consists of two tasks. *(1) Object recognition:* BLV participants explore each tactile graphic without prior information and attempt to identify the depicted object, both with and without braille annotations. *(2) Texture quality rating:* All participants are informed of the target texture and rate how well the tactile surface matches their expectation on a 5-point Likert scale. To mitigate bias, condition labels were not disclosed to participants, the three artifacts per prompt were shown in randomized order, and plate dimensions were standardized so that no condition was identifiable by its form factor.

Results. Our method receives higher ratings than all baselines across both participant groups (Table 2), demonstrating the benefit of explicitly modeling tactile

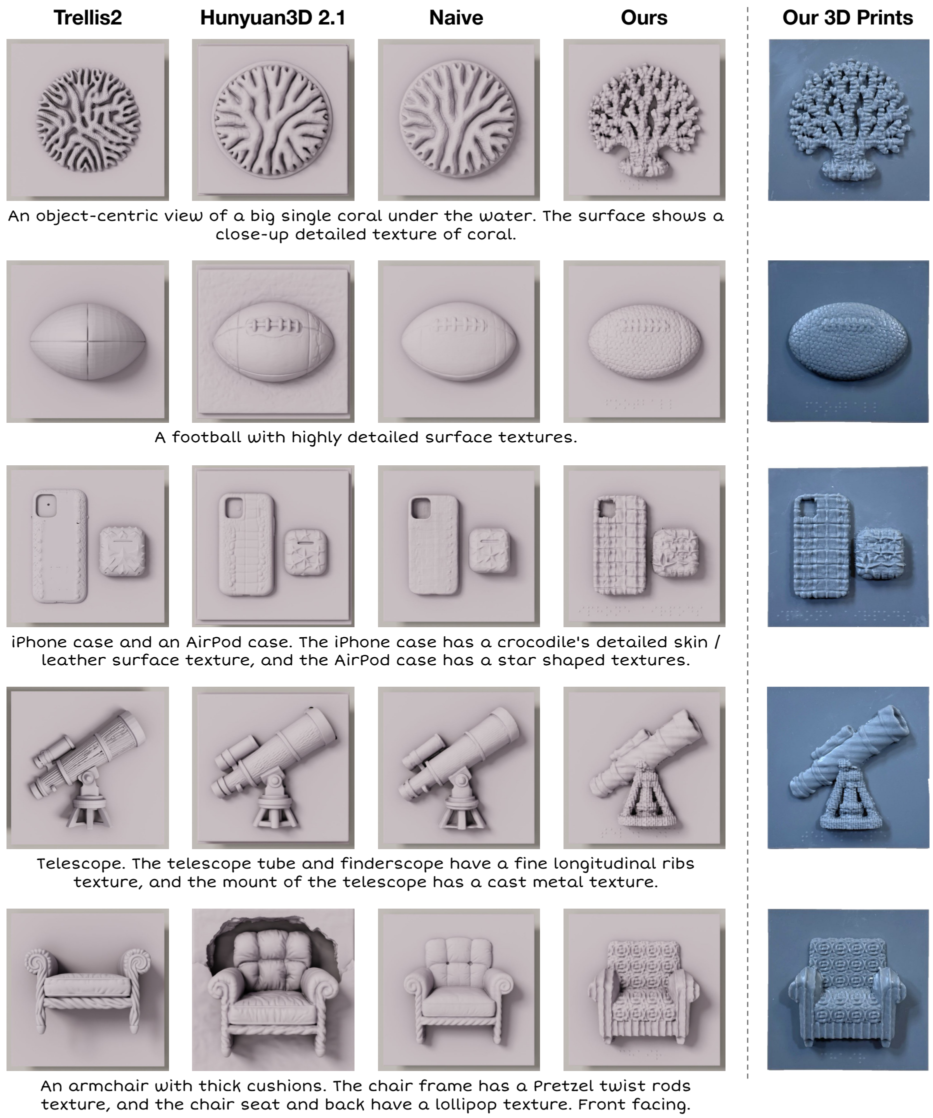

Fig. 6: Qualitative comparison of 2.5D reliefs. Given the same text prompt, our method produces consistent plate geometry with readable braille and fine-grained textures aligned with the prompt. In contrast, baselines often lack texture detail (overly smooth corals in Row 1), produce non-fabricable surfaces (fragmented surface of telescope in Row 4 Col 1), or generate broken geometry (chair intersecting with the base plate in Row 5 Col 2). Please zoom in for finer details, including the braille.

textures. To verify statistical robustness, we ran one-sided paired t-tests over a pool of 240 paired observations (16 prompts $\times$ 15 participants). We confirm a statistically significant preference for our method over both Naive and Hunyuan3D baselines (Ours $>$ Naive: mean diff. $= 0.32$, $p = 0.008 \ll 0.05$, Cohen's $d = 0.16$; Ours $>$ Hunyuan3D: mean diff. $= 0.34$, $p = 0.006 \ll 0.05$, Cohen's $d = 0.16$). Adding braille increases object recognition accuracy from $\sim 20\%$ to nearly perfect, confirming that the printed braille is readable and effective. We additionally compare our 3D-printed outputs against expert-designed swell-paper graphics and provide in-depth qualitative analysis in Appendix C and D.

5 Discussion

We have presented an automated method for generating 3D-printable tactile graphics directly from text prompts. Unlike traditional embossed or swell paper graphics, our approach produces expressive 2.5D representations that integrate geometry, tactile textures, and braille. User studies and quantitative analysis demonstrate the effectiveness of our generative graphics for physical fabrication and tactile exploration. We hope our work enables more accessible and personalized tactile graphics for the blind and low-vision community and inspires new research directions in accessibility for generative models.

Limitations. Object recognition from touch alone is inherently difficult. Our $\sim$20% recognition rate without braille is consistent with prior evidence that passive touch poorly recovers global shape [26, 35]. With standard-compliant braille, however, our generated artifacts function as a self-sufficient platform for exploring fine surface details, experiencing otherwise inaccessible materials, and enabling customized design and artistic expression through touch. Additionally, our in-person perceptual study involves 15 participants (8 BLV, 7 BSP) recruited from the local community; while the results are statistically significant, expanding to a larger and more geographically diverse BLV population remains important future work.

Acknowledgment. We thank Amin Mirzaee, Maxwell Jones, and Nupur Kumari for their helpful comments and discussion. We are also grateful to Sheng-Yu Wang for proofreading the draft. We appreciate the generous help from the Library of Accessible Media for Pennsylvanians (LAMP) and VisAbility Pittsburgh for recruiting blind and low-vision users for our studies. The project is partially supported by the Amazon Faculty Research Award, Cisco Research, Sloan Foundation, the Packard Foundation, and the IITP grant funded by the Korean Government (MSIT) (No. RS-2024-00457882, National AI Research Lab Project). Ruihan Gao is supported by the A*STAR National Science Scholarship (Ph.D.). Joonghyuk Shin and Jaesik Park are supported by the NRF grant (No. RS-2024-00405857, 50%) and IITP grant (No. RS-2021-II211343, 5% and No. RS-2024-00509257, 45%) funded by the Korea government (MSIT).

References

1. American Printing House: APH tactile graphic image library. https://imagelibrary.aph.org/portals/aphb (2026), Accessed on 2026-06-28
2. Braille Authority of North America, Canadian Braille Authority: Guidelines and Standards for Tactile Graphics, 2010. Braille Authority of North America (2011)
3. brick2face: seamless-tile-inpainting: An automatic1111 extension for making seamless tiles using stable diffusion inpainting, https://github.com/brick2face/seamless-tile-inpainting, Accessed on 2026-06-28
4. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P., Feichtenhofer, C.: SAM 3: Segment anything with concepts. arXiv preprint arXiv:2511.16719 (2025)
5. Clepper, G., McDonnell, E.J., Findlater, L., Peek, N.: "What would I want to make? Probably everything": Practices and speculations of blind and low vision tactile graphics creators. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (2025)
6. Cohen, M.F., Shade, J., Hiller, S., Deussen, O.: Wang tiles for image and texture generation. ACM Transactions on Graphics (TOG) **22**(3), 287–294 (2003)
7. Dzhurynskyi, Y., Mayik, V., Mayik, L.: Artificial intelligence-driven text-to-tactile graphics generation for visual impaired people. In: International Workshop of IT Professionals on Artificial Intelligence (2024)
8. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: ACM SIGGRAPH Conference Papers (2001)
9. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., Podell, D., Dockhorn, T., English, Z., Rombach, R.: Scaling rectified flow transformers for high-resolution image synthesis. In: International Conference on Machine Learning (ICML) (2024)
10. Ferro, T.J., Pawluk, D.T.V.: Automatic image conversion to tactile graphic. In: Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS) (2013)
11. Fortin, A., Cobo, L., Vernade, G.: Introducing Veo 3.1 and new creative capabilities in the Gemini API. https://developers.googleblog.com/introducing-veo-3-1-and-new-creative-capabilities-in-the-gemini-api/ (2025), Accessed on 2026-06-28
12. Frühstück, A., Alhashim, I., Wonka, P.: TileGAN: synthesis of large-scale non-homogeneous textures. ACM Transactions on Graphics (TOG) **38**(4), 1–11 (2019)
13. Gao, L., Zhang, Y.W., Wei, M., Liu, H., Chen, Y., Qiu, H., Zhang, C.: MonoRelief: Recovering 2.5D relief from a single image. IEEE Transactions on Visualization and Computer Graphics **31**(10), 7986–7997 (2025)
14. Gao, R., Deng, K., Yang, G., Yuan, W., Zhu, J.Y.: Tactile DreamFusion: Exploiting tactile sensing for 3D generation. In: Advances in Neural Information Processing Systems (2024)
15. Gao, R., Yuan, W., Zhu, J.Y.: Controllable visual-tactile synthesis. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
16. Gardner, J.A.: Tactile graphics: an overview and resource guide. Information Technology and Disabilities **3**(4) (1996)

17. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
18. Google Imagen Team: Imagen 3. arXiv preprint arXiv:2408.07009 (2024)
19. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (2020)
20. Hu, A., Desai, N., Abu Alhaija, H., Kim, S.W., Shugrina, M.: Diffusion texture painting. In: ACM SIGGRAPH Conference Papers (2024)
21. Ji, Z., Feng, W., Sun, X., Qin, F., Wang, Y., Zhang, Y.W., Ma, W.: ReliefNet: Fast bas-relief generation from 3D scenes. Computer-Aided Design **130**, 102928 (2021)
22. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
23. Khan, A., Choubineh, A., Shaaban, M.A., Akkasi, A., Komeili, M.: TactileNet: Bridging the accessibility gap with AI-generated tactile graphics for individuals with vision impairment. arXiv preprint arXiv:2504.04722 (2025)
24. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114 (2013)
25. Krufka, S.E., Barner, K.E.: A user study on tactile graphic generation methods. Behaviour & Information Technology **25**(4), 297–311 (2006)
26. Kuroki, S., Hamazaki, T., Kajimoto, H., Nishida, S.: Passive human touch cannot recognize the shape of a pattern imprinted on the fingertip. iScience **28**(5) (2025)
27. Lai, Z., Zhao, Y., Liu, H., Zhao, Z., Lin, Q., Shi, H., Yang, X., Yang, M., Yang, S., Feng, Y., et al.: Hunyuan3D 2.5: Towards high-fidelity 3D assets generation with ultimate details. arXiv preprint arXiv:2506.16504 (2025)
28. Levin, E., Fried, O.: Differential Diffusion: Giving each pixel its strength. In: Computer Graphics Forum (2025)
29. Lin, C.H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., Kreis, K., Fidler, S., Liu, M.Y., Lin, T.Y.: Magic3D: High-resolution text-to-3D content creation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
30. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2019)
31. Luo, S., Hu, W.: Diffusion probabilistic models for 3D point cloud generation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
32. Madar, O., Fried, O.: Tiled diffusion. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025)
33. Memery, S., Cedron, O., Subr, K.: Generating parametric brdfs from natural language descriptions. In: Computer graphics forum. Wiley Online Library (2023)
34. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: SDEdit: Guided image synthesis and editing with stochastic differential equations. In: International Conference on Learning Representations (ICLR) (2022)
35. Miller, R.L., Sheinberg, D.L.: Evidence for independent processing of shape by vision and touch. eNeuro **9**(3) (2022)
36. OpenAI: Sora 2 is here. `https://openai.com/index/sora-2` (2025), Accessed on 2026-06-28
37. Pakenaite, K., Kamperou, E., Proulx, M.J., Sharma, A., Hall, P.: Pic2Tac: Creating accessible tactile images using semantic information from photographs. In: Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction (TEI) (2024)

38. Panotopoulou, A., Zhang, X., Qiu, T., Yang, X.D., Whiting, E.: Tactile line drawings for improved shape understanding in blind and visually impaired users. ACM Transactions on Graphics (TOG) **39**(4), 89–1 (2020)
39. Perkins: Tactile graphics library (2026)
40. Pesudovs, K., Vision Loss Expert Group of the Global Burden of Disease Study, GBD 2019 Blindness and Vision Impairment Collaborators: Global estimates on the number of people blind or visually impaired by cataract: a meta-analysis from 2000 to 2020. Eye (Lond) **38**(11), 2156–2172 (2024)
41. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: DreamFusion: Text-to-3D using 2D diffusion. In: International Conference on Learning Representations (ICLR) (2023)
42. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning (ICML) (2021)
43. Raisinghani, N.: Introducing Nano Banana Pro. `https://blog.google/innovation-and-ai/products/nano-banana-pro/` (2025), Accessed on 2026-06-28
44. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125 (2022)
45. Reichinger, A., Maierhofer, S., Purgathofer, W.: High-quality tactile paintings. Journal on Computing and Cultural Heritage (JOCCH) **4**(2) (2011)
46. Rodriguez-Pardo, C., Casas, D., Garces, E., Lopez-Moreno, J.: TexTile: A differentiable metric for texture tileability. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
47. Rodriguez-Pardo, C., Garces, E.: SeamlessGAN: Self-supervised synthesis of tileable texture maps. IEEE Transactions on Visualization and Computer Graphics **29**(6), 2914–2925 (2022)
48. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
49. Sartor, S., Peers, P.: Content-aware tile generation using exterior boundary inpainting. ACM Transactions on Graphics (TOG) **43**(6) (2024)
50. Shue, J.R., Chan, E.R., Po, R., Ankner, Z., Wu, J., Wetzstein, G.: 3D neural field generation using triplane diffusion. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
51. Siddiqui, Y., Alliegro, A., Artemov, A., Tommasi, T., Sirigatti, D., Rosov, V., Dai, A., Nießner, M.: MeshGPT: Generating triangle meshes with decoder-only transformers. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
52. Smith, D.W., Smothers, S.M.: The role and characteristics of tactile graphics in secondary mathematics and science textbooks in Braille. Journal of Visual Impairment & Blindness **106**(9), 543–554 (2012)
53. Tanaka, K., Tsutsui, A., Ochiai, Y.: A comparative evaluation of GenAI-assisted tactile graphics from the perspective of totally blind professionals in tactile graphic supervision. In: Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS) (2025)
54. tjm35: asymmetric-tiling-sd-webui: Asymmetric tiling for stable-diffusion-webui, `https://github.com/tjm35/asymmetric-tiling-sd-webui`, Accessed on 2026-06-28

55. Tovar-Batista, B.M., Abellán-Nebot, J.V., Gual-Ortí, J., Serrano, J.: Geometrical evaluation of additive manufacturing techniques for tactile graphic production. The International Journal of Advanced Manufacturing Technology pp. 1–14 (2025)
56. Vecchio, G., Sortino, R., Palazzo, S., Spampinato, C.: Matfuse: controllable material generation with diffusion models. In: CVPR (2024)
57. Wabiński, J., Mościcka, A.: Automatic (tactile) map generation—a systematic literature review. ISPRS International Journal of Geo-Information **8**(7) (2019)
58. Wang, R., Xu, S., Dong, Y., Deng, Y., Xiang, J., Lv, Z., Sun, G., Tong, X., Yang, J.: MoGe-2: Accurate monocular geometry with metric scale and sharp details. arXiv preprint arXiv:2507.02546 (2025)
59. Wang, S., She, Y., Romero, B., Adelson, E.: GelSight wedge: Measuring high-resolution 3D contact geometry with a compact robot finger. In: International Conference on Robotics and Automation (ICRA) (2021)
60. Wei, M., Tian, Y., Pang, W.M., Wang, C.C.L., Pang, M.Y., Wang, J., Qin, J., Heng, P.A.: Bas-relief modeling from normal layers. IEEE Transactions on Visualization and Computer Graphics **25**(4), 1651–1665 (2019)
61. Weyrich, T., Deng, J., Barnes, C., Rusinkiewicz, S., Finkelstein, A.: Digital bas-relief from 3D scenes. ACM Transactions on Graphics (TOG) **26**(3), 32–es (2007)
62. Wu, C., Li, J., Zhou, J., Lin, J., Gao, K., Yan, K., ming Yin, S., Bai, S., Xu, X., Chen, Y., Chen, Y., Tang, Z., Zhang, Z., Wang, Z., Yang, A., Yu, B., Cheng, C., Liu, D., Li, D., Zhang, H., Meng, H., Wei, H., Ni, J., Chen, K., Cao, K., Peng, L., Qu, L., Wu, M., Wang, P., Yu, S., Wen, T., Feng, W., Xu, X., Wang, Y., Zhang, Y., Zhu, Y., Wu, Y., Cai, Y., Liu, Z.: Qwen-Image technical report. arXiv preprint arXiv:2508.02324 (2025)
63. Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J.: Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In: Advances in Neural Information Processing Systems (2016)
64. Xiang, J., Chen, X., Xu, S., Wang, R., Lv, Z., Deng, Y., Zhu, H., Dong, Y., Zhao, H., Yuan, N.J., others: Native and compact structured latents for 3D generation. arXiv preprint arXiv:2512.14692 (2025)
65. Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., Yang, J.: Structured 3D latents for scalable and versatile 3D generation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025)
66. Xiong, B., Liu, J., Hu, J., Wu, C., Wu, J., Liu, X., Zhao, C., Ding, E., Lian, Z.: Texgaussian: Generating high-quality pbr material via octree-based 3d gaussian splatting. In: CVPR (2025)
67. Yang, S., Yang, M., Feng, Y., Huang, X., Zhang, S., He, Z., Luo, D., Liu, H., Zhao, Y., Lin, Q., Lai, Z., Yang, X., Shi, H., Zhao, Z., Zhang, B., Yan, H., Wang, L., Liu, S., Zhang, J., Chen, M., Dong, L., Jia, Y., Cai, Y., Yu, J., Tang, Y., Guo, D., Yu, J., Zhang, H., Ye, Z., He, P., Wu, R., Wei, S., Zhang, C., Tan, Y., Sun, Y., Niu, L., Huang, S., Zheng, B., Liu, S., Chen, S., Yuan, X., Yang, X., Liu, K., Zhu, J., Chen, P., Liu, T., Wang, D., Liu, Y., Linus, Jiang, J., Huang, J., Guo, C.: Hunyuan3D 2.1: From images to high-fidelity 3D assets with production-ready PBR material. arXiv preprint arXiv:2506.15442 (2025)
68. Yang, X., Shi, H., Zhang, B., Yang, F., Wang, J., Zhao, H., Liu, X., Wang, X., Lin, Q., Yu, J., et al.: Hunyuan3D 1.0: A unified framework for text-to-3D and image-to-3D generation. arXiv preprint arXiv:2411.02293 (2024)
69. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)

70. Yuan, W., Dong, S., Adelson, E.H.: GelSight: High-resolution robot tactile sensors for estimating geometry and force. Sensors (2017)
71. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
72. Zhang, Y.W., Luo, P., Zhou, H., Ji, Z., Liu, H., Chen, Y., Zhang, C.: Neural modeling of portrait bas-relief from a single photograph. IEEE Transactions on Visualization and Computer Graphics **29**(12), 5008–5019 (2023)
73. Zhang, Y.W., Zhou, Y.Q., Li, X.L., Liu, H., Zhang, L.L.: Bas-relief generation and shape editing through gradient-based mesh deformation. IEEE Transactions on Visualization and Computer Graphics **21**(3), 328–338 (2015)
74. Zhang, Y., Liu, Y., Xie, Z., Yang, L., Liu, Z., Yang, M., Zhang, R., Kou, Q., Lin, C., Wang, W., et al.: Dreammat: High-quality pbr material generation with geometry-and light-aware diffusion models. TOG **43**(4), 1–18 (2024)
75. Zhao, Z., Lai, Z., Lin, Q., Zhao, Y., Liu, H., Yang, S., Feng, Y., Yang, M., Zhang, S., Yang, X., et al.: Hunyuan3D 2.0: Scaling diffusion models for high resolution textured 3D assets generation. arXiv preprint arXiv:2501.12202 (2025)