MaTe: Images Are All You Need for Material Transfer via Diffusion Transformer

Supplementary Material

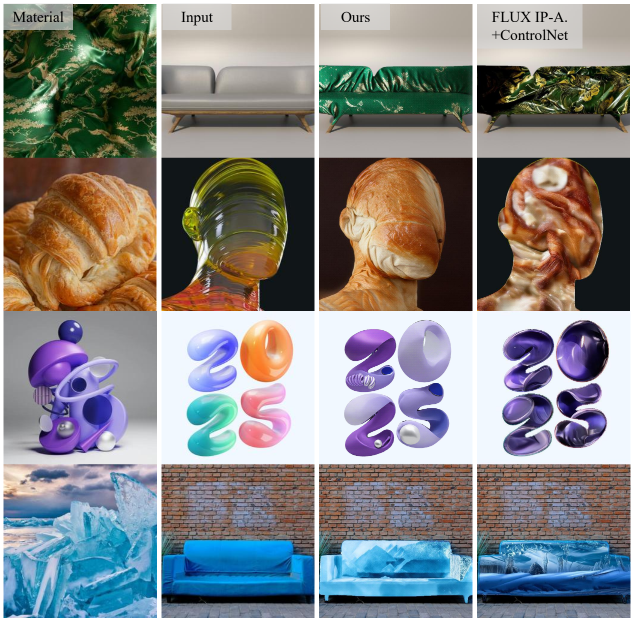

Figure 1. Additional comparisons with the FLUX versions of the IP-Adapter + ControlNet.

A. Comparison with FLUX-ControlNet

We build an IP-Adapter+ControlNet based on FLUX as shown in Fig. 1. For public quantitative results, Flux.1-Depth-dev-lora has a clear advantage over FLUX.1-dev-ControlNet-Depth in terms of model size (↓ 1.24/3.19 GB), performance (ELO score ↑ 1098/931), and user preference (downloads last month ↑ 36,170/1,332). Therefore, we choose Flux.1-Depth-dev-lora to control the depth information from the perspective of comprehensive advantages.

B. User Study Details

To evaluate the effectiveness of our method, we conducted a user study comparing our approach with SOTA material transfer methods, *i.e.*, MaterialFusion [3], ZeST[2], IP-Adapter SDXL [8], IP-Adapter SD1.5 [8], ProSpect [10], and U-VAP [7].

In the user study, the participants were informed that the task of material transfer is to generate a new image where the material of the object is the same as that in the reference artistic image and also adheres to the provided content image. A total of 16 participants with experience in the vision and graphics field took part in the survey that consisted of 34 sets of comparisons. In each set, the participants were presented with a reference image of material (e.g., "a photo of a bread covered with sesame seeds") and a content image (e.g., "a photo of a stone lion"). They were asked to select the option that best met the task objective from 7 randomly arranged options, with each option corresponding to a method.

The experiment involved three evaluation metrics: **texture**, **structure**, and **overall visual result**. Put simply, for each set of comparison images (like the previous example of a stone lion and bread), the users needed to select the following respectively: (1) The image with the best transfer effect (take the texture of the stone lion in the example, the

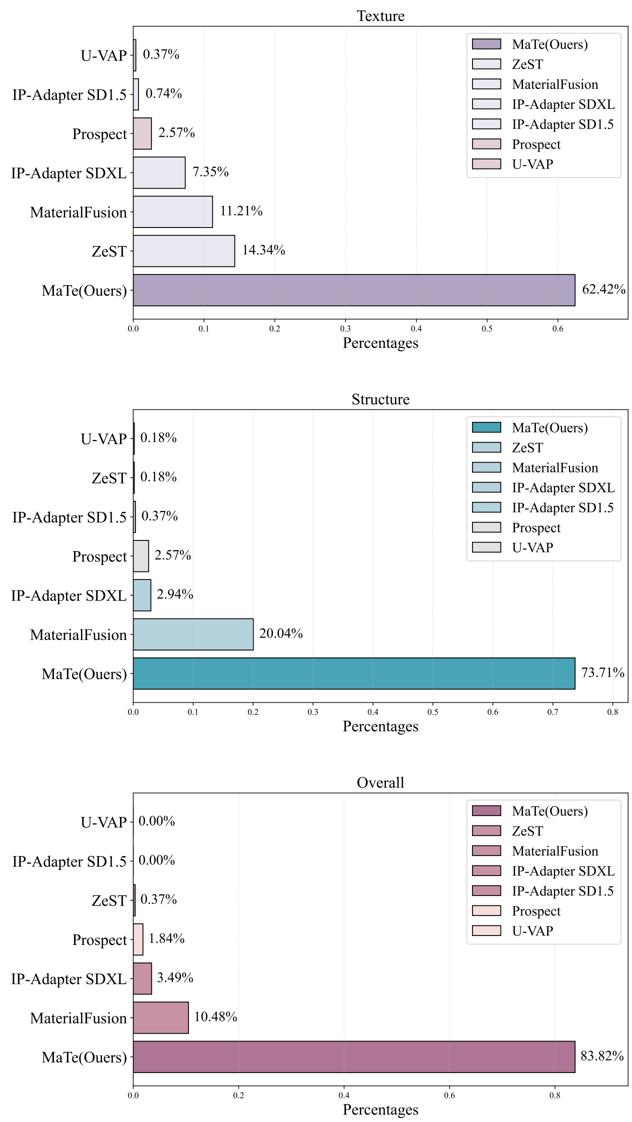

Figure 2. Visualization of vote percentages of different methods in three user-study aspects.

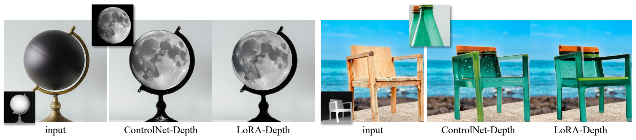

Figure 3. Ablation study with Flux.1 ControlNet depth version.

one that most resembles that of the bread). (2) The image with the best structure preservation effect (in the example, the one that allows people to recognize the original structure of the stone lion most easily). (3) The image with the best overall effect (in the example, the one that best integrates the texture of the bread and maintains the original shape and structure of the stone lion).

In total, we collected 1,632 voting results regarding the evaluation of the three aspects. We visualized the statistical results in Fig. 2, with different colors representing fine-tuning-based methods, ControlNet-and-Adapter-based methods, and our MaTe respectively. The proportion of various methods favored by users is relatively consistent with their performance results in the quantitative experiments. Our MaTe exhibited better performance in all aspects when compared with other methods. In terms of texture, MaTe received a preference rate of 62.42%, in structure it was 73.71%, and for the overall evaluation, it reached 83.82%.

Notably, the settings cover a wide range of material conditions, from smooth and rigid materials such as cola ice cubes to soft materials like the fabric of a pink down jacket. Additionally, they encompass various input condition images, such as sculptures and apples. These diverse settings pose a challenge to the adaptability of these methods and enable a better validation of the generalization ability of different methods across multiple scenarios.

C. More Ablation Study

We test the ablation of replacing the depth control architecture with ControlNet, and the results are shown in Fig. 3. It can be seen that there is no significant difference in the final results when using either ControlNet [9] or LoRA Depth [4]. Therefore, we choose the FLUX.1-Depth-dev-lora [5] model with 622.13M parameters, rather than the FLUX.1-dev-ControlNet-Depth [6] model with 1593.37M parameters. In addition, we have tested the performance of the model on traditional PBR materials, and the results are shown in Fig. 4. This demonstrates the broad adaptability of our method to traditional materials.

D. Cropped Patch Material

We demonstrate the impact of cropping material patches on the results in Fig. 5. The bias matrix design keeps material self-attention weights constant with a zero main diagonal

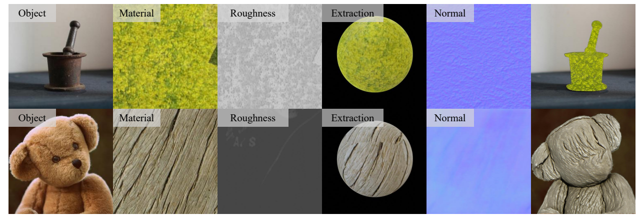

Figure 4. Adaption on traditional material transfer textures.

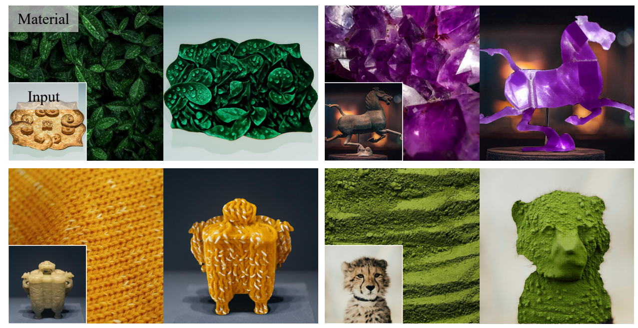

Figure 5. Cropped patch of the material.

block, focusing attention on the material directly and eliminating the need for cropping.

E. More Failure Cases

As shown in Fig. 6 (a), complex environments introduce noise into the mask. Red reflections appear as transparent highlights. (b) Complex materials are challenging to fully display in confined spaces. (c) SAM2's limits may cause incomplete object coverage. (d) Depth maps lacking fine details, like eyes, lead to missing geometric constraints. The challenges you mentioned are also difficult for baseline works [2, 3].

F. MTB Dataset

To assess the effectiveness and performance of material transfer methods, we have created a free open-source dataset called the Material Transfer Benchmark (MTB). This dataset is designed to provide a collection of 30 typical material styles (Fig. 7), including metals, ceramics, stone sculptures, fabrics, fibers, food items, gemstones, crystals, glass, leather, and plastics, covering physical properties such as luster, reflectivity, oxidation marks, softness, transparency, fragility, and smoothness. The diversity of these materials ensures that tests based on the MTB dataset can cover the rendering requirements of various types of materials, thus providing a solid foundation for the study of material transfer methods. Regarding the input objects, we have also constructed a collection of 30 typical object types. As shown in Fig. 8, this includes categories such as an-

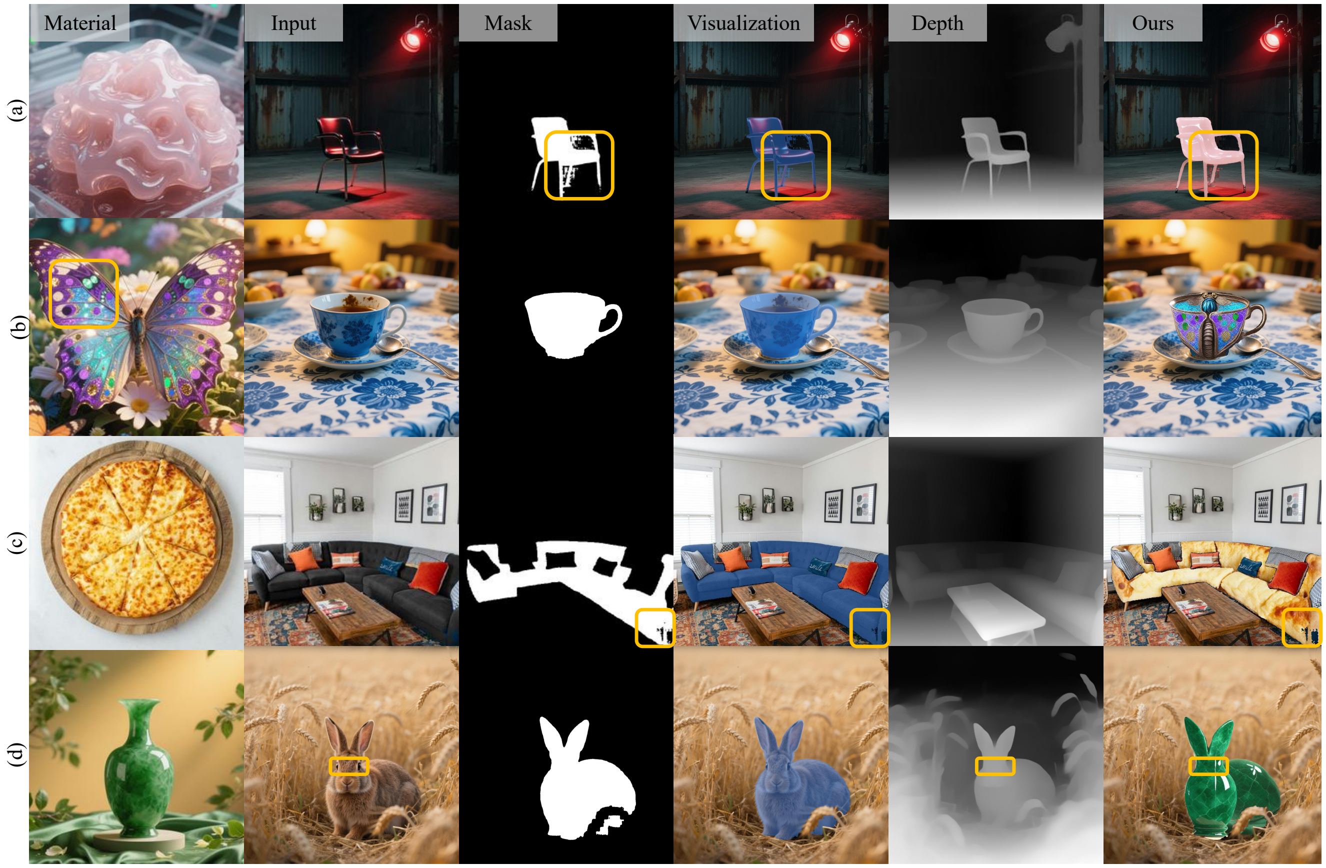

Figure 6. More failure cases about extreme lighting, complex materials, segmentation issue, and structures absence.

tiques, sculptures, furniture, decorations, food items, toys, cameras, fruits, and digital art, thereby testing whether the material transfer method can adapt to various application scenarios.

G. Additional Qualitative Results

To demonstrate the potential applications of MaTe, we conduct extensive experiments across multiple domains, including cultural relic restoration, toy, furniture, and product design. As shown in Figs. 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18, MaTe has provided efficient and creative solutions for practical applications.

For instance, in the field of cultural relic restoration, MaTe can endow faded or slightly damaged cultural relics with new colors and textures, significantly reducing the workload of cultural relic workers during the conceptualization and design phases (Figs. 9 and 10). In toy design, MaTe can quickly apply various material effects to toy models, helping designers iterate and optimize their designs more rapidly (Fig. 11). Additionally, in furniture (Fig. 12) design and product (Figs. 13 and 14) design, MaTe has demonstrated its robust functionality by providing realistic material effects for objects of different shapes and materials. During the experimental process, we paid particular attention to the robustness of the method. To this end, we designed two sets of experiments: one involved transferring different materials to a single object (Figs. 15 and 16), while the other involved transferring a single material to different objects (Figs. 17 and 18). Through these experiments, we verified the stability and reliability of MaTe in handling complex scenarios and diverse requirements.

Through these experiments, we not only showcased the application potential of MaTe across different fields but also demonstrated its robustness when facing a variety of tasks. These results indicate that MaTe, as an innovative material transfer tool, can provide strong support for design and restoration work in relevant fields.

H. Text Prompts for Baselines

To ensure semantic consistency across different methodologies, we have meticulously designed corresponding prompts for each approach while maintaining equivalent semantic expressions, as shown in Table 1. The table of U-VAP only shows the prompts used in the reasoning stage.

Among them, the target description and non-target description are generated by GPT [1]. For example, for "A photo of a chair object made of bread material", the target description is "a chair object made of metal material" and the non-target description is "a dog object made of bread material".

I. Detailed Configuration of the Method and Baselines

All experiments comparing the methods were performed using the official repository provided by the authors. The corresponding code implementations and inference-specific parameters for each method are detailed in Table 2.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 4

[2] Ta-Ying Cheng, Prafull Sharma, Andrew Markham, Niki Trigoni, and Varun Jampani. Zest: Zero-shot material transfer from a single image. In *European Conference on Computer Vision*, pages 370–386. Springer, 2024. 1, 2

[3] Kamil Garifullin, Maxim Nikolaev, Andrey Kuznetsov, and Aibek Alanov. Materialfusion: High-quality, zero-shot, and controllable material transfer with diffusion models. *arXiv preprint arXiv:2502.06606*, 2025. 1, 2

[4] Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations*. 2

[5] Black Forest Labs. Flux.1-depth-dev-lora. https://huggingface.co/black-forest-labs/FLUX.1-Depth-dev-lora, . 2

[6] Black Forest Labs. Flux.1-dev-controlnet-depth. https://huggingface.co/Shakker-Labs/FLUX.1-dev-ControlNet-Depth, . 2

[7] You Wu, Kean Liu, Xiaoyue Mi, Fan Tang, Juan Cao, and Jintao Li. U-vap: User-specified visual appearance personalization via decoupled self augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9482–9491, 2024. 1

[8] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arxiv:2308.06721*, 2023. 1

[9] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2

[10] Yuxin Zhang, Weiming Dong, Fan Tang, Nisha Huang, Haibin Huang, Chongyang Ma, Tong-Yee Lee, Oliver Deussen, and Changsheng Xu. Prospect: Prompt spectrum for attribute-aware personalization of diffusion models. *ACM Transactions on Graphics (ToG)*, 42(6):1–14, 2023. 1

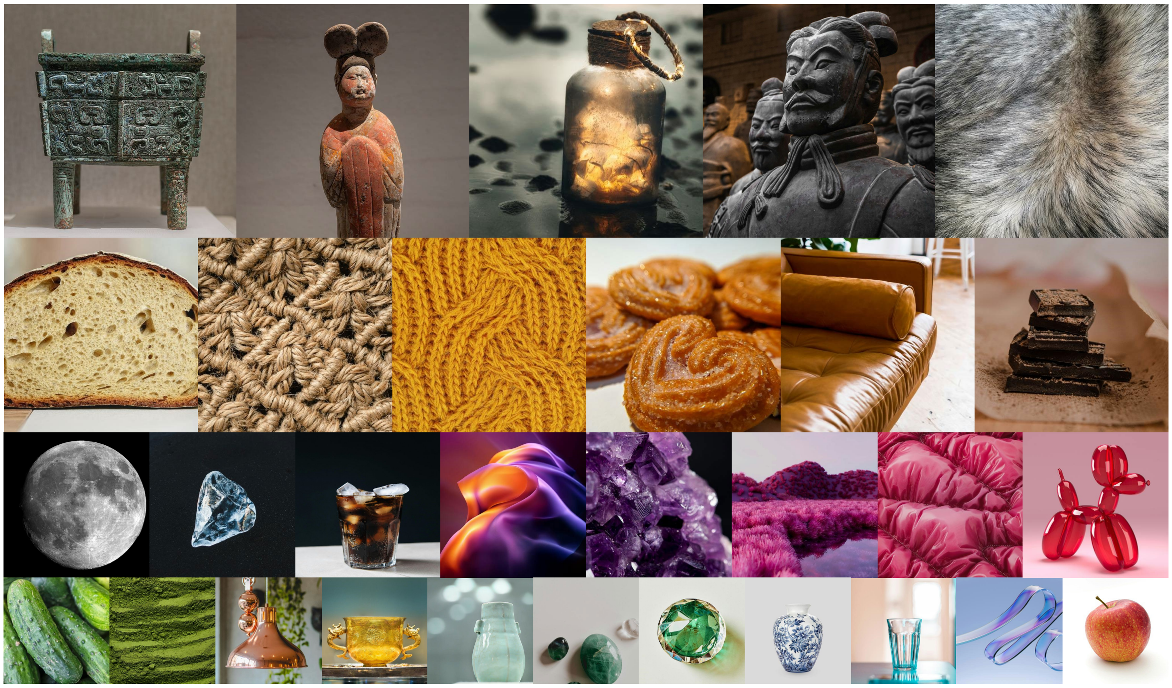

Figure 7. Material images of MTB.

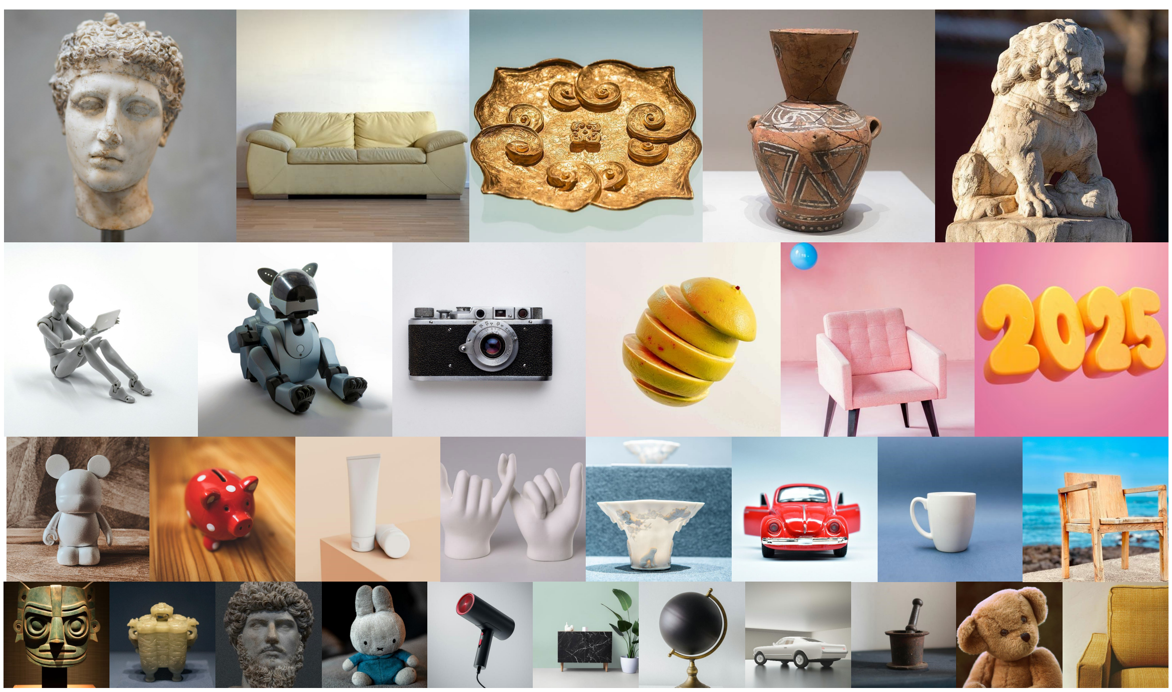

Figure 8. Target objects of MTB.

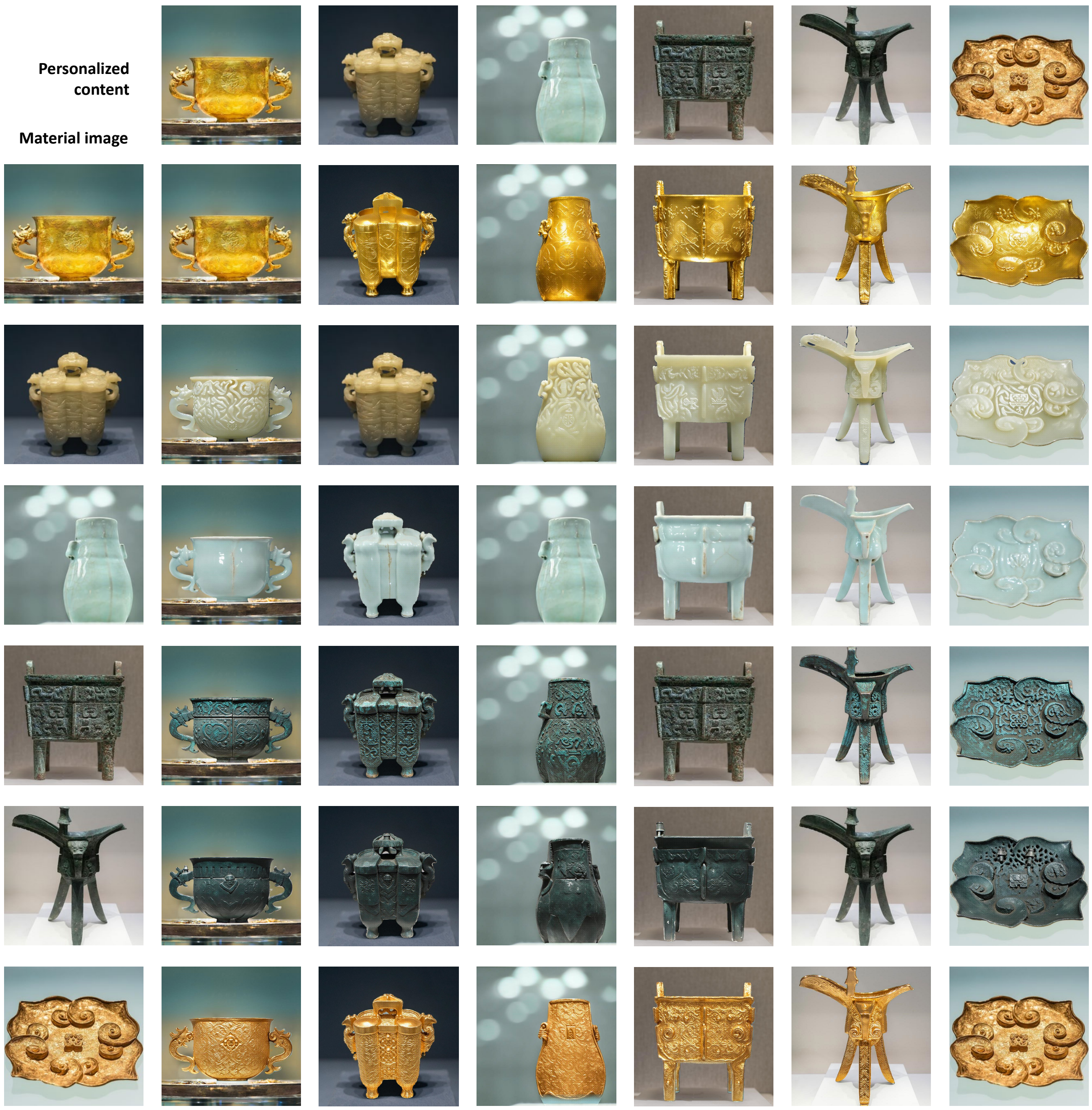

Figure 9. The potential of **MaTe** in cultural relics restoration. Through the material transfer capabilities of **MaTe**, faded or slightly damaged cultural relics can be recolored, significantly reducing the workload of cultural relics workers in conceptualization and design.

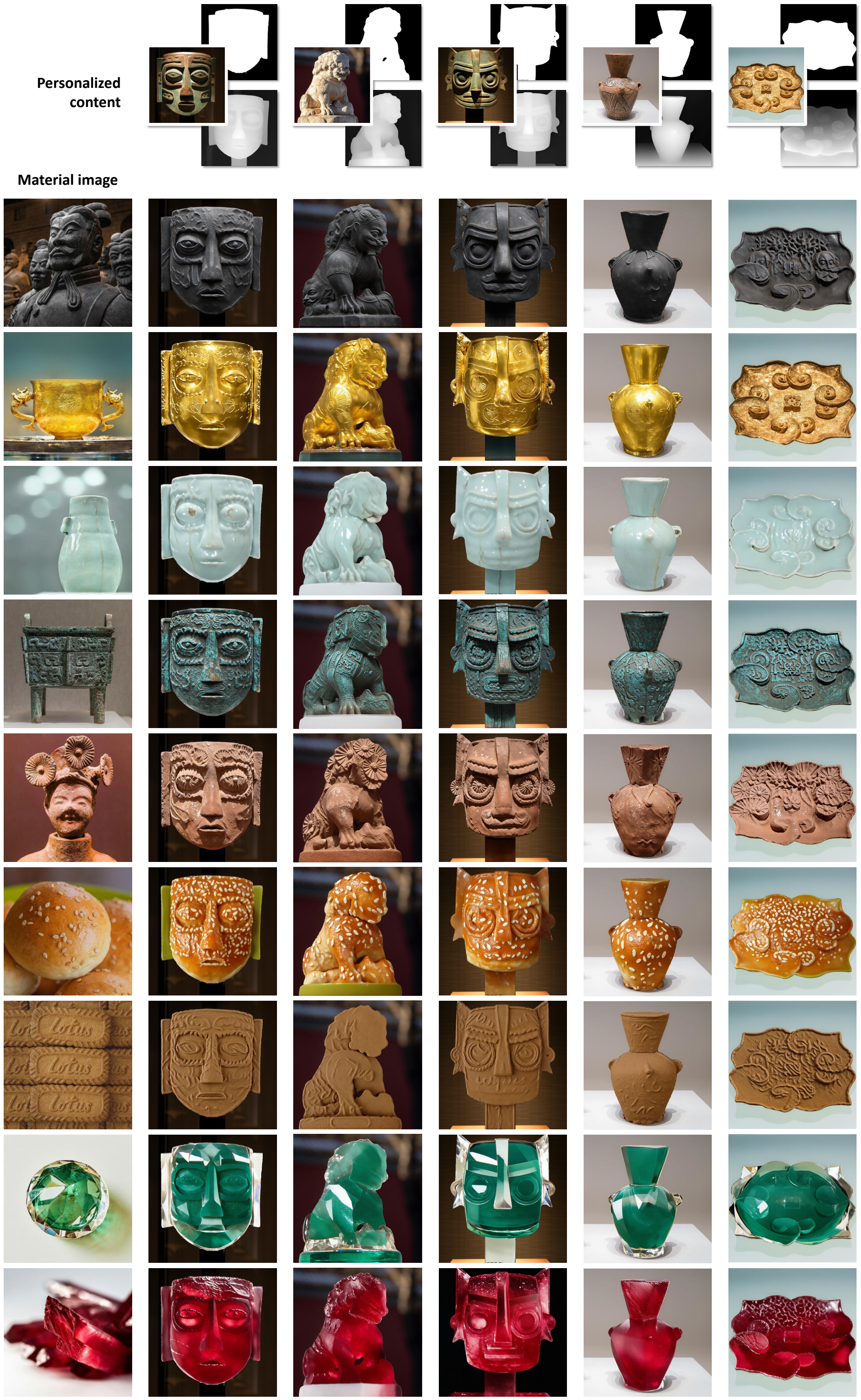

Figure 10. We demonstrate the results of our method **MaTe** on images collected from the internet. Using five cultural relics as input object images, we showcase the possibilities of various material transfers.

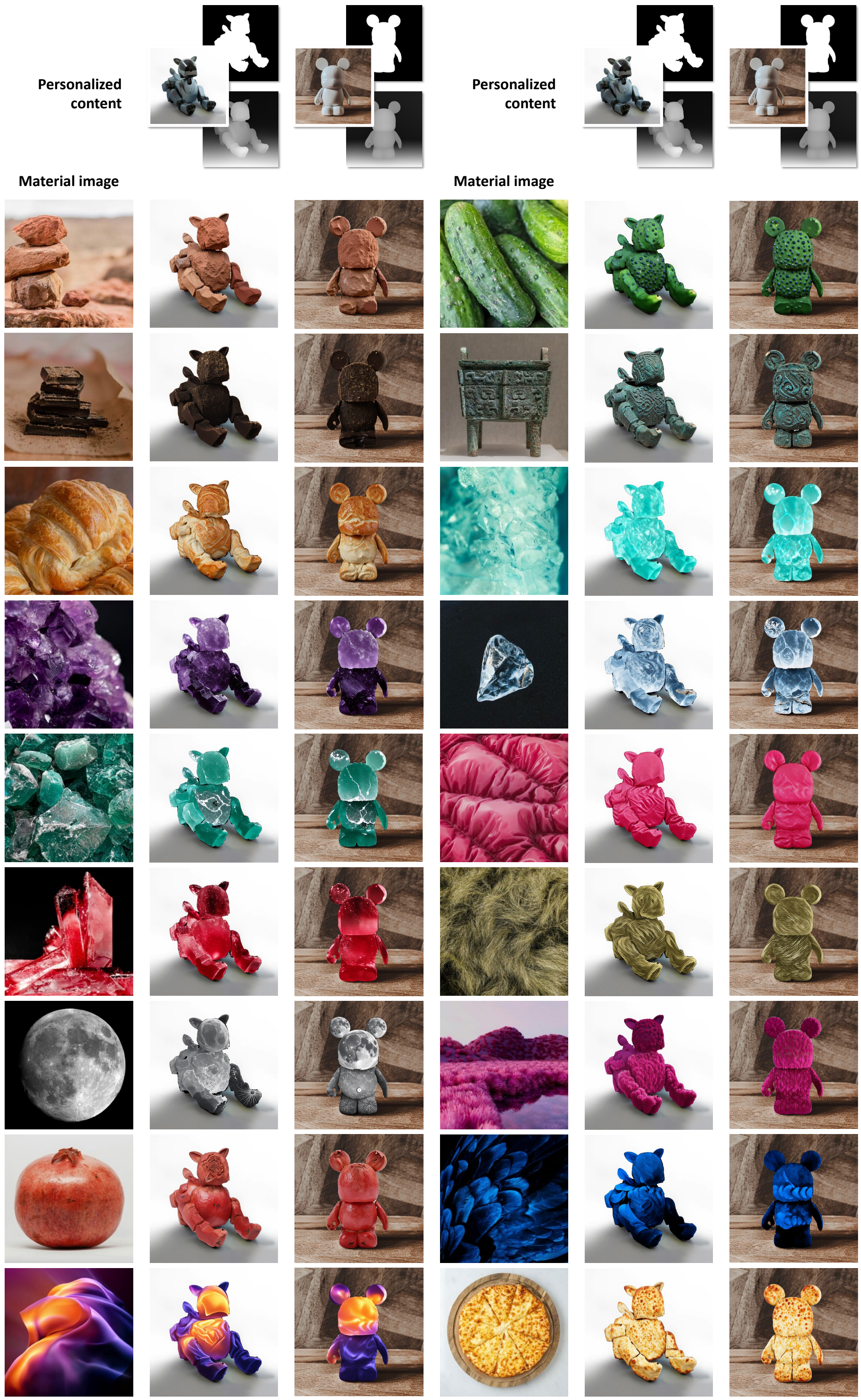

Figure 11. The application of **MaTe**. A demonstration of the potential for assisting in the appearance design of toys.

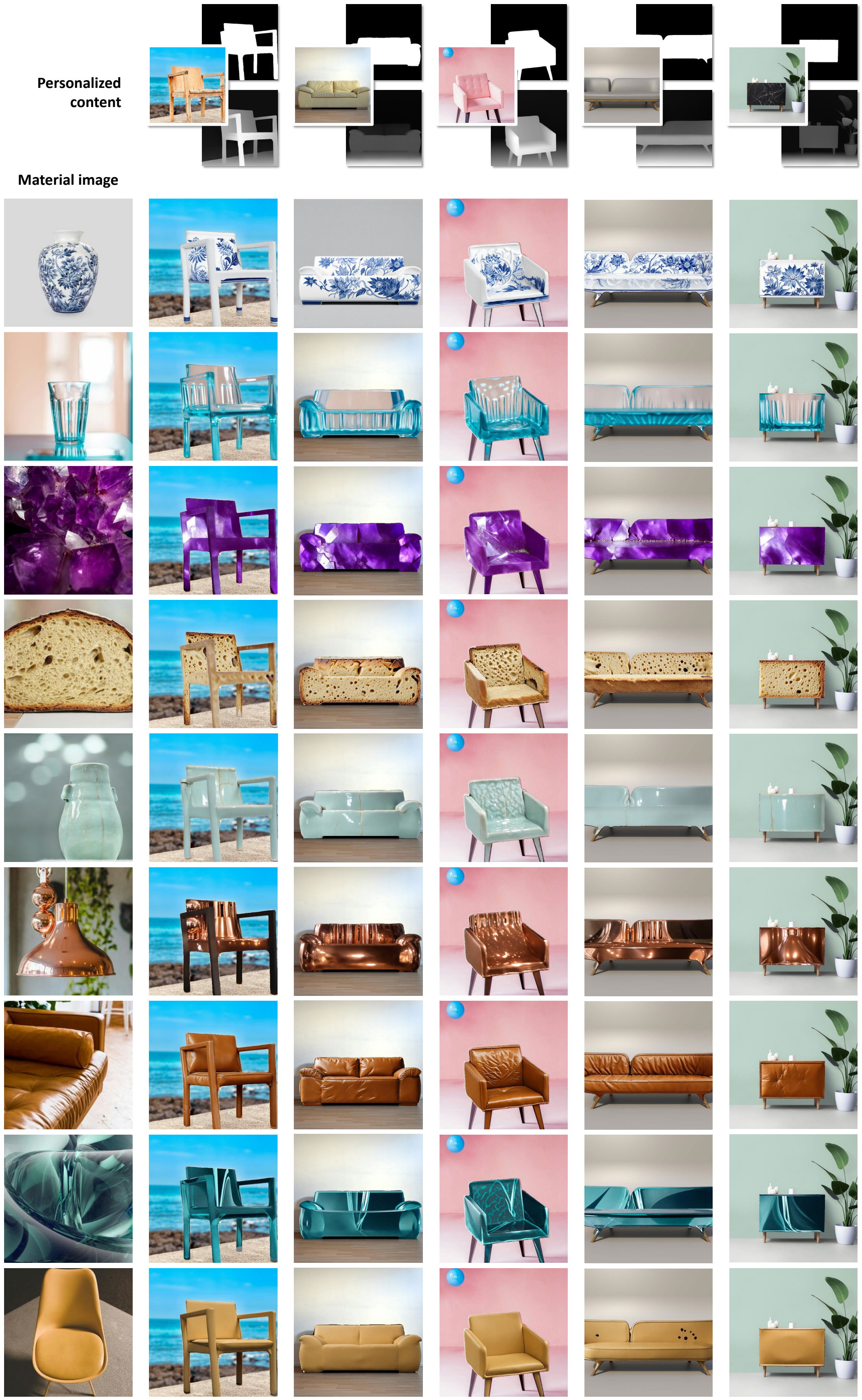

Figure 12. The potential of **MaTe** in assisting furniture design. Leveraging its material transfer capabilities, **MaTe** can endow any furniture surface with new material effects, thereby facilitating interior decoration and design.

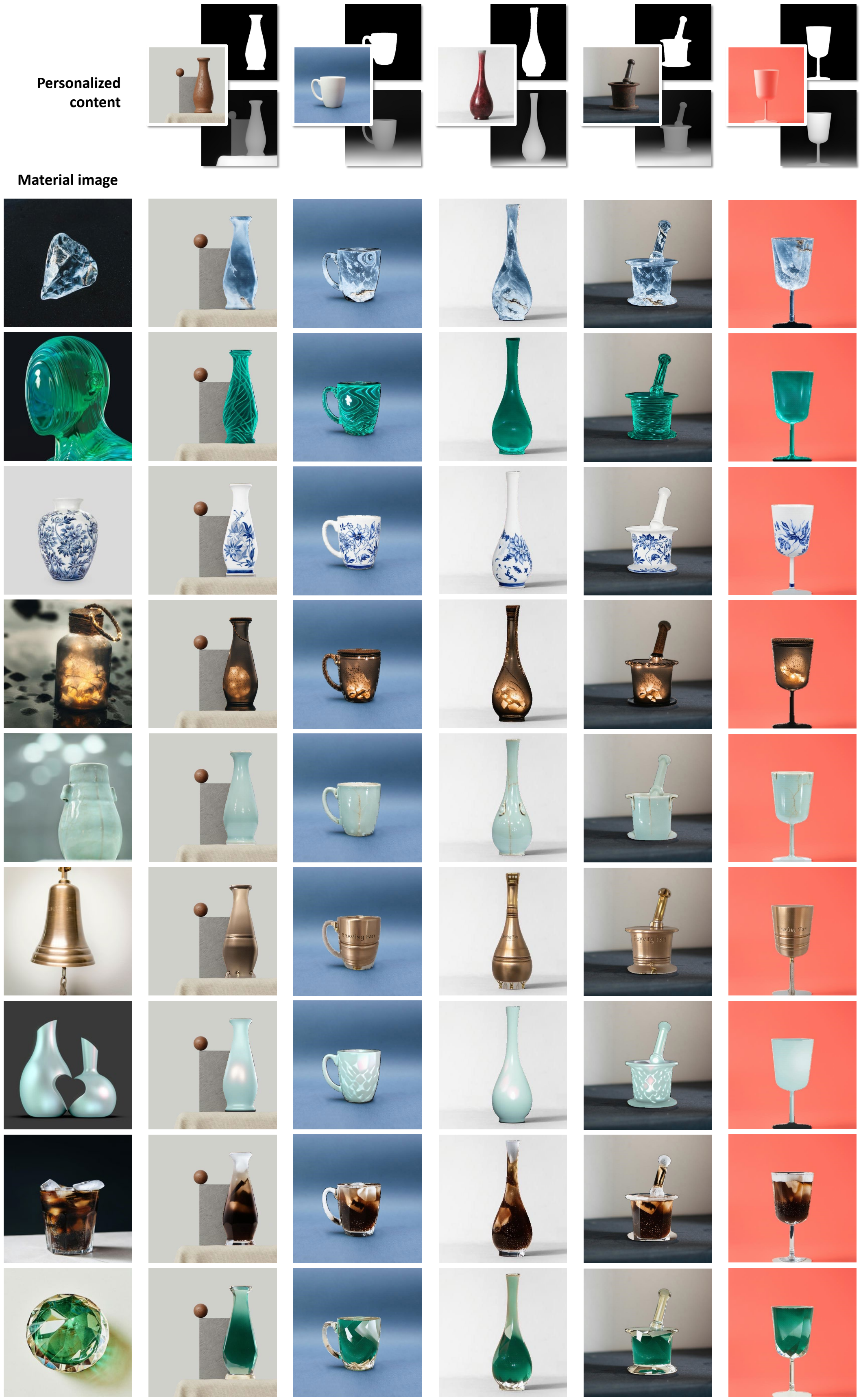

Figure 13. The performance of **MaTe** in transferring textures onto smooth surfaces.

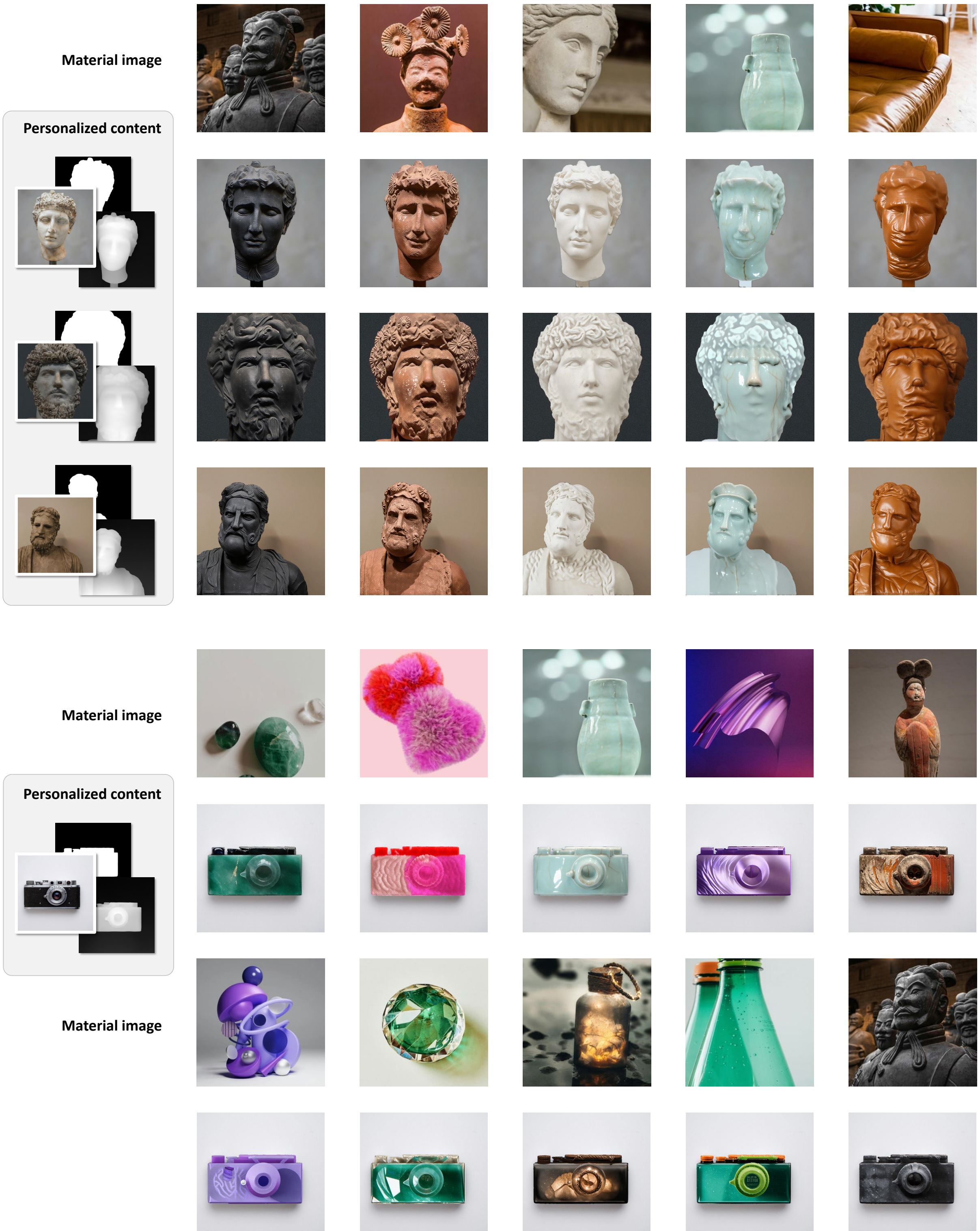

Figure 14. The upper half of the image shows the texture transfer performance on complex content such as human faces. The lower half of the image shows the texture transfer performance on the camera.

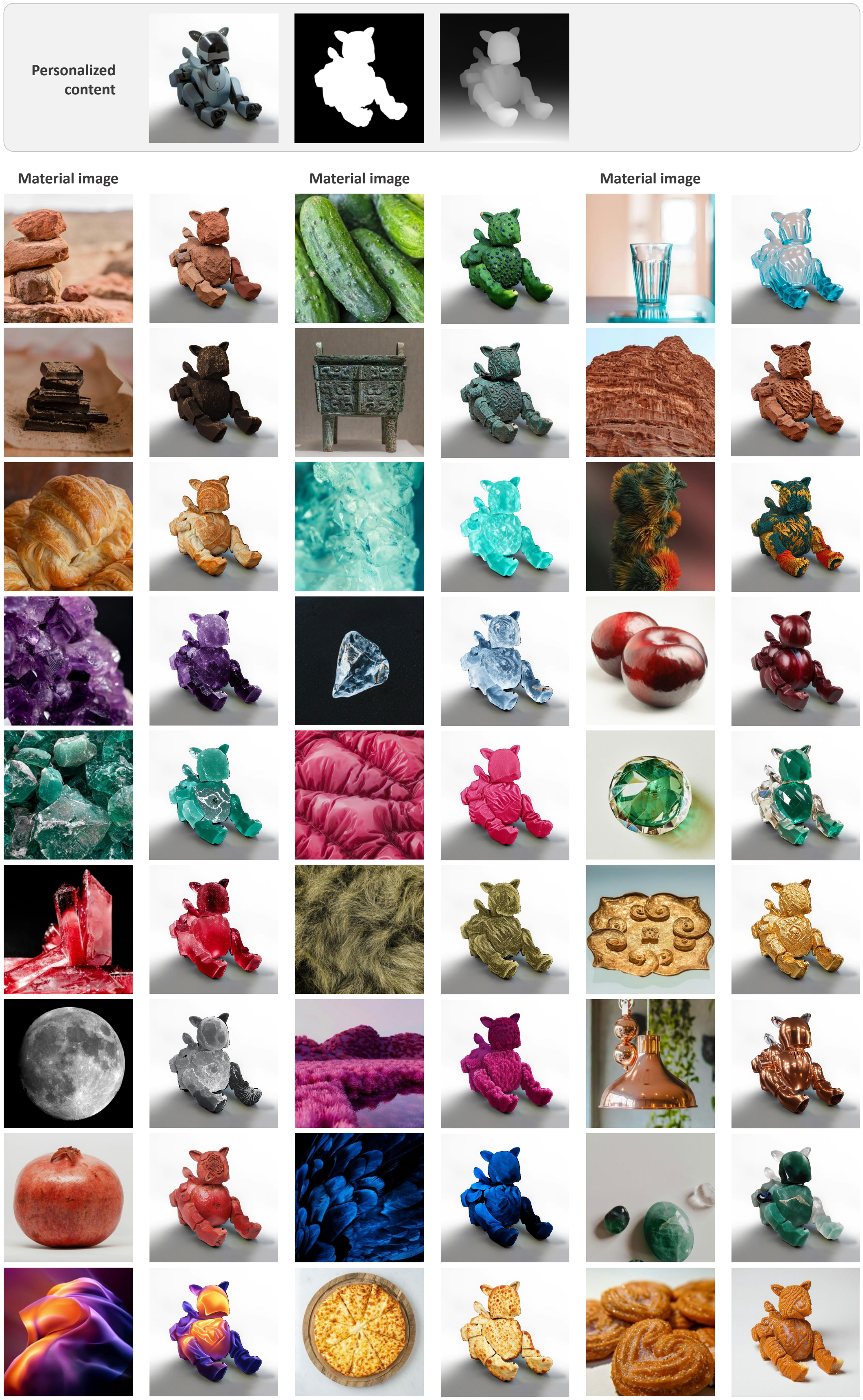

Figure 15. A collection of material transfer results on a single object with various textures.

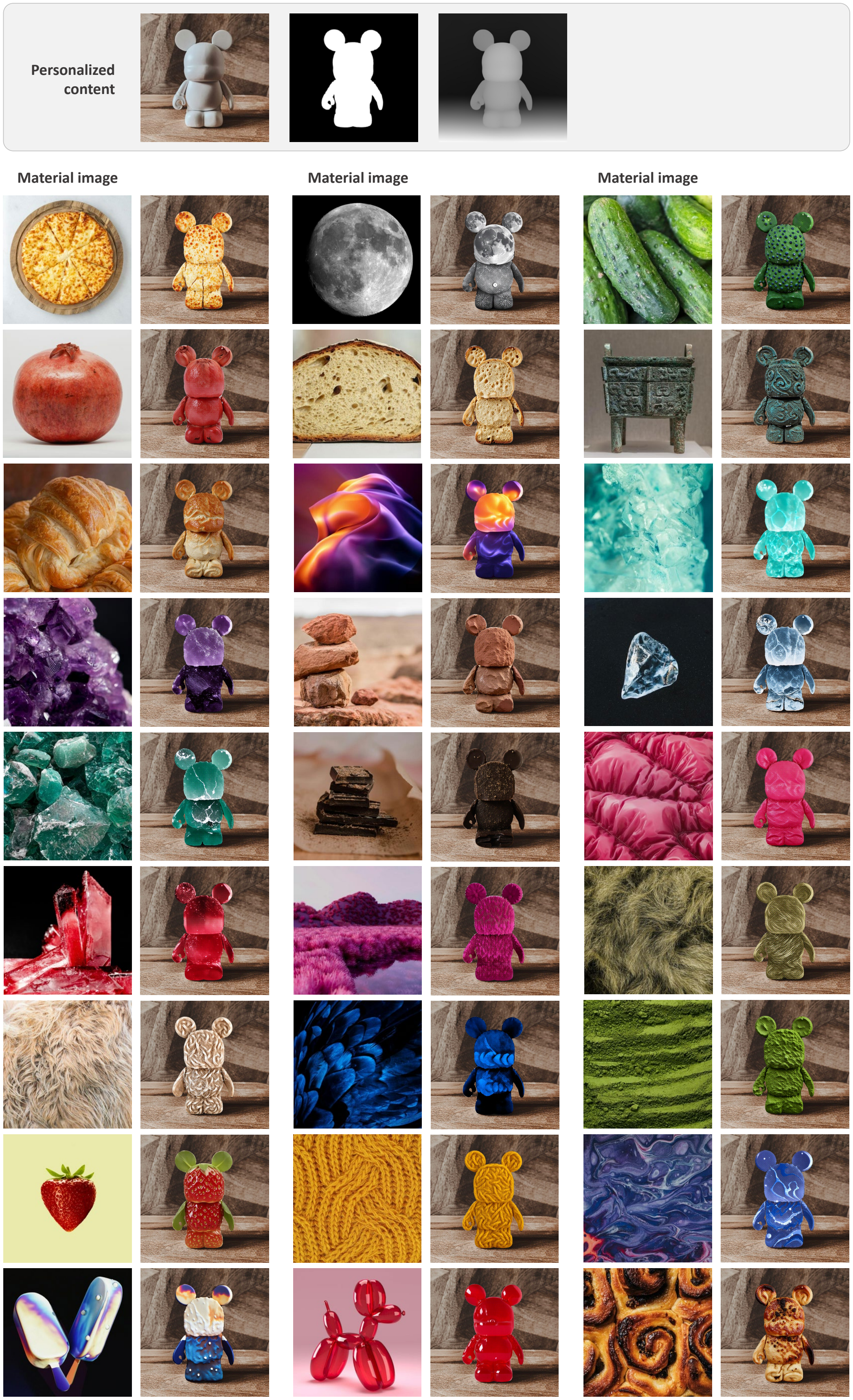

Figure 16. A collection of material transfer results on a single object with various textures.

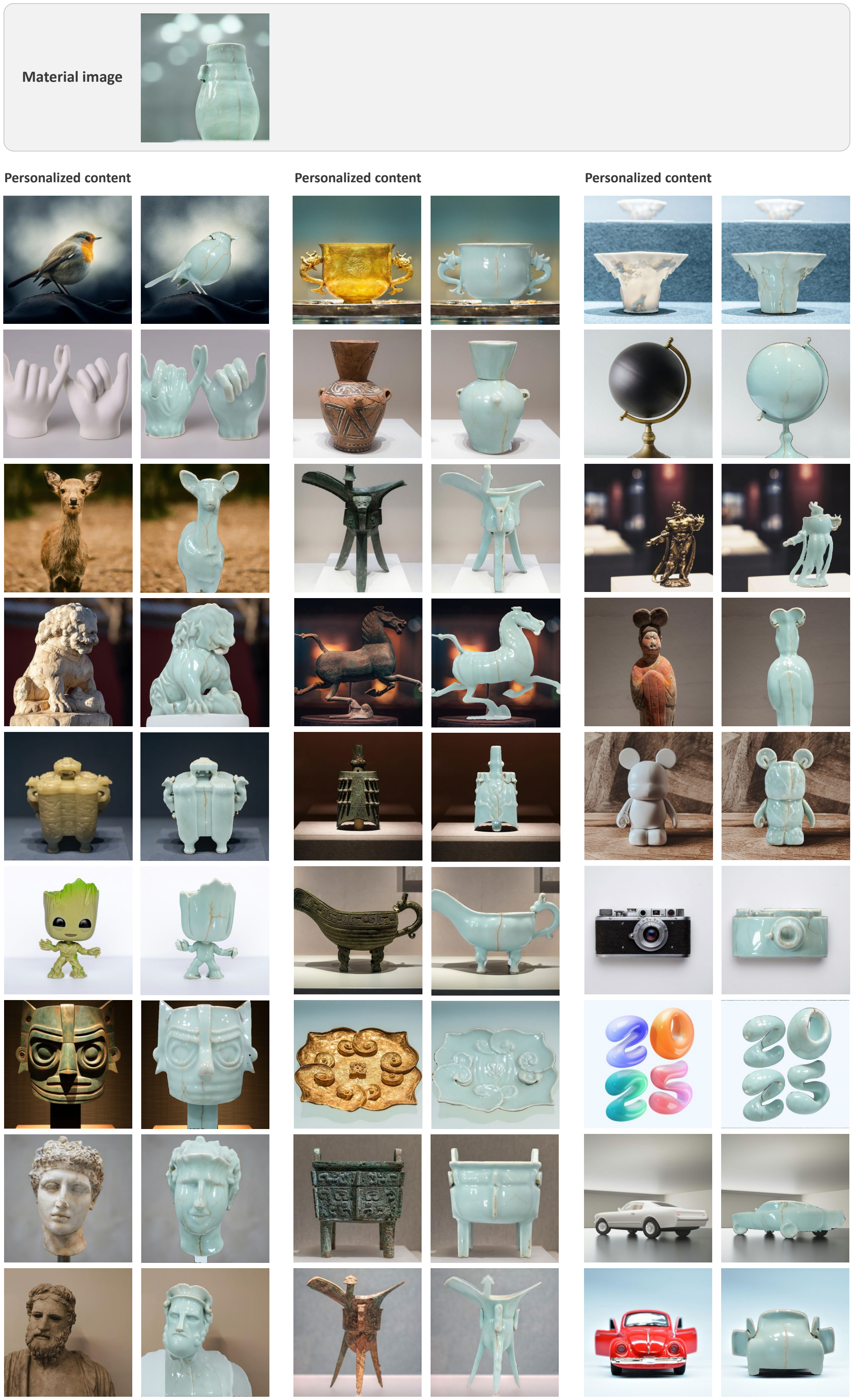

Figure 17. The material transfer effect of a single material on different objects.

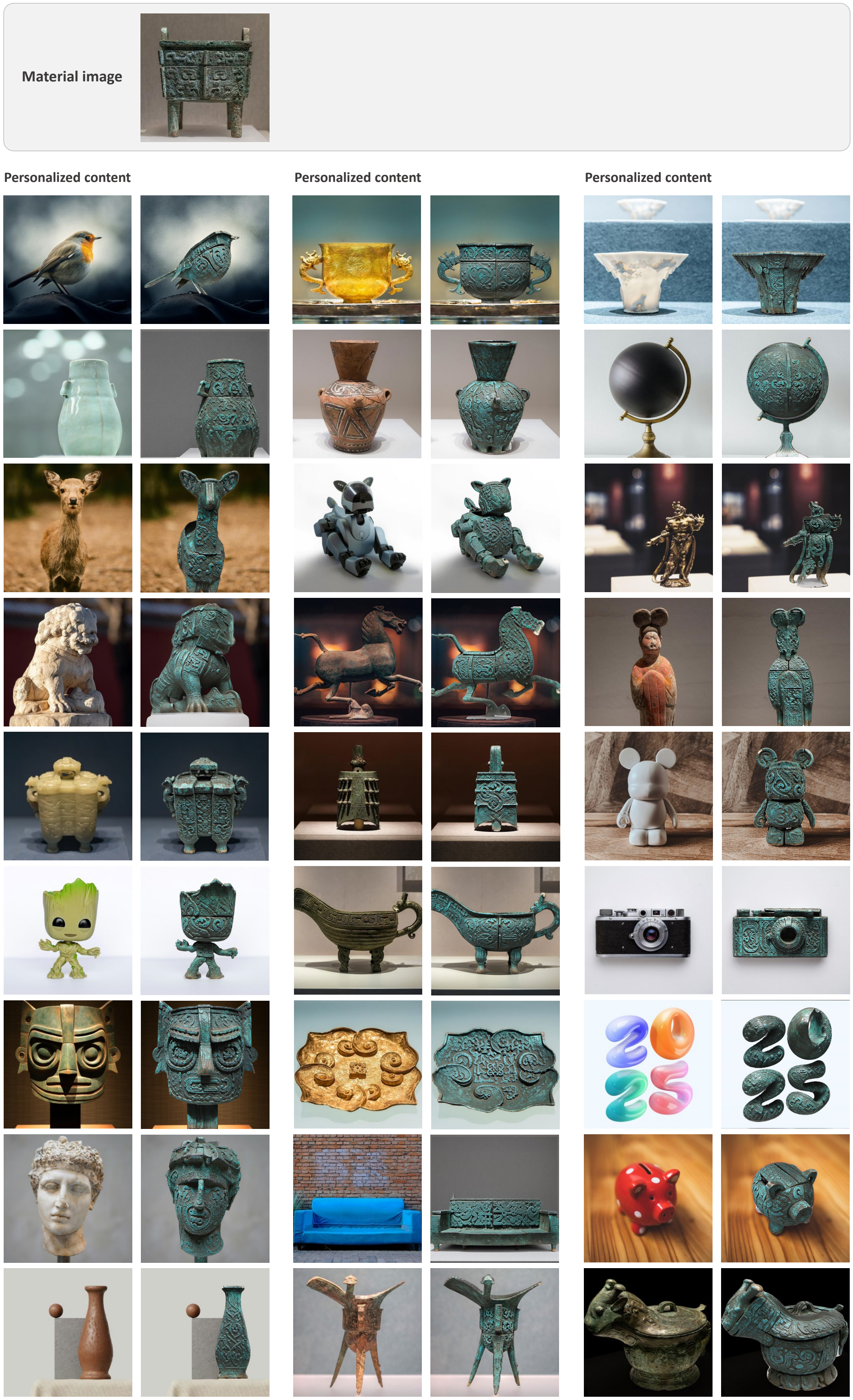

Figure 18. The material transfer effect of a single material on different objects.

MaterialFusion
A photo of a dog-like robot made of blue glass
A photo of a sculpture made of terracotta
A photo of a pottery vessel made of aventurine
A photo of a majestic stone lion statue made of sesame seed-covered pastry
A photo of a figurine with mouse-like ears made of glass filled with cola
A photo of a sliced mango made of apple
A photo of a chair made of bread
A photo of a sofa made of blue and white porcelain
A photo of a poseable mannequin made of glass iridescent ribbon
A photo of the number “2025” made of a pink jacket

ProSpect
A photo of a dog-like robot made of *
A photo of a sculpture made of *
A photo of a pottery vessel made of *
A photo of a majestic stone lion statue made of *
A photo of a figurine with mouse-like ears made of *
A photo of a sliced mango made of *
A photo of a chair made of *
A photo of a sofa made of *
A photo of a poseable mannequin made of *
A photo of the number “2025” made of *

U-VAP
A photo of a dog robot object made of glass material
A photo of a sculpture object made of terracotta material
A photo of a vessel object made of aventurine material
A photo of a stone lion statue object made of pastry material
A photo of a figurine object made of cola material
A photo of a mango object made of apple material
A photo of a chair object made of bread material
A photo of a sofa object made of porcelain material
A photo of a mannequin object made of ribbon material
A photo of a 2025 object made of jacket material

Table 1. Text prompts examples with different methods for inference.

Method	Used Implementation	Training Configuration	Inference Configuration
U-VAP	U-VAP github-repo	Epochs: 100	Inference Steps: 50, Guidance Scale: 7.5
Prospect	Prospect github-repo	Maximum Training Steps: 600	Inference Steps: 50, Strength: 0.6
IP-adapter-sd15	IP-adapter github-repo	N/A	Inference Steps: 50, Strength: 0.7
IP-adapter-sdxl	IP-adapter github-repo	N/A	Inference Steps: 50, Strength: 0.7
Zest	Zest github-repo	N/A	Inference Steps: 30, Strength: 0.9
MaterialFusion	MaterialFusion github-repo	N/A	Inference Steps: 50, Strength: 1.0
Ours	-	N/A	Inference Steps: 8, Strength: 1.0

Table 2. Overview of material transfer methods. The table presents the methods employed for material transfer, along with their respective implementations and configuration settings. Source repositories are included for reference.