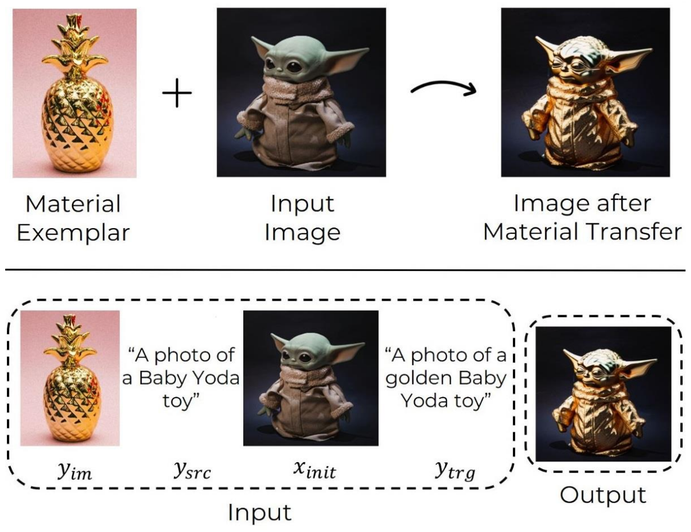
MaterialFusion: High-Quality, Zero-Shot, and Controllable Material Transfer with Diffusion Models

Kamil Garifullin[1,2], Maxim Nikolaev[1,2], Andrey Kuznetsov[2], Aibek Alanov[1,2]

[1] *HSE University* [2] *AIRI*

MAIN TASK

Our task involves **transferring texture** or material from one image y_{im} into an object in the foreground of another image x_{init}, while preserving the background information and fine-grained details of the object. The **goal** is to **generate an image** that corresponds to the target prompt y_{trg}, where the **object** in this image is **imbued with the material** from y_{im}.



PROCESS & METHODS

MODEL SELECTION

- **Base Model:** We utilized the pretrained T2I Stable Diffusion v1.5 model as the foundation for our work. This model is well-known for its robustness and versatility in generating high-quality images from textual descriptions.
- **IP-Adapter Integration:** To enable image prompt capability for the pretrained text-to-image diffusion models, we incorporated the IP-Adapter, an effective and lightweight adapter designed for this purpose.

GUIDE APPLICATION

- **Self-Attention Guider:** This guider helps maintain the background, the geometry of the target object and its pose
- **Feature Guider:** This guider is responsible for saving visual features

SINGLE SAMPLING STEP

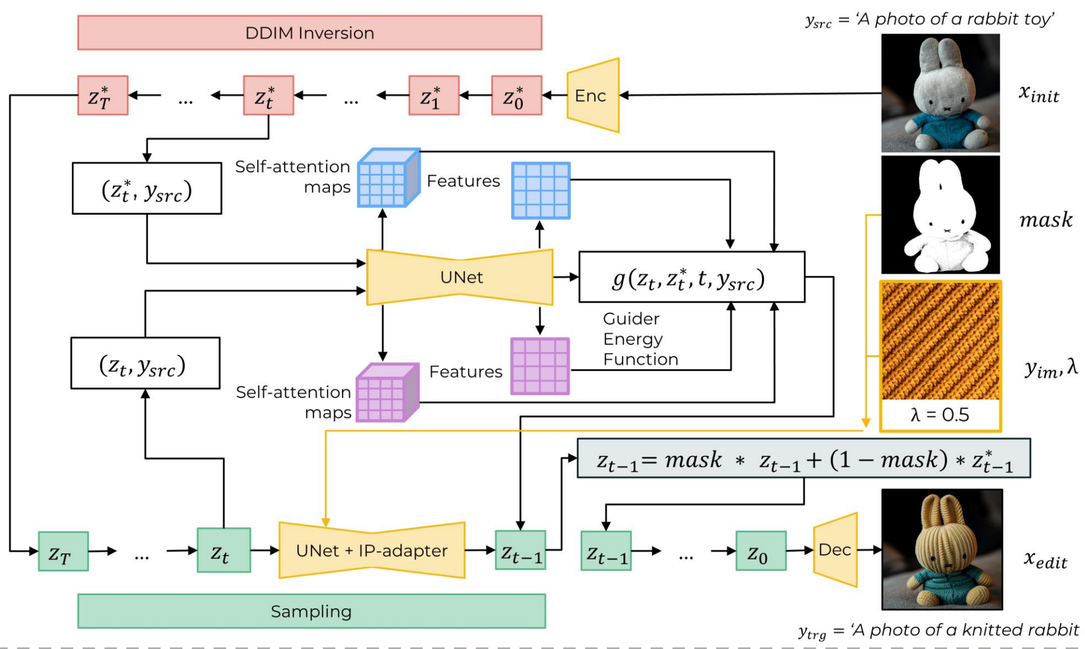
A single sampling step of MaterialFusion is defined by the following formula:

$$\hat{\epsilon}_\theta(z_t, c_t, c_i, t) = w\epsilon_\theta(z_t, c_t, c_i, t) + \\ +(1-w)\epsilon_\theta(z_t, t) + v \cdot \nabla_{z_t} g(z_t, z_t^*, t, y_{src}, I^*, \overline{I})$$

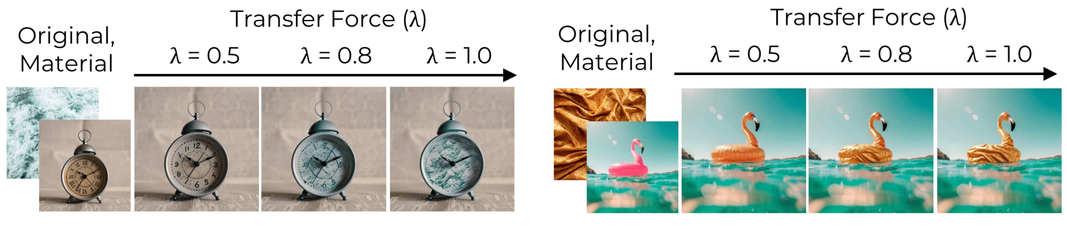
where $\overline{I}$ and I^* are inner representations computed during the forward pass of $\epsilon_\theta(z_t, t, y_{src})$ and $\epsilon_\theta(z_t^*, t, y_{src})$ respectively; v is the self-guidance scale; c_t is the text conditioning, and c_i is the image conditioning; z_t are latent variables from the current generation process, while z_t^* are the time-aligned DDIM latents.

OVERALL PIPELINE

The overall pipeline of **MaterialFusion** for material transfer. Starting with DDIM inversion of the target image x_{init} and material exemplar y_{im}, the framework combines the IP-Adapter with UNet and employs a guider energy function for precise material transfer. A dual-masking strategy ensures material application only on target regions while preserving background consistency, ultimately generating the edited output x_{edit}. The parameter λ, known as the **Material Transfer Force**, controls the intensity of the material application, enabling adjustment of the transfer effect according to user preference.

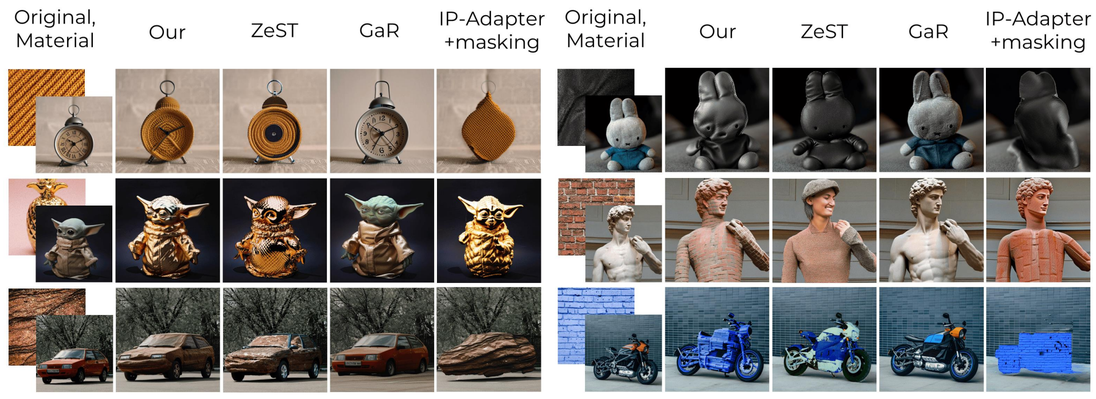


MATERIAL TRANSFER FORCE

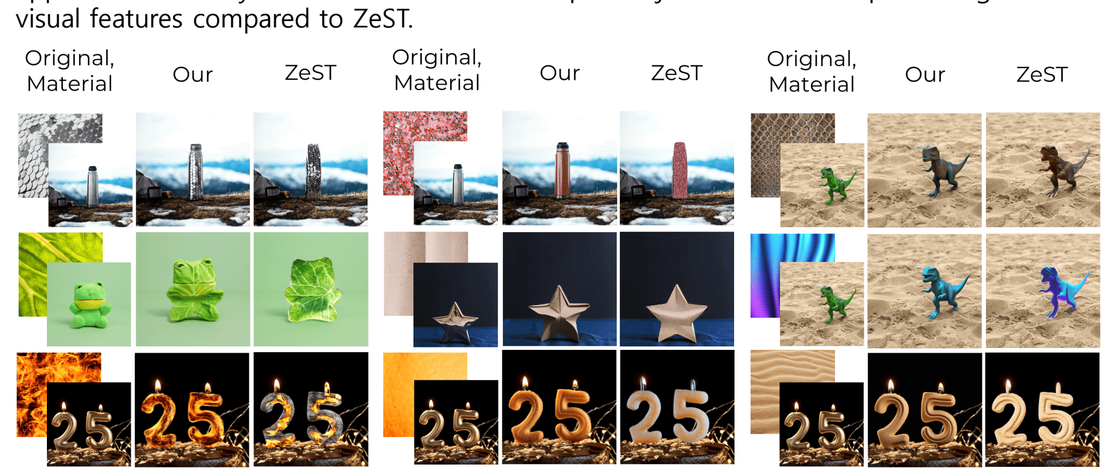


QUALITATIVE ANALYSIS

To compare the qualitative results obtained by different methods: Our method, **ZeST, GaR,** and **IP-Adapter with masking**. Our method demonstrates more realistic material integration, preserving object structure and achieving higher fidelity to the target material

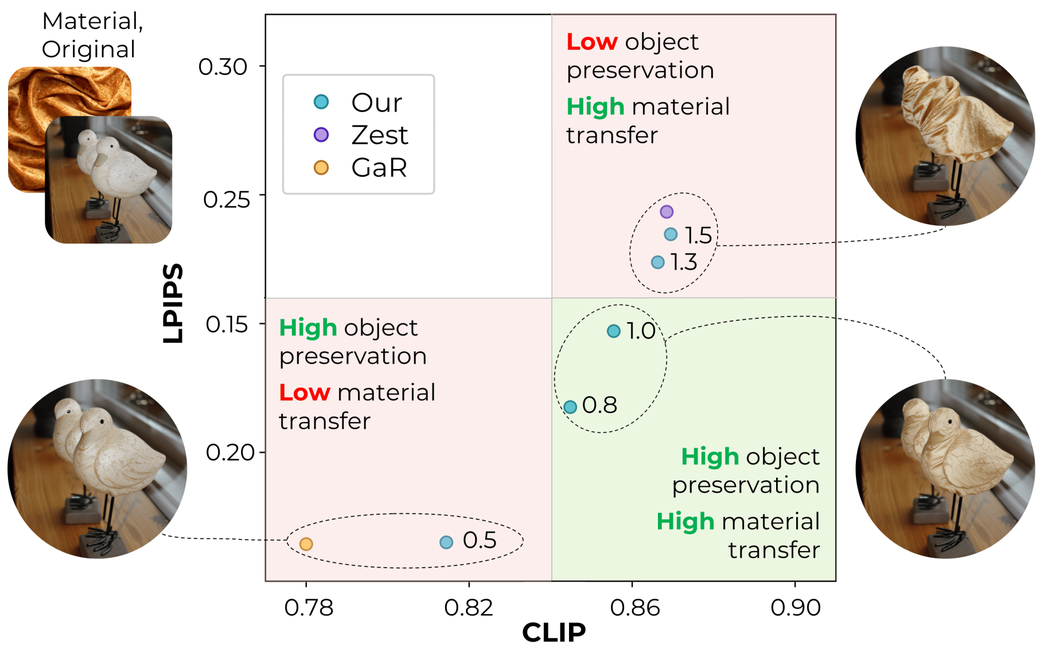


Direct comparison between our method and the current state-of-the-art ZeST. Our approach effectively transfers materials to complex objects while better preserving their visual features compared to ZeST.



QUANTITATIVE ANALYSIS

Quantitative analysis of material transfer and object preservation. The lower right region represents optimal results with high CLIP scores (effective transfer) and low LPIPS values (good detail preservation). The numbers above the dots in the graph represent the material transfer force. **MaterialFusion** achieves the best balance, with results in the optimal zone, while GaR and ZeST show trade-offs between transfer efficiency and detail preservation.



USER STUDY

Our method was preferred overall and rated highly for detail preservation, while ZeST scored better for material fidelity. This balance between material transfer and object fidelity makes our method more effective in delivering coherent and lifelike results.

Questions	Results
Overall Preference (Q1)	68%
Material Fidelity (Q2)	26%
Detail Preservation (Q3)	70%

Examples of comparisons where the vote for Question Q2 (Material Fidelity) was given to ZeST. While ZeST achieves high material transfer, it often overpowers the original object's features, resulting in a "cut-and-paste" effect.

LINKS

Research group

Centre of Deep Learning and Bayesian Methods

Code and Demo

Link to GitHub