MatFuse: Controllable Material Generation with Diffusion Models

Giuseppe Vecchio, Renato Sortino, Simone Palazzo, Concetto Spampinato

University of Catania

Controllable Material Synthesis

Why materials matter?

- Determine how light interacts with surfaces, affecting realism in computer-generated imagery.
- Find applications in architecture, simulation, design, gaming and more…

Challenges

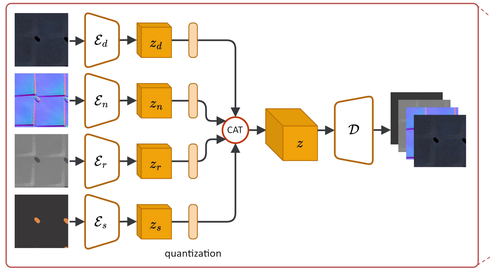
- High expertise and complex tools required for material crafting.
- Learning-based approaches lack the degree of control needed by artists.

Takeaway

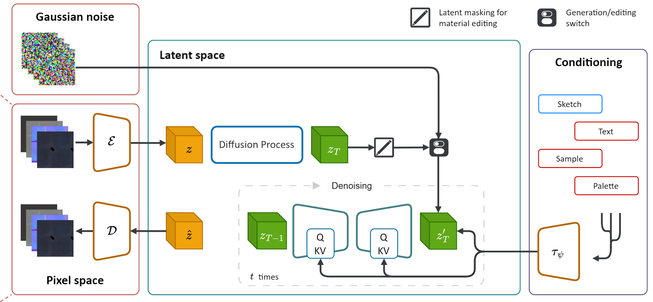
- Compositionality in material domains allow fine grained control over the generation.
- Learning a disentangled latent representation of the material properties enables map-level editing of the material.

MatFuse

Disentangled compression model.

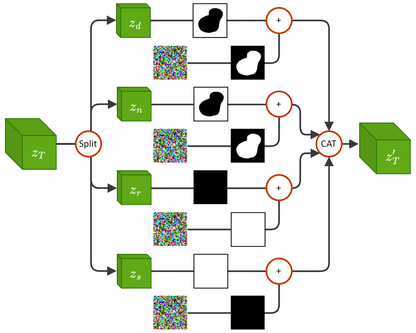


Compression model architecture. The multi-encoder VQ-VAE[2] learns a map-specific latent spaces to enable editing capabilities.



MatFuse architecture. MatFuse is a multi-conditional method leveraging the generation capabilities of Latent Diffusion Models to tackle the task of high-quality material synthesis as a set of SVBRDF maps.

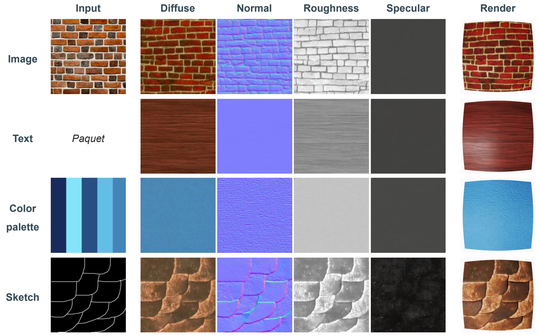
Volumetric inpainting.



Editing via volumetric inpainting. We independently mask the latent channels corresponding to each material property, to enable map level editing.

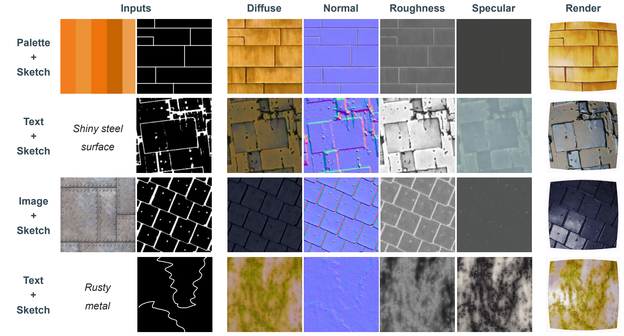
Results

Single-condition generation



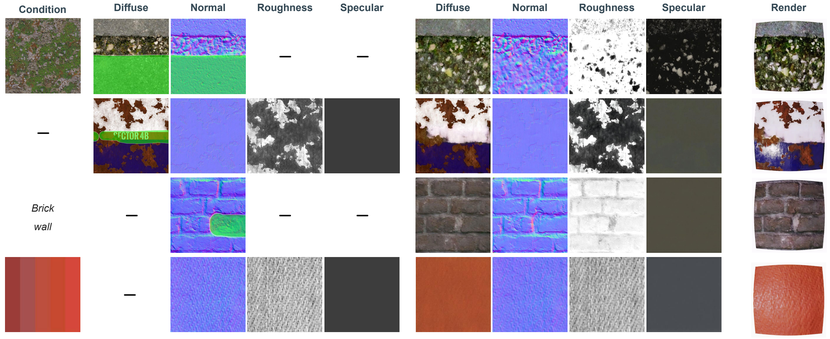
Single condition material generation. MatFuse can be globally conditioned via text prompts, image prompts, and a color palette; and locally through pattern sketches.

Multi-condition generation



Multimodal conditioned material generation. MatFuse is able to combine local and a global conditions for a finer control over both the geometry and the visual features of the material.

Editing via *volumetric inpainting*



Material editing with inpainting. MatFuse is able of materials editing via volumetric inpainting. The masked areas are highlighted in green, while fully masked maps are replaced with the '–' symbol.

References

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10684–10695, 2022.

Patrick Esser, Robin Rombach, and Bjorn Ommer. *Taming transformers for high-resolution image synthesis.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12873–12883, 2021.