SIMULDITEX: A SINGLE IMAGE MULTISCALE AND LIGHTWEIGHT DIFFUSION MODEL FOR TEXTURE SYNTHESIS

Pierrick Chatillon, Julien Rabin, David Tschumperlé

Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, F-14000 Caen, France

name.surname@unicaen.fr https://github.com/PierrickCh/SIMuLDiTex

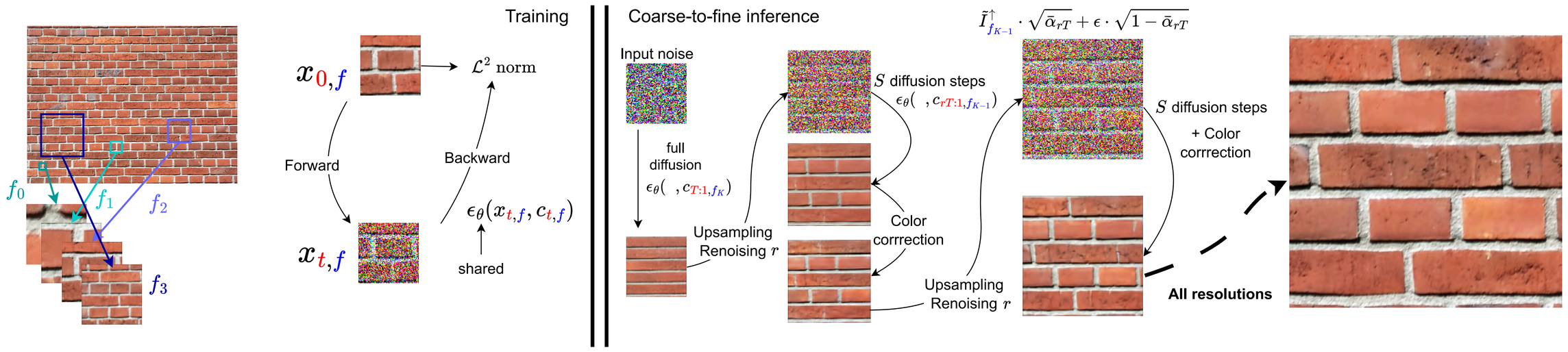

FIGURE 2 – Overview of SIMuLDiTex. Left : The network is trained as a denoiser, conditioned on the noise level t, and the scale factor f. Right : During inference, the image is synthesized in a coarse-to-fine fashion.

1. Motivation

➔ Texture generation :

- ✓ **Realistic**
- ✓ **Structurally consistent**
- ✓ **Diverse**
- ✓ **High-resolution**

➔ Challenges :

- ✘ **Large images = costly**
- ✘ **lightweight models hard to train**

2. Method overview

- ✓ **Multi-scale training on a single image**
- ✓ **Coarse-to-fine inference** : Minimal computation at high resolution ($\sim$ 2s for 2048^2)
- ✓ **Lightweight model** : Only 1M or 4M parameters, no auto-encoder
- ✓ **Fourier modules** : Reproduce large structures at low cost (compared to attention)

Texture interpolation by local linear combination of denoisers

$\omega\epsilon_\theta + (1-\omega)\epsilon_\phi \longrightarrow \nabla log(p_1^\omega p_2^{1-\omega})$

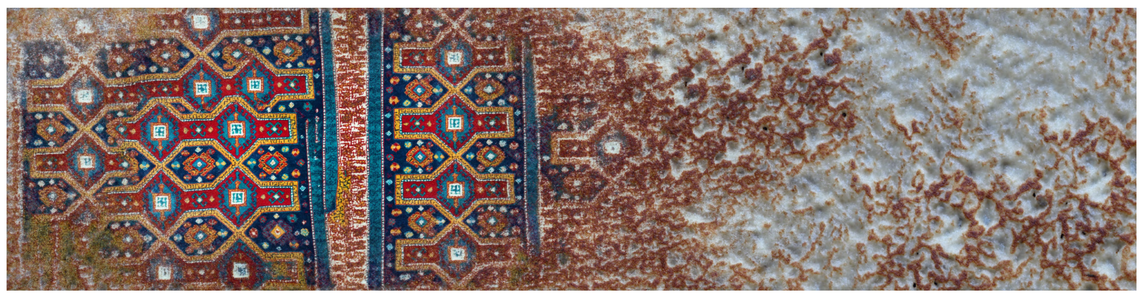

FIGURE 4 – Spatial interpolation

3. Ablation

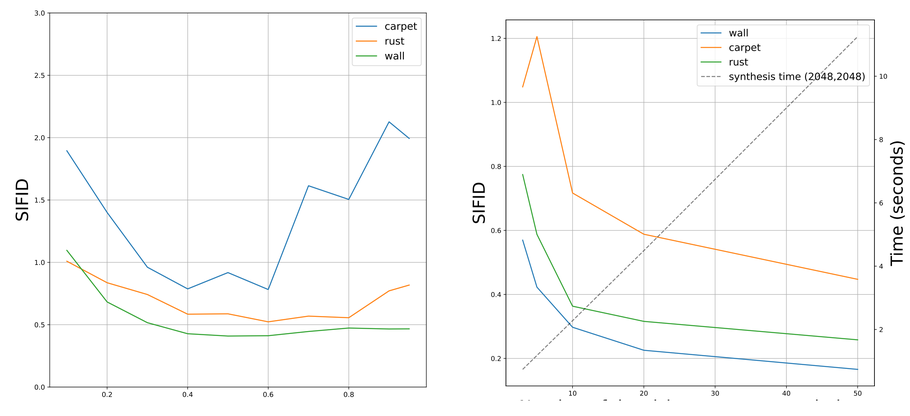

FIGURE 6 – SiFID vs. renoising factor r & Number of steps S.

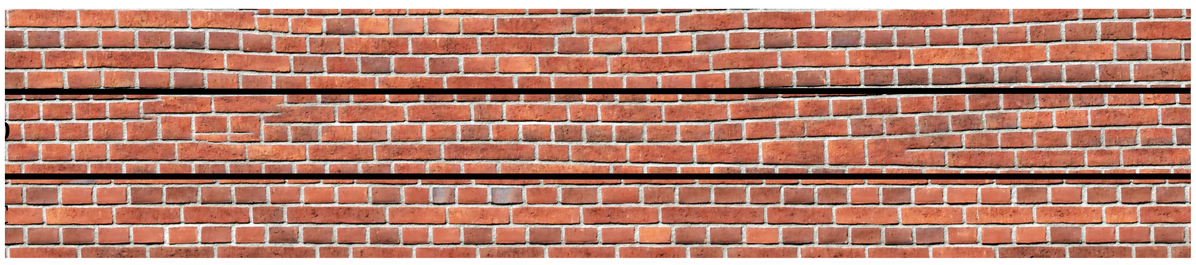

FIGURE 8 – **Top** : bottleneck with attention. **Middle** : Vanilla bottleneck. **Bottom** : Bottleneck with Fourier module

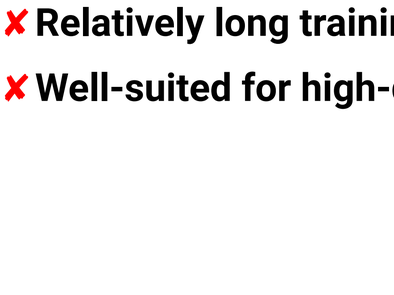

FIGURE 9 – Histogram prescription

FIGURE 11 – Use only the finer resolutions.

4. Conclusion

- ✓ **Patch-based & single-image training ⇒ Efficient lightweight model**
- ✓ **Scale conditioning ⇒ Coarse-to-fine synthesis with one model**
- ✓ **Lightweight + coarse-to-fine ⇒ High-resolution synthesis**
- ✓ **Fourier module ⇒ Low-cost structure generation**
- ✘ **Relatively long training time**
- ✘ **Well-suited for high-quality textures**