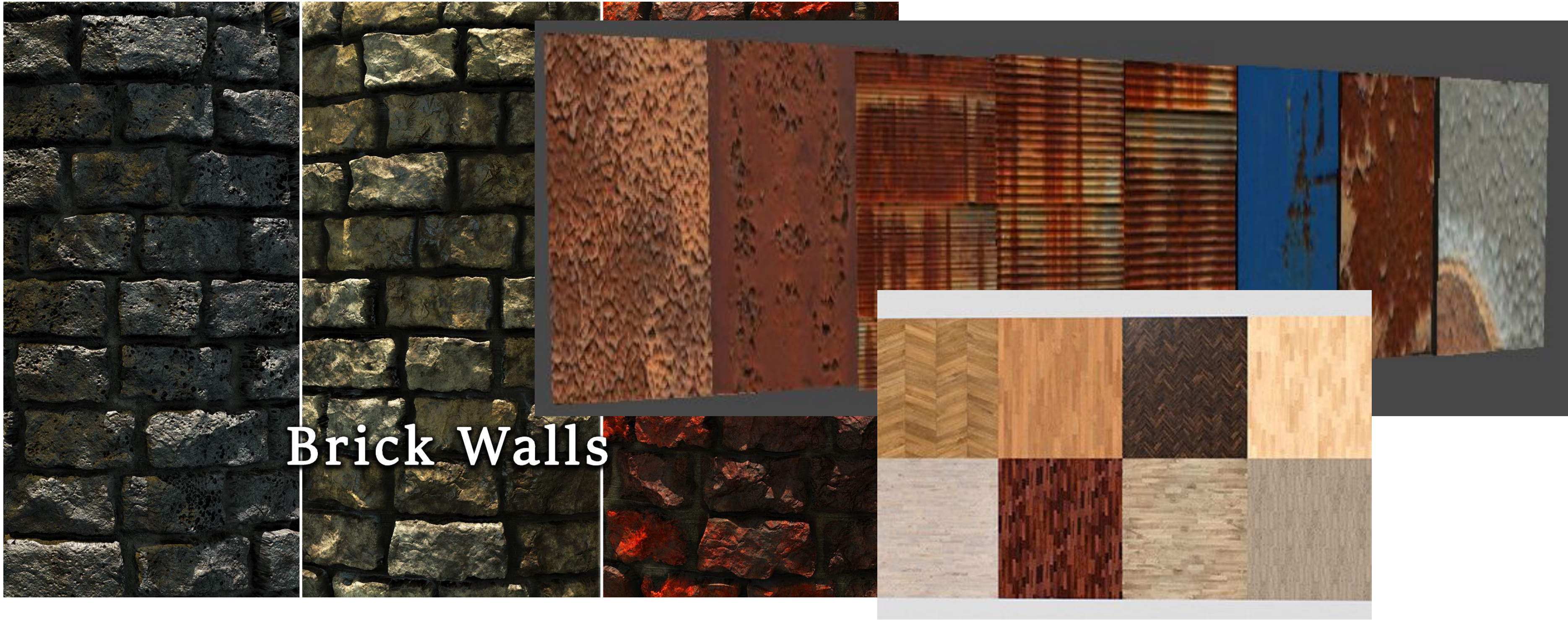
TexSliders:Diffusion Based Editing in CLIP Space

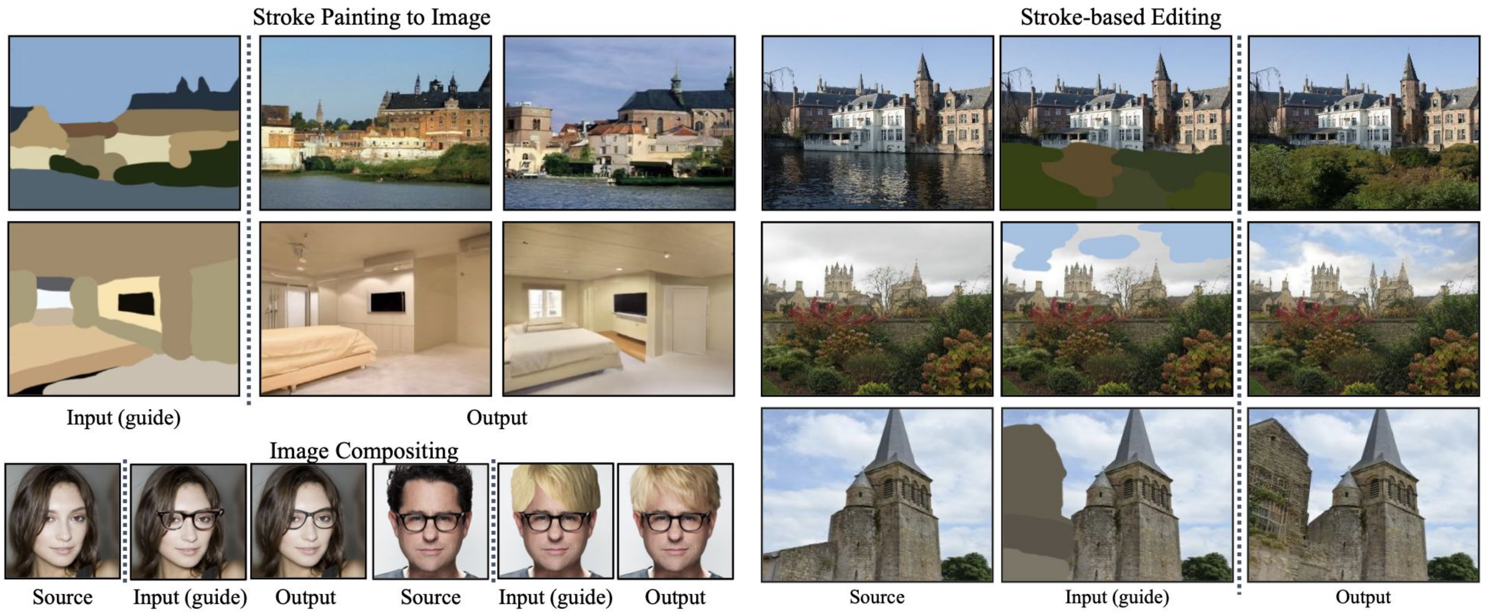
SIGGRAPH 2024, Adobe Research

Team4
Chanryeol Lee, Chanhyuk Lee

Why We Need Texture Editing?



Limitations of previous models



SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, ICLR 2022

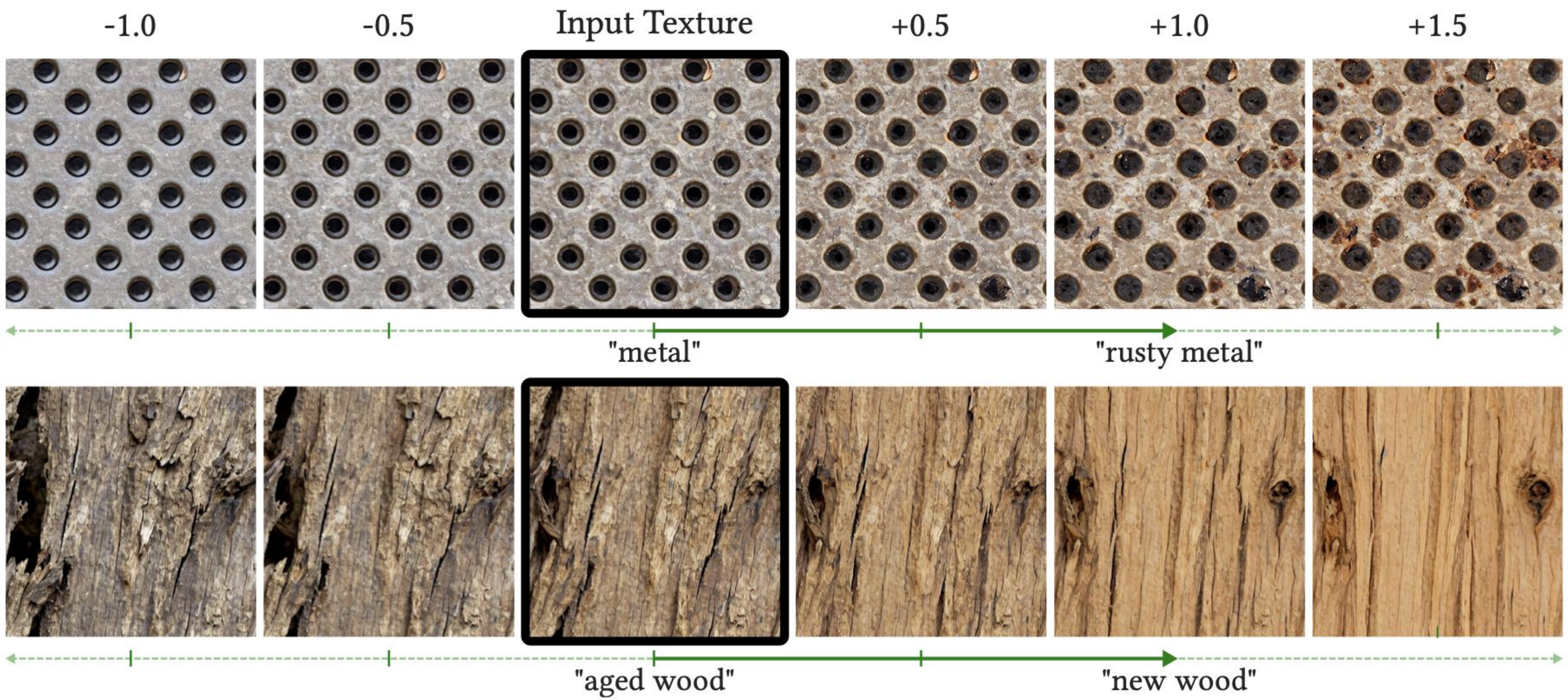
Limitations of previous models

- Focuses on **Global Generation**
  - Hard to capture complex details of texture patterns

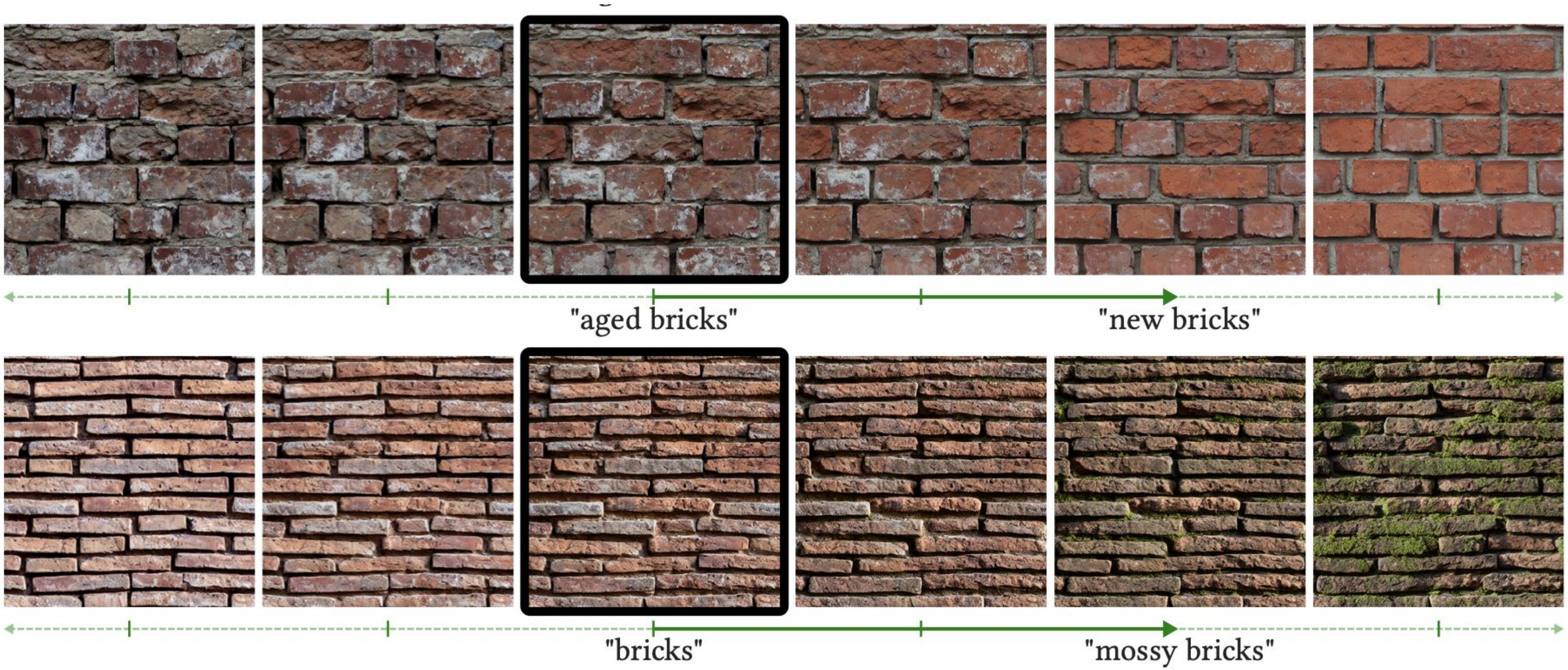
- Datasets lacks of **complex texture patterns**
  - Dataset biased to general scenes, objects, facial expressions..

- Difficulty in **semantic control**
  - Simultaneously maintaining texture pattern and modifying specific feature is hard

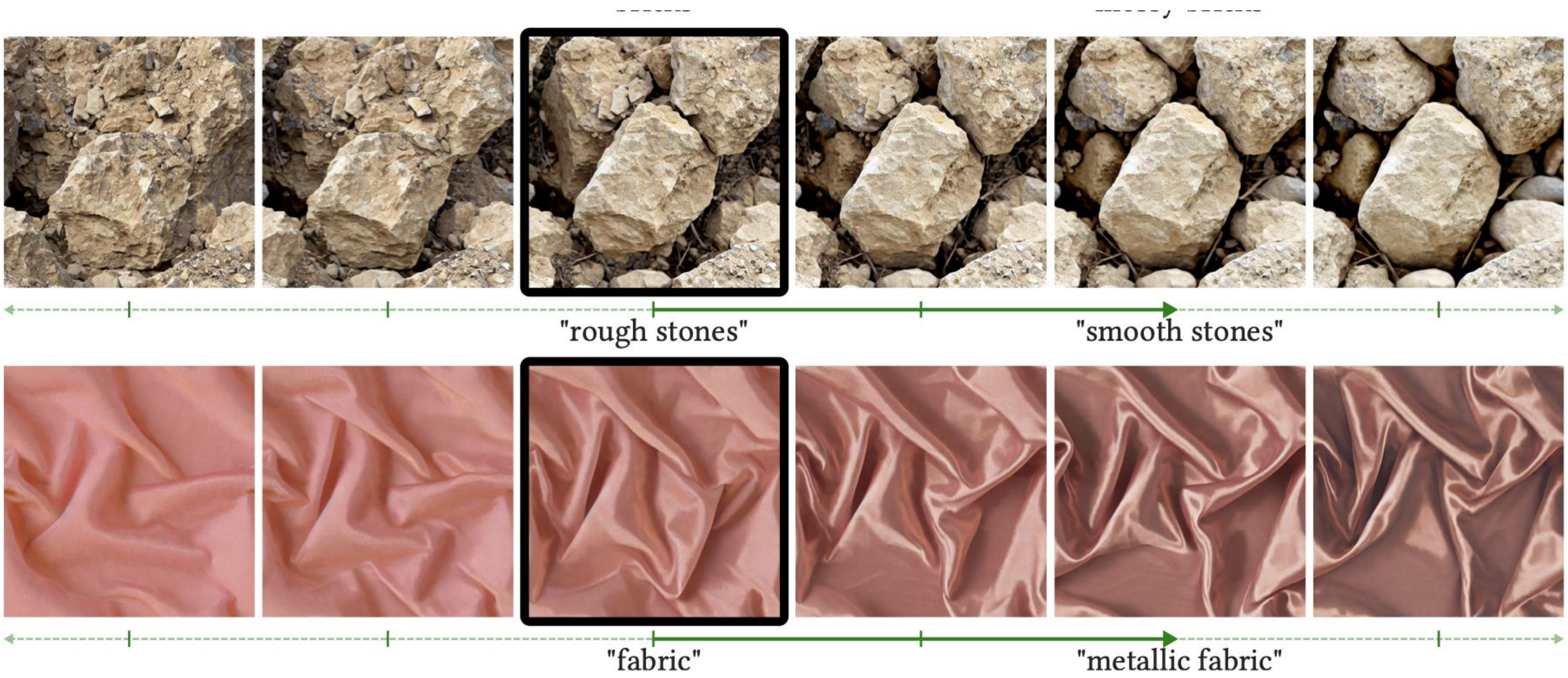
TexSliders: Results



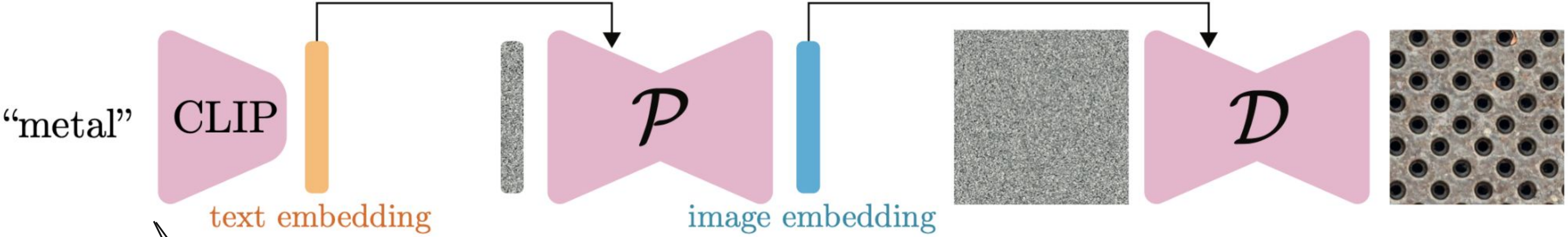
TexSliders: Results



TexSliders: Results

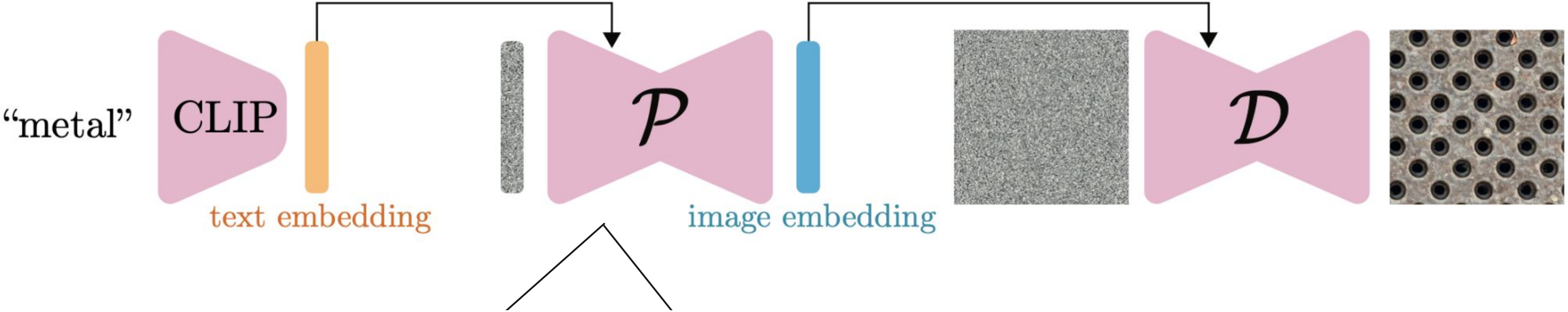


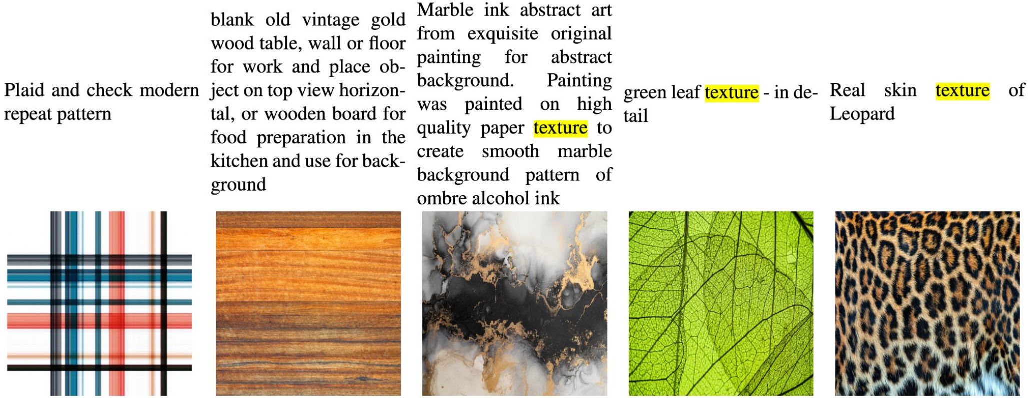
Methods: Step by Step



Encode given text prompt with CLIP text encoder

Methods: Step by Step

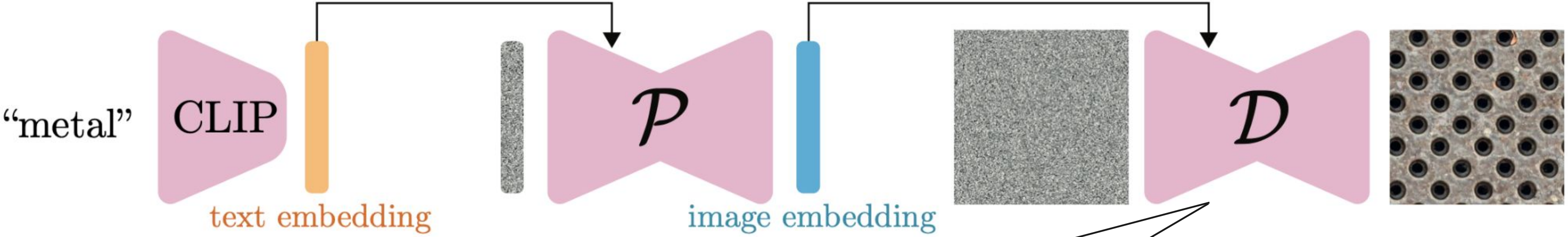


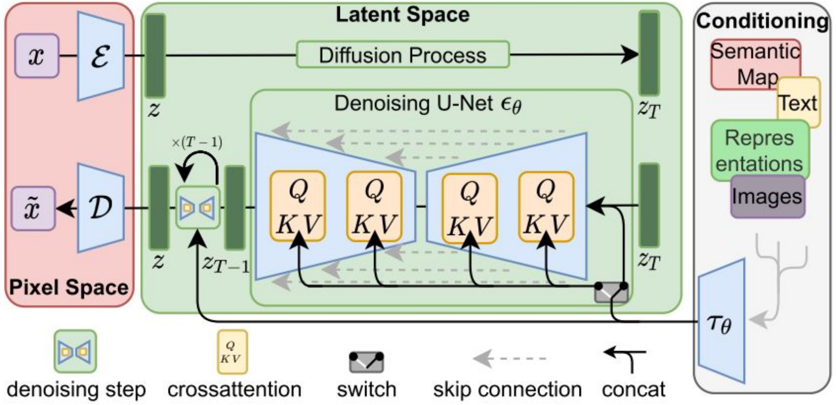


Pre- trained DIffusion Prior
Texture-Text Pair

Input: Text Embedding + Noise
Output: Image Embedding

Methods





Pre- trained Latent Diffusion Model

Input: Noise conditioned with image embedding
Output: Generated Texture

Methods

Focus on Image Embedding space, 768 Dimension

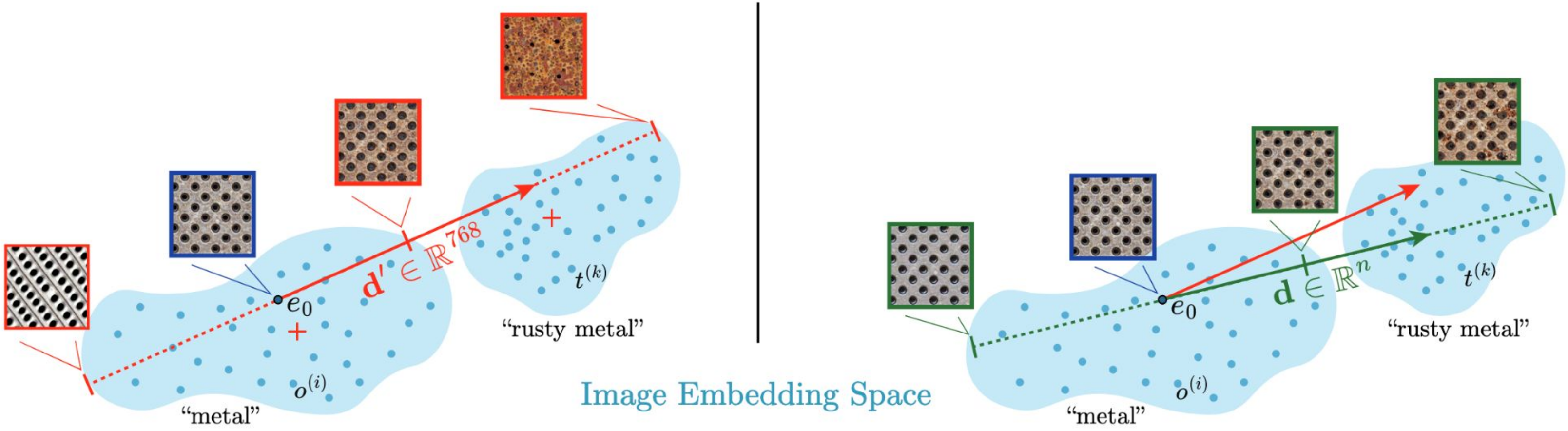
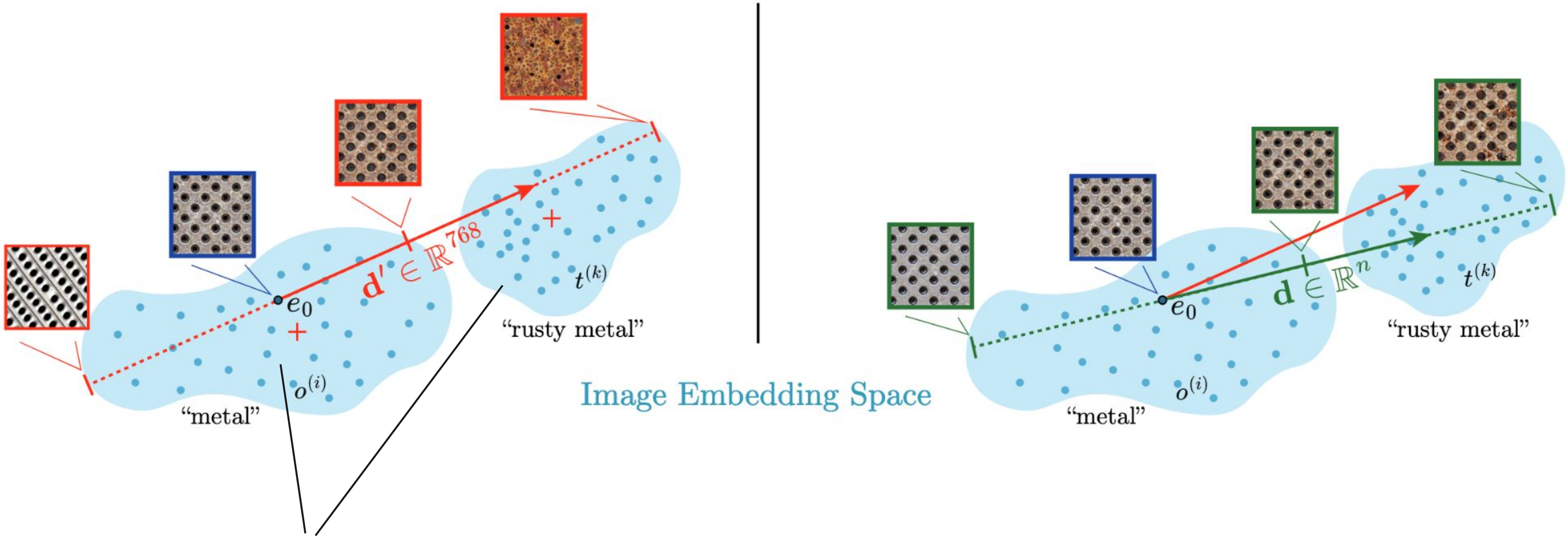


Image embedding space encodes inherent features of the image.

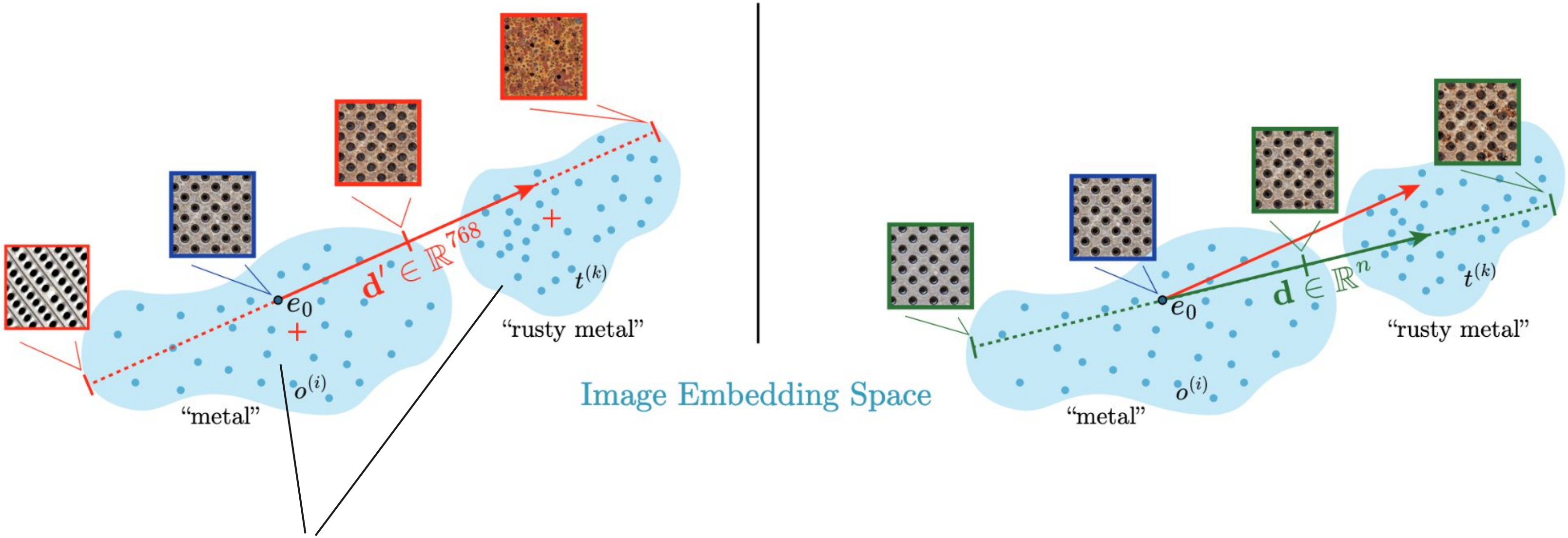
We should finding principal "direction" that changes the feature to modify it.

Methods



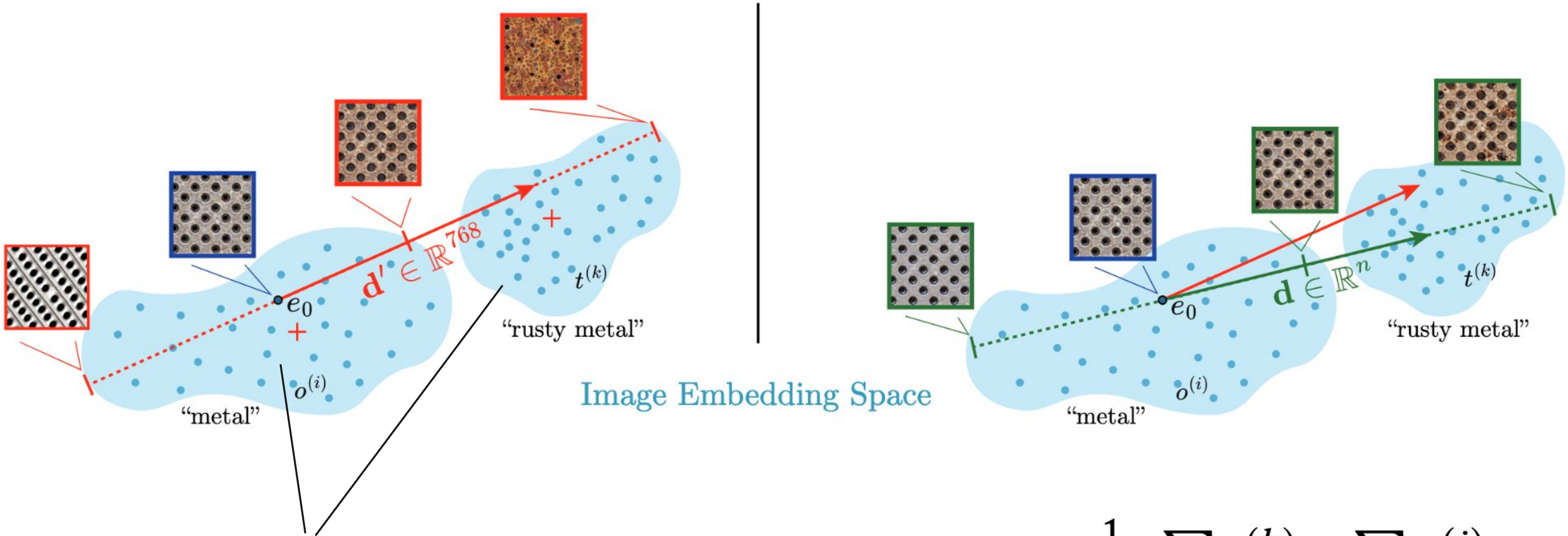
Embeddings obtained from same text prompt are mapped to similar spatial location.

Methods



Obtain vector from "**metal**" embedding cluster to "**rusty metal**" embedding cluster

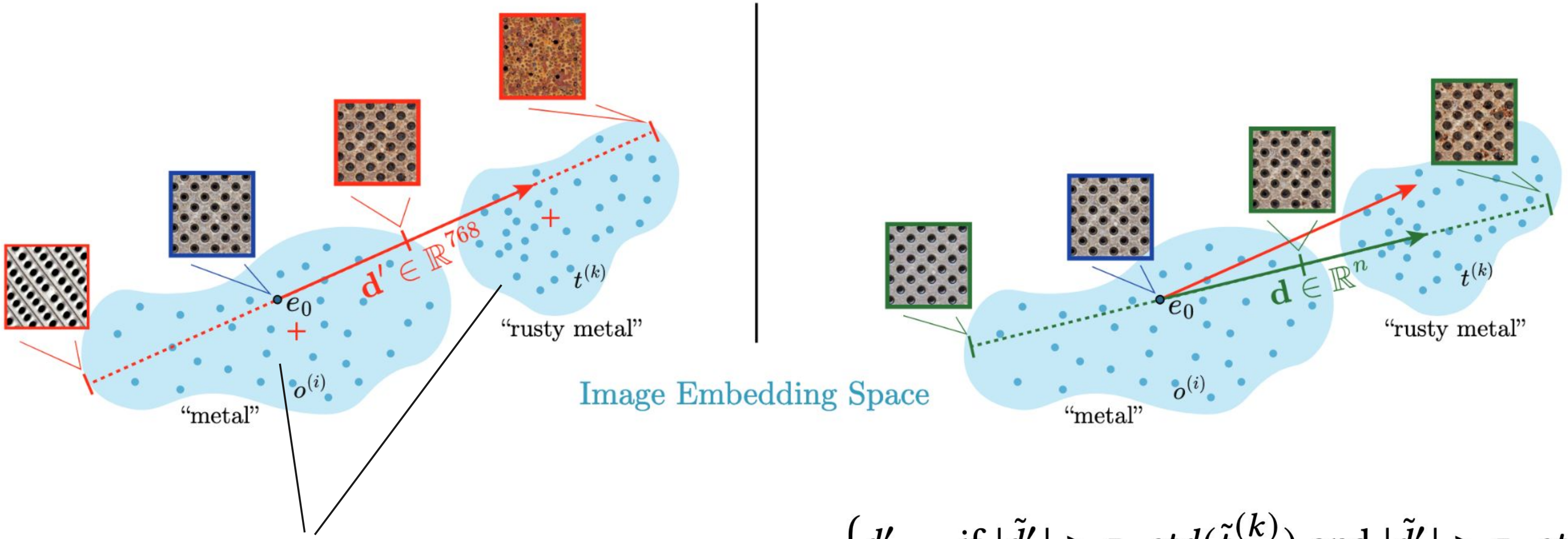
Methods



Naive Approach: Connect centroids of each cluster
-> **Leads to poor result.**

$$d'_j = \frac{1}{n_e}\left(\sum_k t_j^{(k)} - \sum_i o_j^{(i)}\right).$$

Methods



Selectively choose dimensions that actually contributes to desired edit.

$$d_j = \begin{cases} d'_j\,, & \text{if } |\tilde{d}'_j| > \tau \cdot \underset{k}{std}(\tilde{t}_j^{(k)}) \text{ and } |\tilde{d}'_j| > \tau \cdot \underset{i}{std}(\tilde{o}_j^{(i)}) \\ 0\,, & \text{otherwise.} \end{cases}$$

Methods

$$d_j = \begin{cases} d'_j, & \text{if } |\tilde{d}'_j| > \tau \cdot \underset{k}{std}(\tilde{t}_j^{(k)}) \text{ and } |\tilde{d}'_j| > \tau \cdot \underset{i}{std}(\tilde{o}_j^{(i)}) \\ 0, & \text{otherwise.} \end{cases}$$

Distance between cluster centroids: **Inter-Cluster Variability.**

How much does embedding change in following dimension in cluster change.

Methods

$$d_j = \begin{cases} d'_j, & \text{if } |\tilde{d}'_j| > \tau \cdot \underset{k}{std}(\tilde{t}_j^{(k)}) \text{ and } |\tilde{d}'_j| > \tau \cdot \underset{i}{std}(\tilde{o}_j^{(i)}) \\ 0, & \text{otherwise.} \end{cases}$$

Standard deviation within cluster: **Intra-Cluster Variability.**

How much does embedding change in following dimension within same cluster.

Methods

$$d_j = \begin{cases} d'_j, & \text{if } |\tilde{d}'_j| > \tau \cdot \underset{k}{std}(\tilde{t}_j^{(k)}) \text{ and } |\tilde{d}'_j| > \tau \cdot \underset{i}{std}(\tilde{o}_j^{(i)}) \\ 0, & \text{otherwise.} \end{cases}$$

IF

Inter-cluster variability > Intra-cluster variability: The dimension is sensitive to desired edit

Intra-cluster variability > Inter-cluster variability: Dimension is sensitive to other features

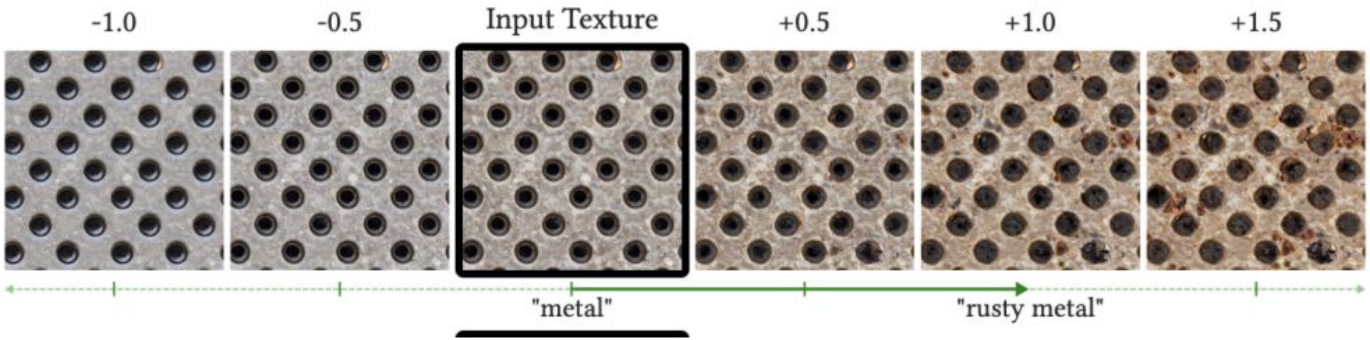
Methods

"rusty metal" embedding

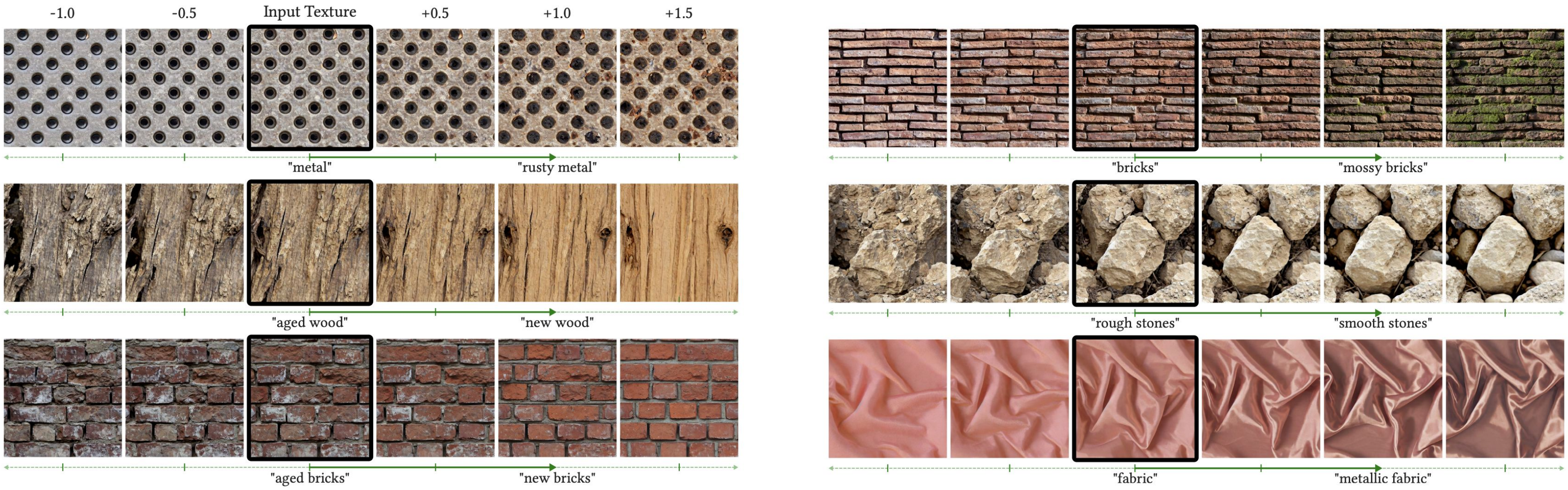
$$\mathbf{e}_{\alpha} = \mathbf{e}_0 + \alpha \cdot \mathbf{d},$$

" metal" embedding

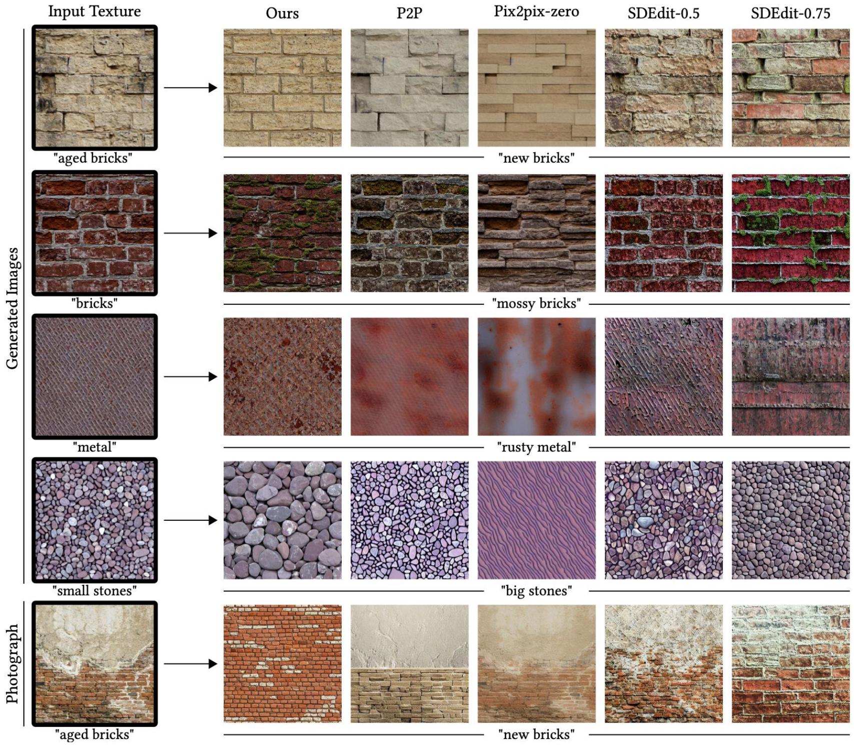
Alpha can work as an "slider" to control amount of "rustiness"



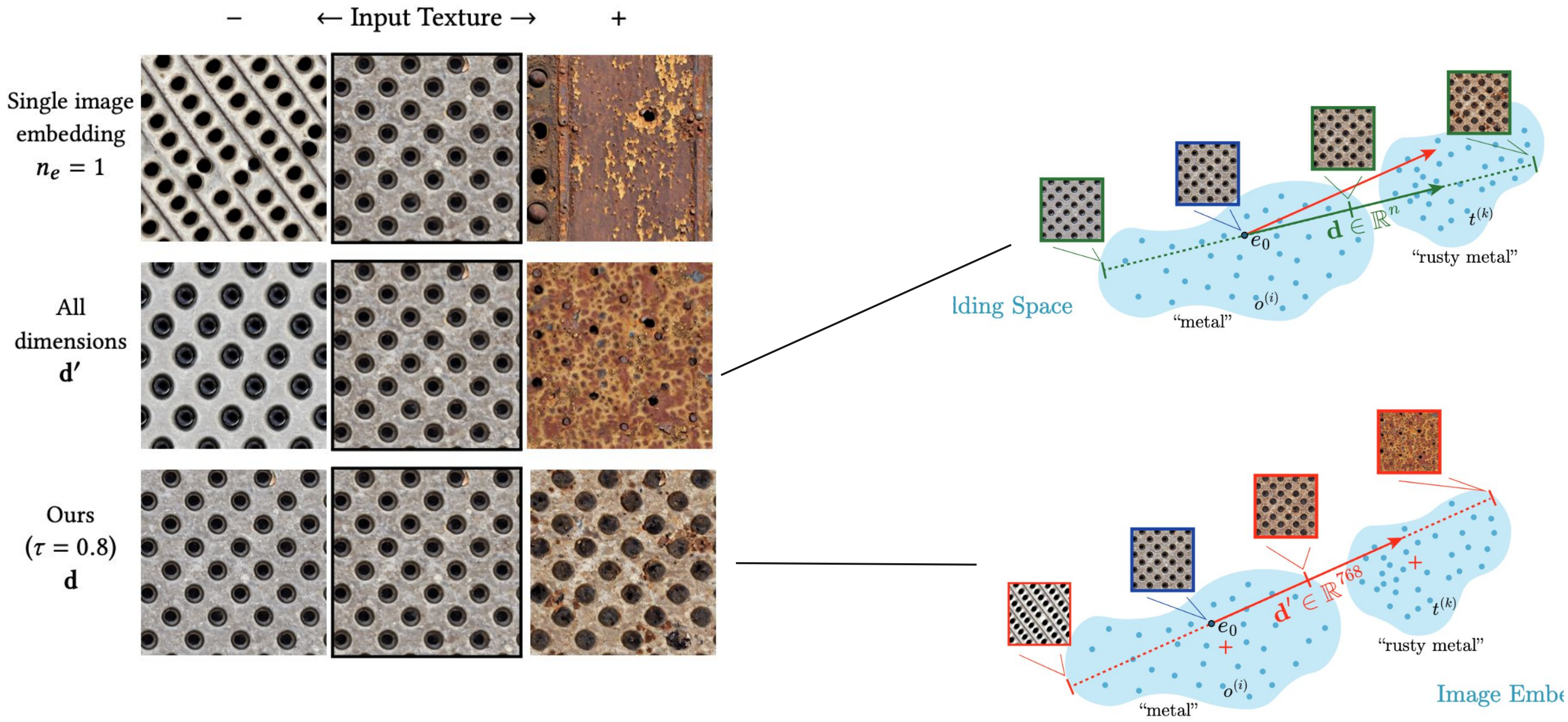
Qualitative Results



Results - Comparison

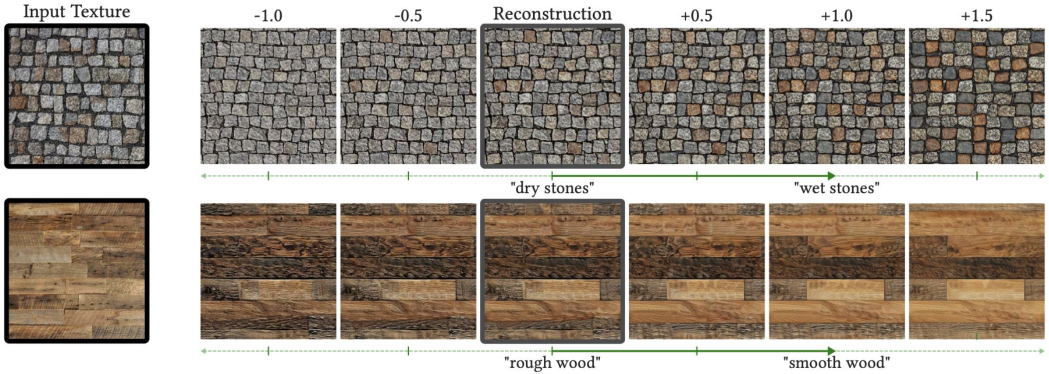


Results - Ablation Study

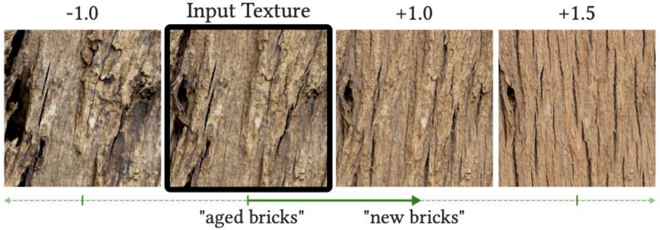


Results - Applications

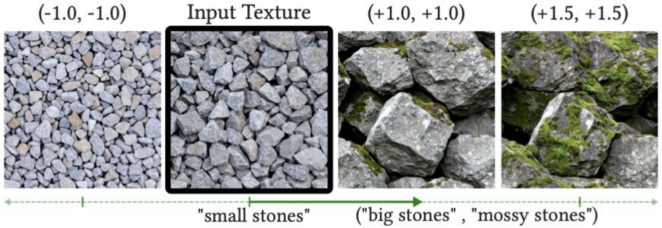
Real Photographs



Generalization



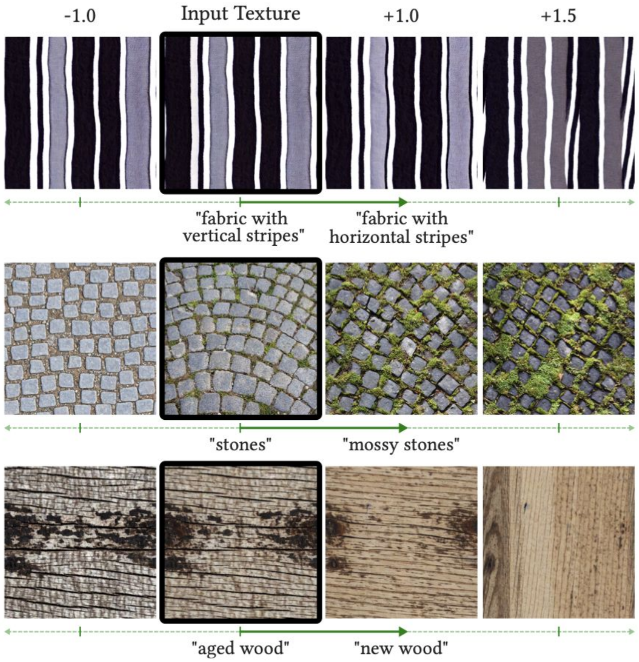
Multi-dimension



Limitations

In some cases:

- CLIP & diffusion model can be **more sensitive** to some concepts
- Identity of the input texture is **not perfectly preserved**
- Extrapolating in the editing direction **too far** from the input texture can hamper identity



Quiz