# Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models

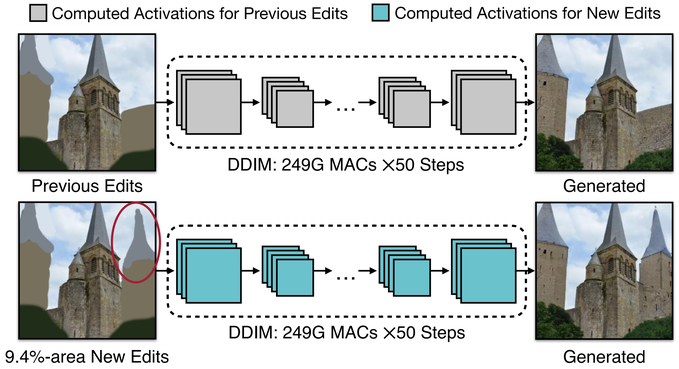
Muyang Li[1], Ji Lin[2], Chenlin Meng[3], Stefano Ermon[3], Song Han[2], and Jun-Yan Zhu[1]

[1]Carnegie Mellon University, [2]Massachusetts Institute of Technology, and [3]Stanford University

## Introduction

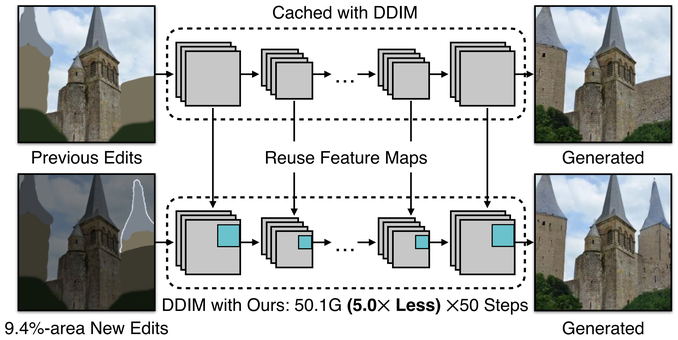
### Background:

Recent models often synthesize the entire image, even for a minor edit, wasting significant computation.

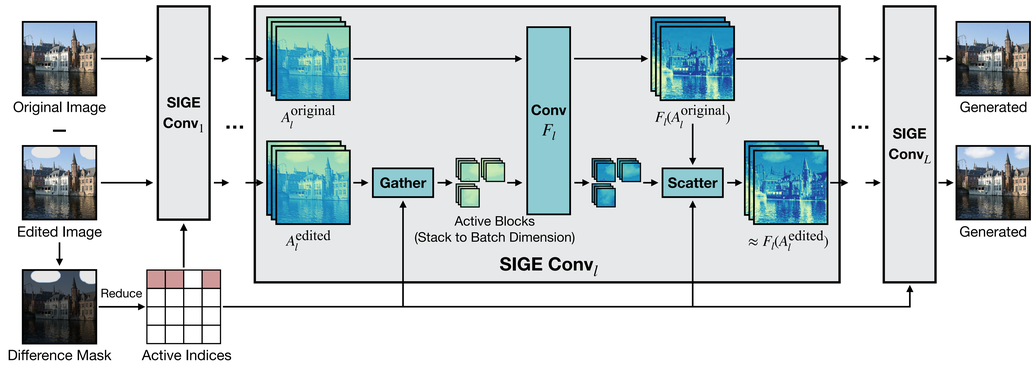



### Idea — Spatially Sparse Inference (SSI):

Cache the features of previous edits. Then reuse them and only apply convolution to the newly edited regions.




## Sparse Incremental Generative Engine (SIGE)



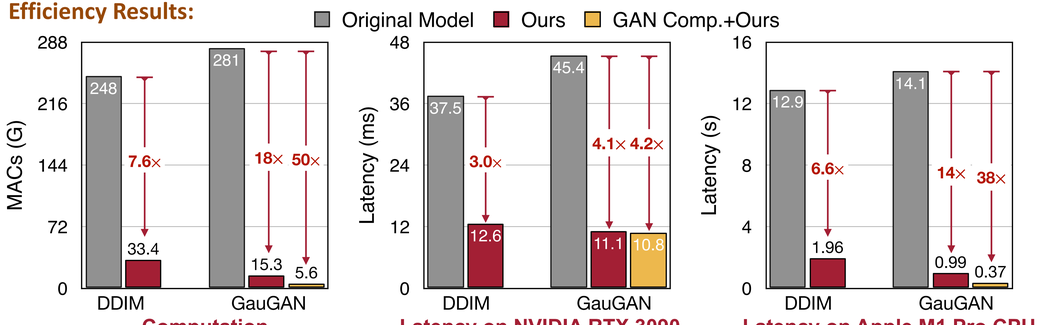

### Tiling-based Sparse Convolution:

- Precompute the features of original image.
- Compute a difference mask between the original and edited image.
- Divide mask to small blocks and reduce it to active block indices.
- Gather the active blocks along batch dimension.
- Scatter the output blocks into the original activation $F_l(A_l^{\text{original}})$.

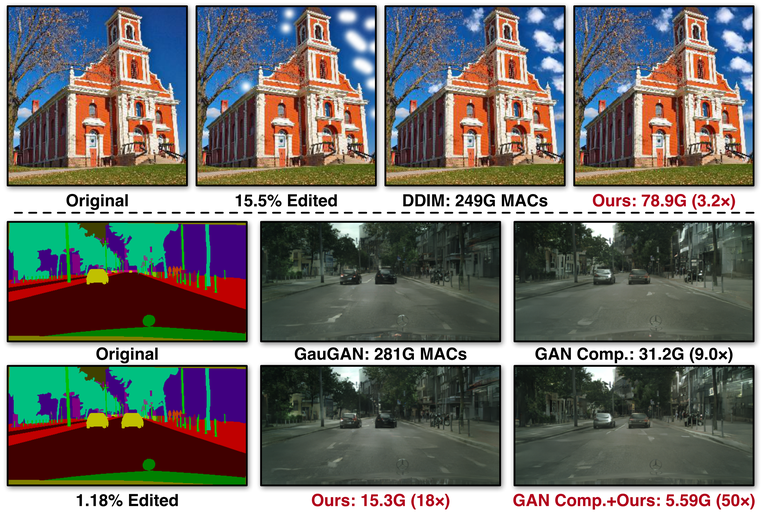
### Optimizations:

- Reuse the original statistics to replace normalization with scale+shift.
- Kernel fusion. (Refer to our paper for more details)

### Efficiency Results:

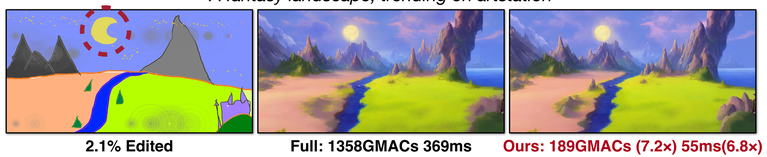



### Quality Results on DDIM and GauGAN




### Extension to Stable-Diffusion

*A fantasy landscape, trending on artstation*




### References:

[1] SBNet: Sparse Blocks Network for Fast Inference, Ren et al., CVPR 2018
[2] Semantic Image Synthesis with Spatially-Adaptive Normalization (GauGAN), Park et al., CVPR 2019
[3] GAN Compression: Efficient Architectures for Interactive Conditional GANs, Li et al., CVPR 2020
[4] Denoising Diffusion Implicit Model (DDIM), Song et al., ICLR 2021
[5] High-resolution image synthesis with latent diffusion models, Rombach et al., CVPR 2022
[6] Progressive Distillation for Fast Sampling of Diffusion Models, Salimans et al., ICLR 2022
[7] SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, Meng et al., ICLR 2022