SCEdit: Efficient and Controllable Image Diffusion Generation via Skip Connection Editing

Zeyinzi Jiang, Chaojie Mao, Yulin Pan, Zhen Han, Jingfeng Zhang

Alibaba

Project page: https://scedit.github.io/
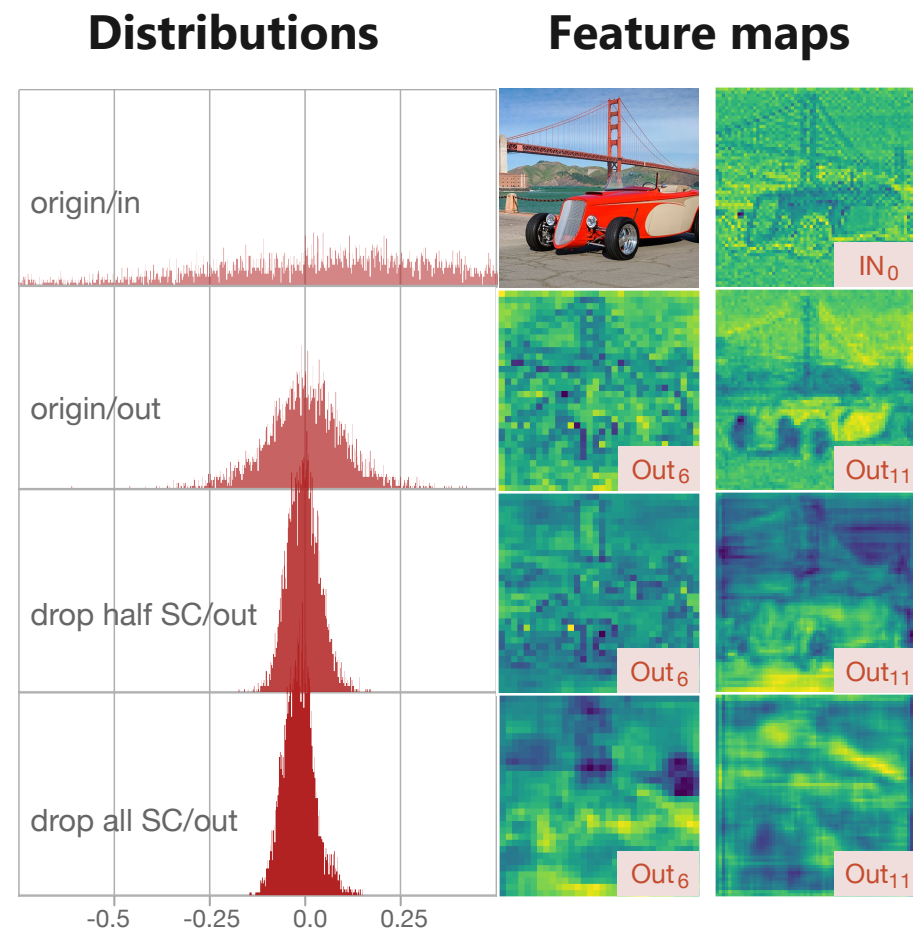
| GitHub Page | SEPTER | HF Demo | MS Demo |

Motivation

Hierarchical features aggregating long-distance information across encoder and decoder make a significant impact on the content and quality of image generation

Distributions   **Feature maps**

Removing the skip connections from different layers markedly affects the overall network output
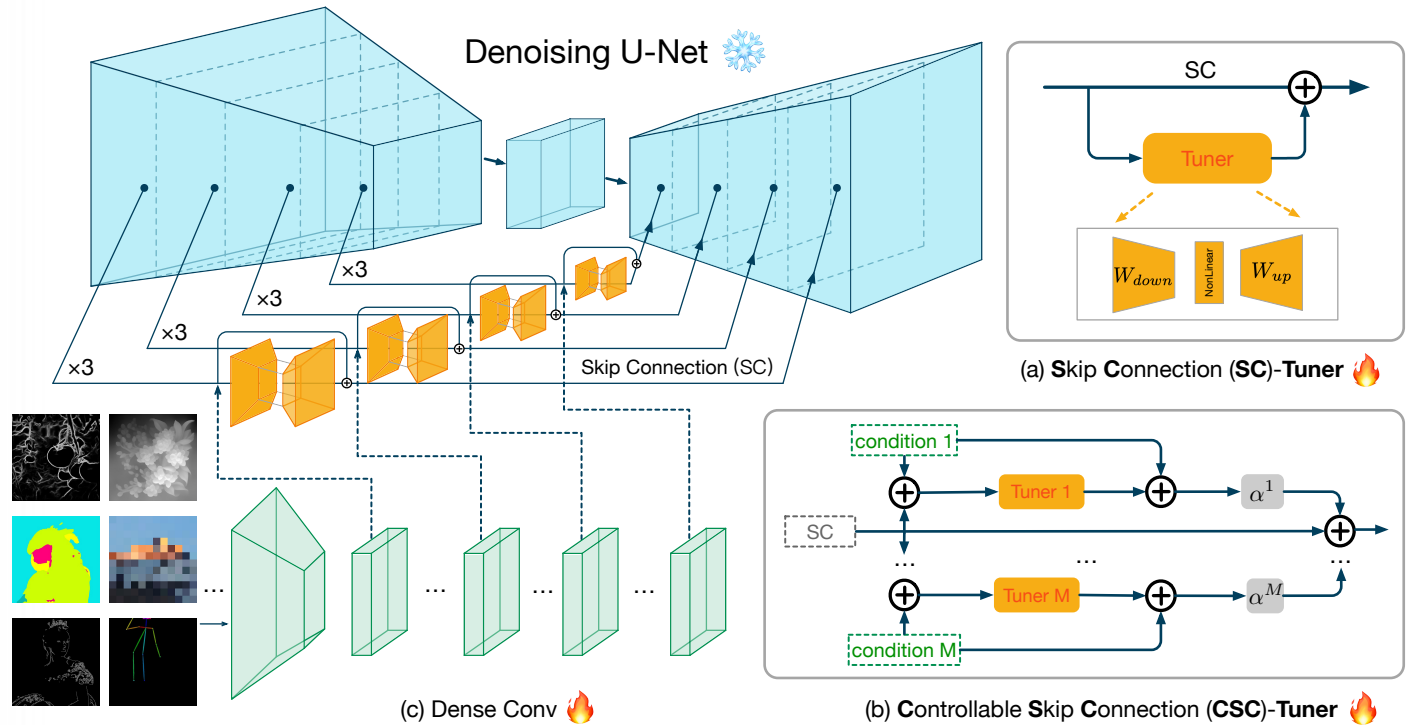
Framework

Efficient

A simple yet highly efficient framework by **S**kip **C**onnection **Edit**ing (**SCEdit**) for image generation with lightweight tuning module
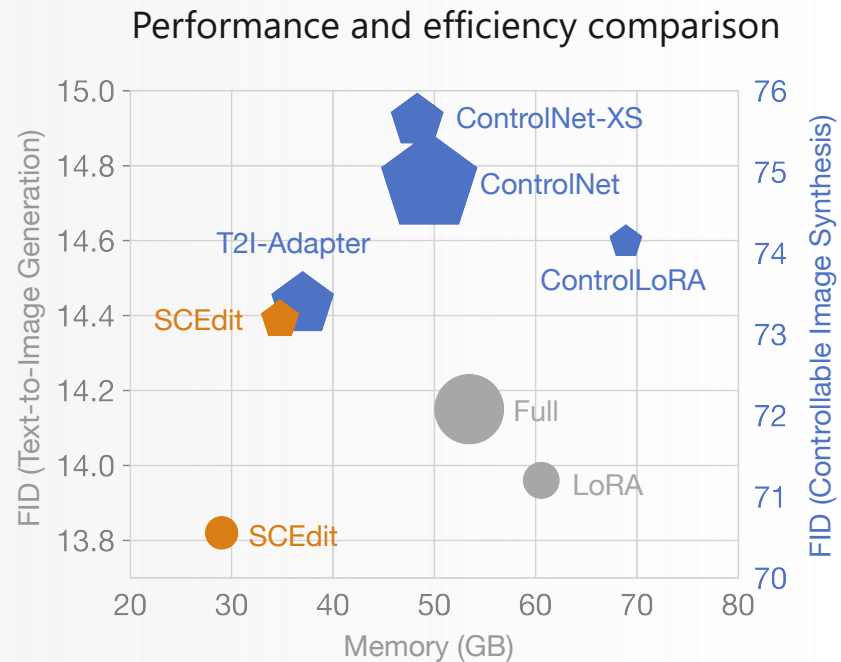
Controllable

Simplify the injection of different conditions for controllable image synthesis extensions, and unify the network design for multi-condition inputs

Denoising U-Net ❄️

×3 ×3 ×3 ×3 ×3

Skip Connection (SC)

(c) Dense Conv 🔥

(a) **S**kip **C**onnection (**SC**)-**Tuner** 🔥

SC

Tuner

W_{down} NonLinear W_{up}

(b) **C**ontrollable **S**kip **C**onnection (**CSC**)-**Tuner** 🔥

condition 1

Tuner 1 α^1

SC

Tuner M α^M

condition M

Text-to-image generation is fine-tuned by **SC-Tuner**, while controllable image synthesis is facilitated by **CSC-Tuner** and Cascade Dense Convolution
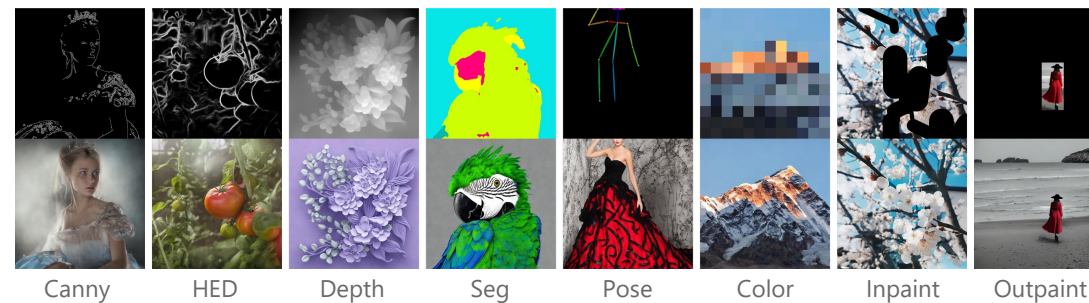
Results

Quantitative Results

Performance and efficiency comparison

- 🟠 : saves **30%~50%** training memory versus LoRA and Full.
- ⬠ : uses **7.9%** of ControlNet's params for a **30%** memory reduction.

Qualitative Results

Text-to-image generation

| Anime | Flat | 3D | Oil | Sketch | Watercolor |

Controllable image synthesis

| Canny | HED | Depth | Seg | Pose | Color | Inpaint | Outpaint |

Composable generation