Minimal Impact ControlNet: Advancing Multi-ControlNet Integration

Shikun Sun[1], Min Zhou, Zixuan Wang, Xubin Li, Tiezheng Ge, Zijie Ye, Xiaoyu Qin, Junliang Xing, Bo Zheng, Jia Jia

[1]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

Email: [1]ssk52839916@gmail.com

Problem

In current ControlNet training, **each control is designed to influence all areas of an image**, which can lead to conflicts when different control signals are expected to manage different parts of the image.

Solution & Case

- **Introduce silent control signals:** First introduce silent control signals that should remain inactive when other control signals are engaged, improving the compactness of the generation.
- **Feature injection and combination:** Employ strategies based on multi-objective optimization principles to improve model performance.
- **Theoretical contribution:** Develop and integrate a conservativity loss function within a large modular network architecture.

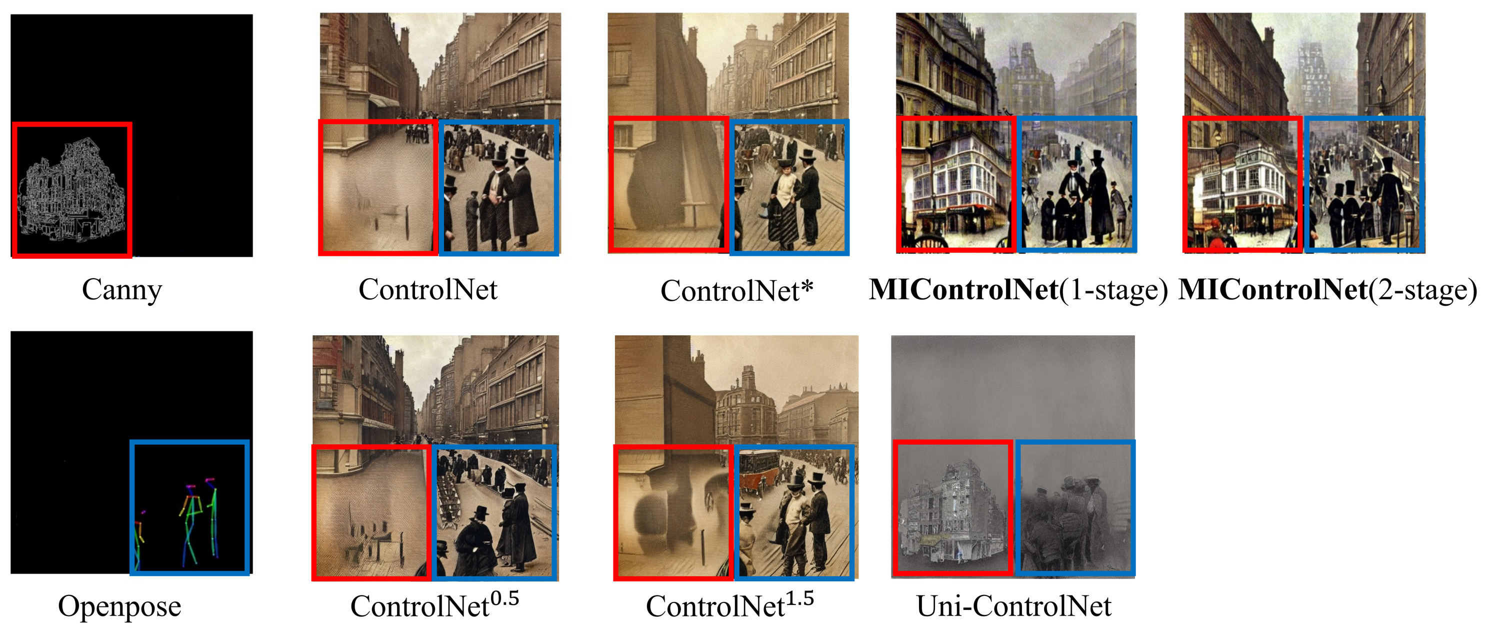
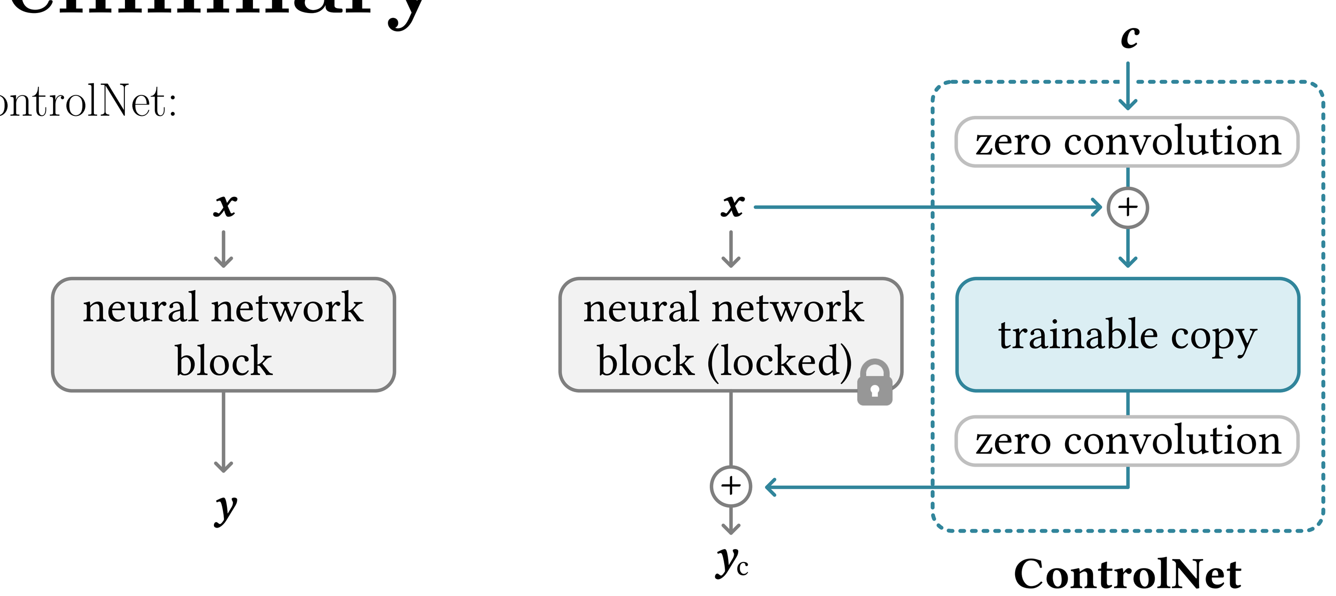


Figure 1: The **silent control signal** from OpenPose ControlNet (outside the blue box) suppresses the high-frequency control signal from Canny ControlNet (inside the red box). The black regions of the control signals represent the silent control signals.

Preliminary

1). ControlNet:



(a) Before (b) After

2). To enforce the conservativity of the score function directly, suppose the Jacobian matrix of $\mathbf{s}_t$ with respect to $\mathbf{x}_t$ is denoted as $\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}$. We propose using the following loss function:

$$\mathcal{L}_{QC} = \frac{1}{2}\mathbb{E}_{t,\mathbf{x}_t}\left\|\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t} - \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{\top}\right\|_F^2, \tag{1}$$

where F represents the Frobenius norm. That formula can be equivalently expressed as (Chao et al., 2022):

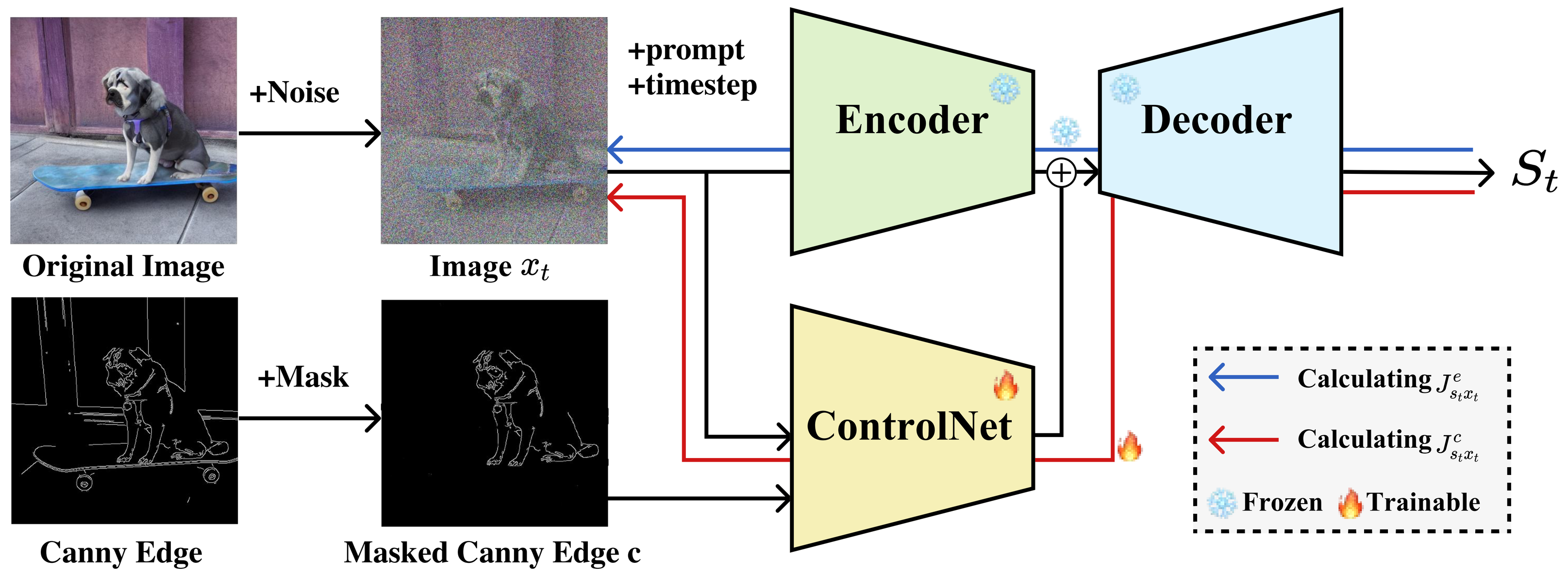
$$\mathcal{L}_{QC} = \mathbb{E}_{t,\mathbf{x}_t}\left[\mathrm{tr}(\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{\top}) - \mathrm{tr}(\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t})\right], \tag{2}$$

Suppose $\mathbf{v}$ fits a distribution whose expectation is $\mathbf{0}$ and variance is $\mathbf{I}$, by the Hutchinson's estimator, we have a unbiased estimation of $\mathcal{L}_{QC}$, which is

$$\mathcal{L}_{QC}^{est} = \mathbb{E}_{\mathbf{v},t,\mathbf{x_t}}\left[\mathbf{v}^{\top}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{\top}\mathbf{v} - \mathbf{v}^{\top}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}\mathbf{v}\right]. \tag{3}$$

Minimal Impact ControlNet

Overall Pipeline:



- **Data Augmentation.**
- **Heuristic Coefficient:**

$$\lambda_i^*(\mathbf{v}_1,\mathbf{v}_2) = \min\left[1, \max\left[\frac{(\mathbf{v}_2-\mathbf{v}_1)^T\mathbf{v}_2}{\|\mathbf{v}_2-\mathbf{v}_1\|_2^2}, 0\right]\right]. \tag{4}$$

- **Simplified Conservatively Loss:**

Proposition 1 (Decomposition of Jacobian Matrix)**.**

$$\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t} = \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{e} + \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}. \tag{5}$$

Assumption 1 (Responsibility for Conservativity)**.**

$$\nabla_{\phi}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{e} = \mathbf{0}. \tag{6}$$

$$\mathcal{L}_{QC}^{c} = \mathbb{E}_{\mathbf{v},t,\mathbf{x_t}}\mathbf{v}^{\top}\left[2\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{e}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c\top} - 2\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{e}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c} + \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c\top} - \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\right]\mathbf{v}. \tag{7}$$

We have

Proposition 2. *Under Assumption 1,*

$$\nabla_{\phi}\mathcal{L}_{QC}^{c} = \nabla_{\phi}\mathcal{L}_{QC}^{est}. \tag{8}$$

Define

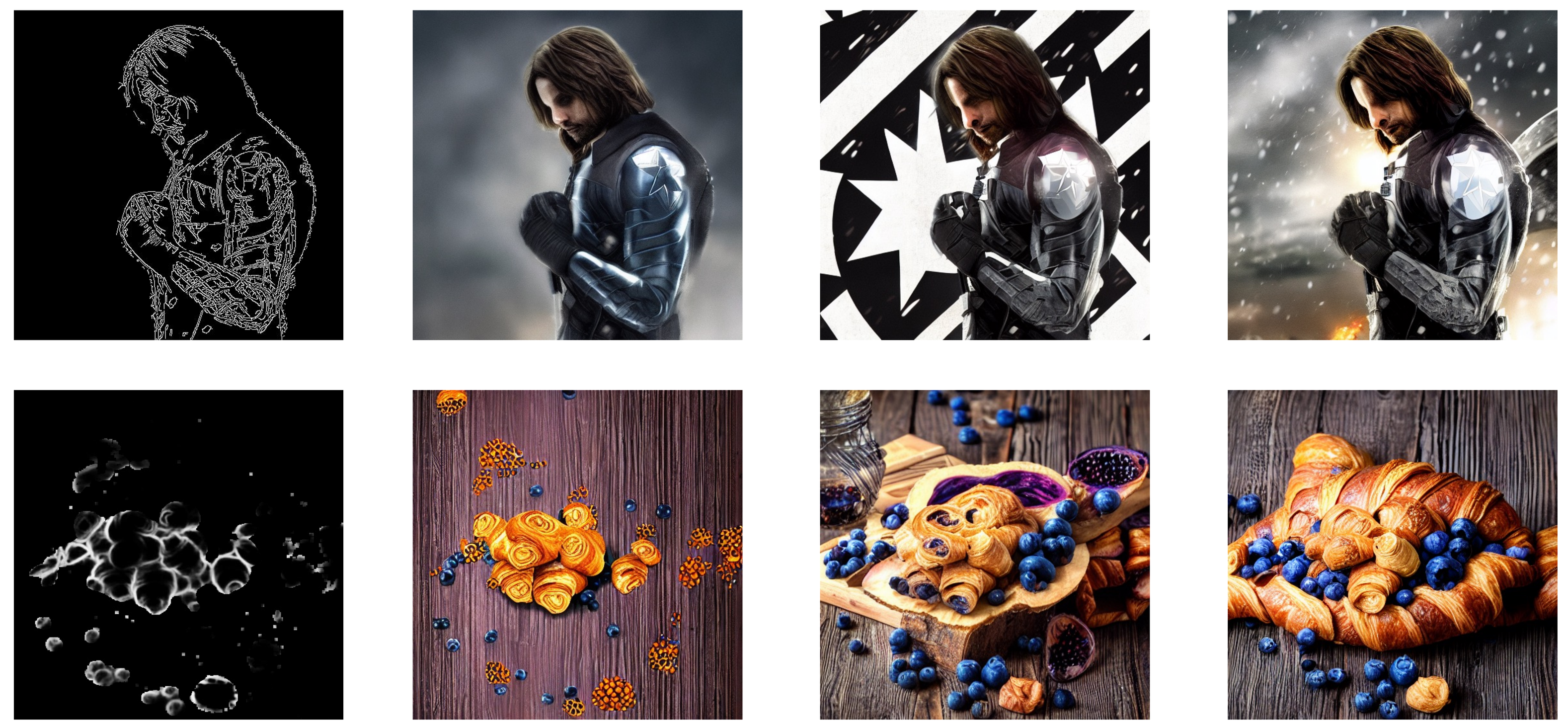
$$\mathcal{L}_{QC}^{simple} = \mathbb{E}_{\mathbf{v},t,\mathbf{x_t}}\mathbf{v}^{\top}\left[\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c\top} - \mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{c}\right]\mathbf{v}. \tag{9}$$

Theorem 1. *Suppose the Frobenius norm of $\mathbf{J}_{\mathbf{s}_t,\mathbf{x}_t}^{e}$ is uniformly bounded by M, we have*

$$\mathcal{L}_{QC}^{c} \leq 2\sqrt{2}M\sqrt{\mathcal{L}_{QC}^{simple}} + \mathcal{L}_{QC}^{simple}, \tag{10}$$

which indicates that if the simplified loss is zero, the original loss is also zero.

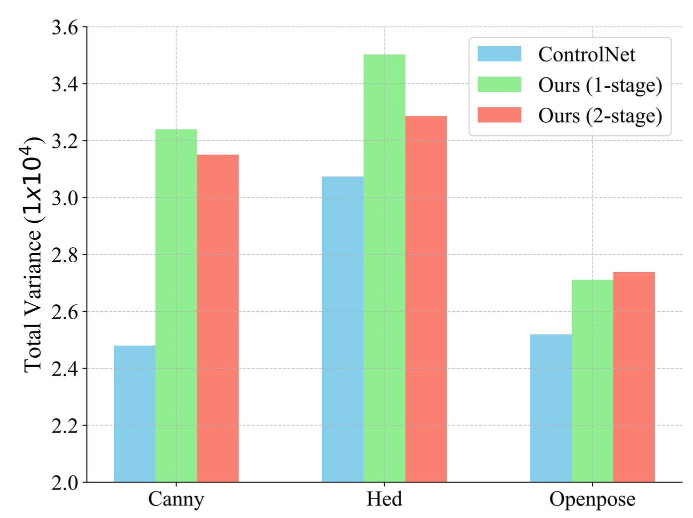
Experiment



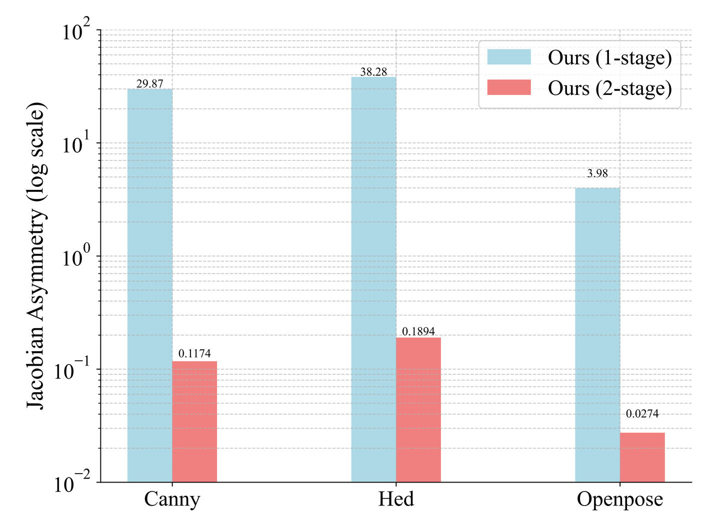
Canny / Hed ControlNet MIControlNet (1-stage) MIControlNet (2-stage)

Table 1: The FID of the multi-condition scenario. Each condition is associated with its own FID. the FID scores are presented with the best result highlighted in bold and the second best underlined.

Methods	Openpose-Canny	Openpose-Hed	Canny-Hed	Hed-Depth
ControlNet	80.37 / 111.30	76.98 / 84.20	123.59 / 86.43	91.98 / 86.25
ControlNet$^{0.5}$	105.86 / 123.13	145.88 / 107.52	143.67 / 106.40	-/-
ControlNet$^{1.5}$	**74.37** / 99.44	74.52 / 86.57	120.84 / 88.38	-/-
ControlNet*	77.43 / 89.57	76.69 / 78.31	122.10 / 85.45	78.14 / 90.65
ControlNet**	92.98 / 84.02	87.33 / 78.49	77.02 / 75.46	74.28 / 81.16
Uni-ControlNet	96.50 / <u>74.55</u>	139.87 / 76.06	88.77 / 75.47	73.68 / 89.94
Ours (1-stage)	76.13 / 77.22	**70.32 / 68.42**	<u>74.19</u> / <u>70.26</u>	<u>71.16</u> / <u>71.93</u>
Ours (2-stage)	<u>75.77</u> / **72.25**	<u>73.45</u> / <u>71.74</u>	**71.34 / 69.35**	**69.68 / 71.18**



(a) Total variance under silent control signals.



(b) Asymmetry in the Jacobian Matrix.