FlowIE: Efficient Image Enhancement via Rectified Flow

CVPR 2024 Oral

Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, Jiwen Lu

Department of Automation, Tsinghua University
Tsinghua Shenzhen International Graduate School, Tsinghua University

Presenter: Junxin Lin
2025.3.16

FlowIE: Background

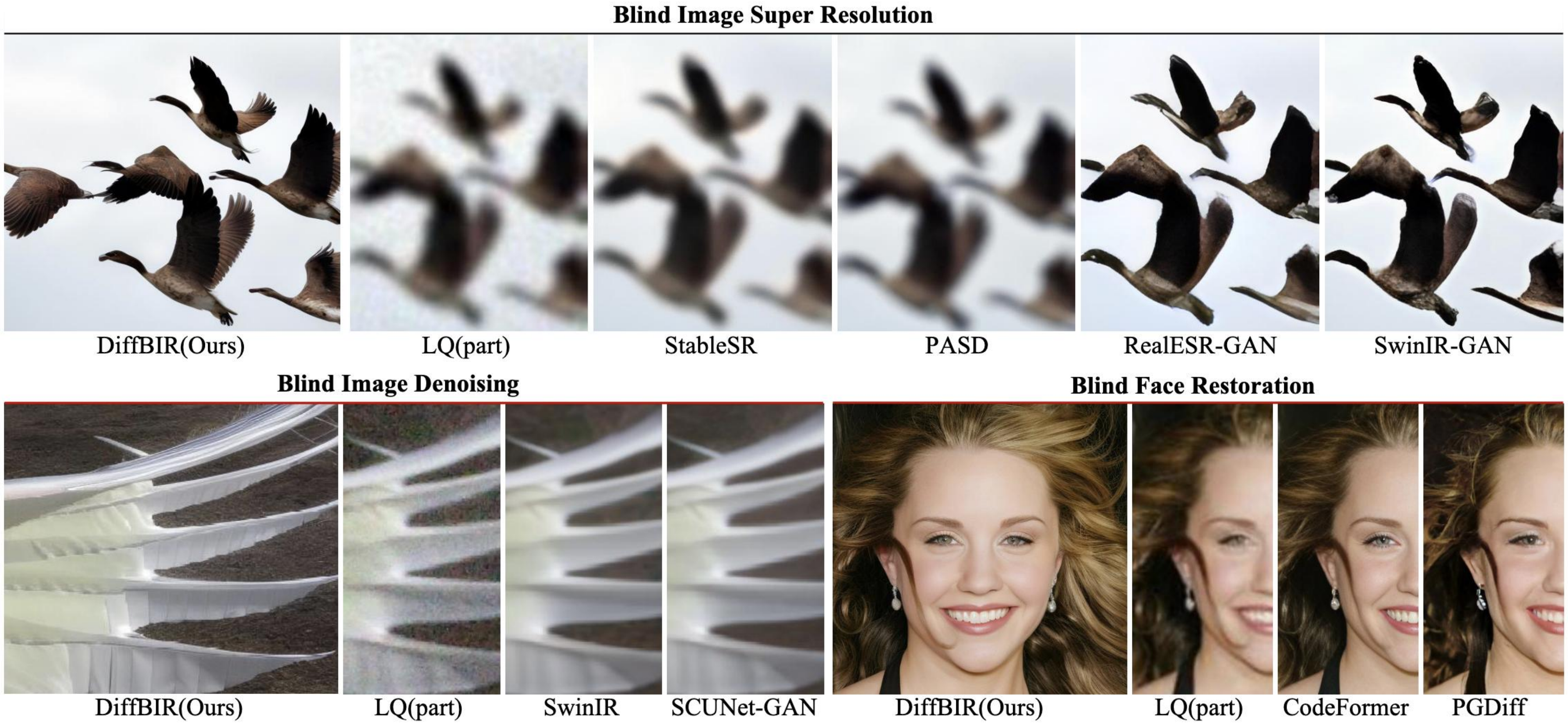
北京大學 PEKING UNIVERSITY

- Deep learning has shown strong power in image enhancement

	Model	Task	Limitation
Conventional Methods	BSRGAN	Blind image super-resolution	Generalization
	GFPGAN	Blind face restoration	Generalization
	RestoreFormer	Blind face restoration	Generalization
	CodeFormer	Blind face restoration	Generalization
Diffusion-based methods	SRDiff	Blind image super-resolution	Generalization
	DiffBIR	Blind face restoration Blind image denoising Blind image super-resolution	Inference speed

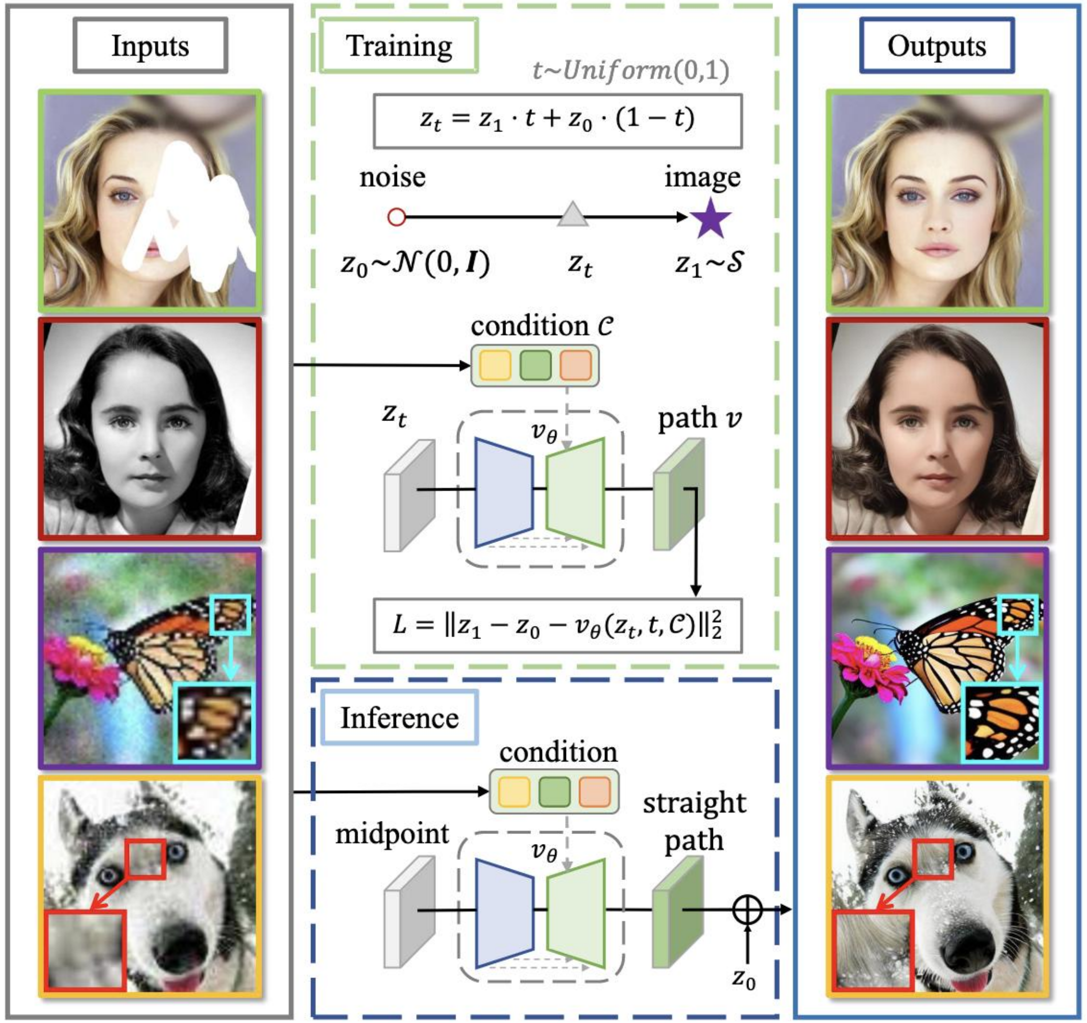
DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, Chao Dong



FlowIE: Overview

- **Task:**
  - Generalization capability
  - Less inference time cost
- **Overview:**
  - Apply conditioned Rectified Flow to diffusion models
  - Use mean value sampling
- **Performance:** surpass DiffBIR in inference speed and image details



Background: Diffusion Model

- Learn a noise predictor to gradually denoise a Gaussian distribution into a given distribution

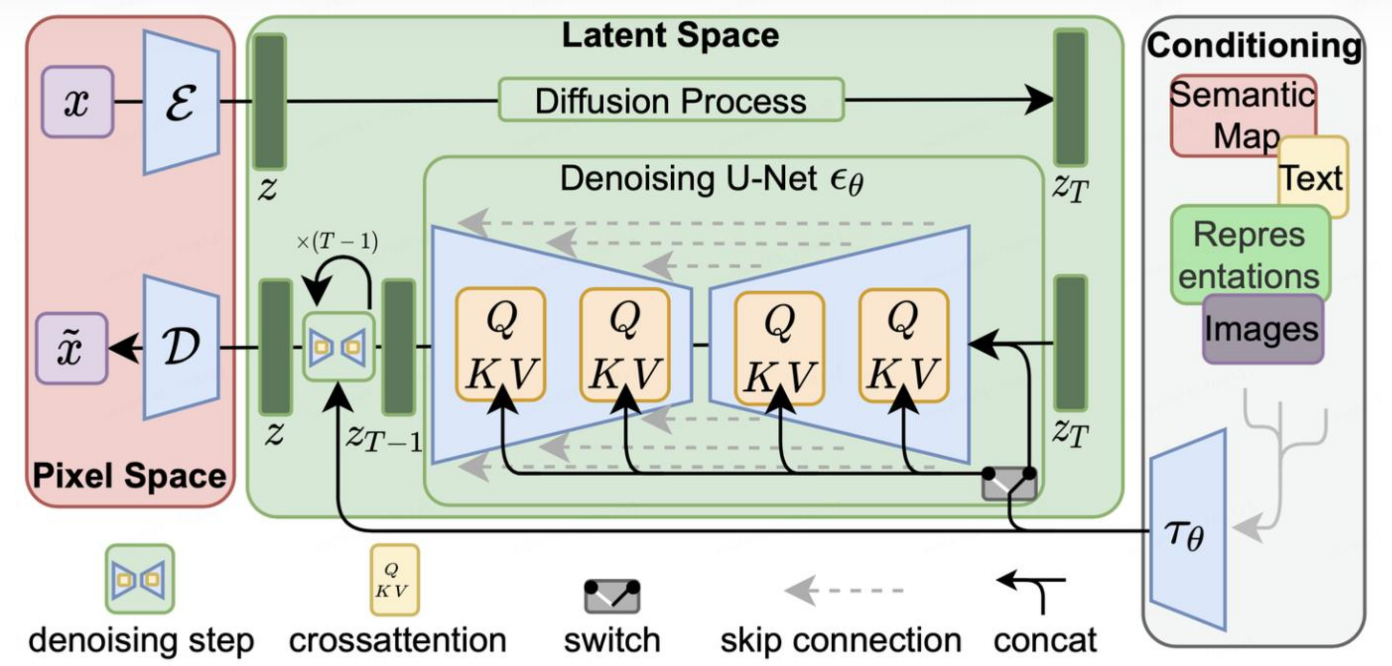
- **Forward:** $x_t = \sqrt{\overline{\alpha_t}} x_0 + \sqrt{1 - \overline{\alpha_t}} \cdot \epsilon$
- **Reverse:** $x_{t-1} = \frac{1}{\sqrt{\alpha}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}}} \cdot \epsilon(x_t, t) \right) + \sigma_t z$

Diffusion models:
Gradually add Gaussian noise and then reverse

Denoising Diffusion Probabilistic Models, NIPS 2020

Background: Diffusion Model

- **Latent diffusion model:**
  - Pixel space → latent space
  - Multimodal condition guidance



High-resolution Image Synthesis with Latent Diffusion Models, CVPR 2022

Background: Rectified Flow

- **Problem:** long-time inference

DDPM

$$
\begin{aligned}
L_{\mathrm{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T)\prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \Big(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \cdot \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}\Big) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)\Big] \\
&= \mathbb{E}_q[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_T))}_{L_T} + \sum_{t=2}^T \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]
\end{aligned}
$$

- **Solution:** one-step inference

Rectified Flow

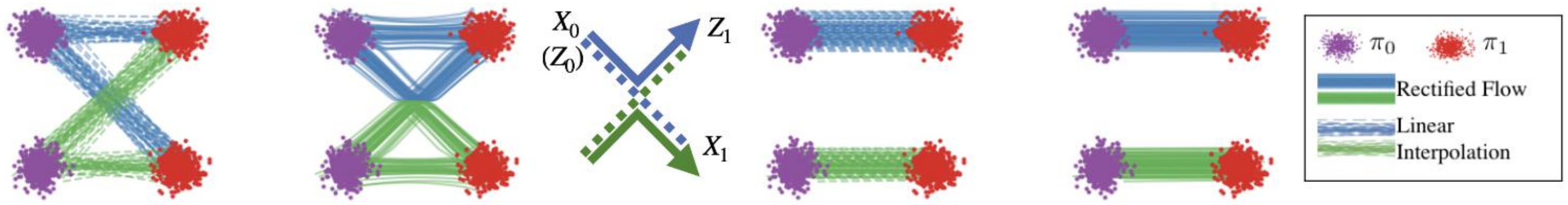
$$
L = \int_0^1 \mathbb{E}_{X_0\sim\text{noise},\, X_1\sim\text{data}} [\|\, (X_1 - X_0) - v(X_t, t)\, \|^2]\, \mathrm{d}t,
$$

$$
\text{with } X_t = t\, X_1 + (1-t) X_0
$$

Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, ICLR 2023

Background: Rectified Flow

- **Question:** find a transport map $T: R^d \to R^d$ from $X_0 \sim \pi_0$ to $X_1 \sim \pi_1$
- Build as a continuous normalized flow problem
  - $dX_t = v(X_t,\ t)dt$
- **Main idea:** Build straight flow from π_0 to π_1
  - $dX_t = (X_1 - X_0)dt \Leftrightarrow X_t = tX_1 + (1-t)X_0,\ t \in [0,1]$



(a) Linear interpolation $X_t = tX_1 + (1-t)X_0$

(b) Rectified flow Z_t induced by (X_0, X_1)

(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$

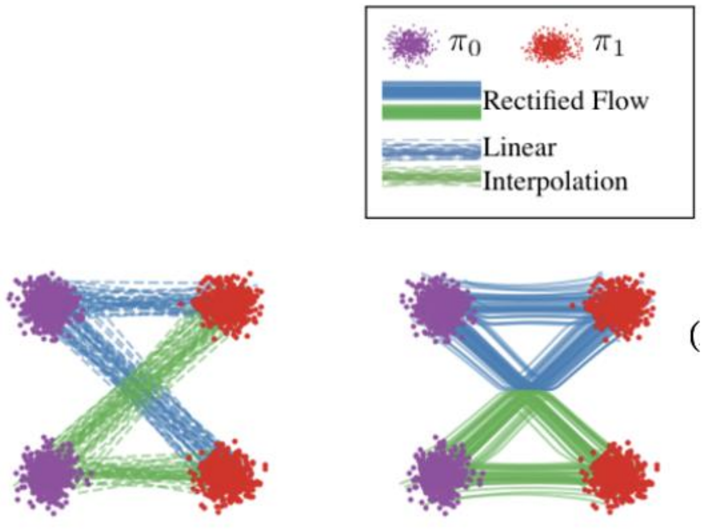
(d) Rectified flow Z_t' induced by (Z_0, Z_1)

Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, ICLR 2023

Background: Rectified Flow

- **1-Rectified:** $\min_{v} \int_0^1 E[\|(X_1 - X_0) - v(X_t, t)\|^2]\, dt, \quad with\; X_t = tX_1 + (1-t)X_0$
  - Randomly sampled pair $\langle X_0, X_1 \rangle$
  - Flows are not straight
  - $dX_t = v(X_t,\ t)dt$ promises the uniqueness

$$
\begin{aligned}
\mathcal{L} &= \int_0^1 \mathbb{E}_{x_0,x_1}\left[\|x_1 - x_0 - v_\theta(x_t,t)\|^2\right] dt \\
&= \int_0^1 \mathbb{E}_{x_0,x_1}\left[\|x_1 - x_0\|^2 + \|v_\theta(x_t,t)\|^2 - 2(x_1 - x_0)^T v_\theta(x_t,t)\right] dt \\
&= \int_0^1 \left\{\mathbb{E}_{x_0,x_1}\left[\|v_\theta(x_t,t)\|^2\right] - 2\mathbb{E}_{x_0,x_1}\left[(x_1 - x_0)^T v_\theta(x_t,t)\right]\right\} dt + C \\
&= \int_0^1 \left\{\mathbb{E}_{x_t}\left[\|v_\theta(x_t,t)\|^2\right] - 2\mathbb{E}_{x_t}\left[\mathbb{E}_{x_0,x_1}\left[x_1 - x_0 \mid x_t\right]\right]^T \mathbb{E}_{x_t}\left[v_\theta(x_t,t)\right]\right\} dt + C \\
&= \int_0^1 \mathbb{E}_{x_t}\left[\|\mathbb{E}_{x_0,x_1}\left[x_1 - x_0 \mid x_t\right] - v_\theta(x_t,t)\|^2\right] dt + C'
\end{aligned}
$$



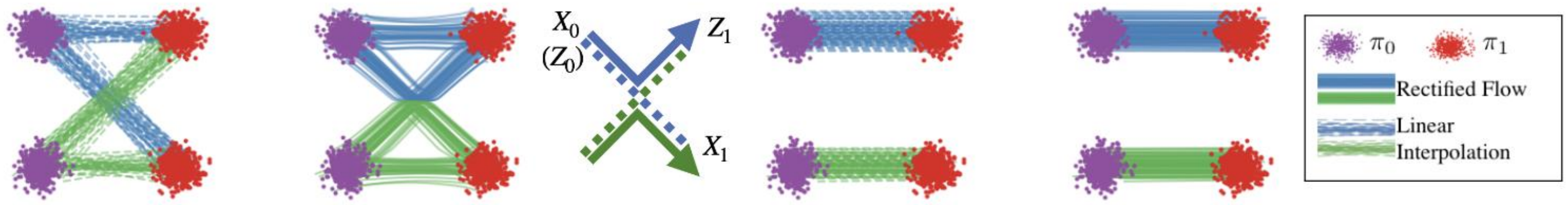
(a) Linear interpolation $X_t = tX_1 + (1-t)X_0$

(b) Rectified flow Z_t induced by (X_0, X_1)

Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, ICLR 2023

Background: Rectified Flow

- **2-Rectified:** $\min_{v} \int_0^1 E[\|(Z_1 - Z_0) - v(Z_t, t)\|^2]\, dt, \quad with\ Z_t = tZ_1 + (1-t)Z_0$
  - Paired data $\langle Z_0, Z_1 \rangle$ from rectified flow, i.e. $T(Z_0) = Z_1$
  - Most of the flows are straightened



(a) Linear interpolation $X_t = tX_1 + (1-t)X_0$

(b) Rectified flow Z_t induced by (X_0, X_1)

(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$

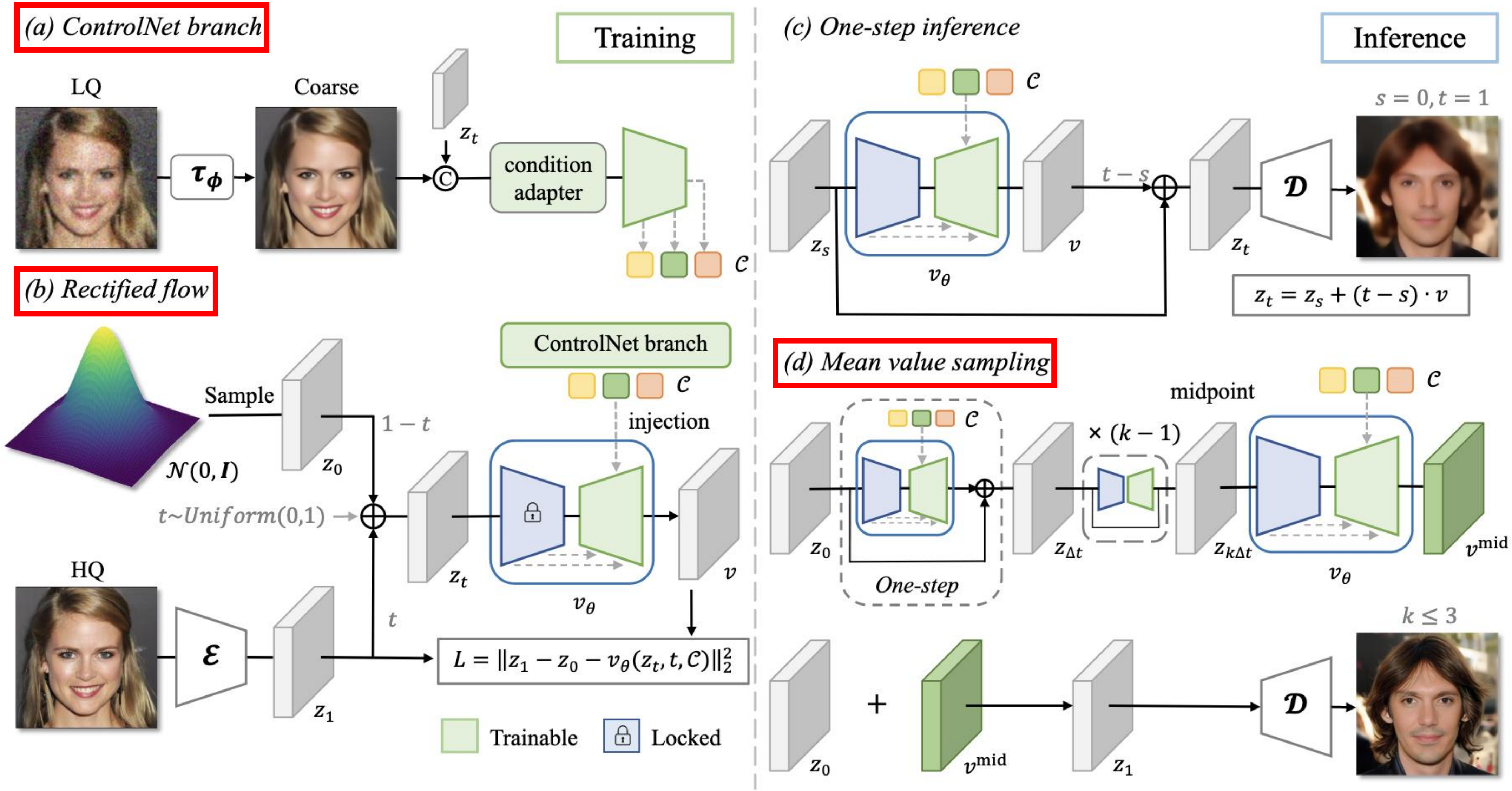
(d) Rectified flow Z'_t induced by (Z_0, Z_1)

Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, ICLR 2023

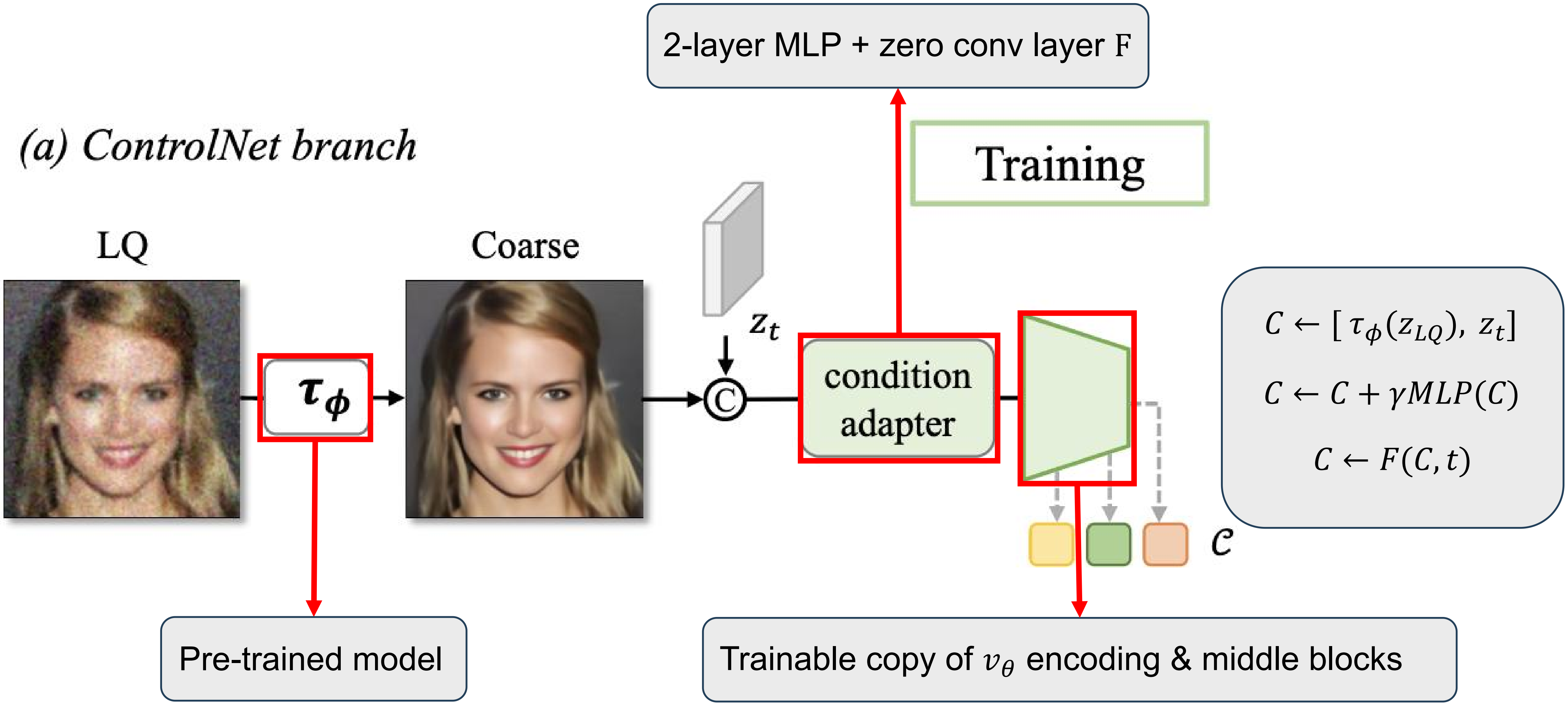
FlowIE: Methods

- **Flow-based image enhancement:**
  - Apply conditioned Rectified Flow to diffusion models
  - U-Net ϵ_θ → velocity predictor v_θ
  - LQ input z_{LQ} → condition guidance C
- **Improve quality via mean value sampling:** more precise v_{pred} from z_0 to z_1

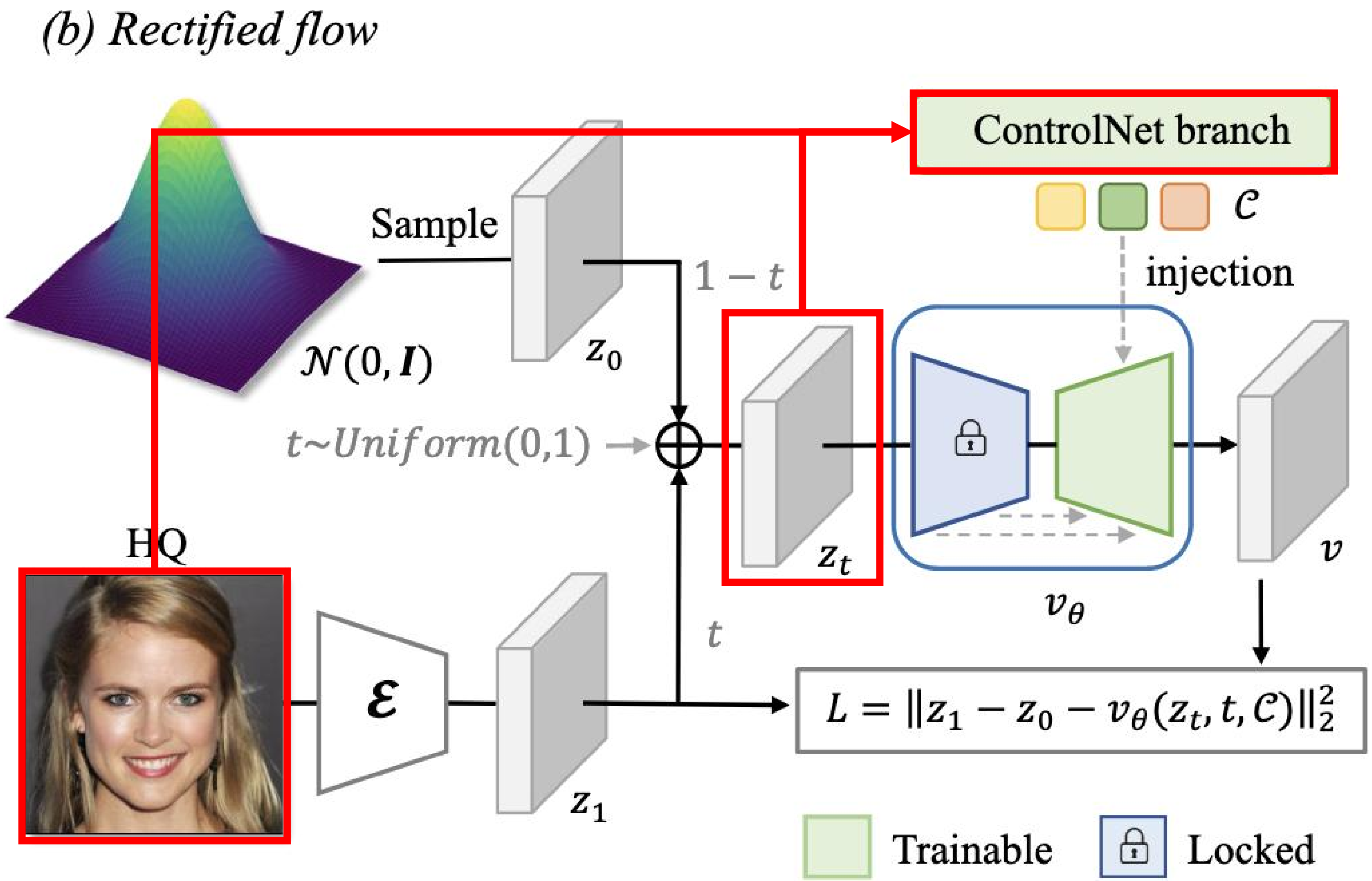
FlowIE: Pipeline



FlowIE: Pipeline



FlowIE: Pipeline



$$Z_{LQ} = D_h(Z_{HQ})$$

$$z_t = t \cdot z_1 + (1-t) \cdot z_0$$

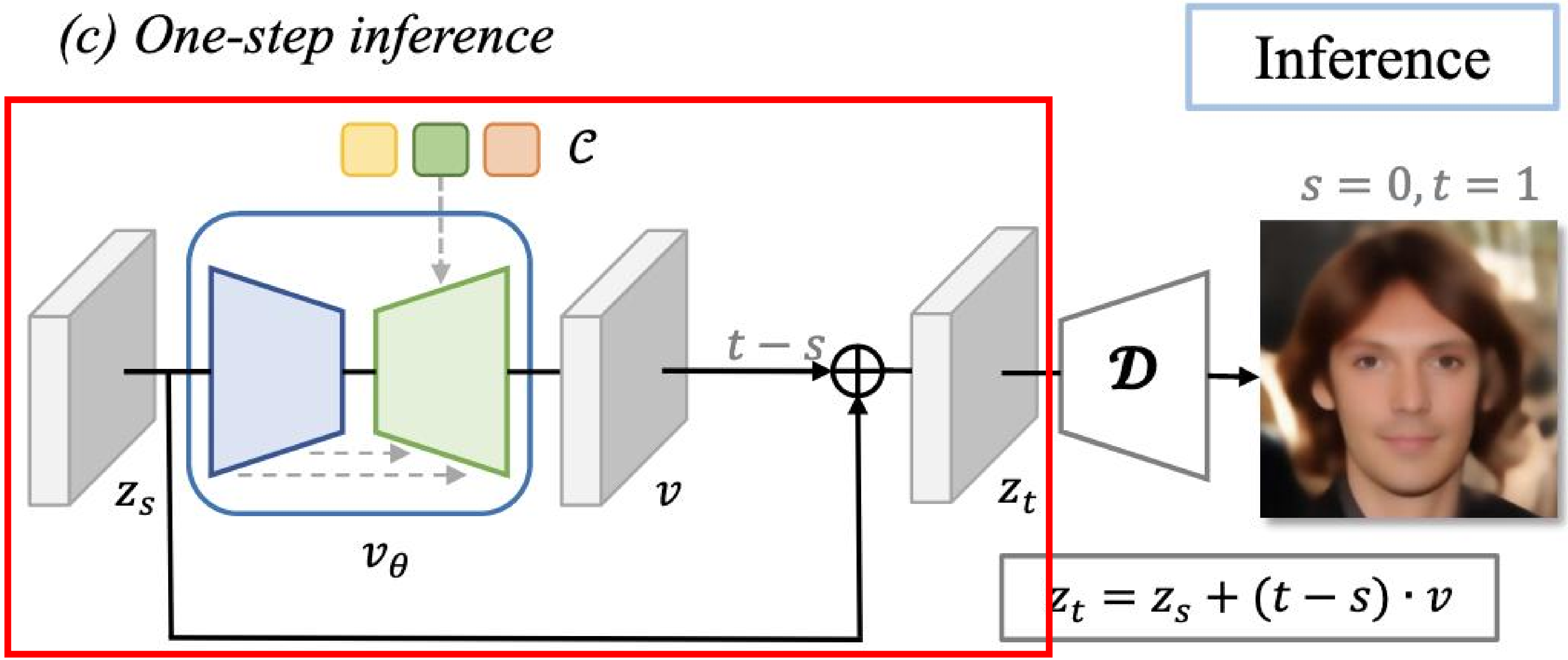
$$C \leftarrow [\,\tau_\phi(z_{LQ}),\ z_t]$$

$$C \leftarrow C + \gamma MLP(C)$$

$$C \leftarrow F(C, t)$$

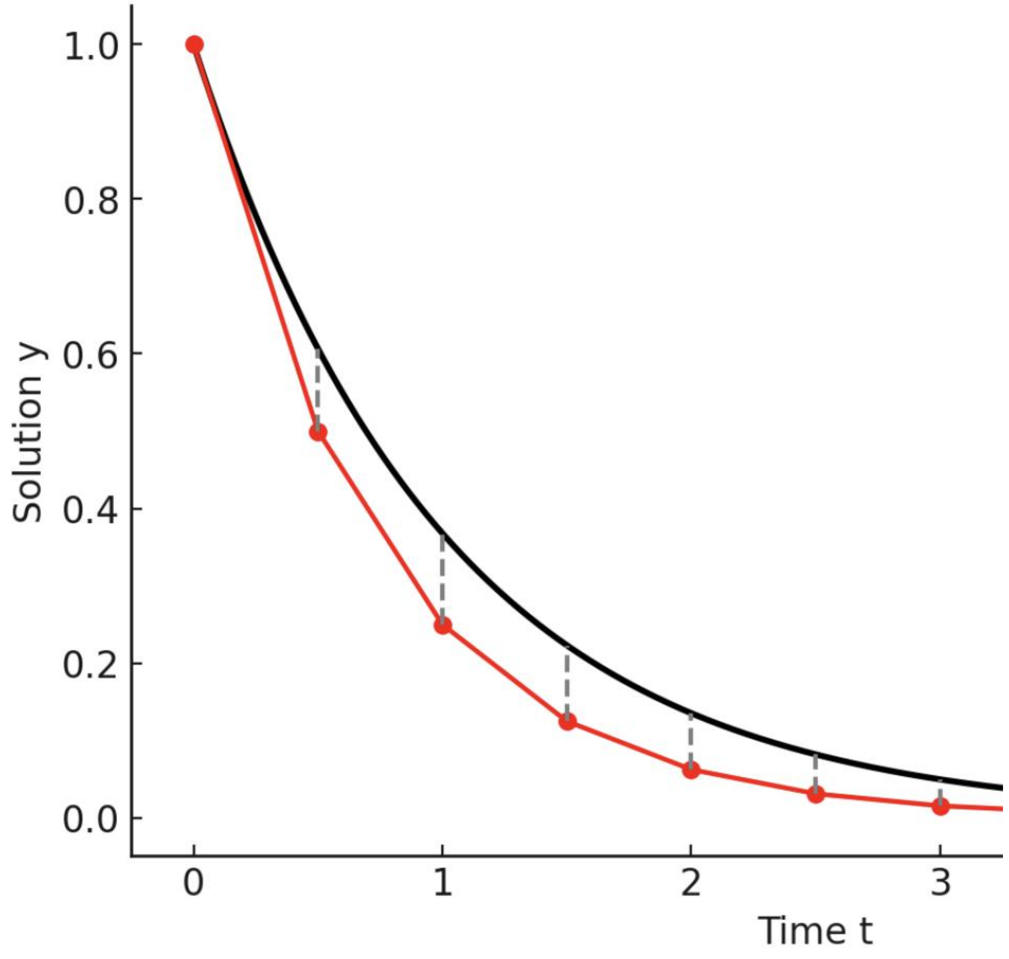
FlowIE: Pipeline

Forward Euler Method
$z_{t+\Delta t} = z_t + \Delta t \cdot v_\theta(z_t, t, C)$



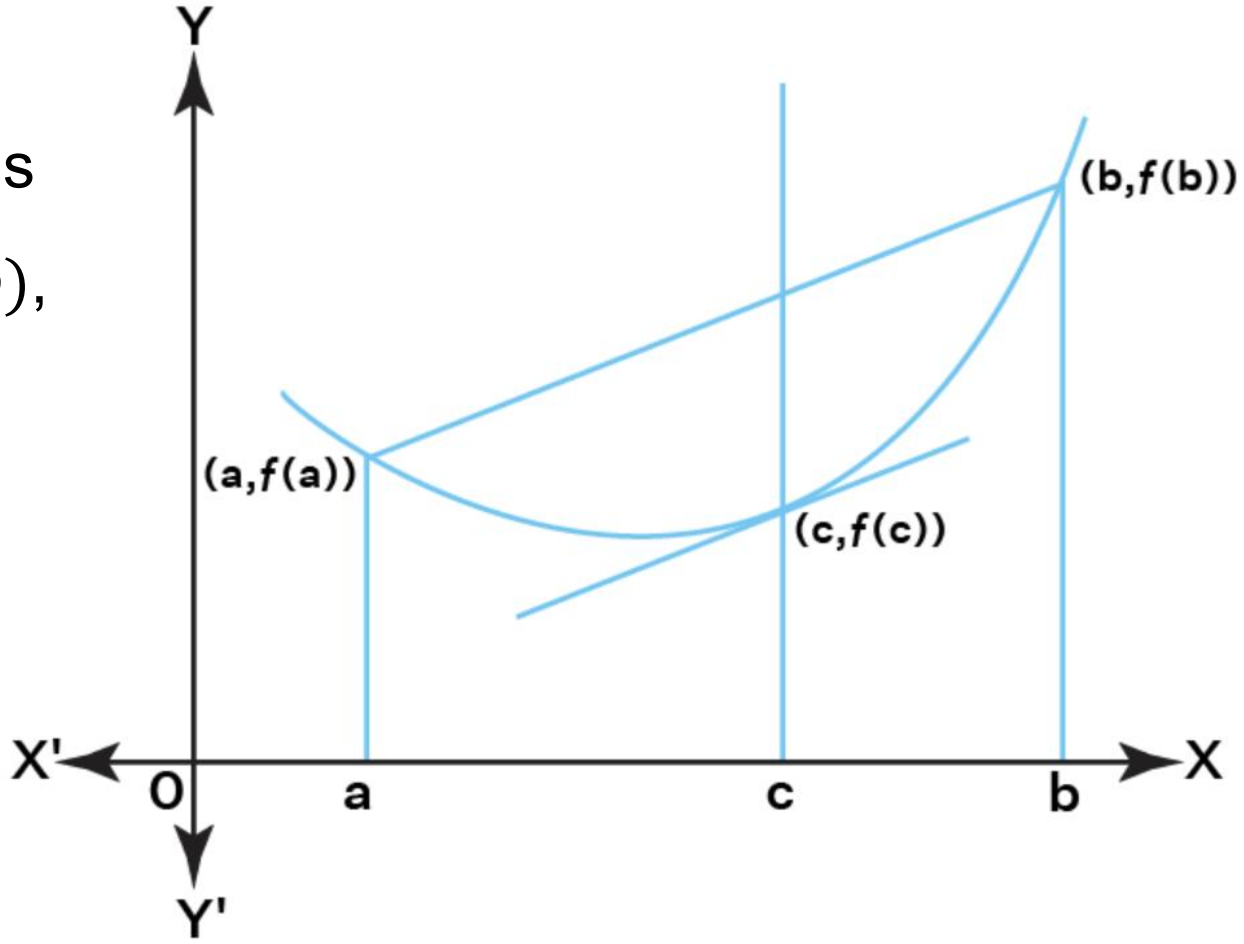
FlowIE: Pipeline

- **Forward Euler Method:**
  - $z_{t+\Delta t} = z_t + \Delta t \cdot v_\theta(z_t, t, C)$
  - may cause error accumulation
- Reduce error accumulation while obtaining correct results
- Lagrange’s Mean Value Theorem



FlowIE: Pipeline

- Lagrange's Mean Value Theorem: for function $f\colon [a,b] \to R$ continuous on $[a,b]$ and differentiable on (a,b), exists $c \in [a,b]$, such that $f'(c) = \frac{f(b)-f(a)}{b-a}$



FlowIE: Pipeline

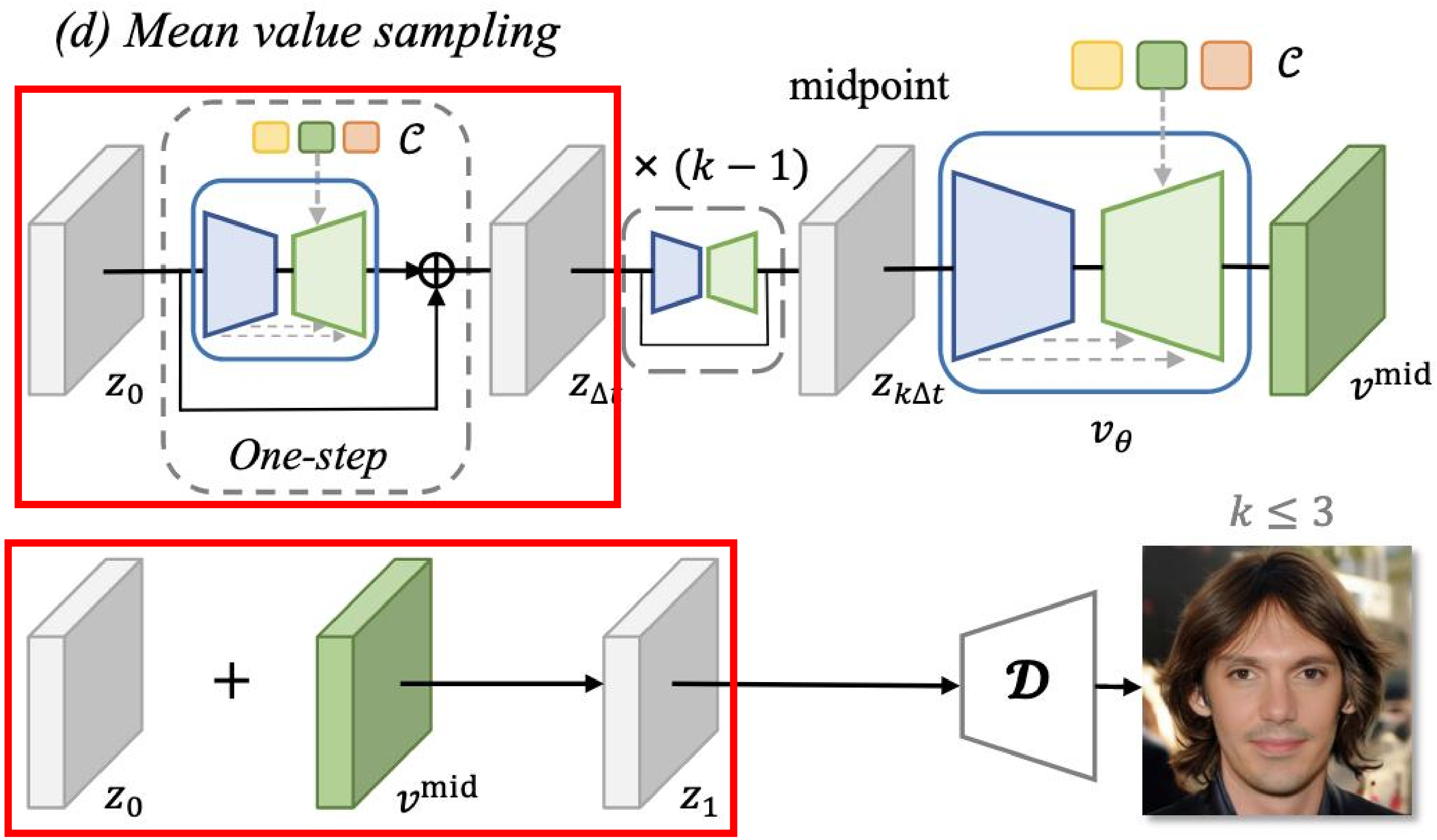
$$z_{(r+1)\Delta t} = z_{r\Delta t} + \Delta t \cdot v_\theta(z_{r\Delta t}, r\Delta t, C)$$
$$\Delta t = 1/N$$

$$C \leftarrow [\,\tau_\phi(z_{LQ}),\ z_t]$$

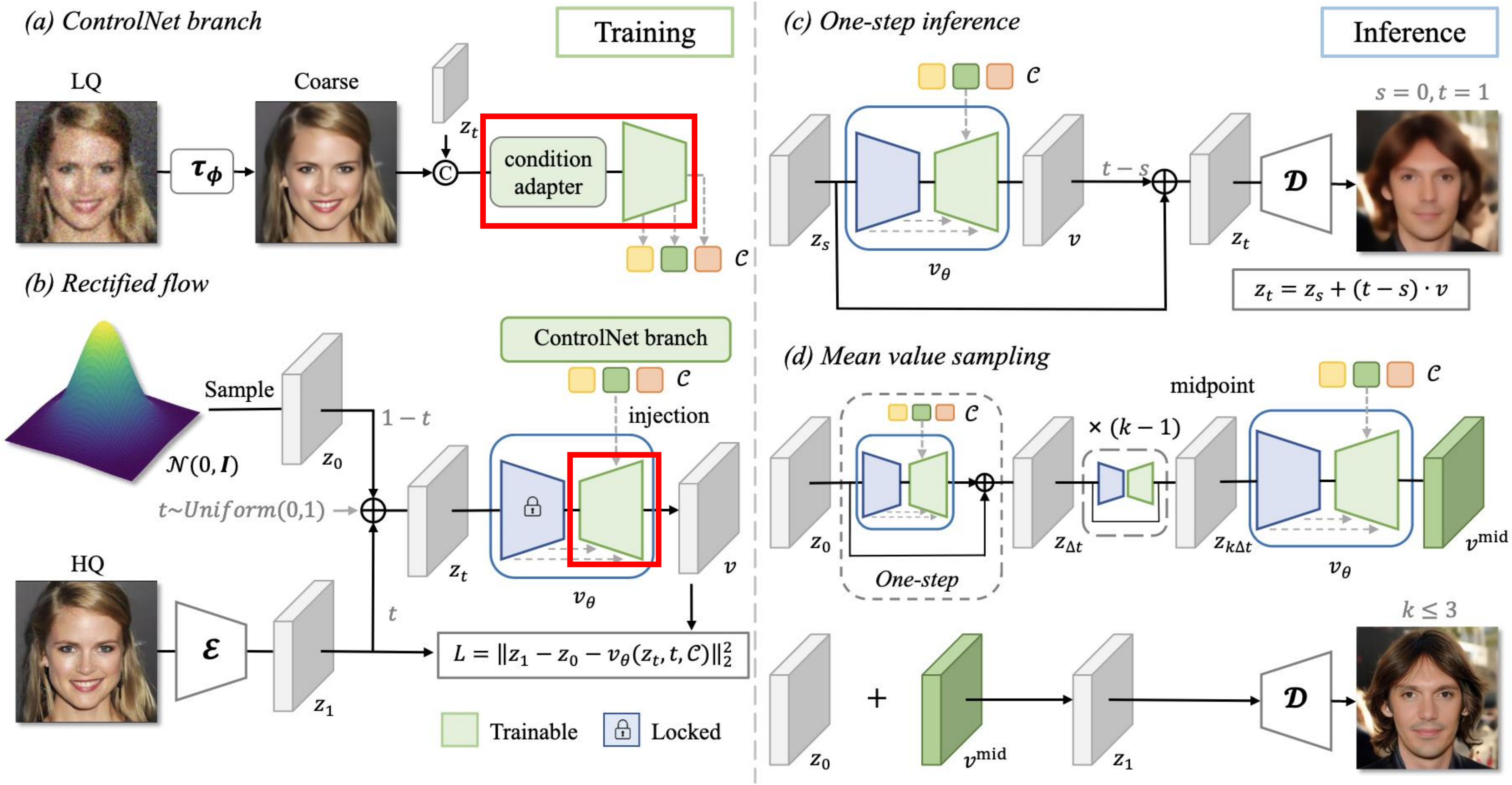
$$C \leftarrow C + \gamma MLP(C)$$

$$C \leftarrow F(C, t)$$

$$z_1 = z_0 + v_\theta(z_{k\Delta t}, k\Delta t, C)$$



FlowIE: Pipeline



FlowIE: Experiments

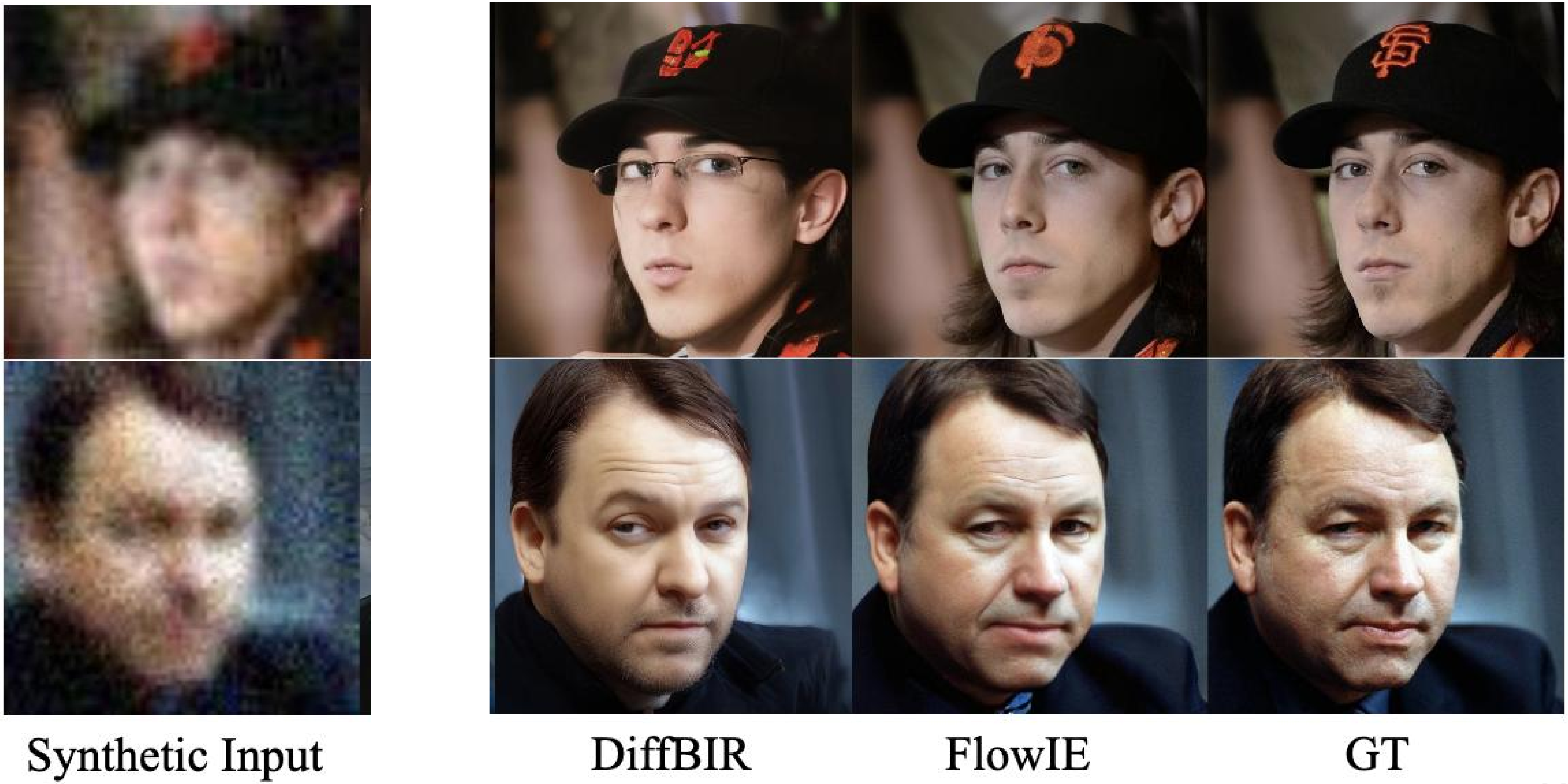
- **Face-related task (blind face restoration, face color enhancement, face inpainting):**
  - Train datasets: FFHQ
  - Test datasets:
    - CelebA-Test (synthetic),
    - LFW-Test & CelebChild-Test & WIDER-Test (real-world)
- **Blind image super-resolution:**
  - Train datasets: ImageNet
  - Test datasets: RealSRSet & collect-100 (newly constructed)
- Use image restoration baseline SwinIR as τ_ϕ

FlowIE: Experiments

- **Blind face restoration:**
  - Tune τ_ϕ on corresponding dataset
  - Lower FID
  - Higher PSNR, higher efficiency (compared with DiffBIR)

Method	Wild Datasets LFW	Wild Datasets WIDER	Wild Datasets CelebChild	Synthetic Dataset CelebA					FPS↑
	FID↓	FID↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	IDS ↑	
GPEN [50]	51.95	46.41	76.62	21.3941	0.5745	0.4685	23.88	0.49	7.278
GCFSR [16]	52.18	40.89	76.32	21.8789	0.6070	0.4579	35.52	0.45	9.243
GFPGAN [42]	52.11	41.70	80.69	21.6953	0.6060	0.4304	21.69	0.49	8.152
VQFR [15]	49.92	37.89	74.75	21.3012	0.6125	0.4127	20.47	0.48	3.837
RestoreFormer [47]	48.41	49.82	71.09	21.0029	0.5289	0.4791	43.76	0.55	4.964
DMDNet [23]	43.38	40.53	79.37	21.6620	0.5997	0.4825	64.21	0.66	3.454
CodeFormer [58]	52.34	38.79	79.58	22.1513	0.5949	0.4057	22.23	0.48	5.188
DiffBIR [25]	39.61	33.51	77.74	21.7512	0.5968	0.4575	20.19	0.52	0.285
FlowIE (Ours)	38.66	32.41	74.25	21.9211	0.6005	0.4367	19.81	0.69	2.846

FlowIE: Experiments



FlowIE: Experiments

Real Input

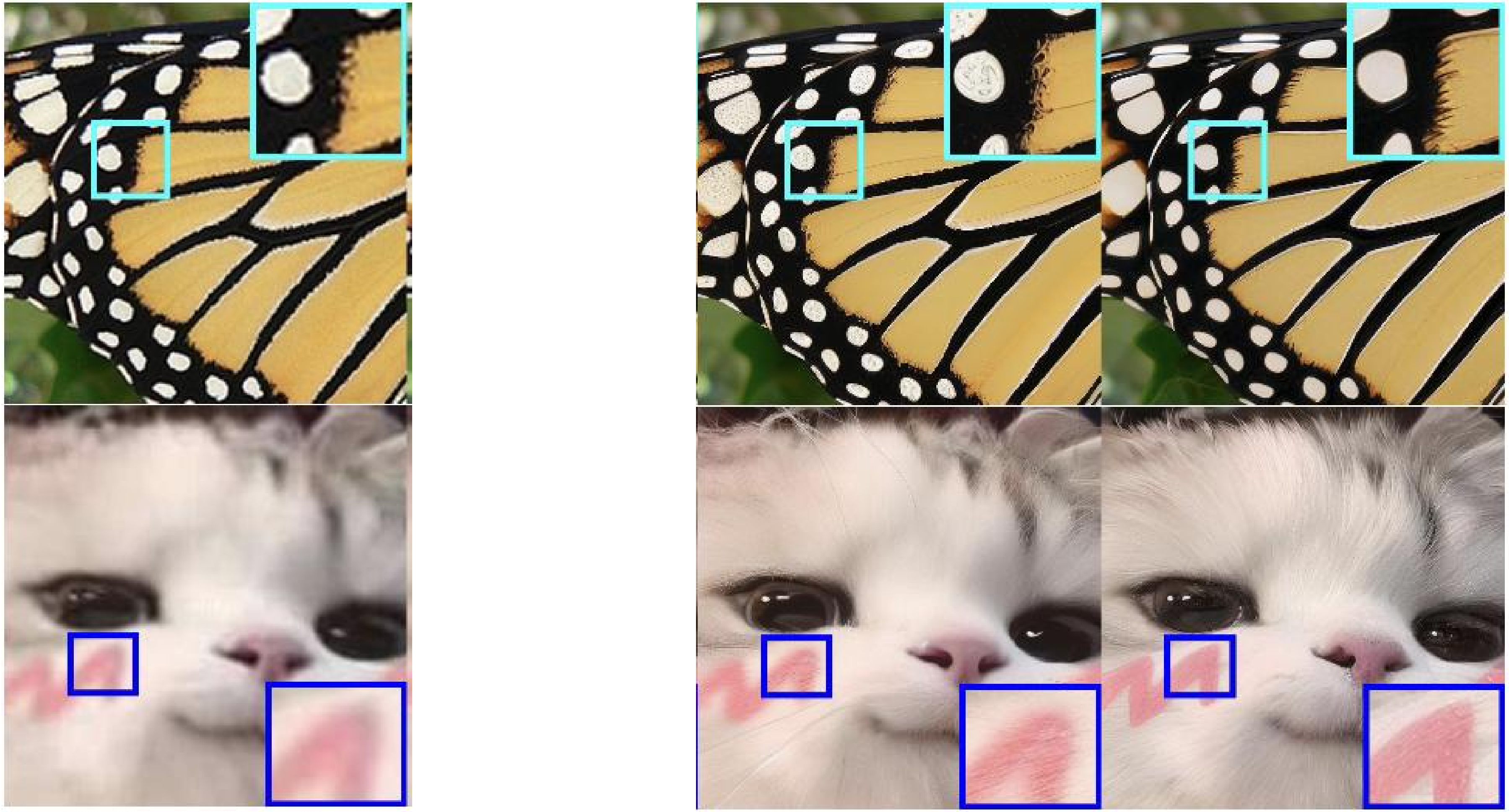
DiffBIR FlowIE

FlowIE: Experiments

- **Blind image super-resolution:**
  - Tune τ_ϕ on corresponding dataset
  - Higher MANIQA, higher efficiency (compared with DiffBIR)

Type	Method	MANIQA↑		FPS↑
		RealSRSet	Collect-100	
GAN	Real-ESRGAN+ [43]	0.5373	0.5901	1.875
	BSRGAN [57]	0.5638	0.5889	1.725
	SwinIR-GAN [24]	0.5296	0.5721	5.978
	FeMaSR [2]	0.5250	0.5718	3.167
Diffusion	DDNM [44]	0.4539	0.4813	0.071
	GDP [10]	0.4583	0.5237	0.016
	DiffBIR [25]	0.5906	0.6022	0.286
Flow	FlowIE (Ours)	**0.5953**	**0.6087**	2.853

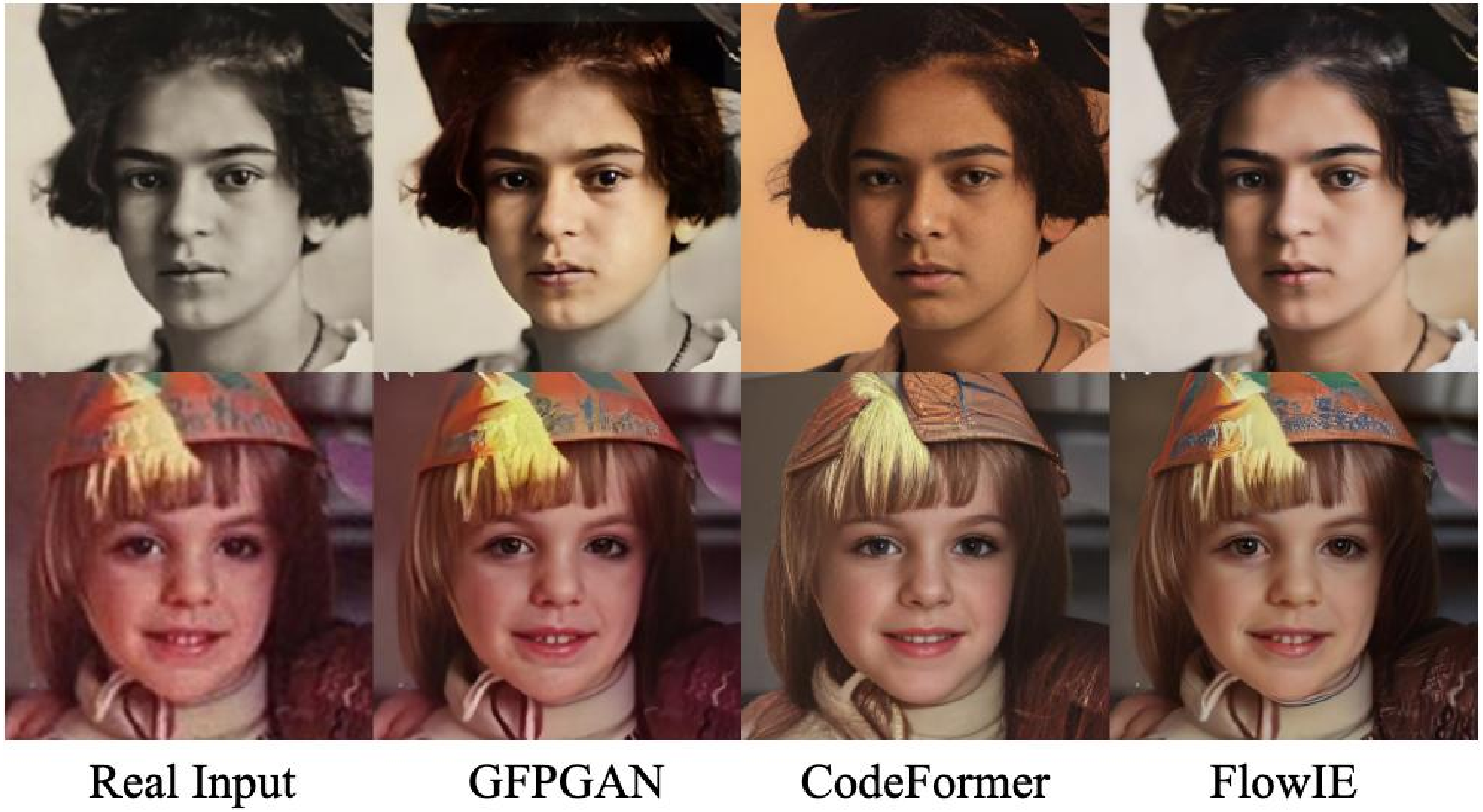
FlowIE: Experiments



LQ DiffBIR FlowIE

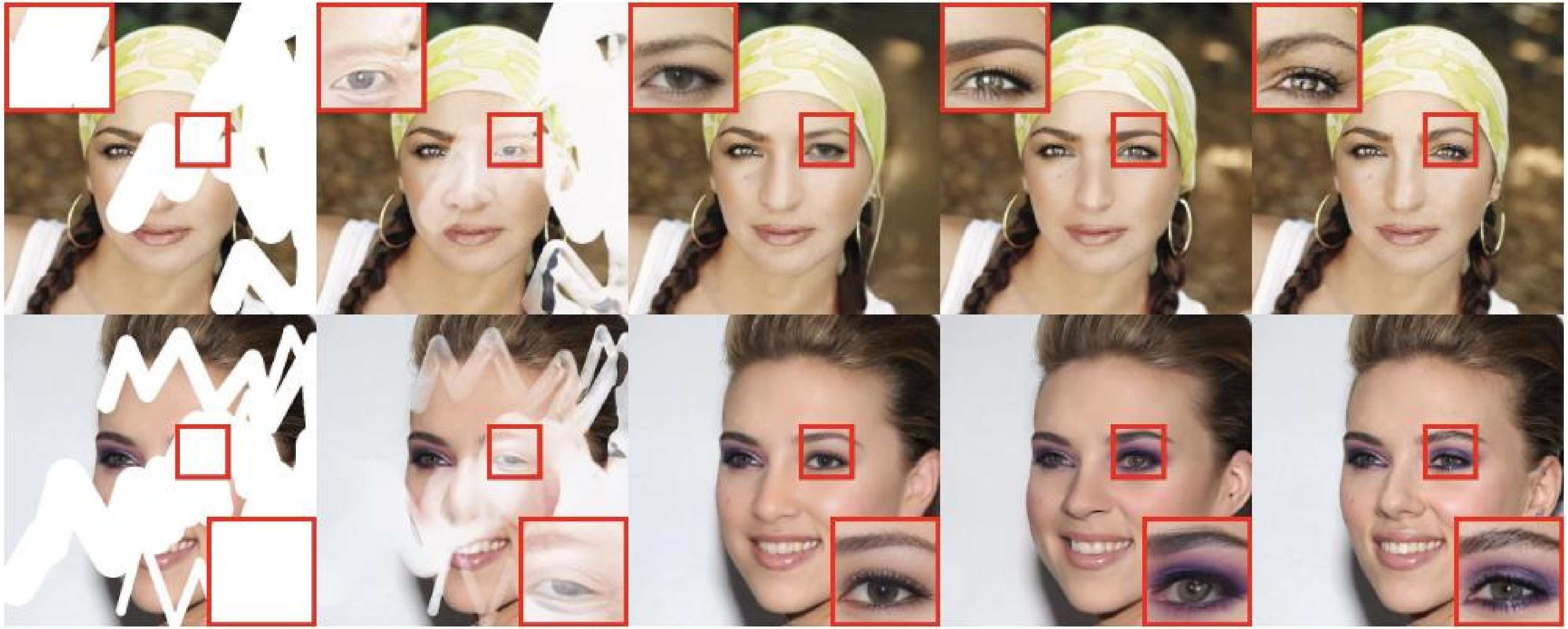
FlowIE: Experiments

- **Face color enhancement**
  - Fine-tune rectified flow



FlowIE: Experiments

- **Face inpainting**
  - Fine-tune rectified flow



Masked Input GPEN CodeFormer FlowIE GT

FlowIE: Ablation Study

- **w/o flow:** direct distillation rather than using Rectified Flow
  - Set the student identical to v_θ during training
  - Fix $t = 0$ during training

Method	FID↓		MANIQA↑	
	CelebA	LFW	RealSRSet	Collect-100
w/o flow	49.74	53.71	0.5311	0.5723
w/o mid sample	25.19	48.95	0.5489	0.5805
w/o init	27.76	52.63	0.5301	0.5698
FlowIE (Ours)	**19.81**	**38.66**	**0.5953**	**0.6087**

FlowIE: Ablation Study

- **w/o mid sample:** always use Forward Euler Method instead of mean value sampling
  - Euler method: struggling to produce HQ images in very few steps (e.g. 5)
  - Mean value sampling: a more efficient process (< 5 steps)

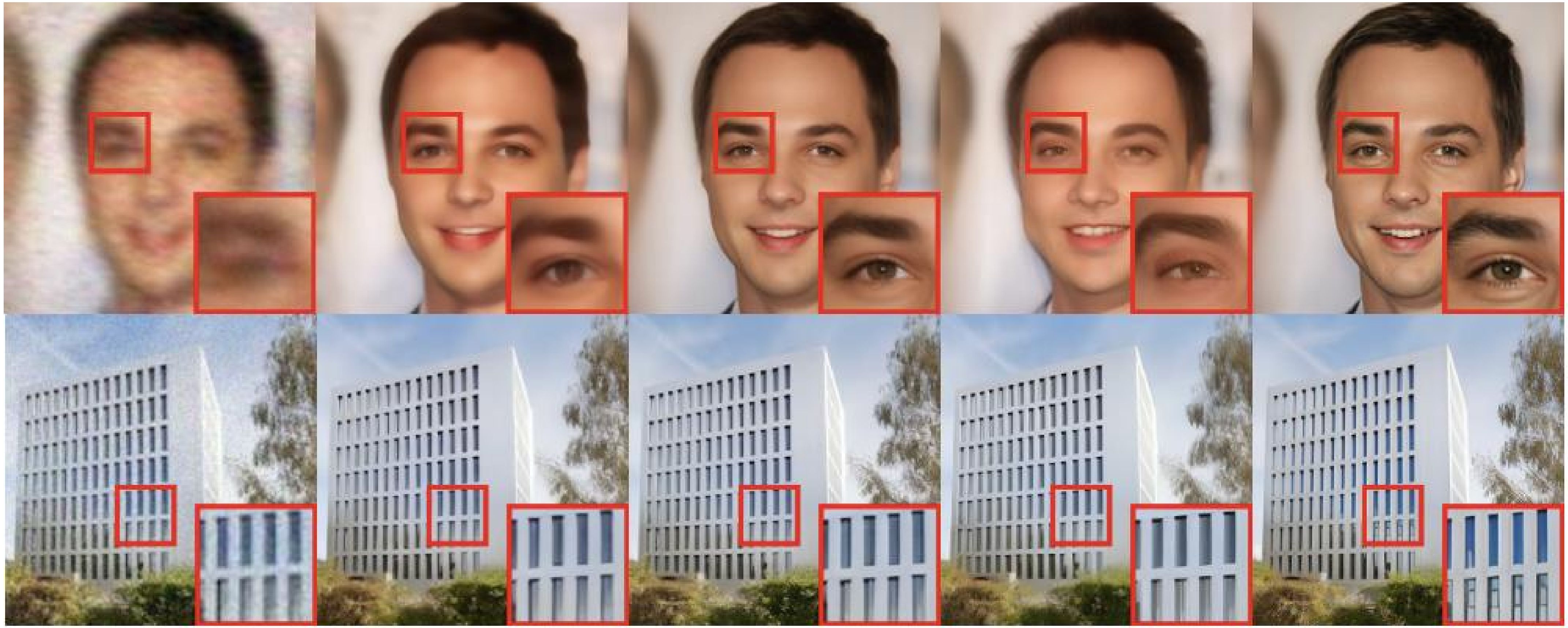
Method	FID↓		MANIQA↑	
	CelebA	LFW	RealSRSet	Collect-100
w/o flow	49.74	53.71	0.5311	0.5723
w/o mid sample	25.19	48.95	0.5489	0.5805
w/o init	27.76	52.63	0.5301	0.5698
FlowIE (Ours)	**19.81**	**38.66**	**0.5953**	**0.6087**

FlowIE: Ablation Study

- **w/o init:** train model without initial stage model τ_ϕ

Method	FID↓		MANIQA↑	
	CelebA	LFW	RealSRSet	Collect-100
w/o flow	49.74	53.71	0.5311	0.5723
w/o mid sample	25.19	48.95	0.5489	0.5805
w/o init	27.76	52.63	0.5301	0.5698
FlowIE (Ours)	**19.81**	**38.66**	**0.5953**	**0.6087**

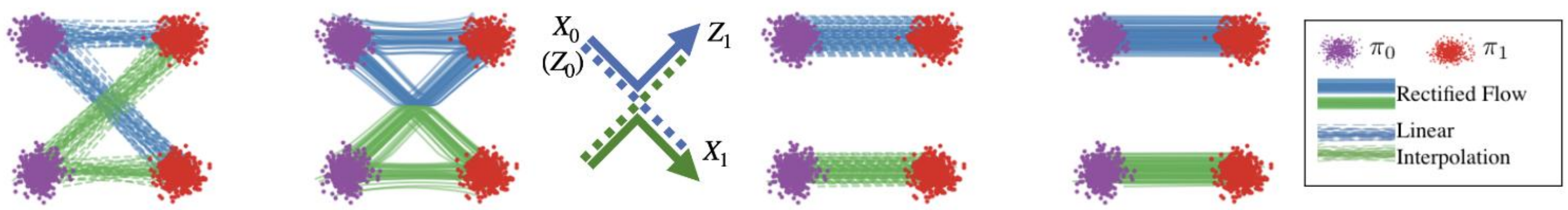
FlowIE: Ablation Study



LQ w/o flow w/o mid sample w/o init FlowIE

FlowIE: Discussion

- Problems of mean value sampling
  - Necessity: 2-Rectified can straighten flows
  - Rationality:
    - FlowIE uses 1-Rectified
    - Some points of flows produced by 1-Rectified may not be differentiable



(a) Linear interpolation $X_t = tX_1 + (1-t)X_0$

(b) Rectified flow Z_t induced by (X_0, X_1)

(c) Linear interpolation $Z_t = tZ_1 + (1-t)Z_0$

(d) Rectified flow Z'_t induced by (Z_0, Z_1)

Thanks for listening!

Presenter: Junxin Lin
2025.03.16