01 背景介绍

- 论文标题截图：
- 论文链接：https://arxiv.org/abs/2112.02475
- 录用信息：CVPR'22

去模糊 diffusion model

Deblurring via Stochastic Refinement

Jay Whang[†*] Mauricio Delbracio[‡] Hossein Talebi[‡] Chitwan Saharia[‡]
Alexandros G. Dimakis[†] Peyman Milanfar[‡]
[†]University of Texas at Austin [‡]Google Research

02 论文摘要

- **提出问题：**
- 目前的图像去模糊问题主要是deterministic的方法，重建的视觉质量不好

只用人眼看到的感受来评价是不够客观的

- **提出解决方案：**
- 提出了一个新框架，基于条件扩散模型
- 在视觉质量上有了很大的提升
- 同时也提出了一个有效的predict-and-refine的方法，并给出了diffusion model在PD曲线上遍历的方法

- **优势&实验结果：**
- 获得了更好的性能

- 论文摘要截图：

非高清图对应的高清图是无穷多的，如何从中选择最优的是非常困难的

Image deblurring is an ill-posed problem with multiple plausible solutions for a given input image. However, most existing methods produce a deterministic estimate of the clean image and are trained to minimize pixel-level distortion. These metrics are known to be poorly correlated with human perception, and often lead to unrealistic reconstructions. We present an alternative framework for blind deblurring based on conditional diffusion models. Unlike existing techniques, we train a stochastic sampler that refines the output of a deterministic predictor and is capable of producing a diverse set of plausible reconstructions for a given input. This leads to a significant improvement in perceptual quality over existing state-of-the-art methods across multiple standard benchmarks. Our predict-and-refine approach also enables much more efficient sampling compared to typical diffusion models. Combined with a carefully tuned network architecture and inference procedure, our method is competitive in terms of distortion metrics such as PSNR. These results show clear benefits of our diffusion-based method for deblurring and challenge the widely used strategy of producing a single, deterministic reconstruction.

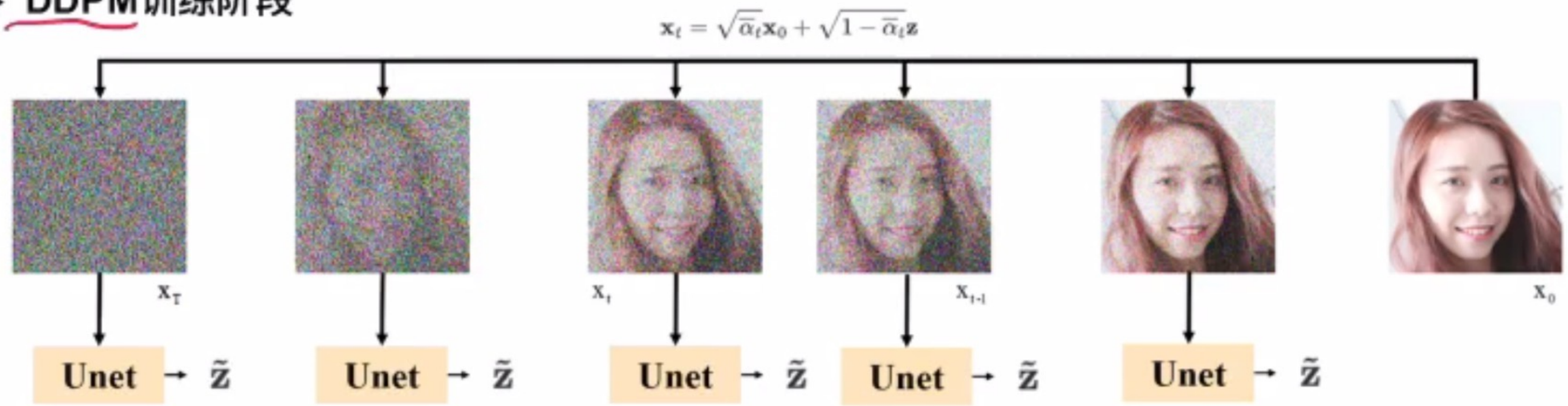
确定性方式

条件

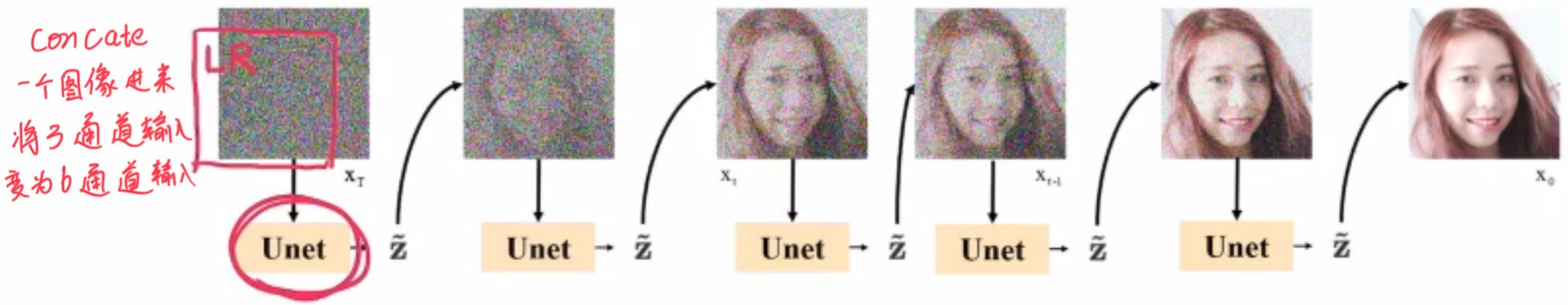
diffusion model

03 相关工作

➤ DDPM训练阶段



➤ DDPM推理阶段



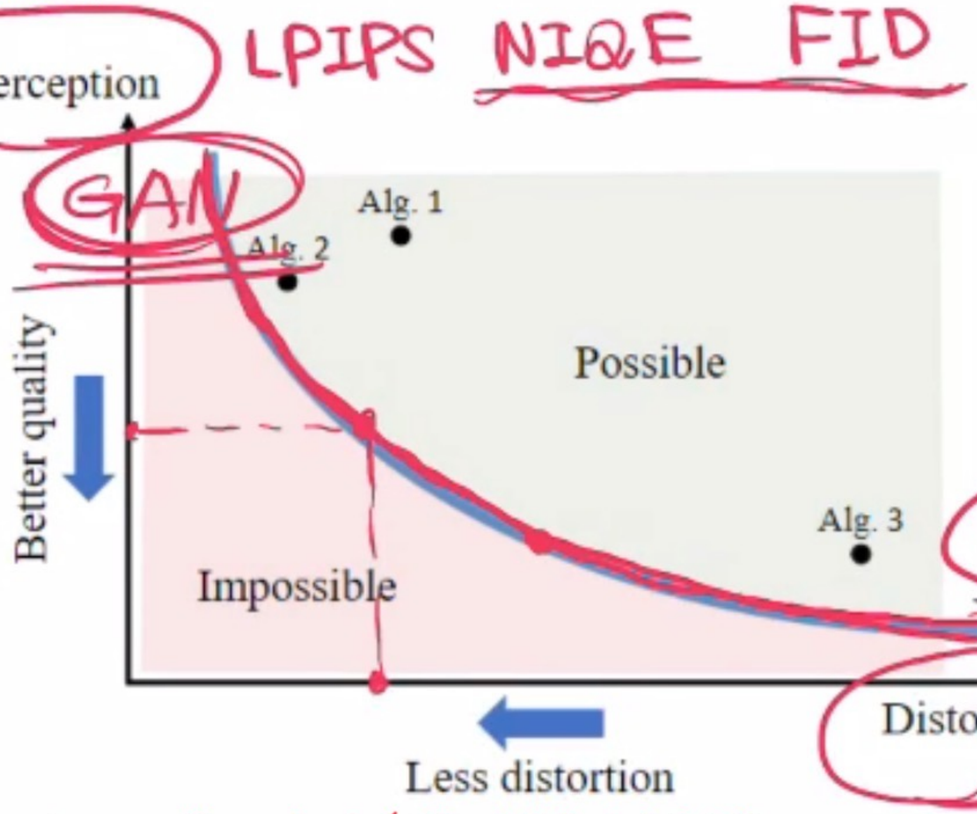
$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\tilde{\mathbf{z}}_\theta(\mathbf{x}_t, t)) + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\mathbf{z}$$

03 相关工作

- (CVPR'18) The Perception-Distortion Tradeoff
- PD曲线 模拟人的主观角度

评价标准

Deblur 目的是
给出同一方法，控制模型在这条线上移动



LPIPS NIQE FID

capacity

非GAN/l1 l2 loss

PSNR SSIM
MSE

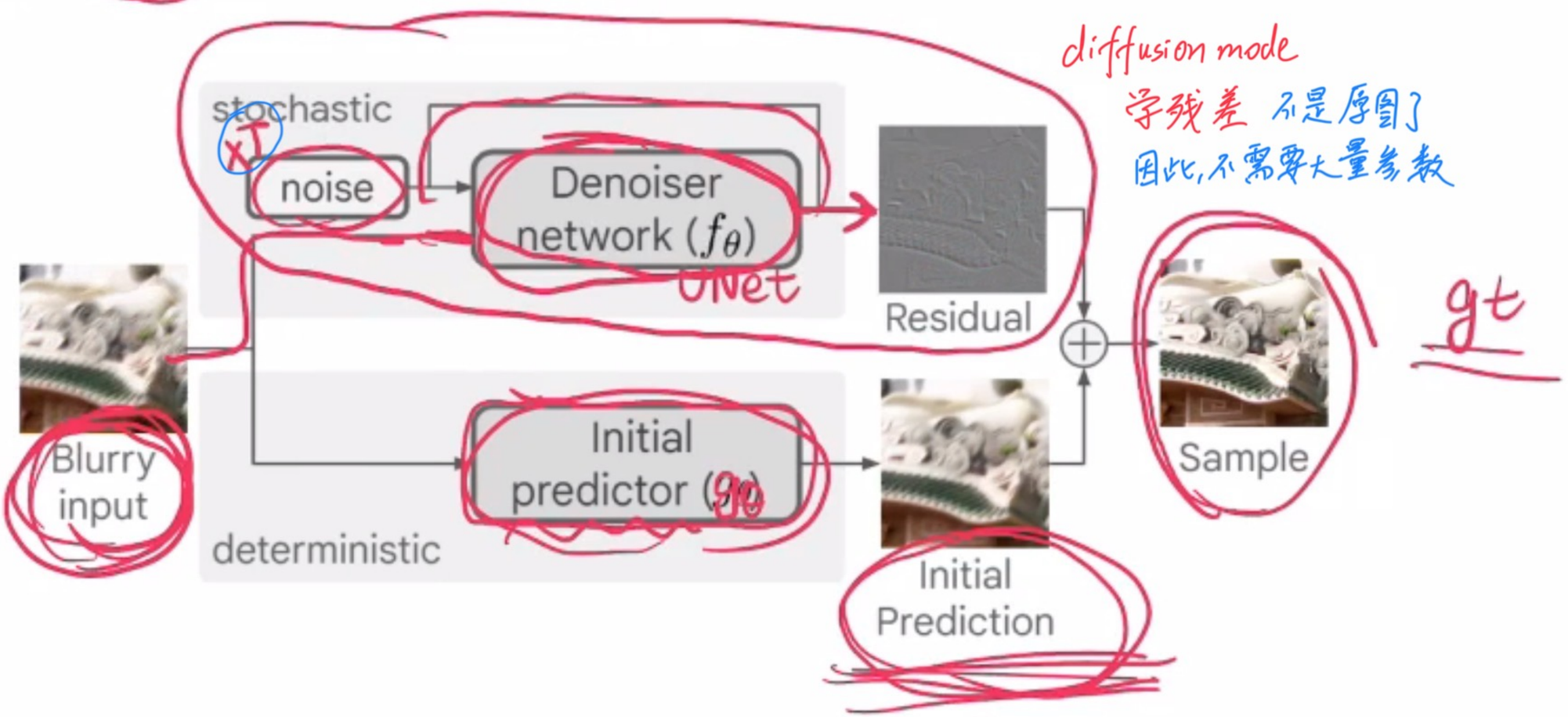
曲线上移动意味着模型的能力没有改变
= 模型的大小 深度

和GT比 逐像素比

由于训练目标的不同，会在这两种指标上呈现不同结果

04 提出方法

- 改进1:
  - predict and refine策略，实际上就是扩散模型的x0不再是原图，而是原图和predictor的残差



04 提出方法

- 改进2：
  - Sample averaging：由于每一次的采样具有随机性，所以可以多重建几次，然后取平均，这是一种比较简单的self-ensemble的方法

- 改进3：
  - Traversing the Perception-Distortion curve：采样的步数越多，则主观质量越好，反之则客观质量越好

- 改进4：
  - Resolution-agnostic Architecture：训练的时候用小patch，测试的时候用整张图 1024×1024

128×128

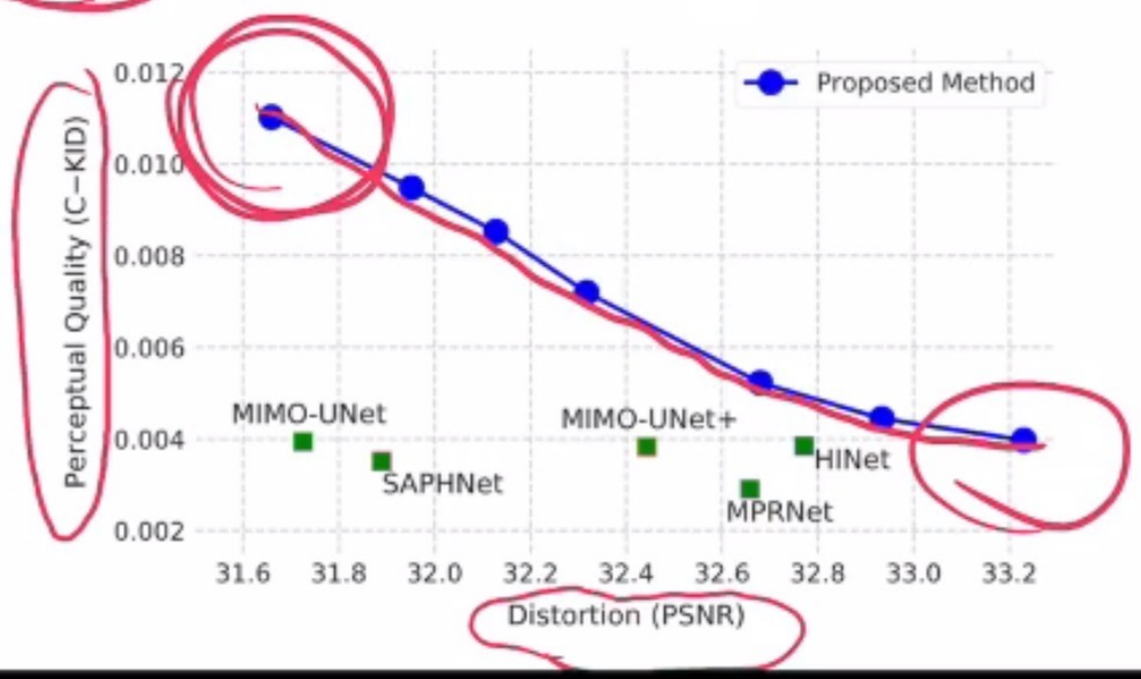
F. Evaluation Details

通过调步数和噪声强度

For all our experiments (on all datasets: GoPro, HIDE, DIV2K), we performed a grid search over the following hyperparameter combinations during inference:

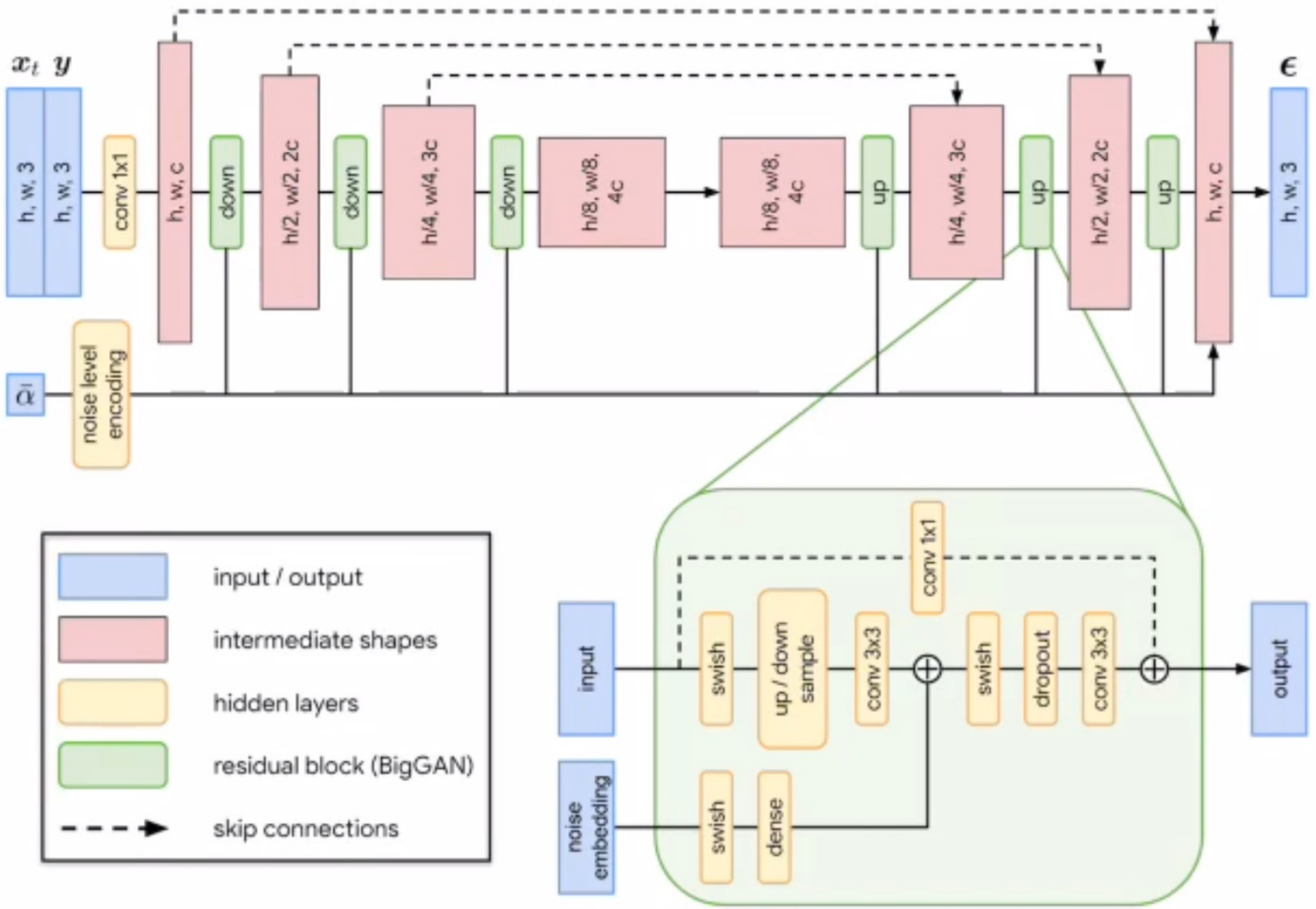
1. Inference steps (T): 10, 20, 30, 50, 100, 200, 300, 500.
2. Noise schedule ($\alpha_{1:T}$): We fixed the initial forward process variance $(1 - \alpha_0)$ to 1×10^{-6}. For the final variance $(1 - \alpha_T)$, we sweep over $\{0.01, 0.02, 0.05, 0.1, 0.2, 0.5\}$. The intermediate values are linearly interpolated.

T=2000
10
20
⋮
500



04 提出方法

- 模型结构要点：
- 作者对网络结构的交代还不算太详细
- initial predictor和denoiser的结构是一样，只是base channel不一样，前者是64，后者是32
- 参数量上，前者是~26M，后者是~7M



05 实验结果

Table 1. Image deblurring results on the GoPro [47] dataset. Our proposed method sets the new Pareto frontier in terms of Perception-Distortion trade-off. Best values and second-best values for each metric are color-coded. KID values are scaled by a factor of 1000 for readability.

	Perceptual				Distortion	
	LPIPS↓	NIQE↓	FID↓	KID↓	PSNR↑	SSIM↑
Ground Truth	0.0	3.21	0.0	0.0	∞	1.000
HINet [11]	0.088	4.01	17.91	8.15	32.77	0.960
MPRNet [71]	0.089	4.09	20.18	9.10	32.66	0.959
MIMO-UNet+ [14]	0.091	4.03	18.05	8.17	32.45	0.957
SAPHNet [61]	0.101	3.99	19.06	8.48	31.89	0.953
SimpleNet [38]	0.108				31.52	0.950
DeblurGANv2 [33]	0.117	3.68	13.40	4.41	29.08	0.918
Ours	0.059	3.39	4.04	0.98	31.66	0.948
Ours-SA	0.078	4.07	17.46	8.03	33.23	0.963

sampling average

Table 2. Image deblurring results on the HIDE [57] dataset, using models trained on GoPro [47]. Our method significantly outperforms the baseline methods under all perceptual metrics while maintaining competitive PSNR and SSIM. Best values and second-best values for each each metric are color-coded.

	Perceptual				Distortion	
	LPIPS↓	NIQE↓	FID↓	KID↓	PSNR↑	SSIM↑
Ground Truth	0.0	2.72	0.0	0.0	∞	1.000
HINet [11]	0.120	3.20	15.17	7.33	30.33	0.932
MIMO-UNet+ [14]	0.124	3.24	16.01	7.91	29.99	0.930
MPRNet [71]	0.114	3.46	16.58	8.35	30.96	0.939
SAPHNet [61]	0.128	3.21	16.77	8.39	29.99	0.930
DeblurGAN-v2 [33]	0.159	2.96	15.51	6.97	27.51	0.885
Ours	0.089	2.69	5.43	1.61	29.77	0.922
Ours-SA	0.092	2.93	6.37	2.40	30.07	0.928

05 实验结果

Table 4. Ablation study on the effects of various hyperparameters for our U-Net architecture, evaluated on the GoPro dataset.

	Hyperparameters			Metrics			
	ch.	batch	EMA	LPIPS	PSNR	MParam.	BFLOPs
More Channels	16	32	No	0.137	29.93	1.63	301
	32	32	No	0.113	31.05	6.52	1200
	64	32	No	0.103	31.63	26.07	4790
+Larger Batch	64	64	No	0.099	31.85	26.07	4790
	64	128	No	0.087	32.56	26.07	4790
	64	256	No	0.086	32.61	26.07	4790
+Use EMA	64	256	Yes	0.0809	33.07	26.07	4790

滑动平均

06 总结与收获

- 本文虽然相比DDPM改动不多，但是有一些思路可以借鉴：
  - predict-and-refine：由于每一步去噪做的非常简单，所以denoiser本身可能其实可以非常**轻量**
  - 如何通过变化采样步数来变换perception和distortion的权衡的，这一点很有启发性