International Conference On Machine Learning

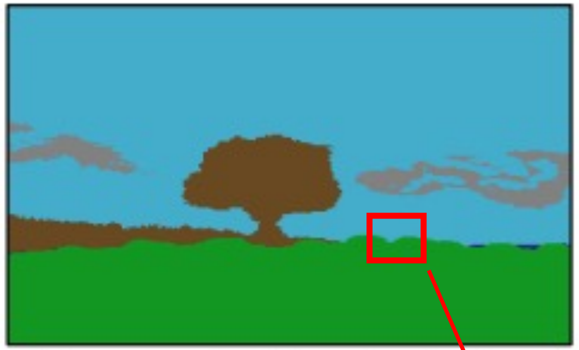
Stochastic Conditional Diffusion Models for Robust Semantic Image Synthesis

Juyeon Ko*, Inho Kong*, Dogyun Park, Hyunwoo J. Kim

Department of Computer Science and Engineering, Korea University

Machine Learning & Vision Lab

Semantic Image Synthesis (SIS)

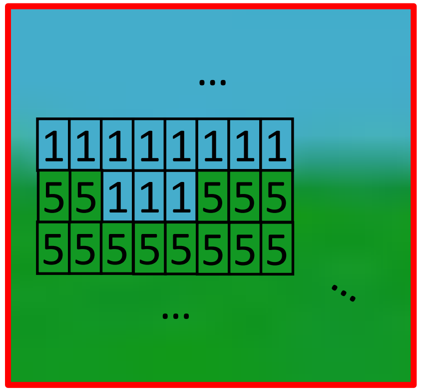


generate

Semantic map $\mathbf{y}$

(Label)

Image $\mathbf{x}$



1: sky
2: tree
...
5: grass
...

approximate

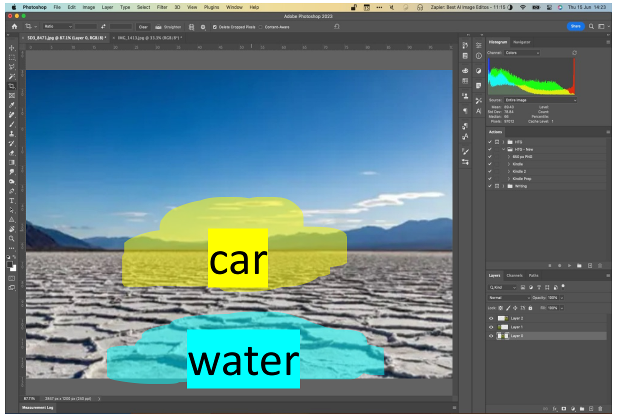
$$q(X|Y)$$

$$p_\theta(X|Y)$$

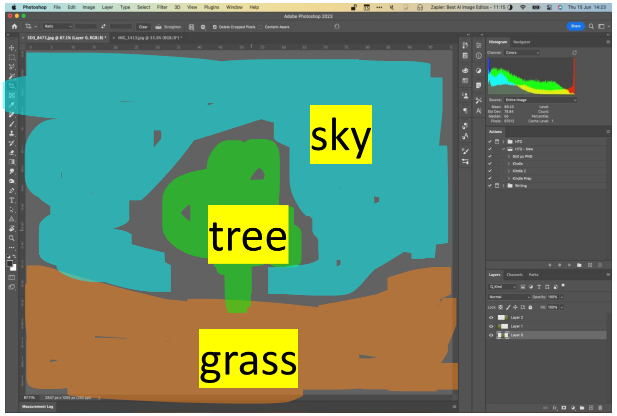
Semantic Image Synthesis with Spatially-Adaptive Normalization, Park et al., CVPR 2019 (Oral)

Motivation

Photo editing



Content creation



Model

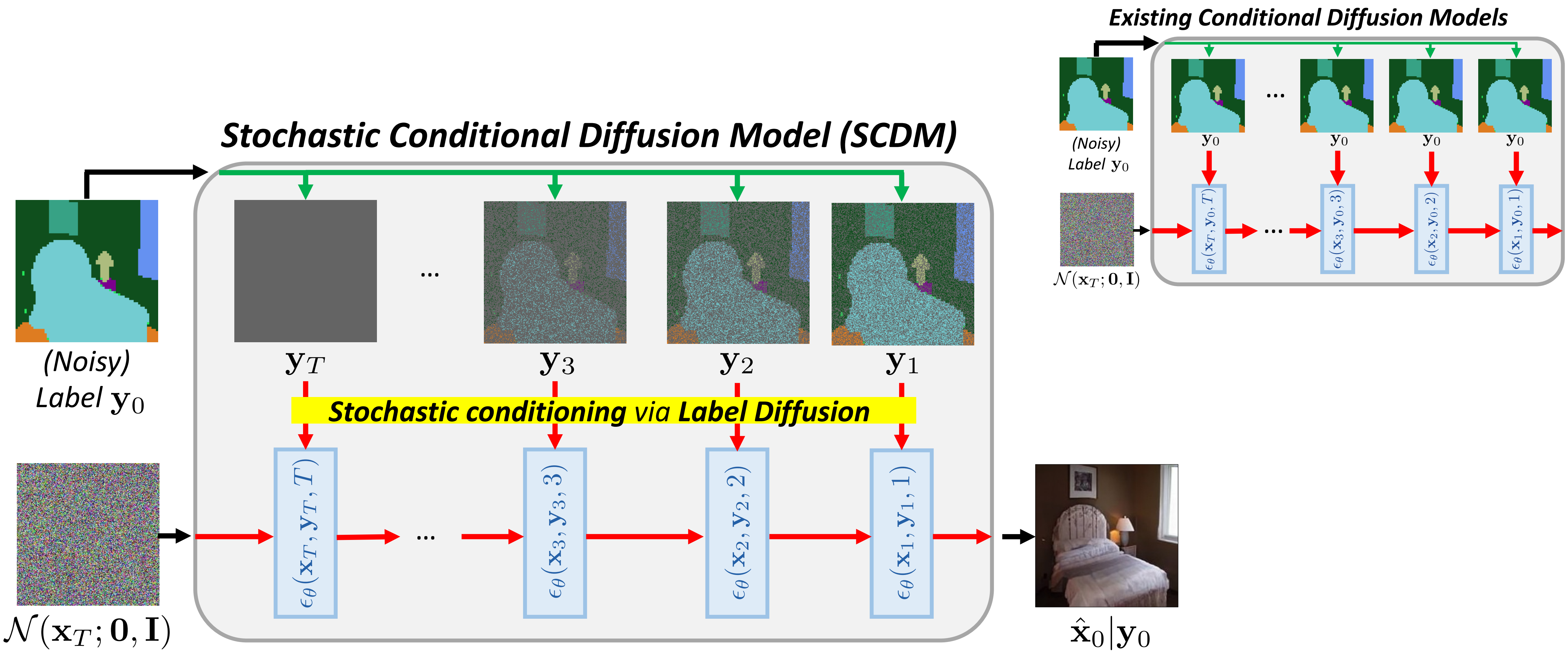
Train
Clean labels from the dataset
$\mathbf{y}$

gap

Inference
Noisy labels from users
$\tilde{\mathbf{y}}$

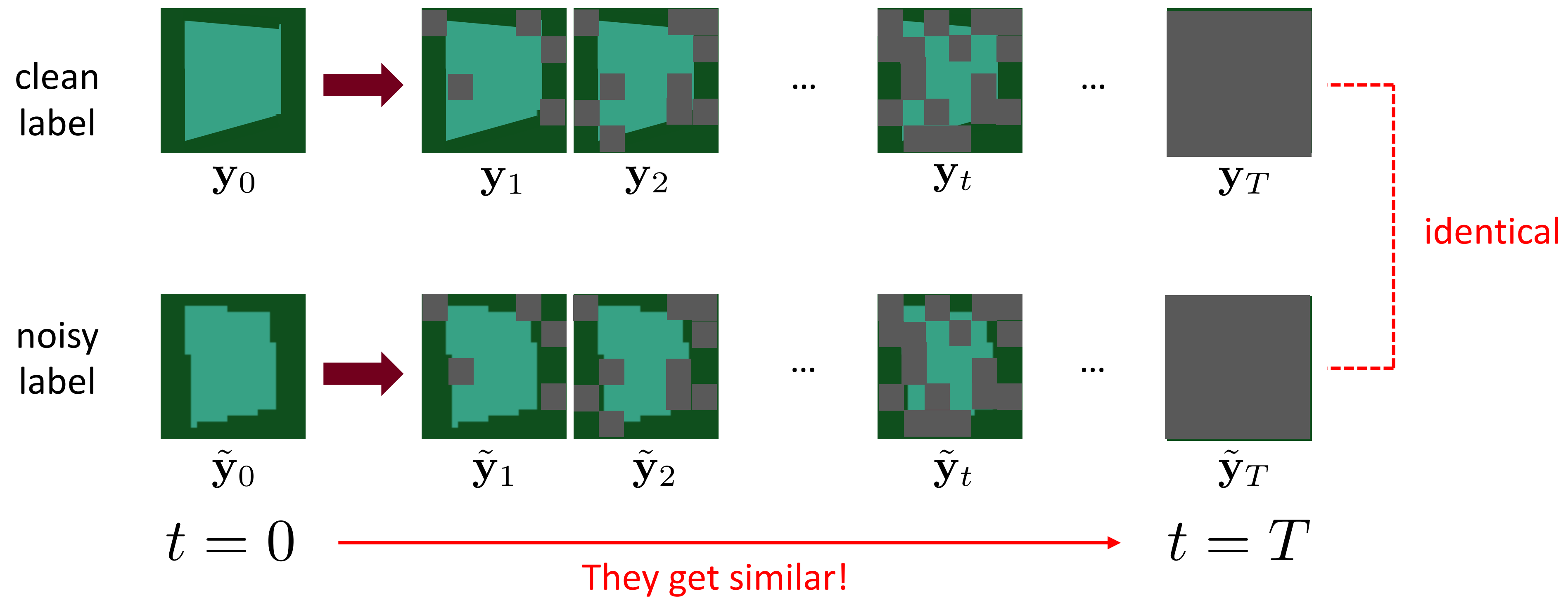
Image credit: https://tech.hindustantimes.com/tech/news/this-photoshop-ai-feature-will-change-the-way-you-edit-photos-know-what-is-generative-fill-71686747616850.html

Stochastic Conditional Diffusion Models (SCDM)

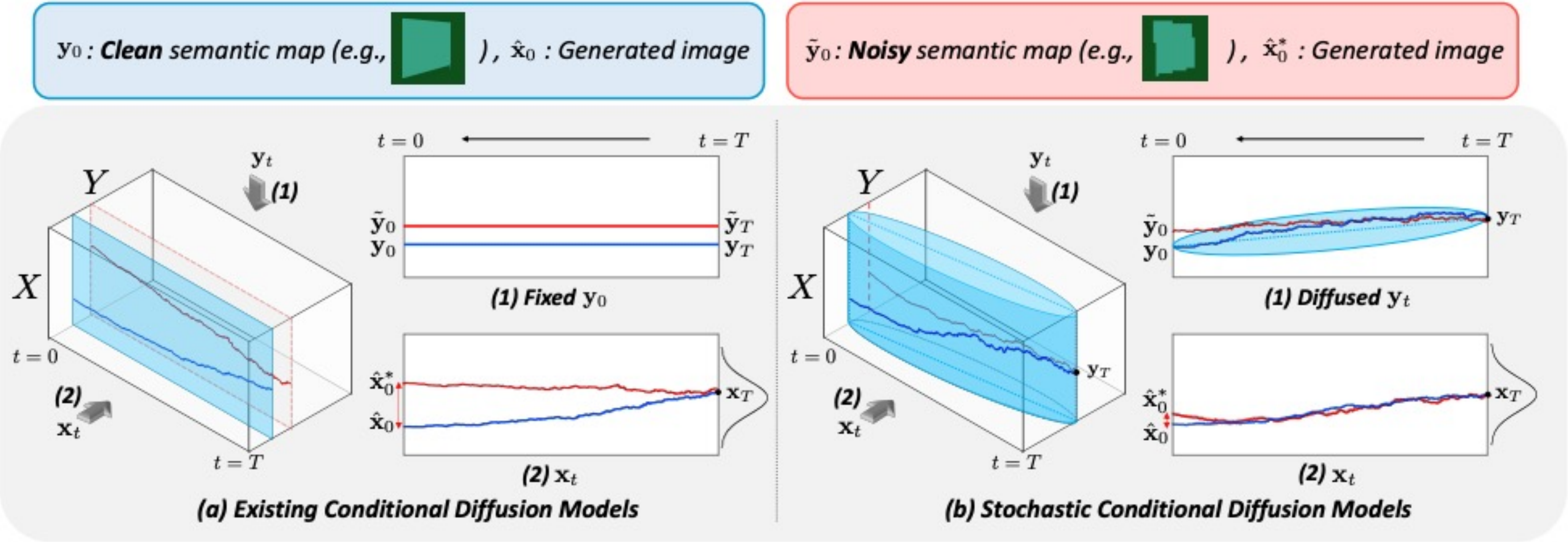


Label Diffusion

- Discrete diffusion for the labels, $q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) := \mathrm{Cat}(\boldsymbol{z}_t; \boldsymbol{p} = \boldsymbol{Q}_t \boldsymbol{z}_{t-1})$
- Noise = [MASK] (absorbing state)



Label Diffusion



- The trajectories $\mathbf{y}_{1:T}$ and $\tilde{\mathbf{y}}_{1:T}$ provided to the model during generation process are similar.
- The generated image is close to the clean image.

Forward process and Generation process

- Forward process

$$q(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_0, \mathbf{y}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t, \mathbf{y}_t | \mathbf{x}_{t-1}, \mathbf{y}_{t-1}),$$

Label Diffusion

$$q(\mathbf{x}_t, \mathbf{y}_t | \mathbf{x}_{t-1}, \mathbf{y}_{t-1}) := \underbrace{q(\mathbf{x}_t | \mathbf{x}_{t-1})}_{\textit{continuous}} \underbrace{\boxed{q(\mathbf{y}_t | \mathbf{y}_{t-1})}}_{\textit{discrete}}.$$

- Generation process

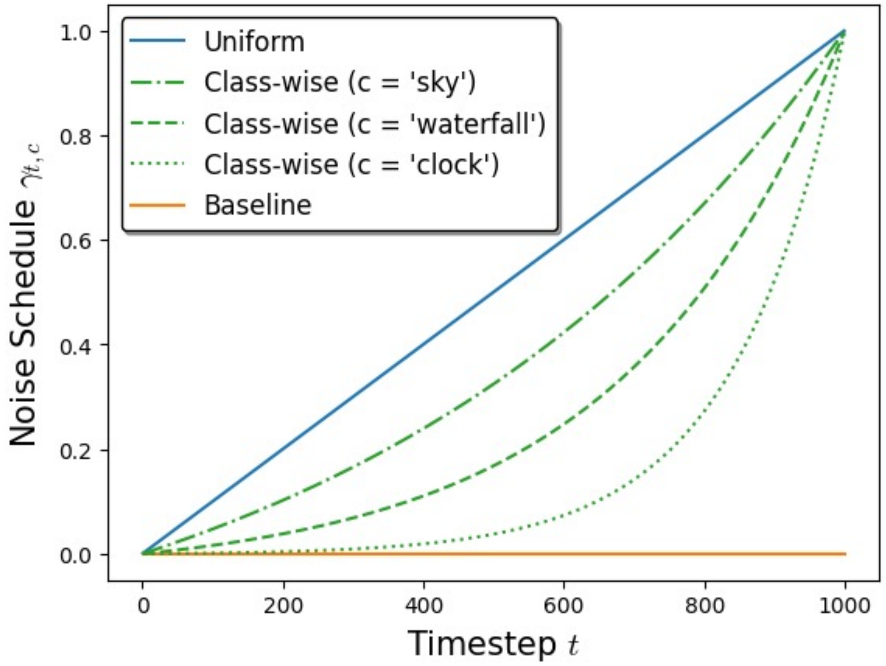
- *discrete*
- *forward*

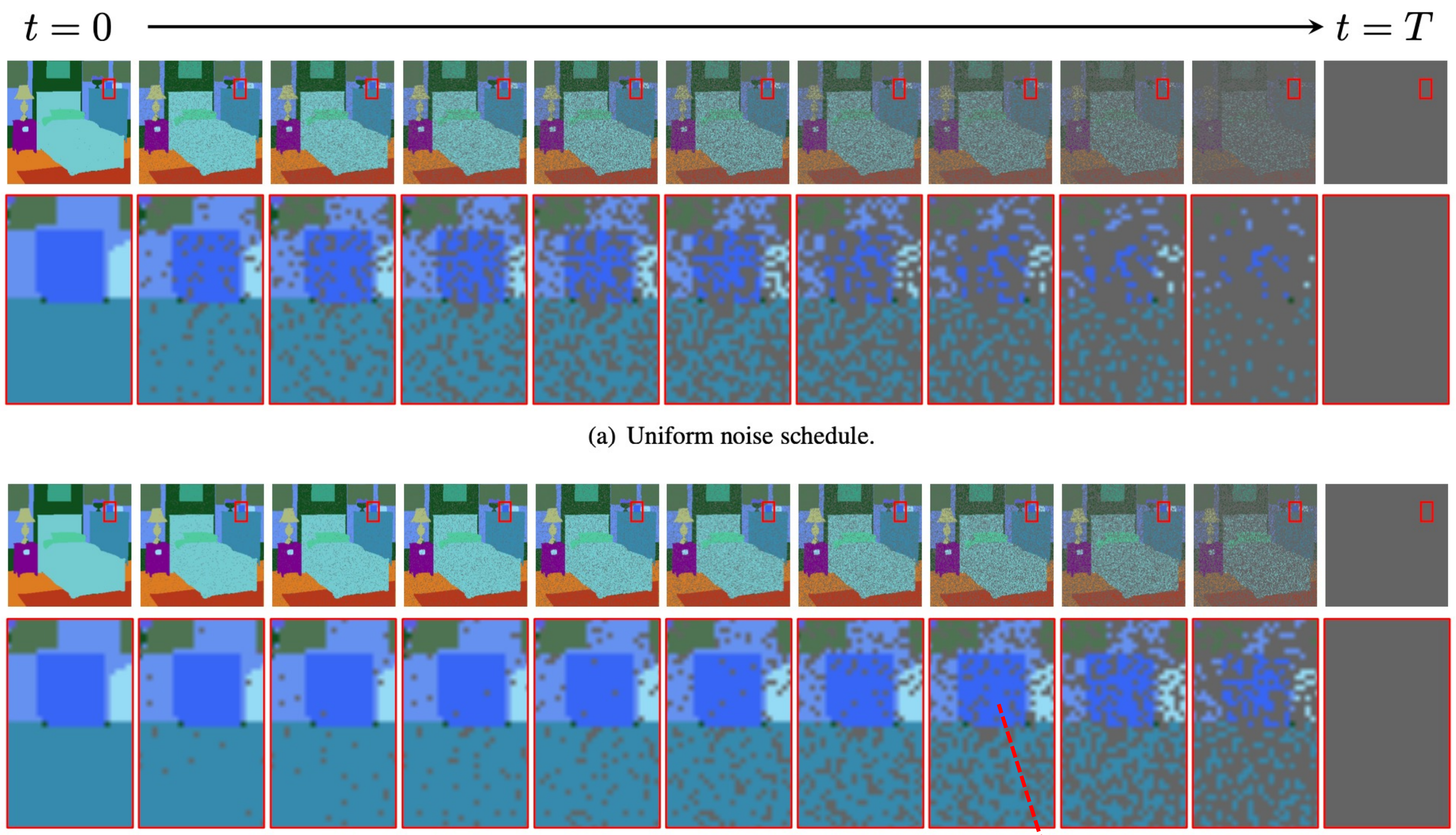
- *continuous*
- *reverse*

$$p_\theta(\mathbf{x}_{0:T}, \mathbf{y}_{1:T} | \mathbf{y}_0) := p(\mathbf{x}_T) \underbrace{\boxed{q(\mathbf{y}_{1:T} | \mathbf{y}_0)}}_{\textit{Label Diffusion}} \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_t),$$

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, \mathbf{y}_t, t), \Sigma_\theta(\mathbf{x}_t, \mathbf{y}_t, t)),$$

Class-wise Noise Schedule





(a) Uniform noise schedule.

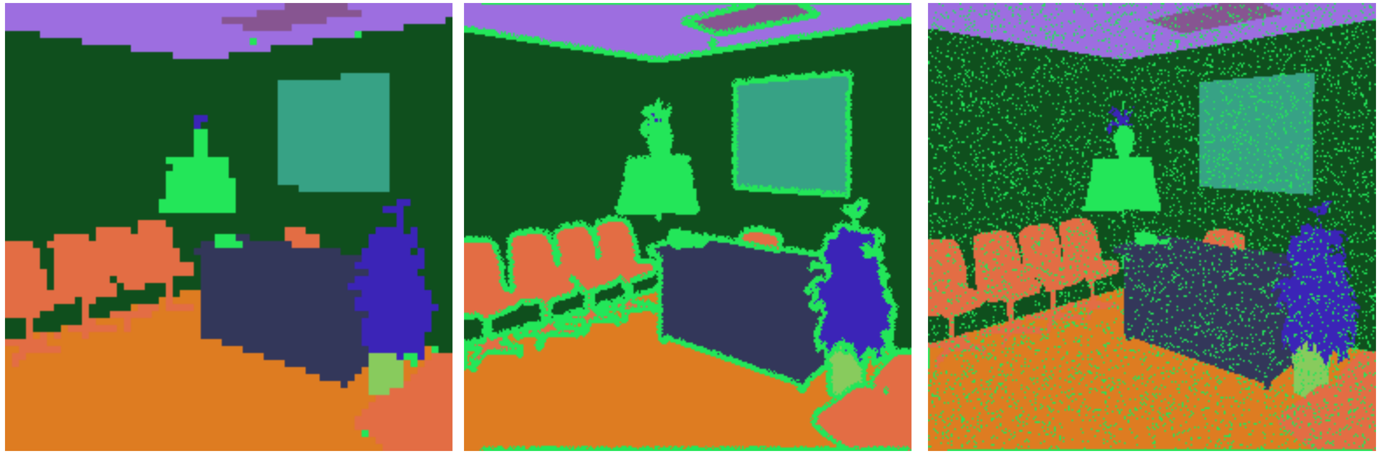
(b) Class-wise noise schedule.

Slowly diffuse small/rare classes

Noisy SIS Benchmark

- We introduce a new benchmark to assess generation performance under noisy conditions, simulating human errors that can occur during real-world applications.
- Three setups:

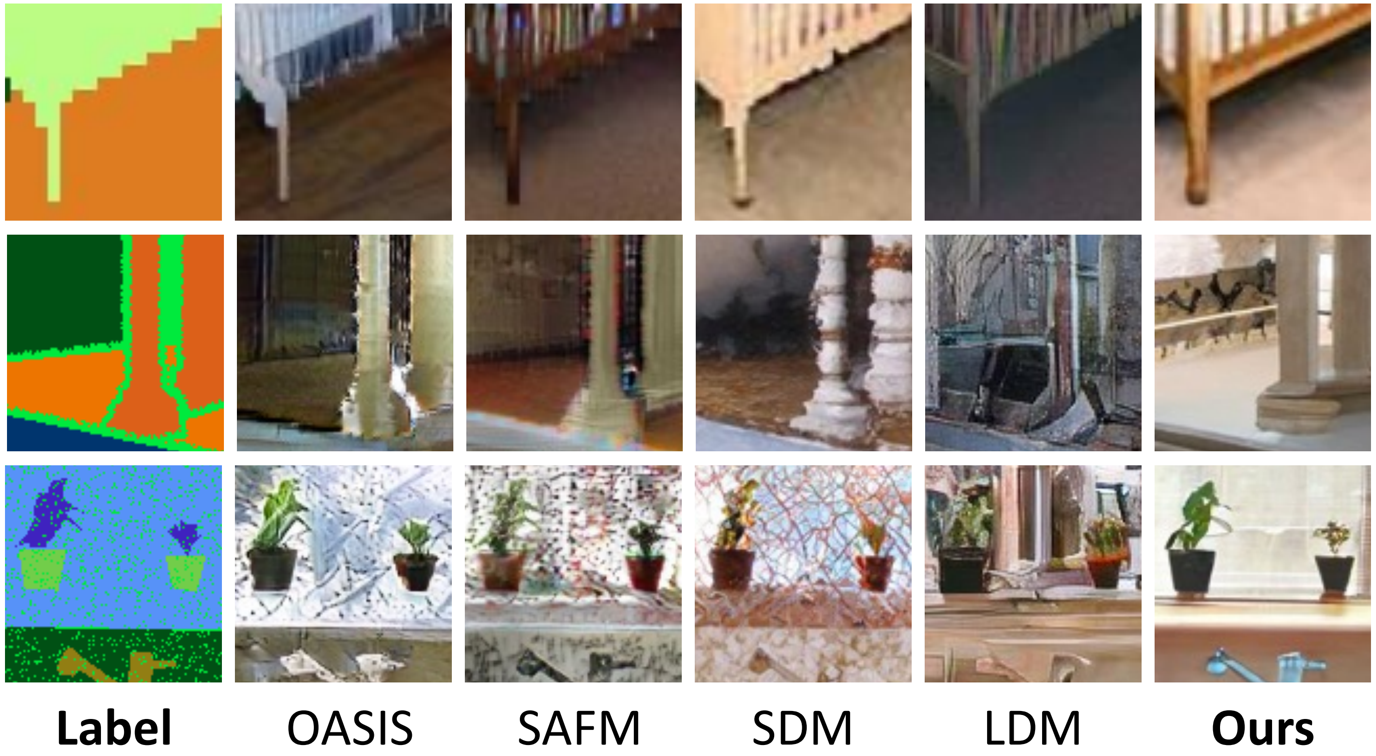
[DS] **[Edge]** **[Random]**



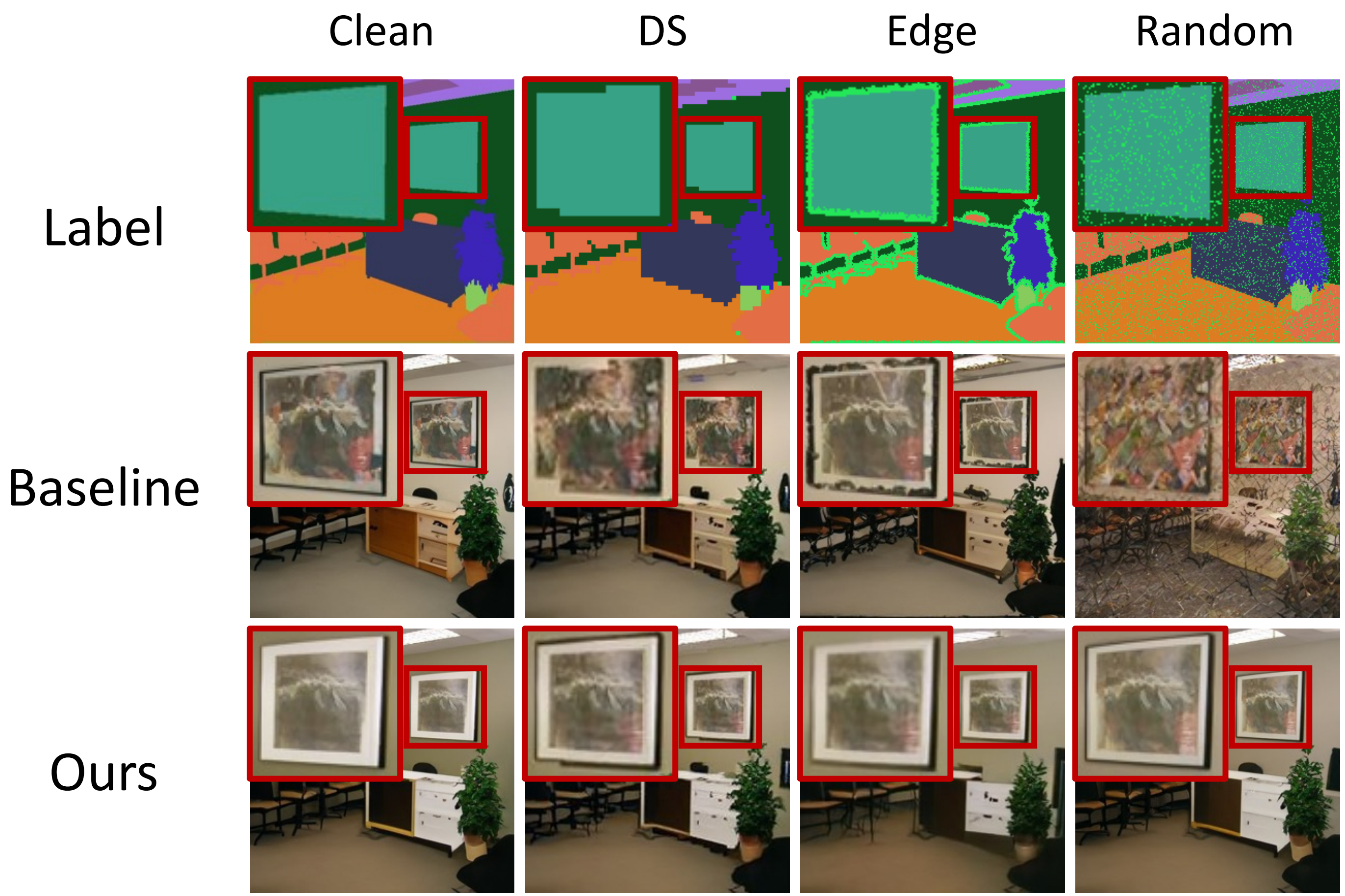
1. **[DS]** downsampled semantic maps
2. **[Edge]** masking the edges of instances
3. **[Random]** randomly adding unlabeled class to the semantic maps (10%)

Experiments - Noisy SIS

Category	Methods	DS F(↓)	DS M(↑)	Edge F(↓)	Edge M(↑)	Random F(↓)	Random M(↑)
GAN	SPADE	36.6	41.0	40.2	39.7	92.7	18.5
	CC-FPSE	42.0	40.4	38.7	41.7	141.3	16.2
	DAGAN	37.6	41.3	42.2	39.0	109.1	16.4
	GroupDNet	45.9	28.9	49.9	28.0	76.0	21.2
	OASIS	34.5	48.1	37.6	<u>47.4</u>	54.2	42.1
	CLADE	37.4	41.8	41.9	40.1	71.8	29.9
	INADE	36.1	38.3	40.3	35.6	61.3	30.2
	SCGAN	38.1	44.0	63.9	43.3	40.9	38.6
	SAFM*	36.7	<u>50.1</u>	40.0	44.9	80.6	34.1
DM	SDM	35.5	43.8	39.4	39.4	141.9	11.8
	LDM	38.9	28.1	39.5	26.0	36.3	27.1
	Ours	**<u>32.4</u>**	**44.7**	**<u>31.2</u>**	**40.1**	**<u>28.1</u>**	**<u>45.2</u>**



Analysis – Label Diffusion and Robustness



Dataset	Method	LPIPS(↓)	SSIM(↑)	PSNR(↑)	FID(↓)
DS	Baseline	0.221	0.823	30.4	24.2
	Ours	**0.180**	**0.865**	**31.3**	**19.0**
Edge	Baseline	0.248	0.771	29.7	32.7
	Ours	**0.223**	**0.825**	**30.2**	**20.0**
Random	Baseline	0.560	0.427	28.1	145.6
	Ours	**0.076**	**0.944**	**32.9**	**10.0**

Conclusion

- We first address a significant challenge for SIS models in real-world applications: dealing with noisy input from users.
- We propose a novel conditional diffusion model, SCDM, specifically designed to enhance robustness on noisy labels.
- We present a new benchmark setting to assess performance under noisy conditions.
- SCDM improves generation quality through its Label Diffusion and class-wise noise schedule.

Thank You

Paper

GitHub