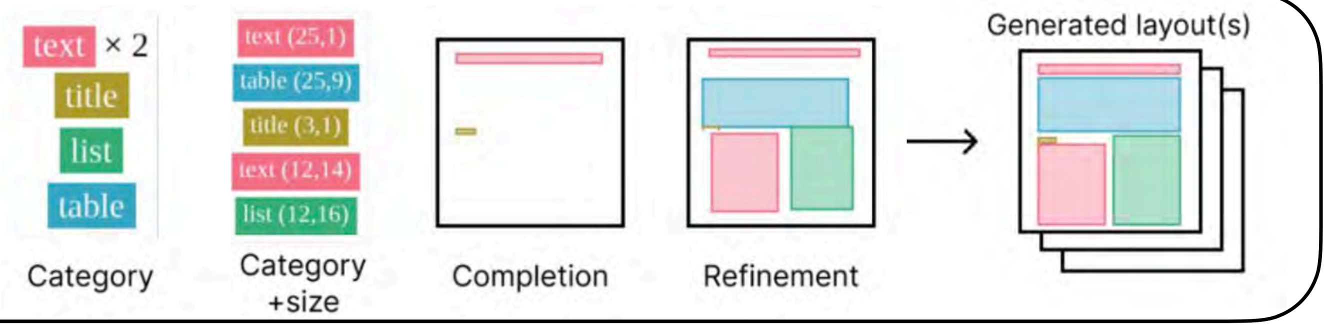
LACE Towards Aligned **LAYOUT** Generation via Diffusion Model with **AESTHETIC CONSTRAINTS**

OpenSource

Jian Chen · Ruiyi Zhang · Yufan Zhou · Rajiv Jain · Zhiqiang Xu · Ryan Rossi · Changyou Chen

LACE is a diffusion model that generate continuous layout coordinates. It can handle both *conditional* and *unconditional* generation tasks.



Continuous diffusion (LACE)

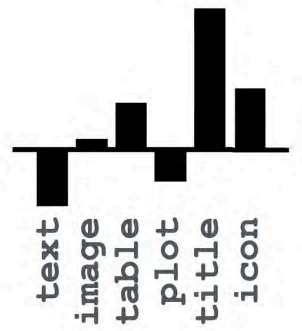
generate sequence of differentiable bounding box coordinate.

Differentiable

Box: $[x, y, w, h] \in [0, 1]$

Numerical vector of box coordinate (4 dim)

Label:



probability logits of category

Noise: Gaussian

$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

- Continuous coordinates adapt to all canvas size.
- No modality-wise noise needed.

Discrete diffusion

generate sequence of probability mass vector for discretized coordinate bins and categorical labels.

Not differentiable

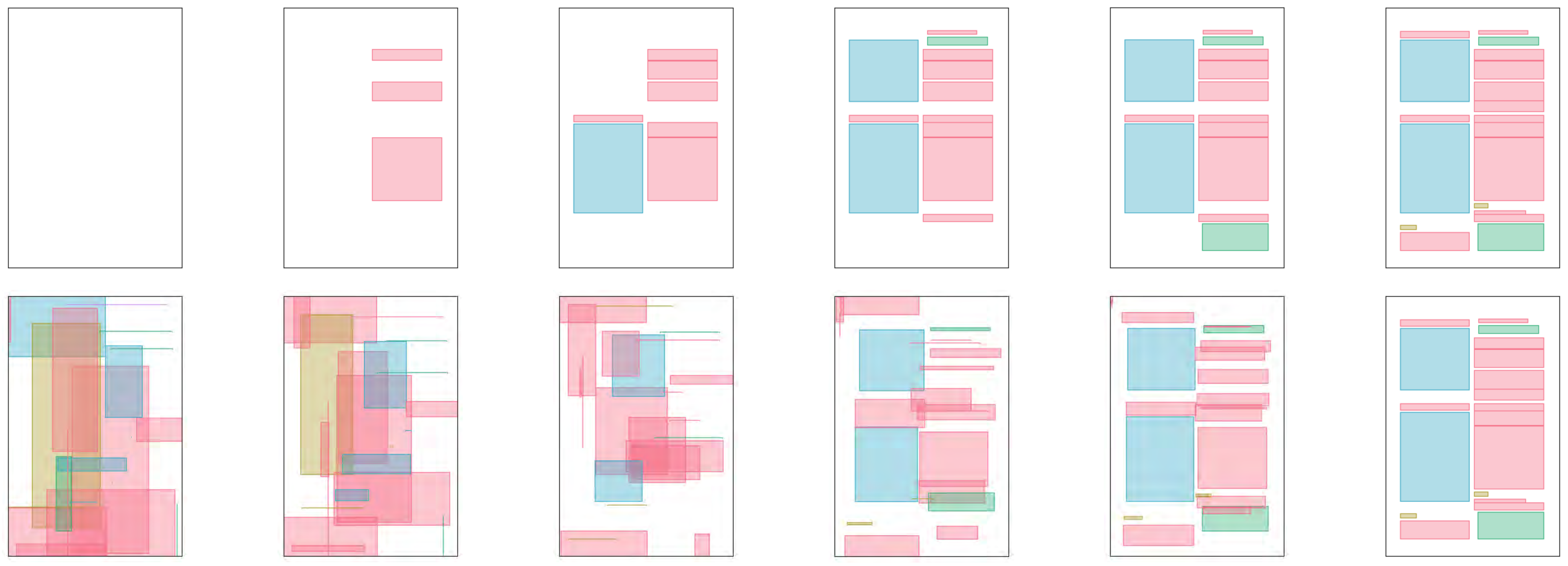
4 probability vectors of discrete bins (e.g., 4 x 128 bins)

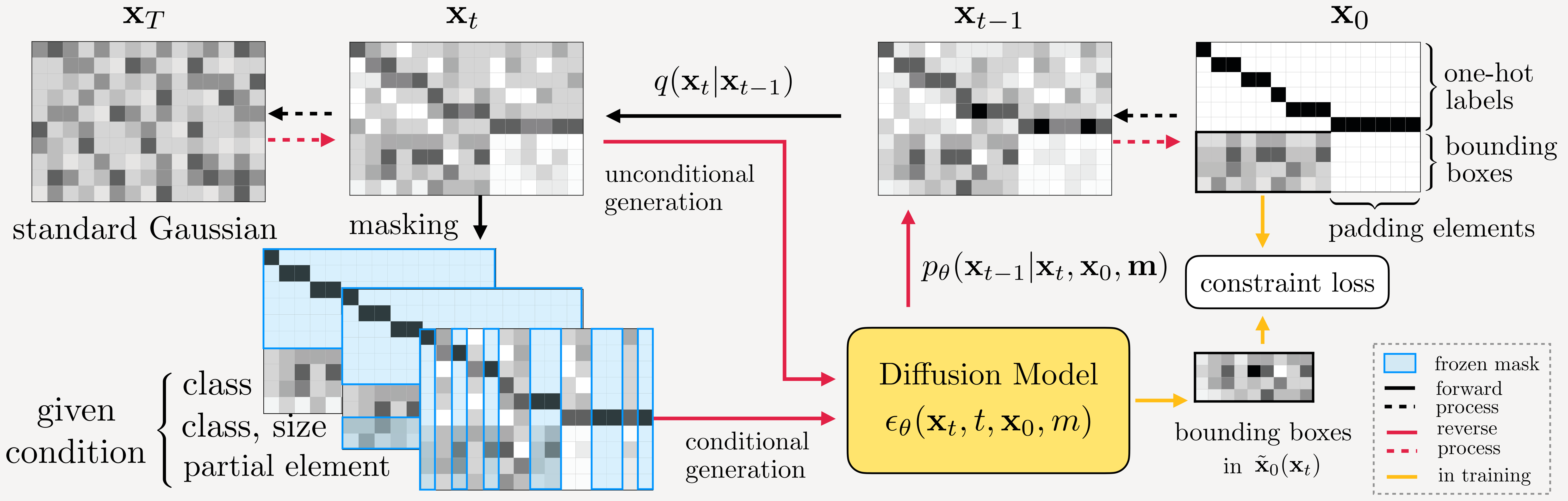
probability of category

Modality-wise discrete noise with absorbing state [mask] with transition probability matrix:

$$\mathbf{Q}_t = \begin{bmatrix} \alpha_t+\beta_t & \beta_t & \cdots & \beta_t & 0 \\ \beta_t & \alpha_t+\beta_t & \cdots & \beta_t & 0 \\ \vdots & \vdots & \ddots & \beta_t & 0 \\ \beta_t & \beta_t & \beta_t & \alpha_t+\beta_t & 0 \\ \gamma_t & \gamma_t & \gamma_t & \gamma_t & 1 \end{bmatrix}$$

This differences result in distinct generation patterns.



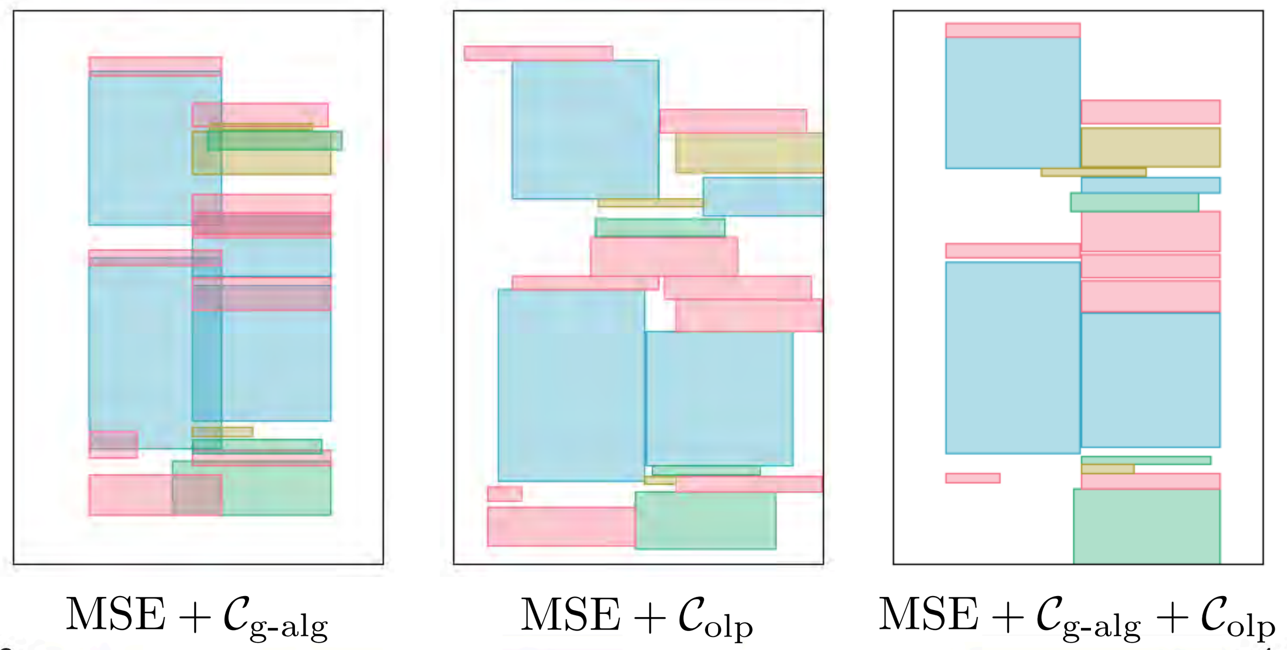


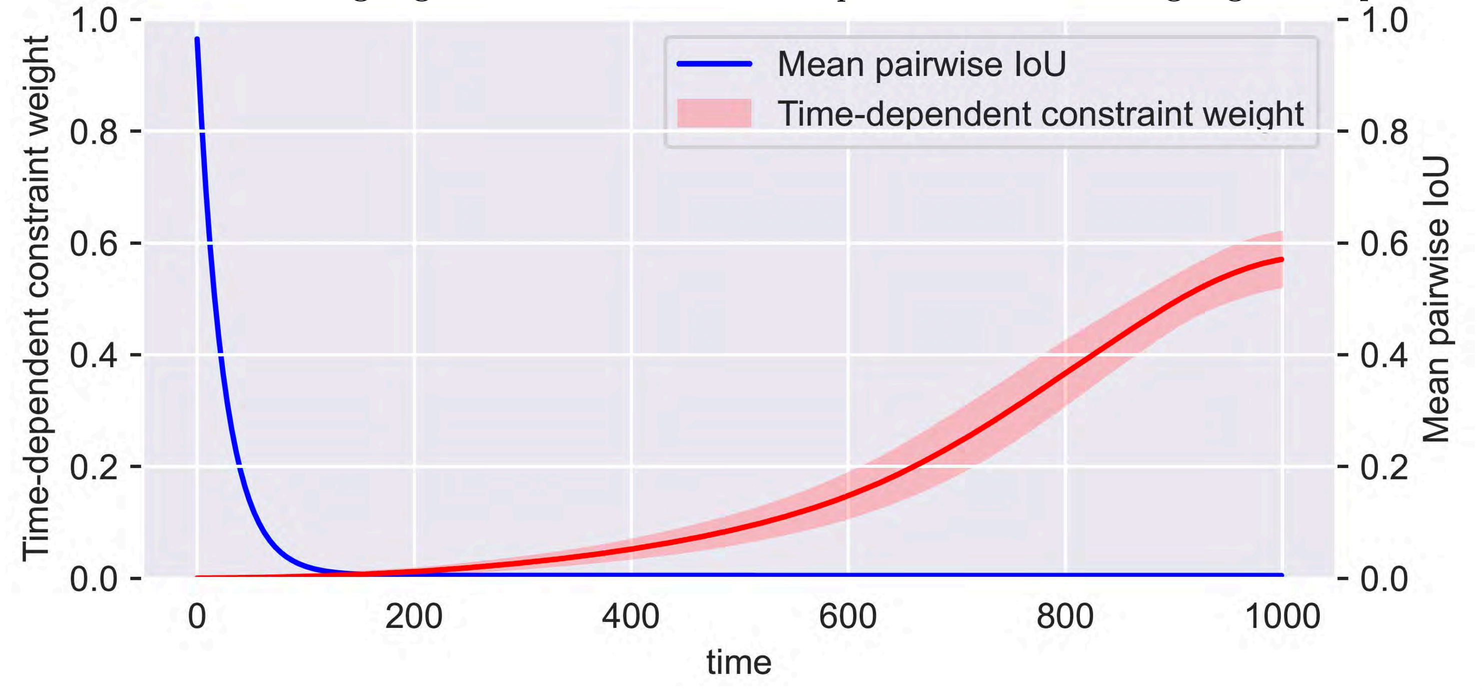
$$\tilde{\mathbf{x}}_0(\mathbf{x}_t) = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\cdot\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}.$$

$$\mathcal{L}_{\text{rec}} = \text{MSE}(\tilde{\mathbf{x}}_0, \mathbf{x}_0) + \omega_t \cdot (\mathcal{C}_{\text{alg}}(\tilde{\mathbf{x}}_0(\mathbf{x}_t), \mathbf{x}_0) + \mathcal{C}_{\text{olp}}(\tilde{\mathbf{x}}_0(\mathbf{x}_t)))$$

- With differentiable box coordinates, LACE can add constraints based on heuristics.
- Alignment & Overlap Constraints Prevent boxes collapse and misalignment.
- Given Partial input using masks to enable conditional generation.

cases converging to a local minimum





Post-processing:

We directly optimize the coordinates during post-processing phase to remove minor defects.

Model	Task / Metric	C→S+P FID↓	Align↓	Overlap↓
Task-specific models				
NDN-none		61.1	0.350	16.5
LayoutGAN++		24.0	0.190	22.80
LayoutGAN++ w/ $\mathcal{C}$		22.3	0.160	14.27
LayoutGAN++ w/ $\mathcal{C}$ & post		26.2	0.160	1.18
Diffusion-based models				
LayoutDM		7.95	0.106	16.43
LayoutDM w/ post		15.2	0.083	6.076
LACE w/o $\mathcal{C}$		6.12	0.054	1.636
LACE (local)		4.88	0.043	1.638
LACE (global)		5.14	0.046	1.791
LACE (local) w/ post		4.63	0.010	1.211
LACE (global) w/ post		4.56	0.009	0.906
Validation data		6.25	0.021	0.117

Applying post-processing to other methods can damage performance