LACE: LAyout Constraint diffusion modEl

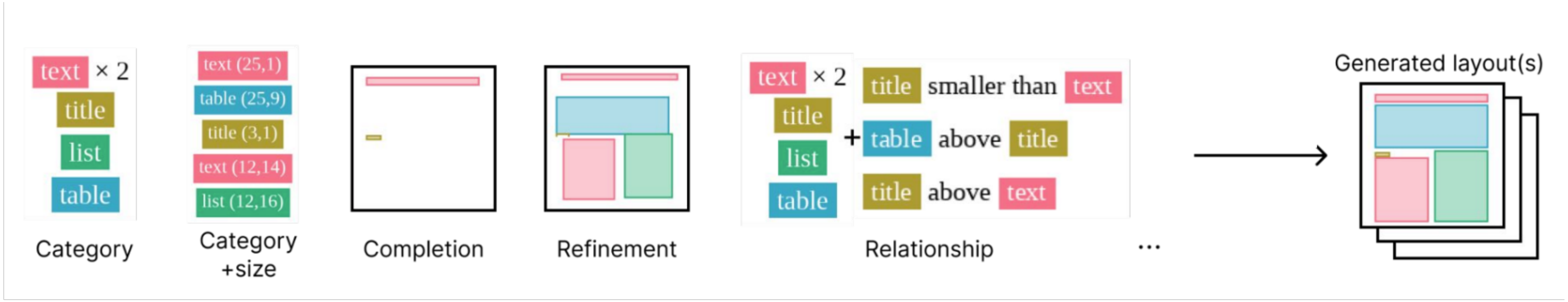
Towards Aligned Layout Generation via Diffusion Model with Aesthetic Constraints

Jian Chen

University at Buffalo Adobe ICLR

Introduction

- LACE is a unified model that can handle unconditional and various conditional generation tasks.
- LACE generates continuous coordinate values, enabling constrained optimization during training.
- LACE surpasses existing baselines across multiple metrics.



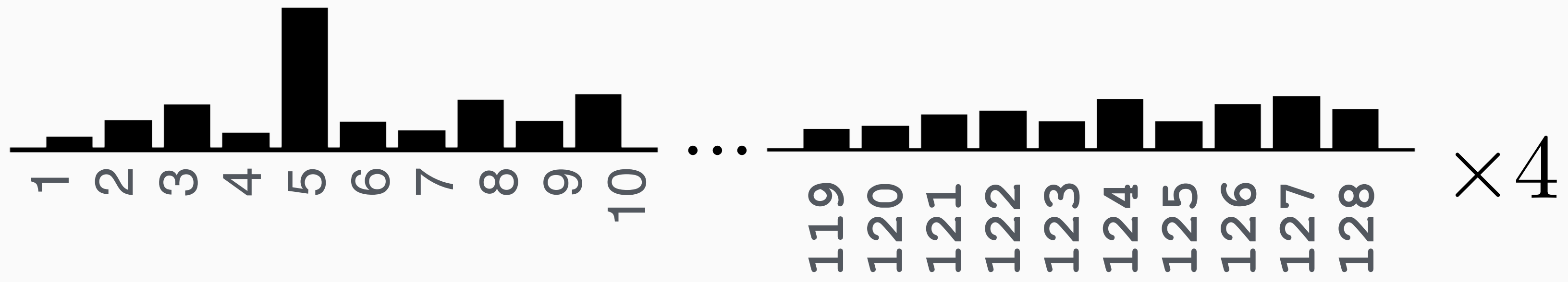
Discrete vs Continuous diffusion for layout generation

Discrete diffusion

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

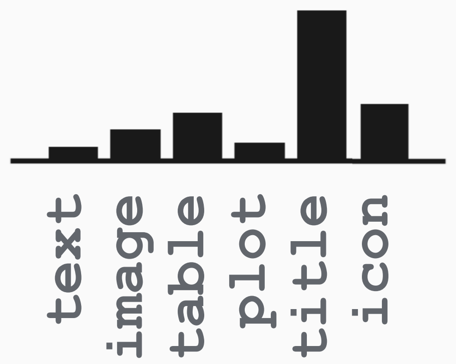
Not differentiable

Box Coordinate:



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

Label:



probability of category

(e.g., 6 labels)

Noise:

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

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

Continuous diffusion

generate sequence of differentiable bounding box coordinate.

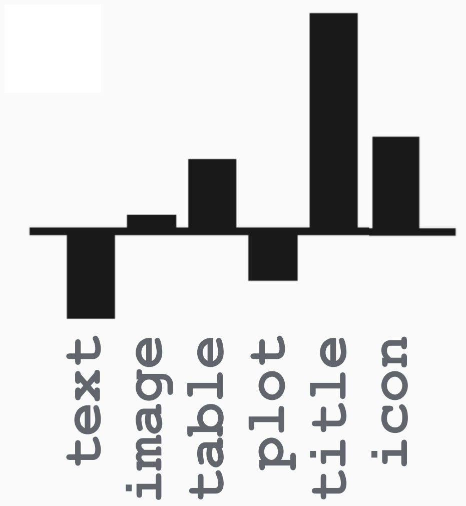
Differentiable

Box Coordinate:

$[x, y, w, h]$

Numerical vector of box coordinate (4 dimension)

Label:



probability logits of category

(e.g., 6 labels)

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 value for box coordinates is aligned with real-world cases.**
- **Multiple tasks can be done with a single model.**

Discrete vs Continuous diffusion for layout generation

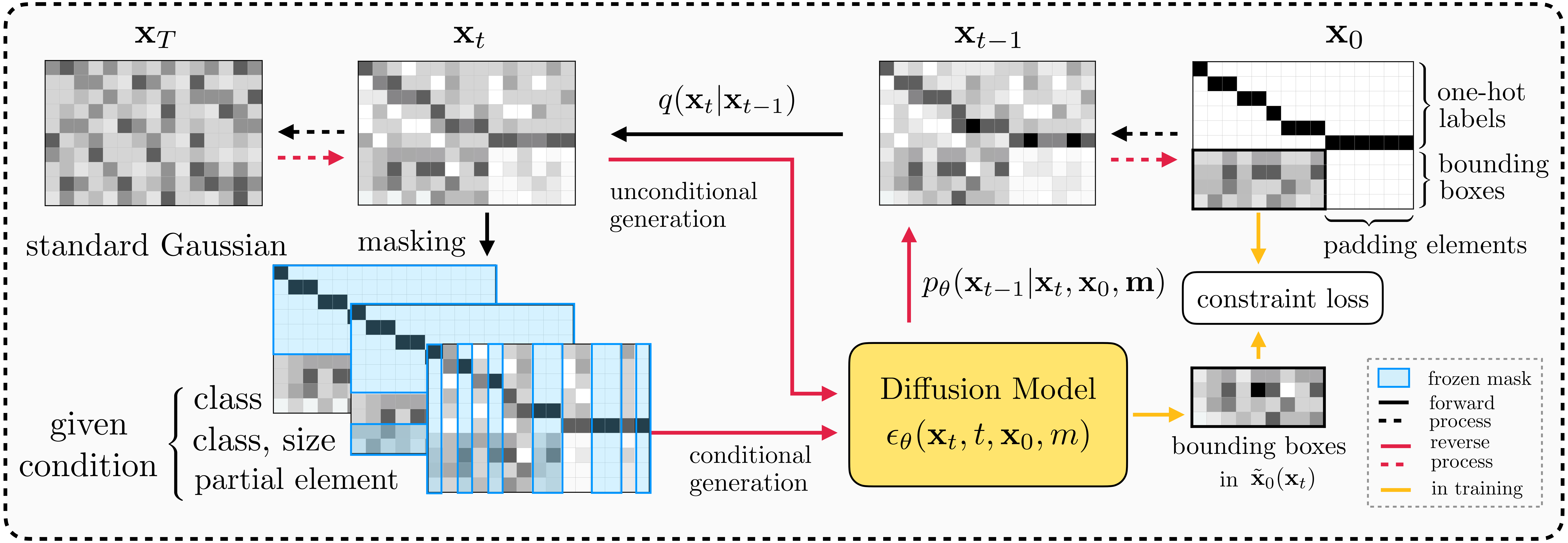
Discrete diffusion

starting from a blank canvas, generates elements incrementally. Emerged elements <u>*tend to*</u> remain static, thus limiting the model's ability to make global adjustments.

Continuous diffusion

starts with a random layout and refines it to an organized one over time, which is more flexible in modeling.

Model overview



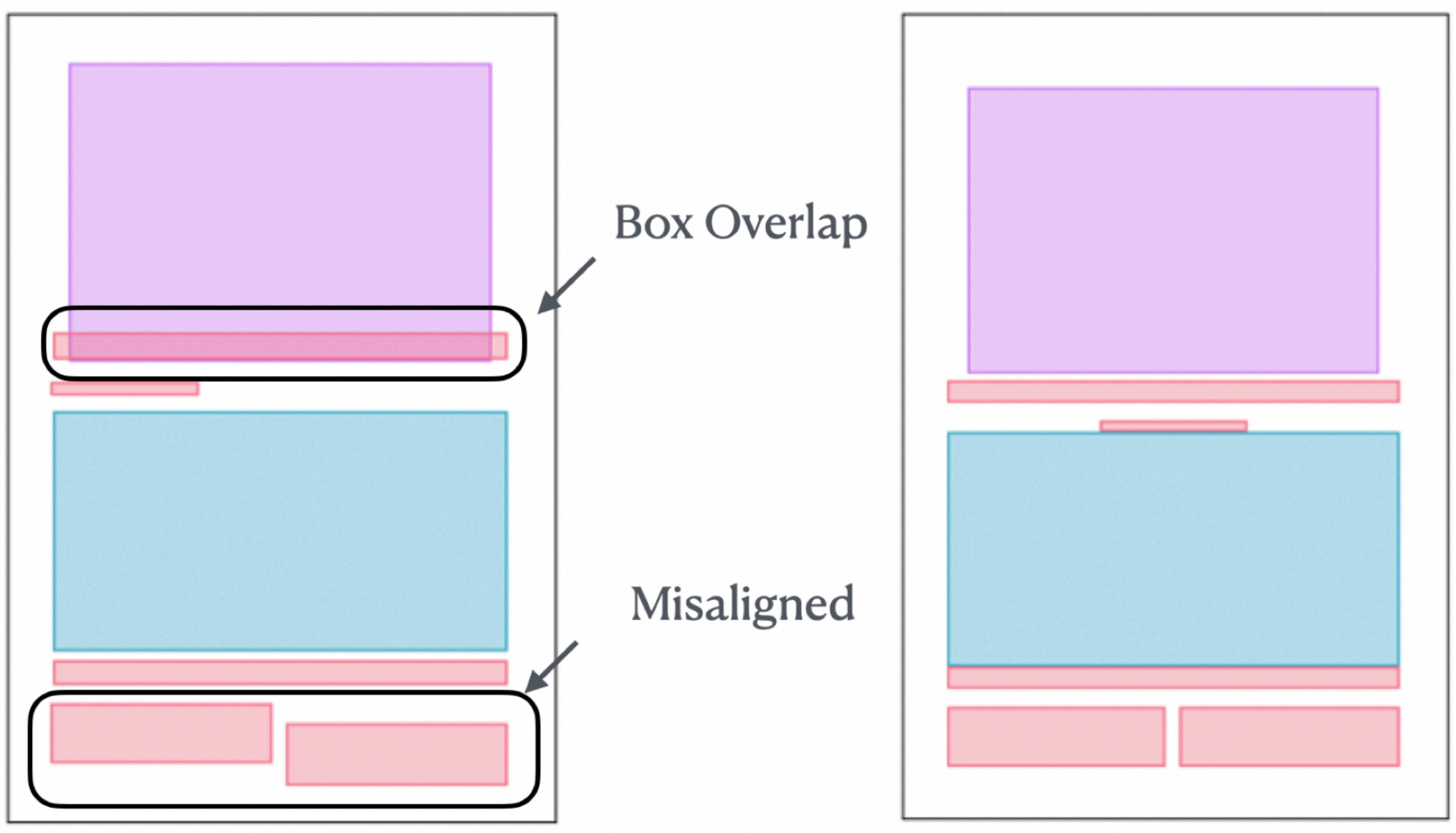
- With differentiable box coordinates, LACE can add constraints based on heuristics.
- Alignment Constraints: encourage elements to be better aligned.
- Overlap Constraints: prevent boxes collapse.

Training with constraints

Since the coordinate is differentiable, we can add constraints to optimize the aesthetic score.

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

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



Raw LayoutDM output | Ground Truth

We modified the alignment and overlap metric to be our constraint loss functions.

local alignment

$$\mathcal{C}_{\text{l-alg}}(\mathbf{x}) = \sum_{i=1}^{l} \min \begin{pmatrix} g(\Delta b_i^{\text{L}}), & g(\Delta b_i^{\text{XC}}), & g(\Delta b_i^{\text{R}}) \\ g(\Delta b_i^{\text{T}}), & g(\Delta b_i^{\text{YC}}), & g(\Delta b_i^{\text{B}}) \end{pmatrix}$$

$$g(x) = -log(1-x)$$

$$\Delta b_i^* = \min_{\forall j \neq i} |b_i^* - b_j^*|$$

global alignment

$$\mathcal{C}_{\text{g-alg}}(\tilde{\mathbf{x}}, \mathbf{x}) = -\log\left(1 - \frac{1}{6} \sum_{* \in \mathcal{A}} \frac{\left\| \mathbf{A}^*(\tilde{\mathbf{x}}) \circ \mathbf{1}_{\mathbf{A}^*(\mathbf{x})=0} \right\|_1}{\left\| \mathbf{1}_{\mathbf{A}^*(\mathbf{x})=0} \right\|_1}\right)$$

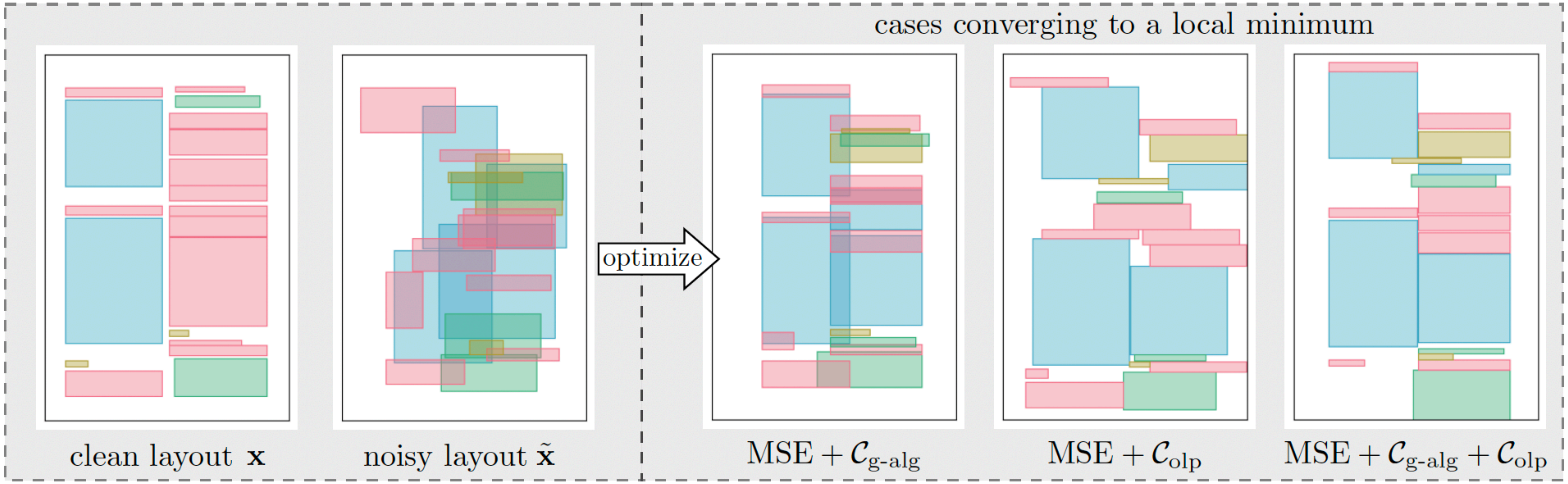
$$\mathbf{A}^*(\mathbf{x})_{i,j} = |b_i^* - b_j^*|,$$

overlap loss

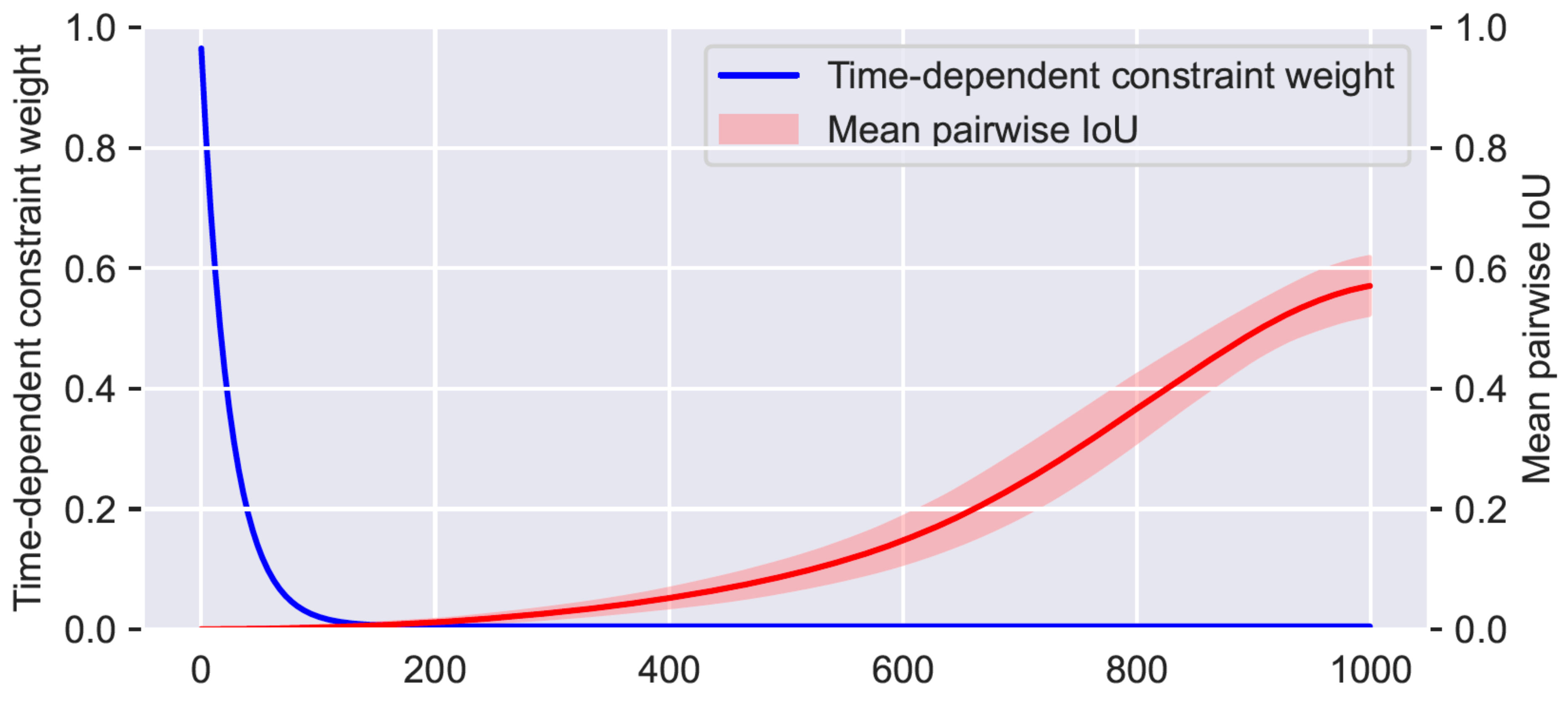
$$\mathcal{C}_{\text{olp}}(\tilde{\mathbf{x}}) = \text{mean}(\mathbf{O}(\tilde{\mathbf{x}}) + \mathbf{D}(\tilde{\mathbf{x}}) \circ \mathbf{1}_{\mathbf{O}(\tilde{\mathbf{x}}) \neq 0})$$

University at Buffalo Adobe ICLR

Training with constraints

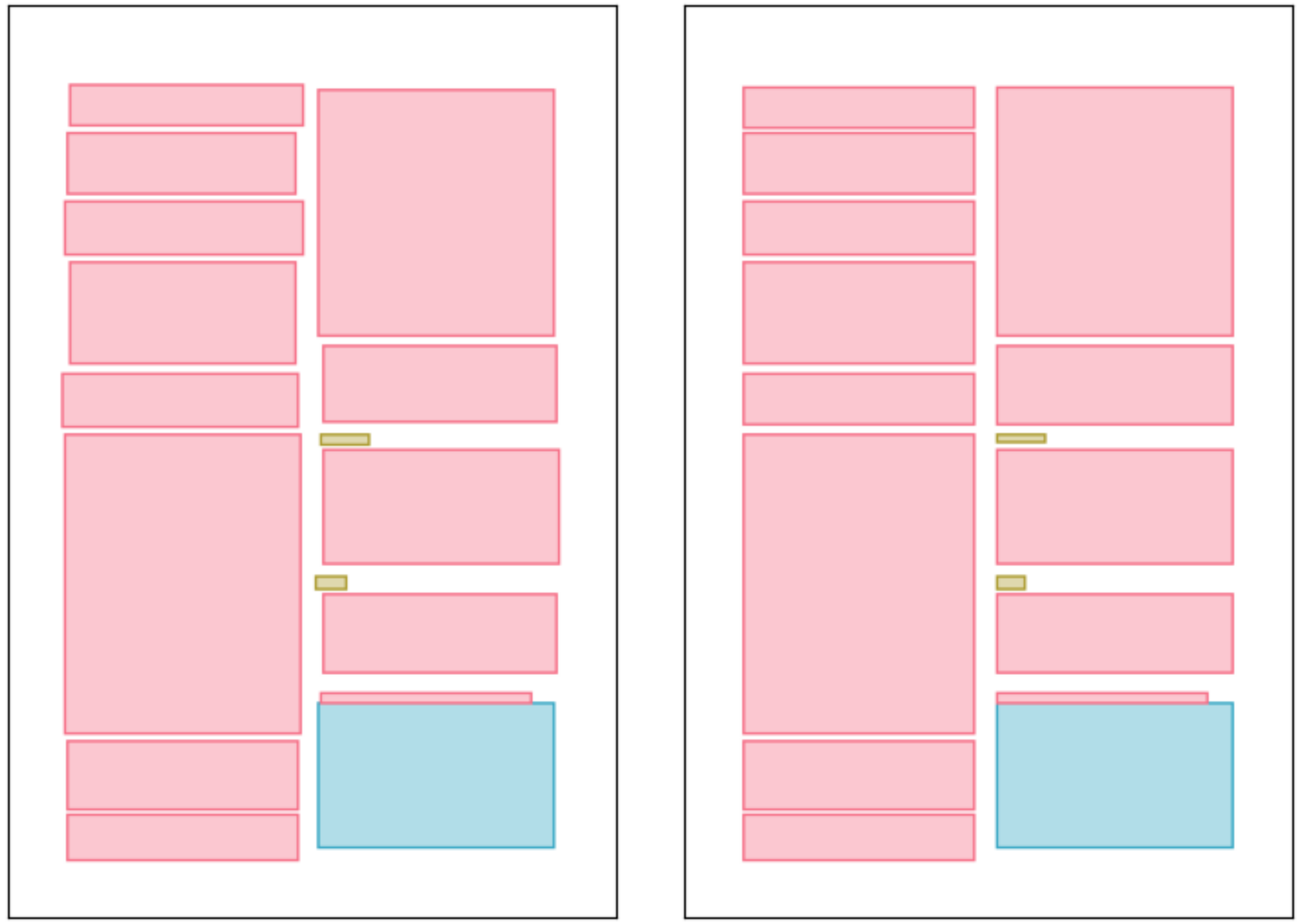


- At larger time steps, the gradient from the constraints can conflict with that of the diffusion loss, potentially collapsing the training process.
- A time-dependent constraint weight is employed to activate constraints only at smaller time steps.



Post-processing

- LACE exhibited enhanced performance across FID, alignment, and Max-IoU metrics when trained with constraints.
- LACE may generate layouts with minor misalignment: simple post-processing works!



Raw LACE output　　post-processed

Model	Task: 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

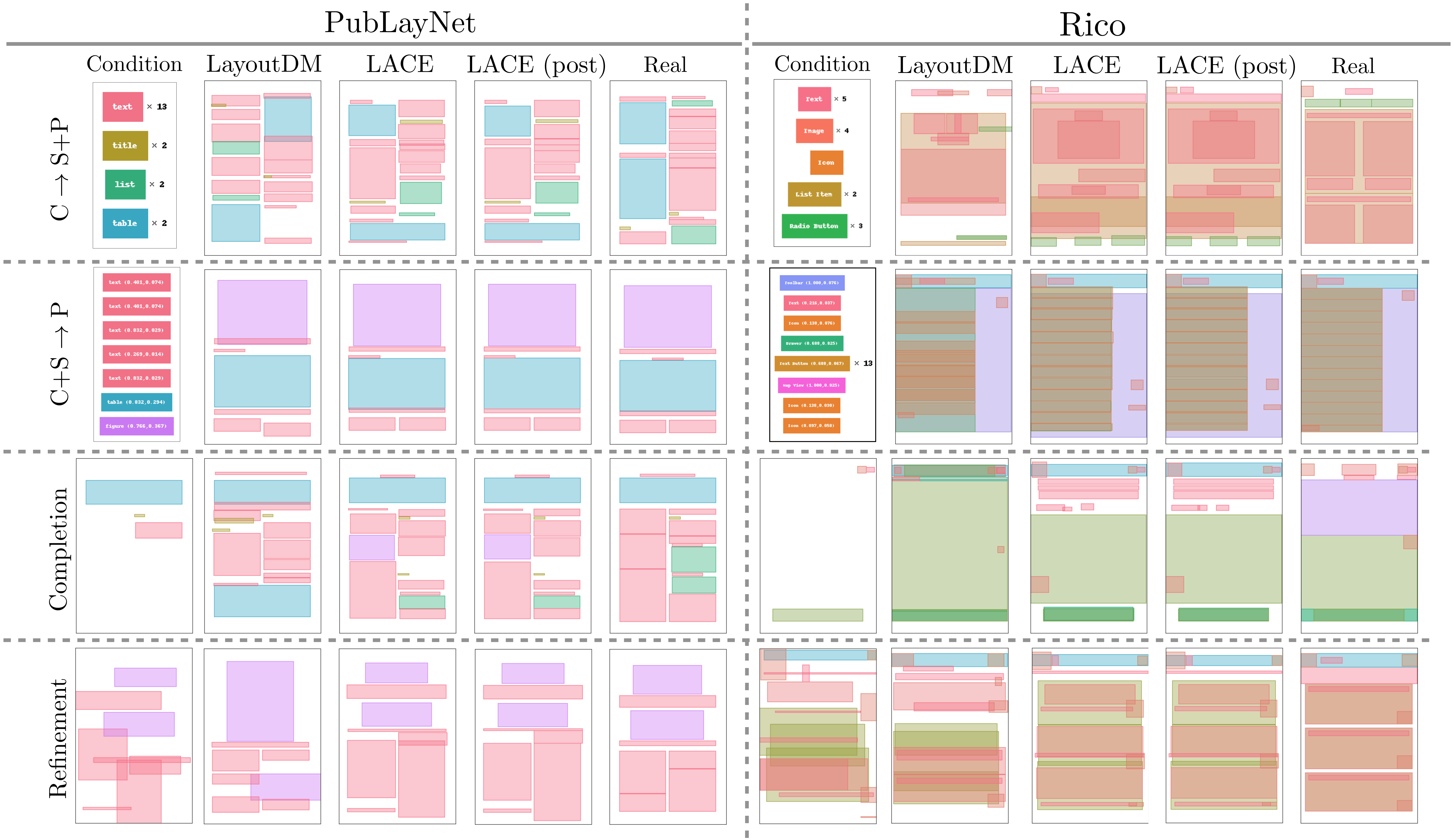
Simply applying post-processing on others damages performance

Experiment results

PubLayNet Task	C→S+P		C+S→P		Completion		U-Cond	
Model / Metric	FID↓	Max.↑	FID↓	Max.↑	FID↓	Max.↑	FID↓	Align.↓
Task-specific models								
LayoutVAE	26.0	0.316	27.5	0.315	-	-	-	-
NDN-none	61.1	0.162	69.4	0.222	-	-	-	-
LayoutGAN++	24.0	0.263	9.94	0.342	-	-	-	-
Task-agnostic models								
LayoutTrans	14.1	0.272	16.9	0.320	8.36	0.451	13.9	0.127
BLT	72.1	0.215	5.10	0.387	131	0.345	116	0.153
BART	9.36	0.320	5.88	0.375	9.58	0.446	16.6	0.116
MaskGIT	17.2	0.319	5.86	0.380	19.7	0.484	27.1	0.101
Diffusion-based models								
VQDiffusion	10.3	0.319	7.13	0.374	11.1	0.373	15.4	0.193
LayoutDM	7.95	0.310	4.25	0.381	7.65	0.377	13.9	0.195
LACE (local)	4.88	0.331	2.80	0.437	5.86	0.401	8.45	0.141
LACE (global)	5.14	0.383	3.07	0.463	6.03	0.396	8.35	0.185
LACE (local) w/ post	4.63	0.390	2.69	0.462	5.90	0.399	8.47	0.032
LACE (global) w/ post	4.56	0.388	2.53	0.463	5.63	0.394	7.43	0.074
Validation data	6.25	0.438	6.25	0.438	6.25	0.438	6.25	0.021

Rico Task	C→S+P		C+S→P		Completion		U-Cond	
Model / Metric	FID↓	Max.↑	FID↓	Max.↑	FID↓	Max.↑	FID↓	Align.↓
Task-specific models								
LayoutVAE	33.3	0.249	30.6	0.283	-	-	-	-
NDN-none	28.4	0.158	62.8	0.219	-	-	-	-
LayoutGAN++	6.84	0.267	6.22	0.348	-	-	-	-
Task-agnostic models								
LayoutTrans	5.57	0.223	3.73	0.323	3.71	0.537	7.63	0.068
BLT	17.4	0.202	4.48	0.340	117	0.471	88.2	1.030
BART	3.97	0.253	3.18	0.334	8.87	0.527	11.9	0.090
MaskGIT	26.1	0.262	8.05	0.320	33.5	0.533	52.1	0.015
Diffusion-based models								
VQDiffusion	4.34	0.252	3.21	0.331	11.0	0.541	7.46	0.178
LayoutDM	3.55	0.277	2.22	0.392	9.00	0.576	6.65	0.162
LACE (local)	3.31	0.336	2.66	0.418	5.09	0.518	4.71	0.107
LACE (global)	3.24	0.340	2.87	0.418	4.45	0.527	4.63	0.117
LACE (local) w/ post	2.88	0.347	2.17	0.430	3.82	0.539	3.99	0.031
LACE (global) w/ post	3.24	0.344	2.16	0.428	4.30	0.540	4.51	0.035
Validation data	1.85	0.691	1.85	0.691	1.85	0.691	1.85	0.109

Experiment results



University at Buffalo Adobe ICLR

Thanks