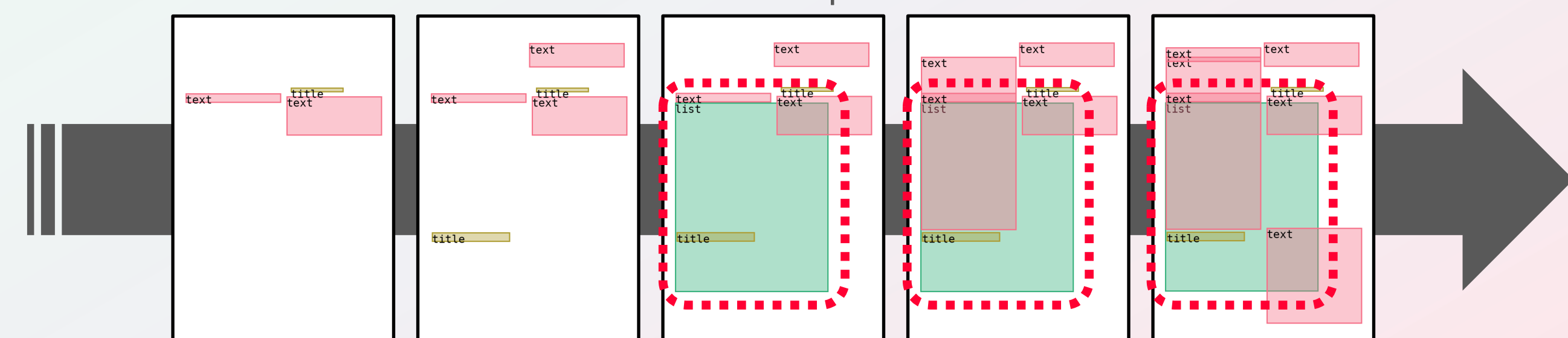


## TL;DR

Existing Discrete Diffusion Model struggles to correct generation errors  
= **Generation errors got stuck! (Layout-Sticking Problem)**

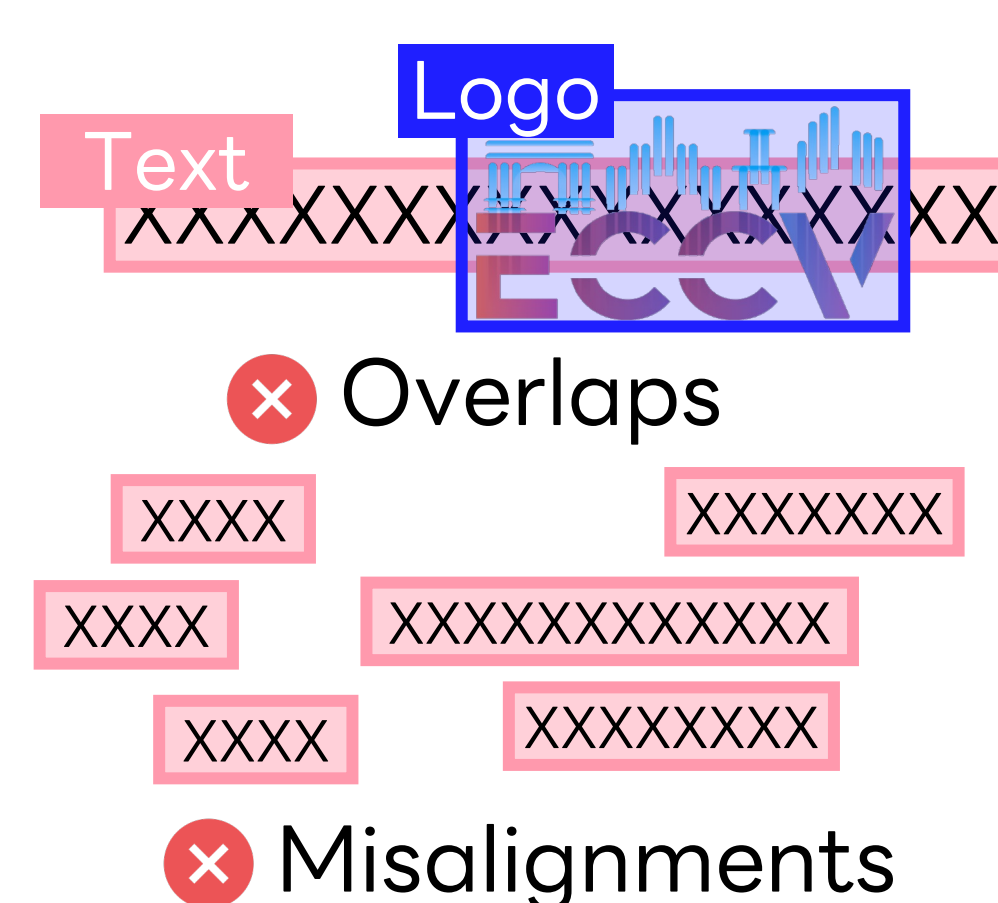
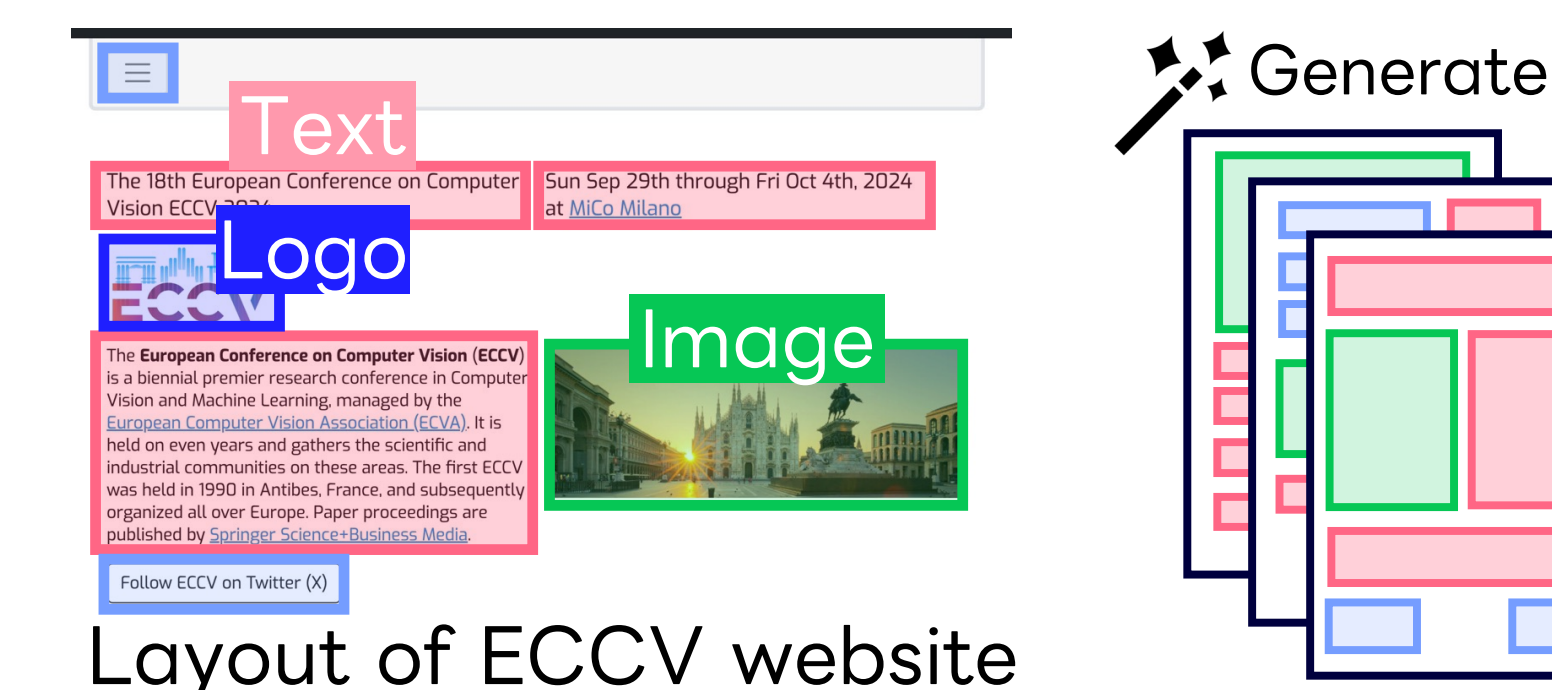
Generation process



→ Our **Layout-Corrector** resets errors and prompts correction

## Background: Layout Generation Goal: Correct Errors

Determine category, position, and size of each element (e.g., text, image)



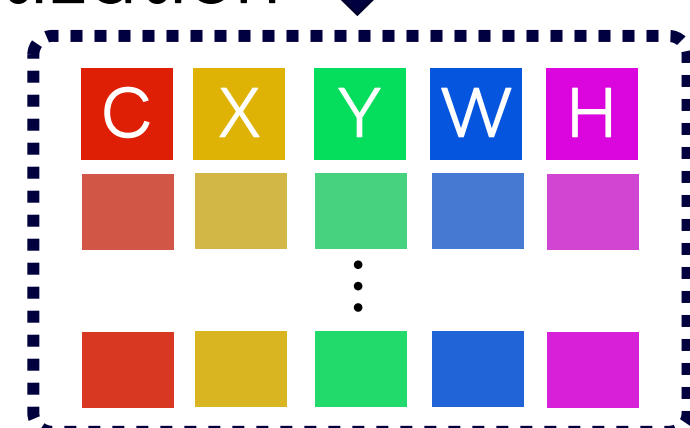
## Preparation: Layout Generation using Discrete Diffusion Model (DDM)

[Inoue+, CVPR2023]

Layout as Discrete Tokens

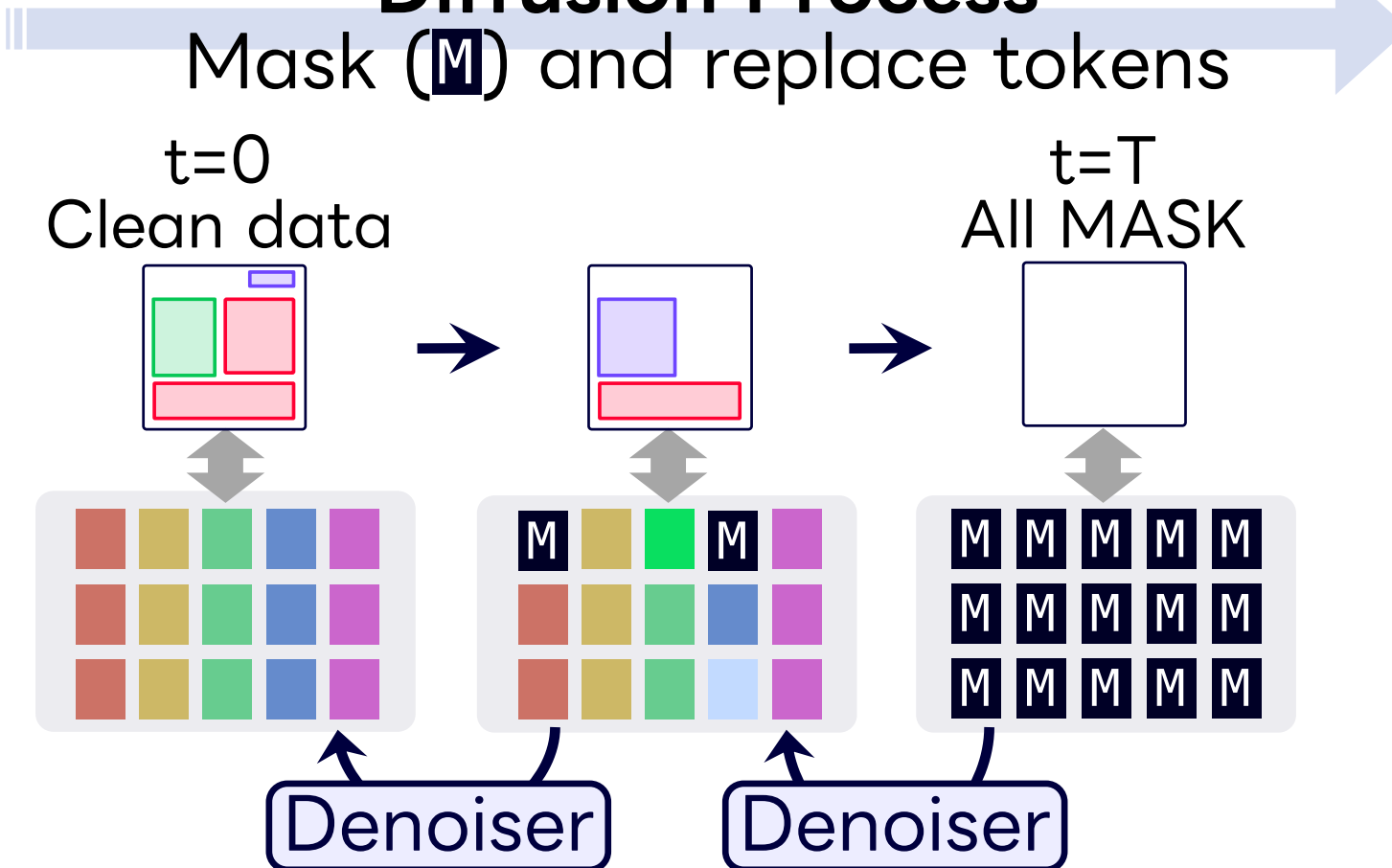
Category	X	Y	W	H
Text	15	10	100	50
...	...	...	...	...
Button	120	200	80	20

Quantization



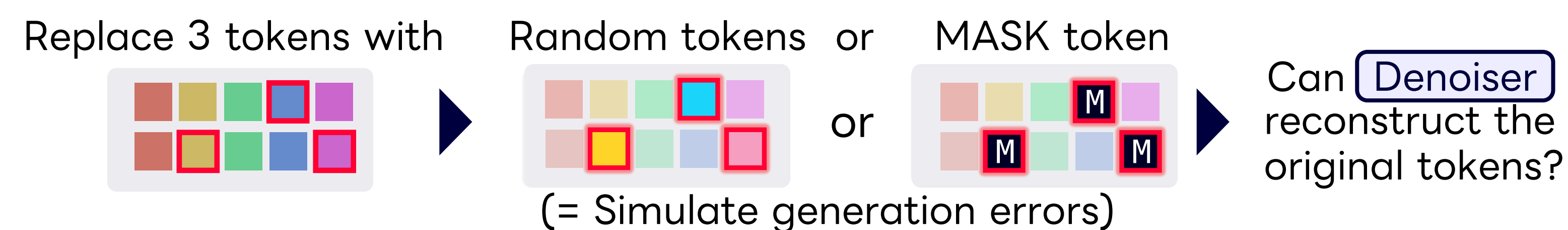
Layout token sequence

Diffusion Process



Backward Diffusion Process  
Unmask and replace tokens

## Preliminary: DDMs Struggle to Correct Error Directly



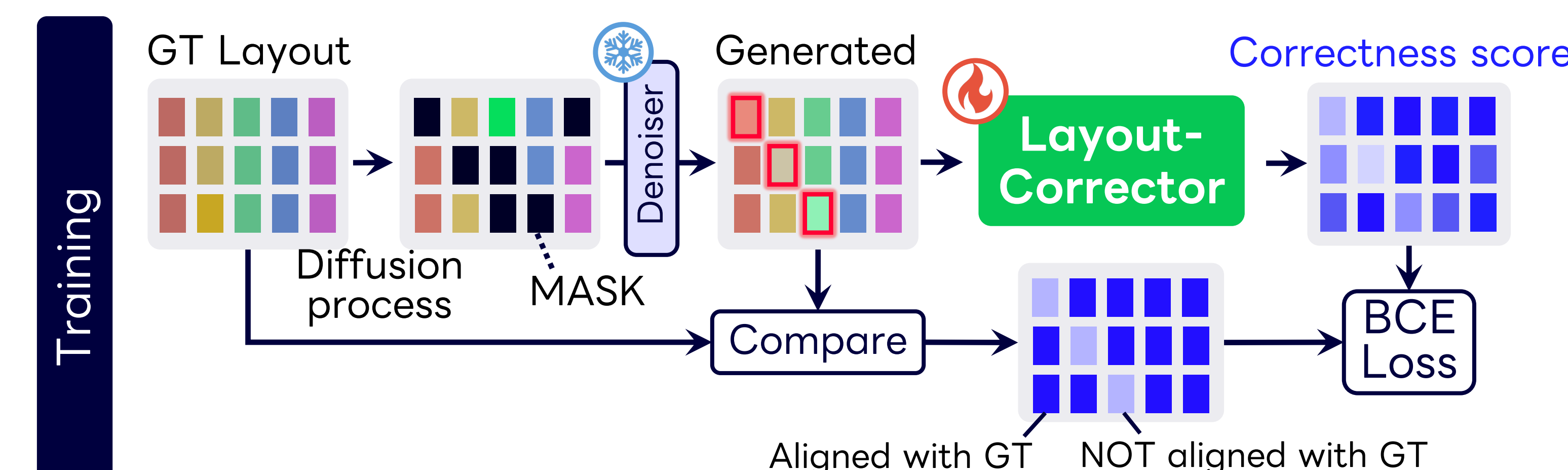
### Reconstruction success rate

Replace tokens with:	$\beta = 0.0$	$\beta = 0.05$	$\beta = 0.1$
Random tokens	0.0%	13.2%	15.1%
<b>MASK</b>	<b>35.4%</b>	<b>30.8%</b>	<b>29.7%</b>

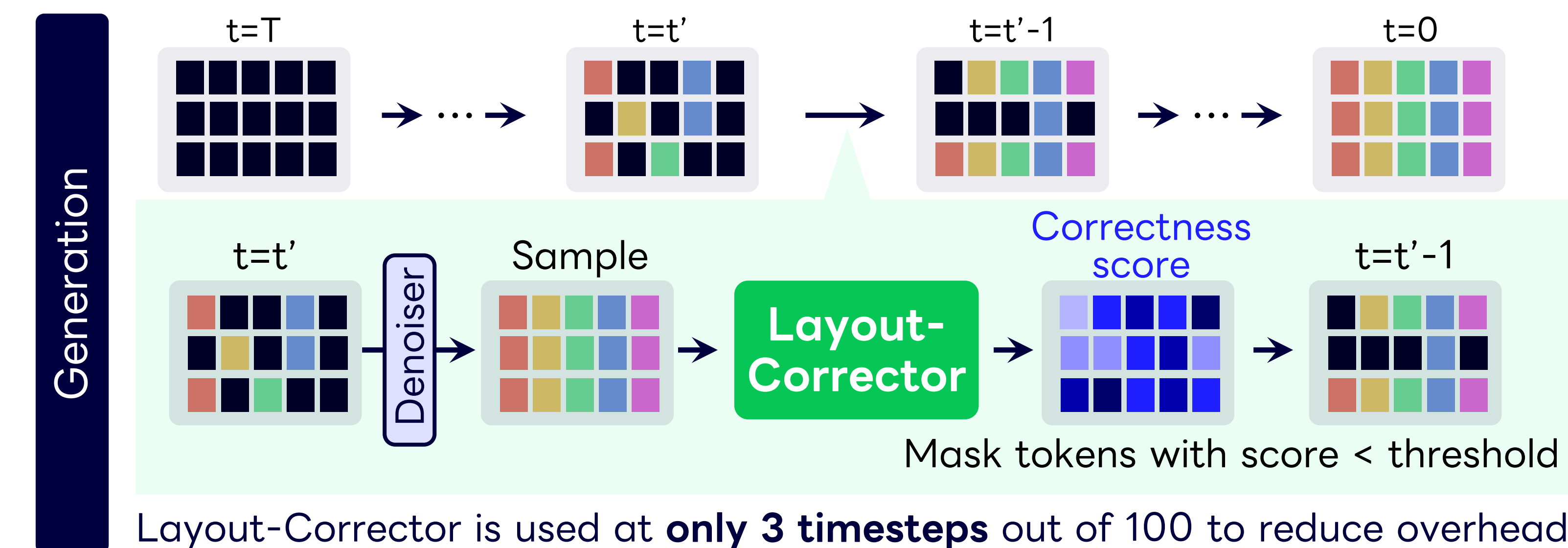
Hyperparameter  $\beta$ : Probability that replacement between non-MASK tokens happens during forward process

- ✗ DDMs struggle to correct erroneous tokens directly. (Success rate: ~15%)
- ✓ **By masking erroneous tokens first, DDMs could correct them.** (Success rate: ~35%)

## Proposed Method: Layout-Corrector Detects erroneous tokens to mask them → Prompt Correction



Predict whether each generated token aligns with the original (GT) token



Layout-Corrector is used at **only 3 timesteps** out of 100 to reduce overhead

## Results

Layout-Corrector improves the performance of multiple baseline models

(Rico dataset)	FID ↓	Precision ↑	Recall ↑
MaskGIT	70.4	0.793	0.437
+Token-Critic	<b>15.7</b>	0.682	<b>0.843</b>
<b>+Layout-Corrector</b>	<b>14.4</b>	<b>0.814</b>	<b>0.744</b>
LayoutDM	6.37	0.759	0.906
+Token-Critic	18.0	<b>0.884</b>	0.670
<b>+Layout-Corrector</b>	<b>4.79</b>	<b>0.811</b>	0.891
VQ-Diffusion	7.83	0.716	0.907
+Token-Critic	15.2	<b>0.842</b>	0.731
<b>+Layout-Corrector</b>	<b>5.29</b>	<b>0.809</b>	0.898

**Bold:** improved compared to the baseline

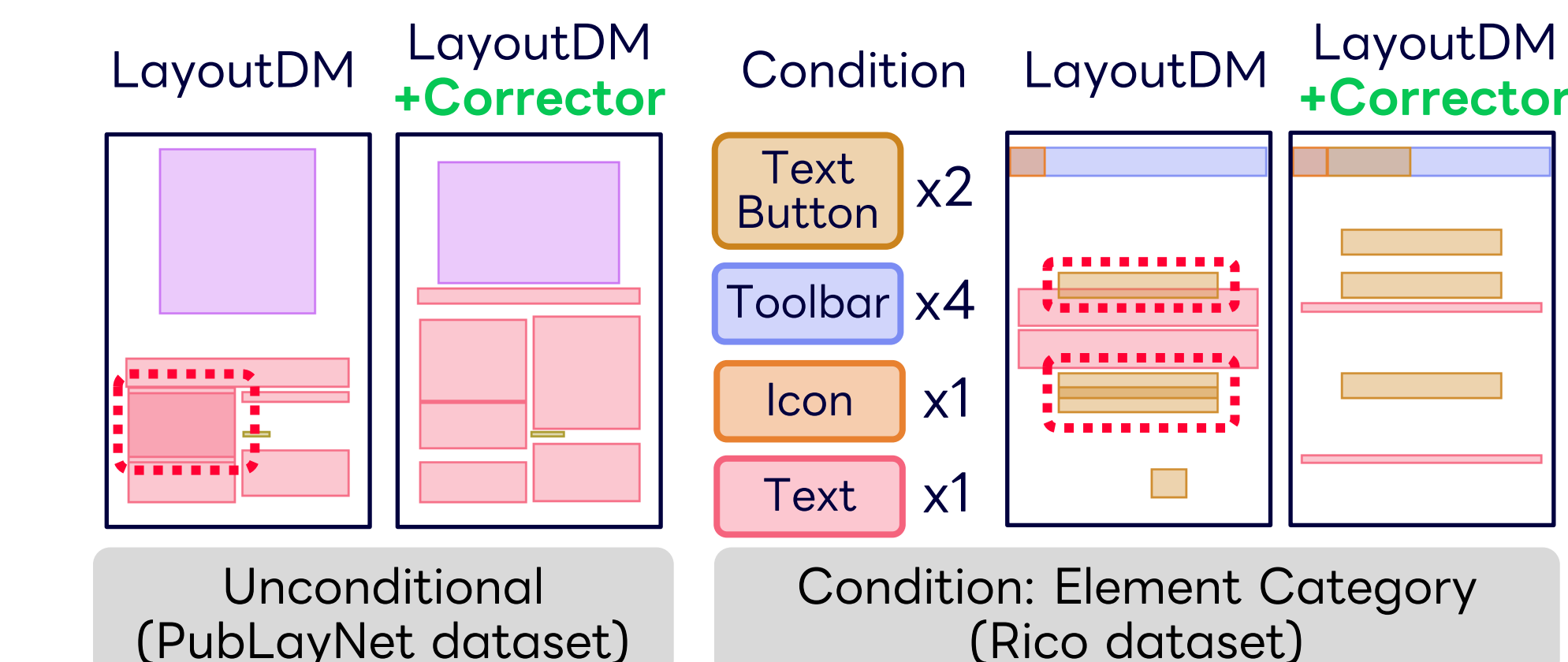
MaskGIT: [Chang+, CVPR2022]

VQ-Diffusion: [Gu+, CVPR2022]

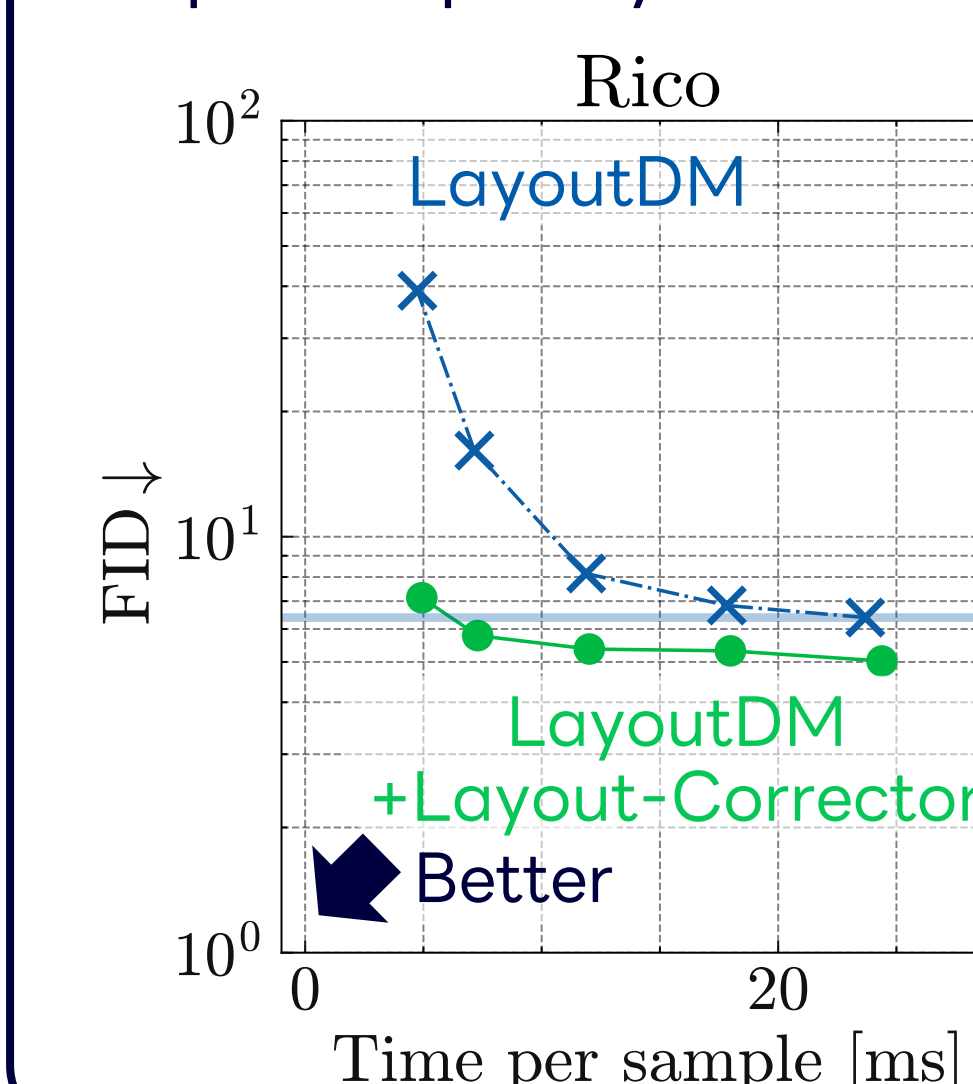
LayoutDM: [Inoue+, CVPR2023]

Token-Critic: [Lezama+, ECCV2022]

Visual samples: Errors are corrected



Speed-quality trade-off



Fidelity-diversity trade-off

