Supplementary Material: Unifying Layout Generation with a Decoupled Diffusion Model

Mude Hui[1]* Zhizheng Zhang[2] Xiaoyi Zhang[2] Wenxuan Xie[2] Yuwang Wang[3] Yan Lu[2]
[1]Xi'an Jiaotong University [2]Microsoft Research Asia [3]Tsinghua University
{zhizzhang, xiaoyizhang, wenxie, yanlu}@microsoft.com
theflood@stu.xjtu.edu.cn wang-yuwang@mail.tsinghua.edu.cn

In this supplementary material, we detail the experimental setup in Section 1, including detailed introduction to datasets in Section 1.1, hyperparameter configurations in Section 1.2, and more implementation details of ablation experiments in Section 1.3. Moreover, we provide more experimental results in Section 2, including *(i)* ablation studies for noise types and decoupling levels in Sections 2.1, 2.2, respectively, and *(ii)* more quantitative and qualitative results in Sections 2.3 and 2.4, respectively.

1. More Details about Experiment Setup

1.1. Detailed Introduction to the Datasets

Three datasets, *i.e.*, Magazine [9], Rico [2], and PubLayNet [10], are adopted in our experiments. Since the data splitting protocols for training and testing are not consistent over different publications, we thus re-implement the proposed methods in them and report the experiment results with the same data splitting protocols introduced below for fair comparison.

- *Magazine* [9] contains total 4K+ images of magazine pages. We use 85% of the dataset for training, 5% for validation, and 10% for testing. The categories in the dataset include *text*, *image*, *headline*, *text-over-image*, *headline-over-image*, and *background*.
- *Rico* [2] is a dataset of mobile app UI that contains 66K+ UI layouts. We randomly select 85% of the dataset for training, 5% for validation, and 10% for testing. Following the common practices in previous works [4–6], we exclude elements whose labels are not in the 13 most frequent labels from using. The adopted categorise are *Toolbar*, *Image*, *Text*, *Icon*, *Text Button*, *Input*, *List Item*, *Advertisement*, *Pager Indicator*, *Web View*, *Background Image*, and *Drawer,Modal*. Following the common practices in previous works [5, 7, 8], we also exclude the layouts with more than 25 elements since these layouts are rare but may lead to low training efficiency.
- *PubLayNet* [10] contains 360K+ document layout examples crawled from the Internet. We adopt the full official training split for training, 33% of the official validation split for validation, and the rest of the validation split for testing. This dataset contains elements from 5 catgories, *e.g.*, *Text*, *Title*, *List*, *Table*, and *Figure*. Following the common practices in previous works [5, 7, 8], similar to that for Rico, we exclude the layouts with more than 25 elements for improving the training efficiency.

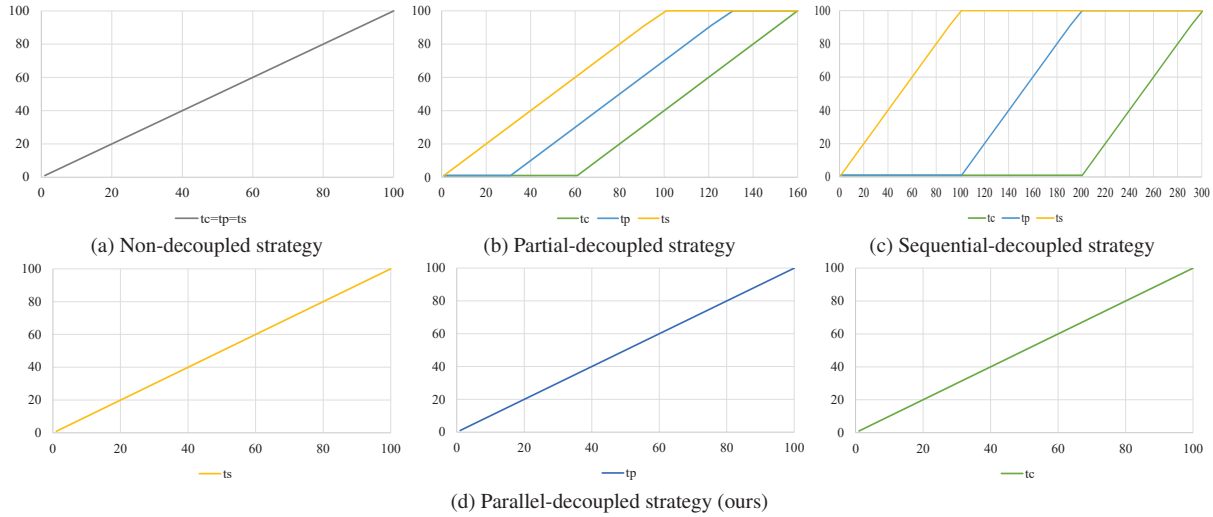
1.2. Hyperparameter Configurations

In our proposed LDGM, the hyperparameter β_t^c controls the transition probability from a category to the other one, which increases from 0 to $0.02/K^c$ as time t increases from 0 to T. Here, K^c is the number of adopted element categories in layouts, which varies for different datasets. For attributes x, y, h, w, σ_t controls the transition probability from one discrete to the other. σ_t^x, σ_t^y, σ_t^h and σ_t^w increase from 0 to 0.02 as time t increases from 0 to T. For all attributes c, x, y, h, w, the hyperparameter γ_t denotes the probability of masking a discrete value as an absorbing status. γ_t^c, γ_t^x, γ_t^y, γ_t^h and γ_t^w increase from 0 to 0.032 as time t increases from 0 to T. We adopt the same hyperparameter configurations in all experiments for fair comparison.

1.3. Implementation Details of Ablation Study on Decoupled Corruption Strategy

In this section, we introduce more implementation details of the four different diffusion strategies in the ablation study for our proposed decoupled diffusion strategy (*i.e.*, the first experiment in Section 5.4 of our main paper). We provide their corresponding illustrations in Figure 1. The algorithm descriptions for *Non-decoupled strategy*, *Partial-decoupled strategy* and *Sequential-decoupled strategy* are placed in Algorithm 1, Algorithm 2 and Algorithm 3, respectively. The algorithm description for our proposed strategy, *i.e.*, *Parallel-decoupled strategy (ours)*, has been placed in Algorithm 1 of our main paper.

*This work was done when Mude Hui was an intern at MSRA.



(a) Non-decoupled strategy (b) Partial-decoupled strategy (c) Sequential-decoupled strategy

(d) Parallel-decoupled strategy (ours)

Figure 1. Illustrations of different decoupled corruption strategies. The x-axis represents the overall timestep t from 1 to R, while the y-axis represents individual timesteps t_c, t_s, and t_p for different attributes respectively in the corresponding algorithms.

Algorithm 1 Non-decoupled strategy

Require: Max diffusion steps T

$l \leftarrow$ sample a layout from the training set
$\hat{l} \leftarrow \text{RandSelect}(l)$ ▷ Select attributes for diffusion.
$\hat{l} \leftarrow [C, P, S]$ ▷ Group $\hat{l}$ upon the semantics.
sample $t \sim \text{Uniform}(\{1, \cdots, T\})$
for x in C **do**
  $t_c \leftarrow t$
  $x = x_{t_c} \leftarrow$ sample from $q(x_{t_c}|x_0)$
end for
for x in P **do**
  $t_p \leftarrow t$
  $x = x_{t_p} \leftarrow$ sample from $q(x_{t_p}|x_0)$
end for
for x in S **do**
  $t_s \leftarrow t$
  $x = x_{t_s} \leftarrow$ sample from $q(x_{t_s}|x_0)$
end for

2. More Experiment Results

We further investigate the effectiveness of our proposed decoupled diffusion strategy by comparing different adopted noise types and decouple degrees/levels for diffusion. All experiments are performed on the Rico dataset and the Gen-PCM task.

2.1. Ablation Study for Noise Types

In our proposed LDGM, we add different types of noises for the diffusion processes of different types of attributes as introduced in our main paper. Specifically, we adopt uniform noise for category c, while adopting Gaussian noise for location (x, y) and size (w, h). In fact, it does not make sense to adopt other types of noises for category c since it is hard to measure the "distance" between different categories. Therefore, here, we validate the effectiveness of adopting Gaussian noise for position (x, y) and size (w, h) by comparing it to adopting another two types of noises to diffusion. One is using the uniform noise that is the same as the one adapted to category. The other is adopting band-diagonal noise [1], with the one for h as an example, which can be formulated as:

Table 1. Experiment results on the Rico dataset by varying the noise type on input tokens.

Model	MaxIoU ↑	FID ↓	Align. ↓	Overlap ↓
Uniform	0.16	61.24	0.53	106.04
Band-diagonal	0.29	47.74	0.49	88.30
Gaussian (Ours)	**0.59**	**21.59**	**0.40**	**54.77**

$$[\alpha_t^h]_{ij} = 1 - \textstyle\sum_{j=0, j\neq i}^{K^h} [Q_t^h]_{ij}, \tag{1}$$

$$[\beta_t^h]_{ij} = \begin{cases} \frac{1}{K}\sigma_t^h & \text{if} \quad 0 < |i-j| \leq v \\ 0 & \text{else} \end{cases} \tag{2}$$

where Q_t^h denotes the transition matrix for height h. K^h is the number of possible values for h. γ_t^h and σ_t^h are two hyperparameters that increase linearly as time t increases.

The comparison results are in Table 2. We can find that adopting Gaussian noise for position (x, y) and size (w, h) as we propose in LDGM is the most effective design choice. This is because adopting Gaussian noise corresponds to a distance-aware noise adding (diffusion) strategy, while

Algorithm 2 Partial decoupled strategy

Require: Max diffusion steps T, overlap $0.3T$
$l \leftarrow$ sample a layout from the training set
$\hat{l} \leftarrow \text{RandSelect}(l)$ ▷ Select attributes for diffusion.
$\hat{l} \leftarrow [C, P, S]$ ▷ Group $\hat{l}$ upon the semantics.
sample $t \sim \text{Uniform}(\{1, \cdots, 1.6T\})$
for x in C **do**
  if $t < 0.6T$ **then**
    $t_c \leftarrow 1$
  else
    $t_c \leftarrow t - 0.6T$
  end if
  $x = x_{t_c} \leftarrow$ sample from $q(x_{t_c}|x_0)$
end for
for x in P **do**
  if $t < 0.3T$ **then**
    $t_p \leftarrow 1$
  else if $t > 1.3T$ **then**
    $t_p \leftarrow T$
  else
    $t_p \leftarrow t - 0.3T$
  end if
  $x = x_{t_p} \leftarrow$ sample from $q(x_{t_p}|x_0)$
end for
for x in S **do**
  if $t < T$ **then**
    $t_s \leftarrow t$
  else
    $t_s \leftarrow T$
  end if
  $x = x_{t_s} \leftarrow$ sample from $q(x_{t_s}|x_0)$
end for

Algorithm 3 Sequential-decoupled strategy

Require: Max diffusion steps T
$l \leftarrow$ sample a layout from the training set
$\hat{l} \leftarrow \text{RandSelect}(l)$ ▷ Select attributes for diffusion.
$\hat{l} \leftarrow [C, P, S]$ ▷ Group $\hat{l}$ upon the semantics.
sample $t \sim \text{Uniform}(\{1, \cdots, 3T\})$
for x in C **do**
  if $t < 2T$ **then**
    $t_c \leftarrow 1$
  else
    $t_c \leftarrow t - 2T$
  end if
  $x = x_{t_c} \leftarrow$ sample from $q(x_{t_c}|x_0)$
end for
for x in P **do**
  if $t < T$ **then**
    $t_p \leftarrow 1$
  else if $T < t < 2T + 1$ **then**
    $t_p \leftarrow t - T$
  else
    $t_p \leftarrow T$
  end if
  $x = x_{t_p} \leftarrow$ sample from $q(x_{t_p}|x_0)$
end for
for x in S **do**
  if $t < T$ **then**
    $t_s \leftarrow t$
  else
    $t_s \leftarrow T$
  end if
  $x = x_{t_s} \leftarrow$ sample from $q(x_{t_s}|x_0)$
end for

adopting uniform noise can not achieve this and adopting band-diagonal noise provides limited distance awareness.

2.2. Ablation Study for Decoupling Levels

In our proposed diffusion strategy of LDGM, we decouple the diffusion processes for different attributes according to attribute types. To validate the effectiveness of this proposed design, we compare it with three other design choices with different decoupling levels: *(i) Non-decoupling*: all attributes are corrupted with a shared timeline without decoupling. *(ii) Element-level decoupling*: attribute tokens corresponding to the same layout element are corrupted with a shared timeline. *(iii) Token-level decoupling*: all attribute tokens are corrupted with their individual timelines (*i.e.*, maximum decoupling). *(iv) Ours*: attribute tokens of the same type are corrupted with a shared timeline as we proposed in our main paper.

The corresponding comparison results are in Table 2. We find that our proposed design in LDGM is the most effective one compared to the other three. This is because we decouple the diffusion processes upon the attribute types, yielding a semantics-aware decoupling strategy.

Table 2. Comparison results for the models with different decoupling levels on the Rico dataset.

Model	MaxIoU ↑	FID ↓	Align. ↓	Overlap ↓
Non-decoupling	0.56	29.24	0.43	60.04
Element-level	0.56	27.71	0.48	64.04
Token-level	0.58	23.58	0.41	57.09
Ours	**0.59**	**21.59**	**0.40**	**54.77**

2.3. More Quantitative Results

The quantitative results in Section 5.2 of our main paper are the mean values averaged over 5 runs with different random seeds. In this section, we supplement these quantitative results by further reporting their corresponding standard deviations. The results on Magazine, Rico and Pin-

LayNet datasets are in Tables 3, 4, and 5, respectively.

2.4. More Qualitative Results

In our main paper, we introduce four more general layout generation task settings which can cover the existing ones defined in previous works. They are *Gen-PM*, *Gen-CM*, *Gen-PC* and *Gen-PCM*. All six available layout generation subtasks defined in previous could be all considered as the special cases of them. In this section, we provide the qualitative generation results on *Gen-PM*, *Gen-CM*, *Gen-PC* and *Gen-PCM* in Figure 2. Note that to the best of our knowledge, no previous works can support them so that we are not allowed to compare our proposed LDGM with others on these newly proposed task settings. We make the first endeavour to achieve such comprehensive versatility.

2.5. Results of Rendered Images

We show the results of rendered images upon the generated layouts by our proposed LDGM in Figure 3. These rendered images help visually demonstrate the quality of our generated layouts.

References

[1] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021. 2

[2] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 845–854, 2017. 1

[3] Kamal Gupta, Justin Lazarow, Alessandro Achille, Larry S Davis, Vijay Mahadevan, and Abhinav Shrivastava. Layouttransformer: Layout generation and completion with self-attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1004–1014, 2021. 5, 6, 7

[4] Zhaoyun Jiang, Huayu Deng, Zhongkai Wu, Jiaqi Guo, Shizhao Sun, Vuksan Mijovic, Zijiang Yang, Jian-Guang Lou, and Dongmei Zhang. Unilayout: Taming unified sequence-to-sequence transformers for graphic layout generation. *arXiv preprint arXiv:2208.08037*, 2022. 1, 5, 6, 7

[5] Kotaro Kikuchi, Edgar Simo-Serra, Mayu Otani, and Kota Yamaguchi. Constrained graphic layout generation via latent optimization. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 88–96, 2021. 1, 5, 6, 7

[6] Xiang Kong, Lu Jiang, Huiwen Chang, Han Zhang, Yuan Hao, Haifeng Gong, and Irfan Essa. Blt: Bidirectional layout transformer for controllable layout generation. *arXiv preprint arXiv:2112.05112*, 2021. 1, 5, 6, 7

[7] Hsin-Ying Lee, Lu Jiang, Irfan Essa, Phuong B Le, Haifeng Gong, Ming-Hsuan Yang, and Weilong Yang. Neural design network: Graphic layout generation with constraints. In *European Conference on Computer Vision*, pages 491–506. Springer, 2020. 1

[8] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. Layoutgan: Generating graphic layouts with wireframe discriminators. *arXiv preprint arXiv:1901.06767*, 2019. 1

[9] Xinru Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics (TOG)*, 38(4):1–15, 2019. 1

[10] Xu Zhong, Jianbin Tang, and Antonio Jimeno Yepes. Publaynet: largest dataset ever for document layout analysis. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1015–1022. IEEE, 2019. 1

Table 3. Experiment results of different layout generation subtasks on the Magazine dataset. *Align.* denotes the alignment metric.

Subtasks	Methods	Magazine			
		MaxIoU ↑	FID ↓	Align. ↓	Overlap ↓
UGen	LayoutTransformer [3]	0.18 ± 0.03	47.84 ± 1.03	0.59 ± 0.03	47.98 ± 0.87
	BLT [6]	0.20 ± 0.04	44.91 ± 1.56	0.55 ± 0.05	55.56 ± 1.05
	UniLayout [4]	0.31 ± 0.01	36.61 ± 1.23	0.49 ± 0.03	**44.50 ± 1.02**
	LDGM (Ours)	**0.38 ± 0.00**	**32.73 ± 0.62**	**0.47 ± 0.01**	46.43 ± 0.98
Gen-T	LayoutGAN++ [5]	0.26 ± 0.01	36.35 ± 0.77	0.54 ± 0.02	58.44 ± 0.73
	BLT [6]	0.22 ± 0.04	48.26 ± 1.96	0.69 ± 0.03	64.01 ± 1.43
	UniLayout [4]	0.32 ± 0.01	28.37 ± 1.26	0.51 ± 0.03	53.56 ± 1.02
	LDGM (Ours)	**0.36 ± 0.01**	**24.67 ± 0.43**	**0.45 ± 0.03**	**45.11 ± 0.79**
Gen-TS	BLT [6]	0.33 ± 0.03	22.72 ± 1.54	0.59 ± 0.01	61.94 ± 1.00
	UniLayout [4]	0.35 ± 0.01	19.35 ± 0.72	0.58 ± 0.03	56.43 ± 1.02
	LDGM (Ours)	**0.37 ± 0.02**	**17.65 ± 0.57**	**0.45 ± 0.02**	**44.25 ± 0.56**
Gen-TR	CLG-LO [5]	0.27 ± 0.01	33.88 ± 1.31	0.59 ± 0.05	59.43 ± 1.39
	UniLayout [4]	0.36 ± 0.01	**19.24 ± 0.88**	0.54 ± 0.03	49.61 ± 0.74
	LDGM (Ours)	**0.39 ± 0.02**	20.58 ± 0.45	**0.48 ± 0.03**	**47.27 ± 0.80**
Refinement	RUITE	0.24 ± 0.03	44.27 ± 1.32	0.64 ± 0.05	54.26 ± 1.41
	UniLayout [4]	0.33 ± 0.01	19.78 ± 0.70	0.49 ± 0.03	49.02 ± 0.96
	LDGM (Ours)	**0.39 ± 0.00**	**14.95 ± 0.68**	**0.42 ± 0.01**	**37.22 ± 0.51**
Completion	LayoutTransformer [3]	0.17 ± 0.03	39.36 ± 1.83	0.67 ± 0.02	55.32 ± 0.76
	UniLayout [4]	0.23 ± 0.01	28.78 ± 1.60	0.52 ± 0.03	46.43 ± 0.64
	LDGM (Ours)	**0.38 ± 0.00**	**24.35 ± 0.43**	**0.49 ± 0.01**	**39.26 ± 0.49**
Gen-PM	LDGM (Ours)	0.38 ± 0.01	27.33 ± 0.54	0.47 ± 0.01	39.02 ± 0.39
Gen-CM		0.37 ± 0.02	28.74 ± 0.79	0.51 ± 0.01	43.25 ± 0.71
Gen-PC		0.37 ± 0.01	22.56 ± 0.62	0.47 ± 0.00	42.95 ± 0.30
Gen-PCM		0.37 ± 0.01	24.45 ± 1.21	0.49 ± 0.01	44.41 ± 0.89
GT	-	0.41	9.89	0.43	34.27

Table 4. Experiment results of different layout generation subtasks on the Rico dataset. *Align.* denotes the alignment metric.

Subtasks	Methods	Rico			
		MaxIoU ↑	FID ↓	Align. ↓	Overlap ↓
UGen	LayoutTransformer [3]	0.46 ± 0.03	46.64 ± 0.97	0.66 ± 0.02	64.10 ± 0.94
	BLT [6]	0.51 ± 0.01	33.81 ± 1.56	0.59 ± 0.04	67.33 ± 0.71
	UniLayout [4]	**0.62 ± 0.01**	26.68 ± 0.74	0.40 ± 0.03	59.26 ± 0.76
	LDGM (Ours)	**0.62 ± 0.01**	**26.06 ± 0.40**	**0.36 ± 0.03**	**56.35 ± 0.71**
Gen-T	LayoutGAN++ [5]	0.46 ± 0.00	34.43 ± 1.13	0.58 ± 0.02	59.85 ± 0.85
	BLT [6]	0.44 ± 0.03	39.64 ± 1.71	0.57 ± 0.01	56.83 ± 1.45
	UniLayout [4]	0.55 ± 0.01	18.06 ± 0.70	0.48 ± 0.03	57.92 ± 0.94
	LDGM (Ours)	**0.58 ± 0.01**	**16.64 ± 0.46**	**0.39 ± 0.03**	**55.87 ± 0.64**
Gen-TS	BLT [6]	0.51 ± 0.03	42.88 ± 1.26	0.46 ± 0.02	57.74 ± 0.47
	UniLayout [4]	0.55 ± 0.01	20.42 ± 0.40	0.49 ± 0.02	58.72 ± 0.36
	LDGM (Ours)	**0.62 ± 0.01**	**12.59 ± 0.37**	**0.35 ± 0.01**	**55.92 ± 0.39**
Gen-TR	CLG-LO [5]	0.38 ± 0.02	38.89 ± 0.79	0.54 ± 0.02	**56.51 ± 0.80**
	UniLayout [4]	0.57 ± 0.01	26.38 ± 0.92	0.46 ± 0.02	66.93 ± 0.52
	LDGM (Ours)	**0.61 ± 0.01**	**16.98 ± 0.40**	**0.39 ± 0.01**	58.75 ± 0.66
Refinement	RUITE	0.46 ± 0.03	36.70 ± 0.70	0.57 ± 0.02	64.13 ± 1.94
	UniLayout [4]	0.56 ± 0.01	24.41 ± 0.57	0.42 ± 0.01	56.04 ± 0.65
	LDGM (Ours)	**0.62 ± 0.00**	**13.19 ± 0.40**	**0.33 ± 0.01**	**52.17 ± 0.58**
Completion	LayoutTransformer [3]	0.46 ± 0.03	36.15 ± 0.64	0.66 ± 0.02	67.10 ± 0.65
	UniLayout [4]	0.59 ± 0.01	25.18 ± 0.42	0.45 ± 0.04	55.99 ± 0.92
	LDGM (Ours)	**0.60 ± 0.01**	**16.42 ± 0.66**	**0.36 ± 0.03**	**53.15 ± 0.64**
Gen-PM	LDGM (Ours)	0.58 ± 0.01	21.64 ± 0.50	0.38 ± 0.03	56.56 ± 0.85
Gen-CM		0.57 ± 0.03	26.15 ± 0.46	0.38 ± 0.03	57.74 ± 0.79
Gen-PC		0.60 ± 0.01	18.13 ± 0.33	0.36 ± 0.00	53.67 ± 0.46
Gen-PCM		0.59 ± 0.01	21.59 ± 0.75	0.40 ± 0.01	54.77 ± 0.70
GT	-	0.66	7.05	0.26	49.86

Table 5. Experimental results of different layout generation subtasks on the PublayNet dataset. *Align.* denotes the alignment metric.

Subtasks	Methods	PubLayNet			
		MaxIoU ↑	FID ↓	Align. ↓	Overlap ↓
UGen	LayoutTransformer [3]	0.32 ± 0.01	49.72 ± 1.81	0.37 ± 0.01	36.63 ± 0.43
	BLT [6]	0.34 ± 0.01	48.24 ± 0.68	0.27 ± 0.04	42.79 ± 0.53
	UniLayout [4]	0.33 ± 0.01	32.29 ± 0.56	**0.22 ± 0.02**	22.19 ± 0.22
	LDGM (Ours)	**0.46 ± 0.01**	**25.94 ± 0.41**	0.25 ± 0.01	**19.83 ± 0.19**
Gen-T	LayoutGAN++ [5]	0.36 ± 0.01	30.48 ± 0.75	0.19 ± 0.02	32.80 ± 0.42
	BLT [6]	0.37 ± 0.03	44.86 ± 0.66	0.21 ± 0.03	38.21 ± 0.43
	UniLayout [4]	0.41 ± 0.01	27.34 ± 0.96	0.20 ± 0.02	20.98 ± 0.37
	LDGM (Ours)	**0.44 ± 0.01**	**20.69 ± 0.36**	**0.15 ± 0.02**	**16.88 ± 0.39**
Gen-TS	BLT [6]	0.40 ± 0.00	24.32 ± 0.73	0.16 ± 0.02	31.06 ± 0.58
	UniLayout [4]	0.43 ± 0.01	27.47 ± 0.85	**0.16 ± 0.01**	23.82 ± 0.66
	LDGM (Ours)	**0.47 ± 0.01**	**19.02 ± 0.35**	**0.16 ± 0.01**	**10.09 ± 0.38**
Gen-TR	CLG-LO [5]	0.38 ± 0.02	31.87 ± 0.82	0.21 ± 0.03	34.39 ± 0.59
	UniLayout [4]	**0.46 ± 0.01**	27.73 ± 0.84	0.17 ± 0.02	27.35 ± 0.61
	LDGM (Ours)	0.44 ± 0.01	**19.54 ± 0.43**	**0.16 ± 0.01**	**21.28 ± 0.54**
Refinement	RUITE	0.32 ± 0.01	41.72 ± 0.99	0.49 ± 0.01	35.74 ± 1.89
	UniLayout [4]	0.44 ± 0.01	22.34 ± 0.87	0.11 ± 0.01	27.23 ± 0.73
	LDGM (Ours)	**0.48 ± 0.01**	**15.28 ± 0.59**	**0.10 ± 0.01**	**13.05 ± 0.42**
Completion	LayoutTransformer [3]	0.32 ± 0.01	41.72 ± 0.81	0.37 ± 0.01	39.81 ± 0.20
	UniLayout [4]	0.41 ± 0.01	32.04 ± 0.55	0.19 ± 0.02	22.90 ± 0.29
	LDGM (Ours)	**0.44 ± 0.01**	**25.31 ± 0.60**	**0.10 ± 0.00**	**19.45 ± 0.28**
Gen-PM	LDGM (Ours)	0.46 ± 0.01	23.58 ± 0.18	0.10 ± 0.00	14.11 ± 0.29
Gen-CM		0.44 ± 0.01	24.94 ± 0.24	0.11 ± 0.01	16.26 ± 0.47
Gen-PC		0.50 ± 0.01	16.42 ± 0.18	0.09 ± 0.01	12.51 ± 0.11
Gen-PCM		0.42 ± 0.00	25.76 ± 0.59	0.14 ± 0.01	19.68 ± 0.58
GT	-	0.64	9.38	0.008	5.18

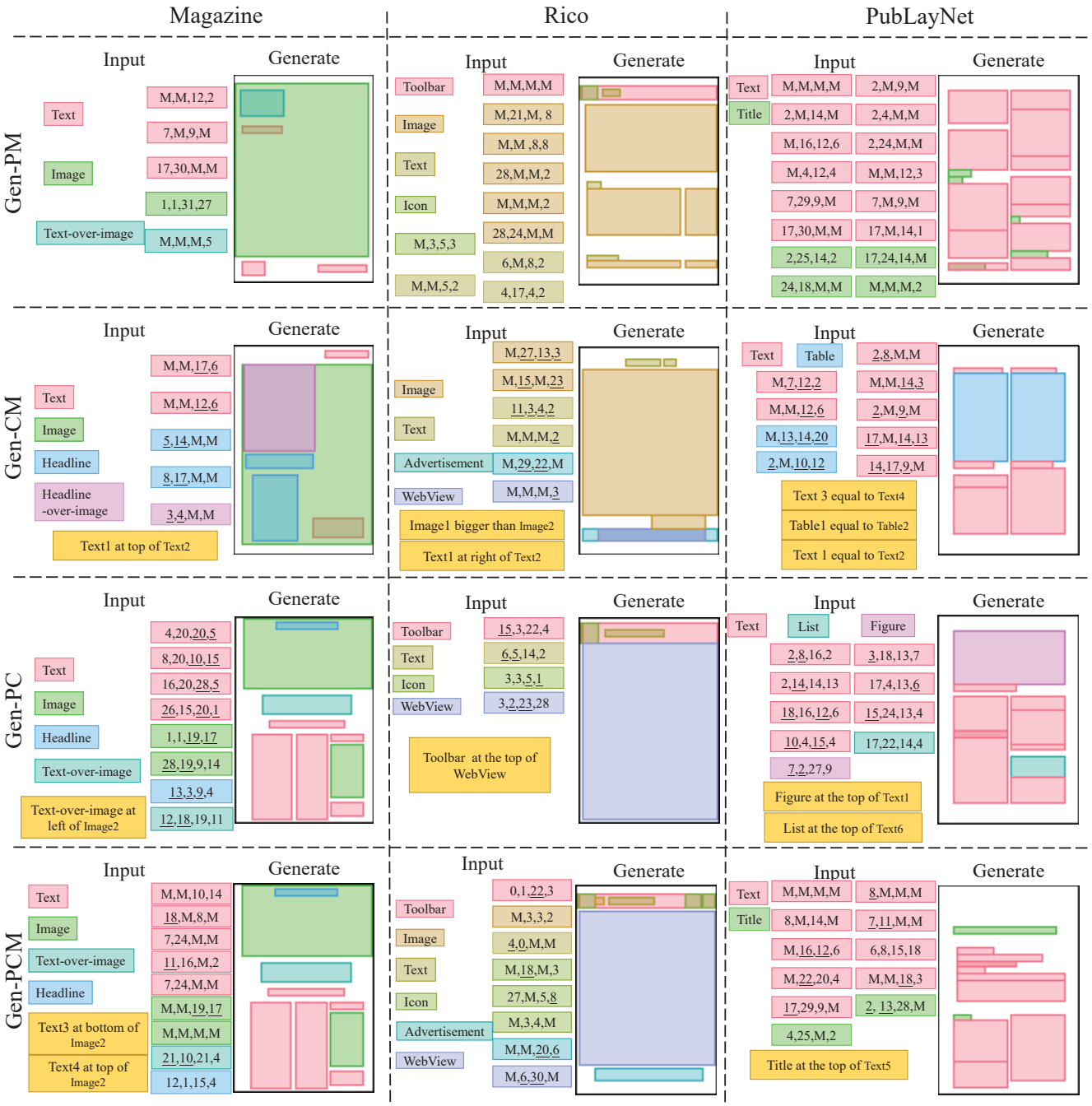


Figure 2. Qualitative results on general layout generation tasks(Gen-CM,Gen-PC,Gen-PC,Gen-PCM).M represents missing attribute, x represents coarse attribute, and x represents precise attribute.

Figure 3. Rendered Images. The generated layouts are from three models trained on Magazine, Rico, and PubLayNet, respectively.