DLT: Conditioned layout generation with Joint Discrete-Continuous Diffusion Layout Transformer
Supplementary Material

Appendix

A. Additional results

We provide additional experimental results. In Table 1 we provide the std of all the experiments in Table 1 in the main text. Each of the experiments was tested with four trials on the test data of the relevant dataset. As can be seen, our model outperforms the tested methods by a statically significant margin. We also provide an additional qualitative comparison on the Magazine dataset in Figure 1. It is important to note that this dataset is a small dataset which results in degraded performances (compared to a bigger dataset such as PubLaynet). However, our model still performs better also on the Magazine dataset compared to the tested model.

Dataset	**Publaynet**											
	Conditioned on Category				Category + Size				Uncoditioned			
Model	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID
LT [2]	0.2	0.3	0.06	1.2	0.3	0.4	0.01	1.4	0.02	0.2	0.01	1.5
BLT [3]	0.1	0.7	0.03	2.3	0.3	0.7	0.01	0.6	0.01	0.1	0.01	2.7
VTN [1]	0.2	0.5	0.03	1.1	0.4	0.4	0.01	1.0	0.01	0.1	0.01	1.9
DLT	0.08	0.2	0.02	0.9	0.09	0.3	0.01	0.8	0.01	0.1	0.01	0.9
Dataset	**Rico**											
	Conditioned on Category				Category + Size				Uncoditioned			
Model	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID
LT [2]	0.9	1.6	0.02	0.5	1.0	1.1	0.06	0.3	1.5	0.7	0.01	0.4
BLT [3]	1.2	1.6	0.01	0.7	0.9	0.9	0.08	0.4	1.8	0.8	0.01	0.9
VTN [1]	1.0	1.4	0.02	0.4	1.1	0.9	0.05	0.3	1.9	0.7	0.01	1.0
DLT	1.0	1.3	0.01	0.5	0.8	0.6	0.07	0.3	1.6	0.9	0.01	0.6
Dataset	**Magazine**											
	Conditioned on Category				Category + Size				Uncoditioned			
Model	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID	pIOU	Overlap	Alignment	FID
LT [2]	0.8	2.9	0.4	3.1	1.1	2.4	0.3	2.7	1.8	1.6	0.3	2.7
BLT [3]	1.1	3.1	0.3	2.7	1.7	4.6	0.4	2.0	2.3	2.6	0.3	3.6
VTN [1]	0.7	2.7	0.3	2.6	1.0	2.1	0.3	1.9	1.6	0.9	0.5	4.4
DLT	0.2	1.7	0.2	2.2	0.7	1.3	0.2	1.6	0.4	0.5	0.4	2.6

Table 1: Std results of all tested experiments in Table 1 in the main text

B. Conditioning mechanism experiments

In this section, we provide additional information on the conditioning ablation study. The three initial scenarios described in the main text subsection 4.3, were chosen according to [3] evaluation protocol. It is important to note that in this setting each component attribute is either condition on **all** the components, or predicting the attributes for **all** the components. In these

Setting	Publaynet Category+Size			
Model	pIOU	Overlap	Alignment	FID
Edit only in inference	3.3	11.5	0.24	20.1
Without condition embedding	0.86	4.8	0.15	5.55
DLT	0.85	5.3	0.14	5.73

Table 2: Quantitative comparison of different conditioning approaches on the Publaynet dataset in the category+size condition setting. The values of Alignment, Overlap, and pIOU are multiplied by 100 for clarity.

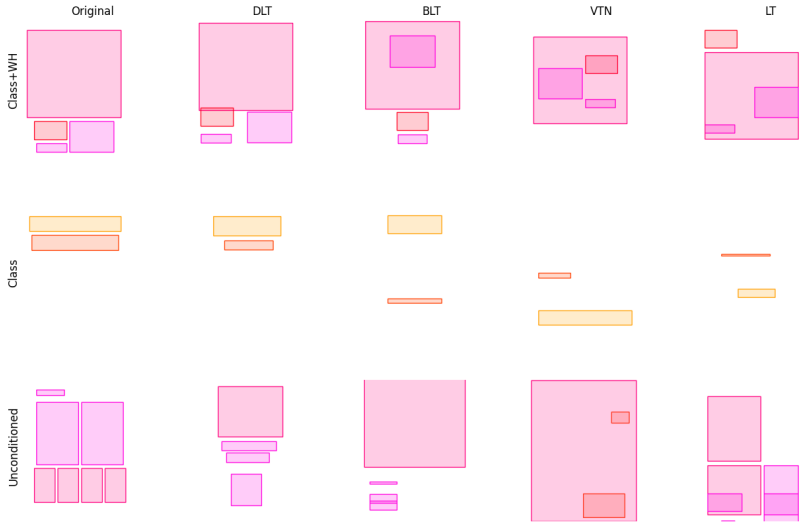

Figure 1: Qualitative comparison of generated layouts between all tested methods on the Magazine dataset with the three tested conditioning settings.

settings, the condition embedding does not have an impact on the model's performance as can be seen in Table 2. However, when the conditioning state of the attributes is not uniform across all components, the condition embedding provides the model with important information that can resolve the alignment ambiguity in the generation process. For example, if we look at the location attributes, as can be seen in Figure 2, without the condition embedding the model shifts the locations of all the components. However, when using the condition embedding the model aligns the unconditioned component according to the condition component. In the main text Table 3, we indeed see that in this setting the condition embedding boosts the model performances.

References

[1] Diego Martín Arroyo, Janis Postels, and Federico Tombari. Variational transformer networks for layout generation. *CoRR*, abs/2104.02416, 2021. 1

[2] Kamal Gupta, Alessandro Achille, Justin Lazarow, Larry Davis, Vijay Mahadevan, and Abhinav Shrivastava. Layout generation and completion with self-attention. *CoRR*, abs/2006.14615, 2020. 1

[3] Xiang Kong, Lu Jiang, Huiwen Chang, Han Zhang, Yuan Hao, Haifeng Gong, and Irfan Essa. BLT: bidirectional layout transformer for controllable layout generation. *CoRR*, abs/2112.05112, 2021. 1

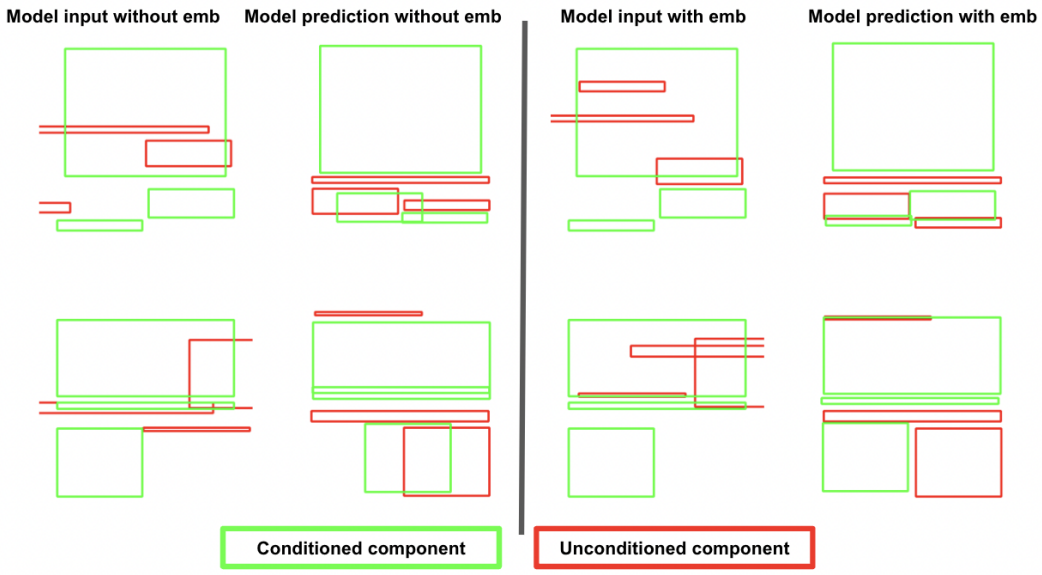

Figure 2: A comparison of model prediction with and without the condition embedding at the second reverse diffusion step. **Green:** Components which are part of the conditioning. **Red:** Components which are part of the generation process. As can be seen, without the embedding the model shifts the conditioned component location and organize the unconditioned part accordingly. With the embedding, the model organizes the unconditioned components according to the conditioned components' locations.