SciPostLayout: A Dataset for Layout Analysis and Layout Generation of Scientific Posters

Hao Wang[1,2*]
conan1024hao@akane.waseda.jp

Shohei Tanaka[2]
shohei.tanaka@sinicx.com

Yoshitaka Ushiku[2]
yoshitaka.ushiku@sinicx.com

[1]Waseda University
1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo, Japan

[2]OMRON SINIC X Corporation
Nagase Hongo Building 3F, 5-24-5 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Scientific posters are used to present the contributions of scientific papers effectively in a graphical format. However, creating a well-designed poster that efficiently summarizes the core of a paper is labor intensive and time-consuming. A system that can automatically generate well-designed posters from scientific papers would reduce the workload of authors and help readers understand the outline of the paper visually. Despite the demand for poster generation systems, generating posters from papers has a limited research population due to the lack of publicly available datasets. Thus, in this paper, we built the SciPostLayout dataset, which consists of 7,855 scientific posters and manual layout annotations for layout analysis and generation. All of the posters in our dataset are under the CC-BY license and publicly available. As benchmark tests for the collected dataset, we conducted experiments for layout analysis and generation using existing computer vision models. Experimental results show that layout analysis and generation of posters using SciPostLayout are more challenging than with scientific papers. The dataset is publicly available at https://huggingface.co/datasets/omron-sinicx/scipostlayout_v1.

1. Introduction

Scientific posters are used to efficiently present the contributions of a scientific paper in a graphical format. A well-designed scientific poster conveys the essential elements of the research and requires less time to read than a paper. Nevertheless, creating a scientific poster that efficiently summarizes a paper is labor-intensive and time-consuming. Although automating this task using ML models is promising for applications, research on scientific poster generation remains scarce due to the high complexity and multimodality of the task. Previous studies [26, 31] have built datasets to evaluate poster generation systems for scientific papers. However, these datasets are either not publicly available or the data license is unclear, leaving this research field without gold-standard benchmarks.

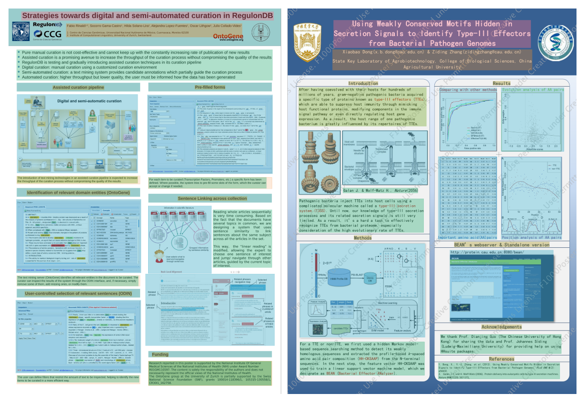

Figure 1. Example posters and annotations in SciPostLayout.

Previous studies [4, 35] have published publicly available datasets for layout analysis [7, 11, 22, 35] and layout generation [4, 6, 16, 32, 34]. However, these datasets focus on scientific papers or mobile application design, whereas the layout of scientific posters has not been studied yet. Layout analysis and layout generation are important tasks for identifying the characteristics of effective layout design in a particular domain. Developing layout analysis models helps to identify layout patterns common to well-designed posters. In addition, by developing layout generation models, we can generate layouts that follow the revealed patterns of attractive posters. In other words, building a layout dataset of scientific posters that can be used for these tasks would contribute greatly to studying scientific poster generation.

In this paper, we introduce SciPostLayout, the first scien-

*Work done during internship at OMRON SINIC X.

tific poster layout dataset for layout analysis and generation. The dataset includes 7,855 scientific posters with manual layout annotations. All posters included in the dataset are under the CC-BY license [1]. Figure 1 shows examples of posters and annotations of SciPostLayout. SciPostLayout can be used to evaluate layout analysis and generation systems in the same way as existing datasets. The layout analysis and generation on SciPostLayout are more challenging than scientific papers because of the various positions of elements such as figures and tables.

We evaluated existing models for layout analysis and generation on SciPostLayout. In the layout analysis task, we found that although the existing models can recognize certain elements with high accuracy, they are less effective than on the scientific paper dataset, indicating that layout analysis in SciPostLayout is more complicated. In the layout generation task, a diffusion-based model can generate layouts most similar to existing layouts, but we found that layout generation on SciPostLayout is also a more challenging task compared to generation on the scientific paper dataset.

2. Related Work

In this section, we describe related work on layout analysis and layout generation.

2.1. Layout Analysis

Layout analysis is a task in which a system detects the layouts of unstructured documents by predicting bounding boxes and categories such as tables or figures. This task is generally categorized as an object detection problem [7, 22, 35].

The PubLayNet dataset [35] is one of the well-known layout analysis datasets. This dataset contains scientific paper images automatically annotated with bounding boxes and polygonal segmentation across five categories: text, title, list, figure, and table. Previous studies have demonstrated that deep neural networks, including UDoc [7], DiT [22], and LayoutLMv3 [11], which are trained on the PubLayNet dataset, can recognize the layouts of scientific papers with high accuracy. These models use Faster R-CNN [28], Mask R-CNN [10], and Cascade R-CNN [2] as object detection models and ResNet [9] and Vision Transformer [5] as visual backbones.

Other studies have also proposed layout analysis datasets, such as Article Regions [29], which is a dataset of region-annotated scientific papers from PubMed Central, TableBank [23], an image-based table detection dataset built with Word and LaTeX documents, and DocBank [24], a document-level dataset with token annotations.

[1] https://creativecommons.org/licenses/by/4.0/, some of the older posters are under older licenses, such as CC-BY 2.0 and CC-BY 3.0.

In contrast to the previous datasets, SciPostLayout proposed in this study focuses on scientific posters. The layout analysis of scientific posters is more challenging than that of scientific papers because of the variety of fonts and positions of figures and tables.

2.2. Layout Generation

Layout generation is a task that arose from the needs of design applications, including magazine covers, UI interfaces, presentation slides, and banner advertising [4, 6, 16, 32, 34]. Layout generation can be categorized into unconditional generation [1, 8, 14], which generates layouts without any constraints from the user, and conditional generation, which enables the user to create their desired layout with constraints. There are also several subcategories within conditional generation, each giving the system different constraints for generation, such element types [17–19], element types and sizes [18], relationships between elements [17, 19], completion [8], and refinement [27].

Early studies explored generative adversarial networks (GANs) [17, 20] and variational autoencoders (VAEs) [1, 14, 32] for layout generation. Motivated by the success of Transformer architecture [30] in NLP, some studies treated layout generation as a sequence-to-sequence problem [8, 15, 18, 25]. Moreover, discrete diffusion models have achieved notable results [3, 12, 13, 33]. Datasets including Rico [4], Magazine [34], and PubLayNet [35] are used to evaluate layout generation models.

We evaluate existing models for layout generation in the scientific poster domain. Generating layouts for scientific posters is challenging, as are other layout generation tasks that require the positioning of figures and tables.

3. Dataset

Previous studies [26, 31] have built datasets to evaluate poster generation systems for scientific papers. However, these datasets are either not publicly available or the data licensing is unclear. In this paper, we construct SciPostLayout, a fully public dataset with all data under the Creative Commons license, which allows unrestricted dissemination, adaptation, and re-use.

First, we downloaded posters in PDF format from F1000Research [2]. Among these, we kept 7,943 posters under the CC-BY license. Posters under non-distributable and noncommercial licenses were excluded. The PDF files were converted into PNG format at DPI=100 for the following annotations. We excluded posters with file sizes below 200KB as they mainly consisted of text, which was unsuitable for layout analysis. After this exclusion, 7,855 posters remained. We investigated the word trends in the titles and

[2] https://f1000research.com/browse/posters

Category	Contents
Title	paper title
Author Info	author, author affiliation
Section	section title[a]
Text	paragraph[b]
List	nested list[c]
Table	main body of table
Figure	main body of figure[d]
Caption	caption of table and figure
Unknown	advertising information, logo of affiliation[e]

[a] Only the highest level of sections are annotated as Section. Subsections are annotated as Text because of the small font size.
[b] When multiple paragraphs are connected, they are annotated as a single object. Footnotes are ignored during annotation.
[c] Including reference block.
[d] When sub-figures exist, the whole figure panel is annotated as a single object.
[e] Only elements with big areas are annotated.

Table 1. Annotation criteria of SciPostLayout.

found that most of the collected posters were in the biomedical field.

Next, we recruited professional data annotators to manually annotate the document layout of the posters. The layouts of scientific posters are more diverse than the layouts of the papers in PubLayNet [35] because of the variety of fonts, size and position of figures and tables, and typography. We expanded PubLayNet's five-category annotation standard to nine categories to acquire fine-grained annotations of the layouts. The annotation criteria are shown in Table 1. Finally, we obtained our dataset SciPostLayout, which includes 7,855 posters and manual layout annotations. Table 2 shows the statistics of the dataset. The train, dev, and test sets contain 7,755, 50, and 50 posters, respectively. SciPostLayout is the first scientific poster layout dataset that can be used for evaluating layout analysis and layout generation systems, similar to Rico [4] and PubLayNet [35].

4. Experiments

Using the collected dataset, we conducted two experiments: layout analysis and layout generation. We compared existing models for each experiment.

4.1. Layout Analysis

We used LayoutLMv3 [11] and DiT [22] for the layout analysis task. We started from the base size checkpoints with Cascade R-CNN detectors and fine-tuned on SciPostLayout train data for both models. The checkpoint with the highest performances on the dev set was used for evaluation.

We measured the performance using the mean average precision (mAP) @ intersection over union (IoU) [0.50:0.95] of bounding boxes, the results of which are reported in Table 3. The *Unknown* category was omitted due to an insufficient number of elements. LayoutLMv3 outperformed DiT in all categories except for *Table*. In addition, both models showed high performances in the *Title* and *Author Info* categories. We attribute this result to the regularity of the *Title* and *Author Info* blocks since they are always at the top of the posters, and there is usually only one of each block per poster. However, compared to the results on PubLayNet [11, 22], in which mAP@IoU is over 90, both models showed a performance drop, indicating the complexity of our dataset.

4.2. Layout Generation

We conducted layout generation experiments with various settings for the information to be input into the model [15]. Generation conditioned on types (Gen-T) aims to generate layouts from the number of categories. Generation conditioned on types and sizes (Gen-TS) aims to generate layouts from the number and size of categories. Generation conditioned on relationships (Gen-R) aims to generate layouts from the number of categories and position relationships between the categories. Completion means generating a complete layout from a part of the layout. Refinement means generating a new layout from a layout that needs improvement.

We used LayoutDM [13], LayoutFormer++ [15], and LayoutPrompter [25] for the layout generation task. Unlike the layout analysis models, the layout generation models were trained from randomly initialized parameters. We did not train LayoutPrompter because it uses GPT-4 via API.

We evaluated the model performances using maximum IoU (mIoU), Alignment, Overlap, and Fréchet Inception Distance (FID). mIoU is a measure of the highest IoU between a generated layout and a real layout [17, 25]. Alignment indicates how well the elements in a layout are aligned with each other [21, 25]. Overlap is the overlapping area between two arbitrary elements in a layout [21, 25]. FID measures how similar the distribution of the generated layouts is to that of real layouts [17, 25]. A higher mIoU value means higher performance; for the other metrics, a lower value means higher performance. Note that both mIoU and FID are evaluation metrics based on similarity to the real layouts, but mIoU is calculated from intersection between layouts, while FID is an embedding-based evaluation metric. Thus, the hierarchical order of mIoU and FID performances across models is not always consistent.

The results are shown in Table 4. mIoU was low for all models, less than half that in PubLayNet. In contrast, all models performed effectively on Alignment, indicating that they can generate aligned layouts. LayoutPrompter was the most effective for Overlap, indicating that it generates layouts with the least overlap. LayoutDM was the most effective in terms of FID, indicating that it can generate layouts most similar to real layouts on each setting. Layout-

Split	Title	Author Info	Section	Text	List	Table	Figure	Caption	Unknown	Total
Train	7,742	7,595	40,793	52,076	23,523	5,589	38,034	14,872	10	190,234
Dev	50	50	273	308	109	30	259	90	1	1,170
Test	50	50	272	327	163	26	277	128	0	1,293

Table 2. Statistics of train, dev, and test data in SciPostLayout.

Model	Title	Author Info	Section	Text	List	Table	Figure	Caption	Overall
LayoutLMv3	**87.32**	**85.21**	**65.45**	**65.80**	**65.93**	64.97	**71.48**	**53.88**	**70.00**
DiT	84.99	84.49	63.98	59.07	58.28	**68.86**	70.82	50.14	67.58

Table 3. Layout analysis performance (mAP@IoU[0.50:0.95]) on SciPostLayout test set. **Bold** numbers indicate highest performances.

Task	Model	mIoU↑	Alignment↓	Overlap↓	FID↓
Gen-T	LayoutDM	0.071	0.001	0.534	**3.933**
	LayoutFormer++	0.074	0.001	**0.224**	23.839
	LayoutPrompter	**0.086**	**0.000**	0.283	11.664
Gen-TS	LayoutDM	**0.114**	0.001	0.607	**3.876**
	LayoutFormer++	0.098	0.002	0.230	22.851
	LayoutPrompter	0.102	**0.000**	**0.172**	12.385
Gen-R	LayoutDM	0.063	0.001	0.731	**6.427**
	LayoutFormer++	0.060	0.002	1.153	30.837
	LayoutPrompter	**0.068**	**0.000**	**0.311**	12.117
Completion	LayoutDM	-	0.001	0.568	**5.824**
	LayoutFormer++	-	0.001	0.225	12.383
	LayoutPrompter	-	**0.000**	**0.012**	16.779
Refinement	LayoutDM	0.115	**0.001**	0.537	3.650
	LayoutFormer++	0.331	**0.001**	0.221	11.807
	LayoutPrompter	**0.536**	**0.001**	**0.127**	**0.780**

Table 4. Layout generation performance on SciPostLayout test set. **Bold** numbers indicate the highest performances. ↑ indicates larger values are ideal, ↓ indicates smaller values are ideal. mIoU is omitted from the Completion setting because the number of elements of generated layouts differs from that of real layouts.

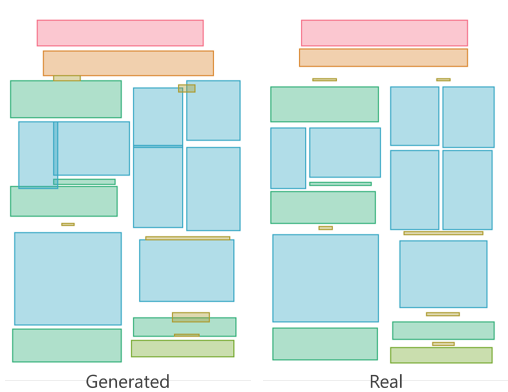

Figure 2. Example of LayoutPrompter generated layout and real layout in Refinement setting.

Prompter outperformed the other models in the Refinement setting, indicating that it can generate layouts similar to real layouts from noisy layouts. Figure 2 shows an example of a LayoutPrompter generated layout and a real layout in the Refinement setting.

5. Conclusion

In this paper, we built a new dataset called SciPostLayout, which consists of 7,855 scientific posters downloaded from a website. All posters in SciPostLayout are manually annotated with categories of layout elements, such as titles and figures, to be used in layout analysis and generation tasks. In addition, SciPostLayout is available for commercial research because all of the posters are under the CC-BY license. We conducted layout analysis and generation experiments to evaluate the performances of existing models on SciPostLayout. For layout analysis, although some elements could be recognized with high accuracy, we found that layout analysis on SciPostLayout was more difficult than on a scientific paper dataset. For layout generation, although existing models could generate aligned layouts, we found it difficult to generate layouts that are similar to real layouts. Our future work will involve developing a model architecture to improve layout analysis and generation. In addition, we will investigate end-to-end methods for generating posters from scientific papers.

Acknowledgements

This work was supported by JST Moonshot R&D Program, Grant Number JPMJMS2236. We would like to thank Naoto Inoue, Ph. D. of CyberAgent Inc. AILab for useful comments about layout generation models.

References

[1] Diego Martin Arroyo, Janis Postels, and Federico Tombari. Variational transformer networks for layout generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13642–13652, 2021. 2

[2] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018. 2

[3] Shang Chai, Liansheng Zhuang, and Fengying Yan. LayoutDM: Transformer-based Diffusion Model for Layout Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18349–18358, 2023. 2

[4] 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, 2, 3

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[6] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. DOC2PPT: automatic presentation slides generation from scientific documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 634–642, 2022. 1, 2

[7] Jiuxiang Gu, Jason Kuen, Vlad I Morariu, Handong Zhao, Rajiv Jain, Nikolaos Barmpalios, Ani Nenkova, and Tong Sun. Unidoc: Unified pretraining framework for document understanding. *Advances in Neural Information Processing Systems*, 34:39–50, 2021. 1, 2

[8] 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. 2

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2

[11] Yupan Huang, Tengchao Lv, Lei Cui, Yutong Lu, and Furu Wei. LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2022. 1, 2, 3

[12] Mude Hui, Zhizheng Zhang, Xiaoyi Zhang, Wenxuan Xie, Yuwang Wang, and Yan Lu. Unifying Layout Generation with a Decoupled Diffusion Model. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1942–1951, 2023. 2

[13] Naoto Inoue, Kotaro Kikuchi, Edgar Simo-Serra, Mayu Otani, and Kota Yamaguchi. LayoutDM: Discrete Diffusion Model for Controllable Layout Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10167–10176, 2023. 2, 3

[14] Zhaoyun Jiang, Shizhao Sun, Jihua Zhu, Jian-Guang Lou, and Dongmei Zhang. Coarse-to-fine generative modeling for graphic layouts. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1096–1103, 2022. 2

[15] Zhaoyun Jiang, Jiaqi Guo, Shizhao Sun, Huayu Deng, Zhongkai Wu, Vuksan Mijovic, Zijiang James Yang, Jian-Guang Lou, and Dongmei Zhang. LayoutFormer++: Conditional Graphic Layout Generation via Constraint Serialization and Decoding Space Restriction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18403–18412, 2023. 2, 3

[16] K. Kikuchi, M. Otani, K. Yamaguchi, and E. Simo-Serra. Modeling Visual Containment for Web Page Layout Optimization. *Computer Graphics Forum*, 40(7):33–44, 2021. 1, 2

[17] Kotaro Kikuchi, Edgar Simo-Serra, Mayu Otani, and Kota Yamaguchi. Constrained Graphic Layout Generation via Latent Optimization. In *ACM International Conference on Multimedia*, pages 88–96, 2021. 2, 3

[18] Xiang Kong, Lu Jiang, Huiwen Chang, Han Zhang, Yuan Hao, Haifeng Gong, and Irfan Essa. BLT: bidirectional layout transformer for controllable layout generation. In *European Conference on Computer Vision*, pages 474–490. Springer, 2022. 2

[19] 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, Cham, 2020. Springer International Publishing. 2

[20] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. Layoutgan: Generating graphic layouts with wireframe discriminators. *arXiv preprint arXiv:1901.06767*, 2019. 2

[21] Jianan Li, Jimei Yang, Jianming Zhang, Chang Liu, Christina Wang, and Tingfa Xu. Attribute-Conditioned Layout GAN for Automatic Graphic Design. *IEEE Transactions on Visualization and Computer Graphics*, 27(10): 4039–4048, 2021. 3

[22] Junlong Li, Yiheng Xu, Tengchao Lv, Lei Cui, Cha Zhang, and Furu Wei. DiT: Self-supervised Pre-training for Document Image Transformer. In *ACM Multimedia*, 2022. 1, 2, 3

[23] Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming

Zhou, and Zhoujun Li. TableBank: A Benchmark Dataset for Table Detection and Recognition, 2019. 2

[24] Minghao Li, Yiheng Xu, Lei Cui, Shaohan Huang, Furu Wei, Zhoujun Li, and Ming Zhou. DocBank: A Benchmark Dataset for Document Layout Analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 949–960, Barcelona, Spain (Online), 2020. International Committee on Computational Linguistics. 2

[25] Jiawei Lin, Jiaqi Guo, Shizhao Sun, Zijiang James Yang, Jian-Guang Lou, and Dongmei Zhang. LayoutPrompter: Awaken the Design Ability of Large Language Models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 2, 3

[26] Yu-Ting Qiang, Yan-Wei Fu, Xiao Yu, Yan-Wen Guo, Zhi-Hua Zhou, and Leonid Sigal. Learning to Generate Posters of Scientific Papers by Probabilistic Graphical Models. *Journal of Computer Science and Technology*, 34(1):155–169, 2019. 1, 2

[27] Soliha Rahman, Vinoth Pandian Sermuga Pandian, and Matthias Jarke. Ruite: Refining ui layout aesthetics using transformer encoder. In *26th International Conference on Intelligent User Interfaces-Companion*, pages 81–83, 2021. 2

[28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2015. 2

[29] Carlos Soto and Shinjae Yoo. Visual Detection with Context for Document Layout Analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3464–3470, Hong Kong, China, 2019. Association for Computational Linguistics. 2

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[31] Sheng Xu and Xiaojun Wan. Neural Content Extraction for Poster Generation of Scientific Papers, 2021. 1, 2

[32] Kota Yamaguchi. Canvasvae: Learning to generate vector graphic documents. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5481–5489, 2021. 1, 2

[33] Junyi Zhang, Jiaqi Guo, Shizhao Sun, Jian-Guang Lou, and Dongmei Zhang. LayoutDiffusion: Improving Graphic Layout Generation by Discrete Diffusion Probabilistic Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 2

[34] Xinru Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics*, 38(4):1–15, 2019. 1, 2

[35] Xu Zhong, Jianbin Tang, and Antonio Jimeno Yepes. PubLayNet: largest dataset ever for document layout analysis. In *International Conference on Document Analysis and Recognition*, pages 1015–1022. IEEE, 2019. 1, 2, 3