SciPostGen: Bridging the Gap between Scientific Papers and Poster Layouts

Supplementary Material

A. Dataset Details

Overview. Figure 1 illustrates an example of the annotated components in SciPostGen, a dataset comprising 18,097 pairs of scientific papers and their corresponding posters.

In the main text, we focused on the paper content annotations (e.g., OCR text and figure/table bounding boxes) and the poster layout annotations. In practice, however, SciPostGen also includes automatically derived poster content annotations. We omitted these details from the main text to maintain clarity, as they are not directly used in our poster layout generation task. These poster OCR texts were extracted automatically using Tesseract[1], an open-source OCR engine.

Resources. SciPostGen is constructed from open-access papers and their associated posters released by four machine learning and computer vision conferences, as follows:

- CVPR: https://cvpr.thecvf.com/
- ICLR: https://iclr.cc
- ICML: https://icml.cc
- NeurIPS: https://neurips.cc

Comparison with Existing Paper–Poster Datasets. Table 1 summarizes existing datasets that pair scientific papers with their corresponding posters or poster layouts. While SciPostLayout contains 7,855 poster layouts in total, only 100 of them are paired with their corresponding papers, and thus only these pairs are included in the comparison. SciPostGen is larger than prior datasets, offering a more substantial foundation for benchmarking poster layout generation and potentially supporting broader poster generation tasks.

Additional Statistics. Table 2 reports additional statistics of SciPostGen, which consists of 15,710, 399, and 1,988 pairs in the train, valid, and test splits, respectively. Word and unique-word statistics are computed from OCR text after lowercasing.

Correlation Analysis on the Train Split. To complement the analyses in the main text, we additionally report the correlation results on the SciPostGen train split. This allows us to examine whether the trends observed in the gold layouts (test split) are consistent with those derived from the silver layouts (train split).

Figure 2a shows the correlation results between paper and layout features on the train split. Overall, the trends are consistent with those observed in Section 3.3 based on the gold layouts.

Figure 2b shows the correlation results between layout features on the train split. Compared with the gold layout results in Section 3.3, the silver layouts yield slightly different correlation strengths for certain elements. In particular, caption–figure correlations increase ($\rho = 0.28$), and section–text correlations become moderately positive ($\rho = 0.48$). These differences may reflect noise in the automatically extracted caption and section annotations.

[1] https://github.com/tesseract-ocr/tesseract

Dataset	#Paper-Poster Pairs
NJU-Fudan [5]	85
Paper2Poster [4]	100
P2P eval [7]	121
SciPostLayout [8]	100
SciPostGen	18,097

Table 1. Comparison of datasets pairing scientific papers with corresponding posters or poster layouts

Paper	Train	Valid	Test
# Sections	109,418	2,818	14,059
# Figures	81,465	2,079	10,572
# Tables	56,182	1,459	7,149
Total Chars. (M)	470.5	11.8	59.2
Total Words (M)	102.4	2.5	12.7
Uniq. Words (k)	314.2	39.0	98.8

(a) Paper statistics

Poster	Train	Valid	Test
# Sections	128,263	2,155	10,605
# Figures	88,773	2,317	12,511
# Tables	25,825	661	3,436
Total Chars. (M)	56.1	1.4	7.3
Total Words (M)	14.0	0.3	1.8
Uniq. Words (k)	278.7	27.6	78.1

(b) Poster statistics

Table 2. Statistics of SciPostGen

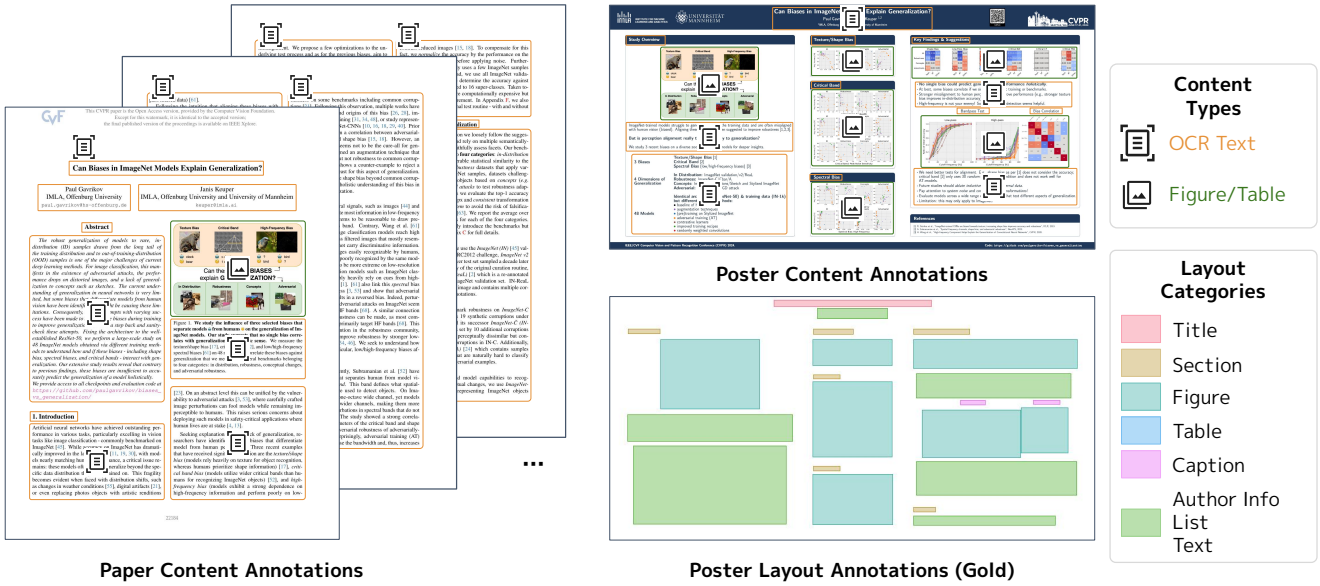
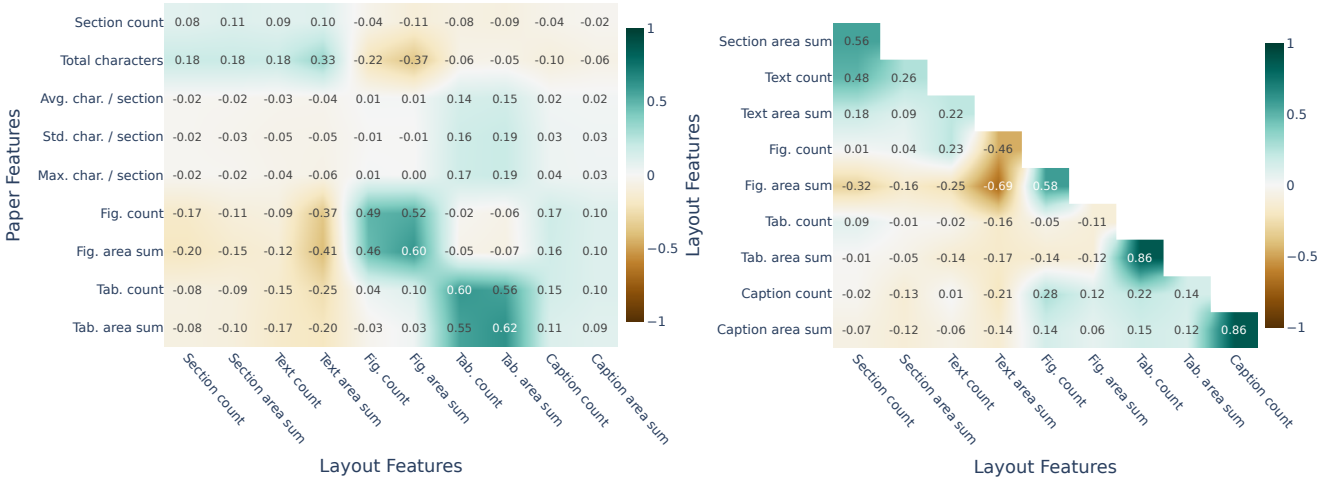


Figure 1. Example of annotations in SciPostGen, including automatically extracted paper and poster annotations and manually corrected poster layout annotations: Paper and poster from [1], licensed under CC BY-SA 4.0.



(a) Spearman's ρ between paper and layout features

(b) Spearman's ρ between layout features

Figure 2. Correlation analyses between paper structures and poster layouts in SciPostGen train split

B. Experimental Details

Split Usage. The train split was used to train the retrieval module. During inference, we queried the trained retriever with papers from the test split and retrieved candidate poster layouts from the train split. The valid split was used both for epoch-wise evaluation and hyperparameter selection during the retriever training, and for ranking layout candidates in the generator.

Data Preprocessing. Because the train split does not include the "Author Info" and "List" categories, we merged them into the "Text" category to ensure consistency with the valid and test splits. This unified annotation scheme was used both for training the retriever and for evaluating our framework.

Implementation Details. We use DiT-base[2] as the backbone encoder for both paper pages and layout images. Figure 3 illustrates the architecture of the paper encoder, which applies patch-level and page-level pooling over the backbone outputs to obtain a paper embedding. The layout encoder follows a similar structure but does not include page-level pooling. Table 3 summarizes the number of trainable parameters in the retriever, which contains 180M parameters in total.

The retriever settings are as follows:

- Paper and layout image sizes: $H = W = 224$
- Number of paper pages: $n_p = 8$
- Paper and layout embedding dimension: $d = 256$

We trained the retriever on 4×NVIDIA A100 GPUs in 12 hours by AdamW [2]. The training hyperparameters are as follows:

- Batch size: $N = 128$
- Temperature parameter: $\tau = 0.07$
- Number of epochs: 20
- Learning rate: $\{1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$
- Weight decay: 0.01
- Scheduler: Linear Warmup Cosine Annealing
- Warmup ratio: 0.1

We used two large language models for layout generation: GPT-5-mini[3] and GPT-5[4]. For cost considerations, we set the reasoning effort to low for GPT-5-mini and minimal for GPT-5. Under these settings, the API cost for generating layouts for the 1,988 samples in the SciPostGen test split was $8–10 with GPT-5-mini and $35–40 with GPT-5.

Table 4 and Table 5 show the example prompts used in the automatic and semi-automatic poster generation settings, respectively. In both settings, the generator receives the following paper structures:

- Number of sections, captions, figures, and tables
- Characters of title, authors, abstract, and each section
- Aspect ratios of figures and tables

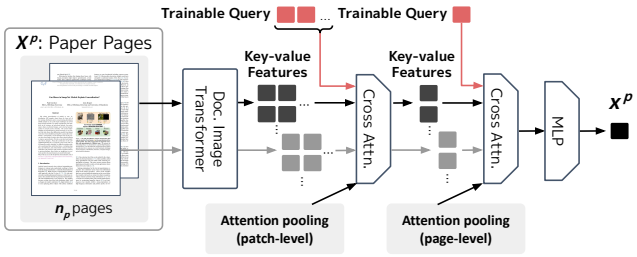


Figure 3. Detailed architecture of the paper encoder

Component	Backbone	Other Params	Total
Paper Encoder	85M	6M	91M
Layout Encoder	85M	4M	89M

Table 3. Number of trainable parameters in each encoder of the retriever: The retriever contains 180M trainable parameters in total

[2] dit-base: https://huggingface.co/microsoft/dit-base

[3] gpt-5-mini-2025-08-07: https://platform.openai.com/docs/models/gpt-5-mini

[4] gpt-5-2025-08-07: https://platform.openai.com/docs/models/gpt-5

Prompt for Automatic Generation with Paper Structure

```
Task: Layout generation conditioned on paper statistics.
Please use layout element types: Title, Section, Text, Figure, Table, Caption.

Here is example 1.
Paper Info:
    #Paper element counts
        Section: 9, Caption: 6, Figure: 7
    #Paper element lengths in characters
        Title length: 69, Author length: 625, Abstract length: 961, 1 Introduction length: 7242 ...
    #Paper figure/table aspect ratios (width/height)
        Figure 1: 1.92, Figure 2: 4.86 ...
Input:
    <html><body>
        <div class='canvas' style='left: 0px; top: 0px; width: 5120px; height: 2560px'></div>
    </body></html>
Output:
    <html><body>
        <div class='canvas' style='left: 0px; top: 0px; width: 3456px; height: 2304px'></div>
        <div class='Title' style='left: 57px; top: 61px; width: 3306px; height: 76px;'></div>
        ...
    </html></body>

Here is example 2.
...

Please generate a layout.
Paper Info: ...
Input: ...
Output:
```

Table 4. Example prompt used in the automatic poster generation setting

Prompt for Semi-Automatic Generation with Paper Structure and Layout Constraints

```
Task: Layout completion conditioned on partial layout and paper statistics.
Please use layout element types: Title, Section, Text, Figure, Table, Caption.

Here is example 1.
Paper Info:
    #Paper element counts
        Section: 9, Caption: 6, Figure: 7
    #Paper element lengths in characters
        Title length: 69, Author length: 625, Abstract length: 961, 1 Introduction length: 7242 ...
    #Paper figure/table aspect ratios (width/height)
        Figure 1: 1.92, Figure 2: 4.86 ...
Input:
    <html><body>
        <div class='canvas' style='left: 0px; top: 0px; width: 5120px; height: 2560px'></div>
        <div class='Figure' style='left: 1522px; top: 1306px; width: 1761px; height: 778px;'></div>
        <div class='Text' style='left: 57px; top: 1944px; width: 2298px; height: 544px;'></div>
    </body></html>
Output:
    <html><body>
        <div class='canvas' style='left: 0px; top: 0px; width: 5120px; height: 2560px'></div>
        <div class='Title' style='left: 1265px; top: 26px; width: 2561px; height: 299px;'>
        ...
        <div class='Figure' style='left: 1522px; top: 1306px; width: 1761px; height: 778px;'></div>
        <div class='Text' style='left: 57px; top: 1944px; width: 2298px; height: 544px;'></div>
        ...
    </html></body>

Here is example 2.
...

Please generate a layout.
Paper Info: ...
Input: ...
Output:
```

Table 5. Example prompt used in the semi-automatic poster generation setting

C. Additional Results

Comparison under the Oracle Setting. We compare the predicted layouts with the oracle setting, which consists of the gold layouts and retriever upper bounds computed for each test sample by selecting the best-matching silver layout from the training pool in terms of mIoU or TC_{std}. For reference-free evaluation, we report Overlap and Alignment. Overlap measures the extent to which layout elements undesirably intersect, while Alignment measures how well their edges and centers are aligned.

Figure 6 shows that the layouts generated by GPT-5 show higher Alignment scores while achieving lower Overlap than the retrieved layouts. The gold layouts, however, do not show high Alignment (0.065), suggesting that strong layout alignment is not necessarily a characteristic of human-designed scientific posters.

The retriever upper bound corresponds to 0.347 in mIoU and 0.390 in TC_{std}, computed by selecting the best-matching training layout for each test sample. Although our predicted layouts do not reach this upper bound, these oracle values show the potential of Retrieval-Augmented Poster Layout generation.

Ablation Study of the Retriever. Figure 7 shows the effect of replacing layout images with full poster canvases as the input to the layout encoder. We observe that using posters results in consistently lower performance across all metrics compared with using the layout images. This suggests that visual elements such as colors and decorative styling do not help the retriever estimate layouts that align with the corresponding papers.

We reduce the number of input pages n_p from the default 8 to 6, 4, and 2, training and evaluating the retriever for each setting. Figure 4a shows that retrieval performance consistently improves as n_p increases: mIoU increases while TC_{std} decreases. These results indicate that incorporating more pages is beneficial for retrieval and suggest that the attention pooling components effectively aggregate global information across the entire paper.

Figure 4b shows how retrieval performance changes as the pool size is reduced, based on the retriever trained with $n_p = 8$. The mIoU remains stable across different pool sizes, whereas TC_{std} improves as the pool size increases but saturates around a pool size of 10k.

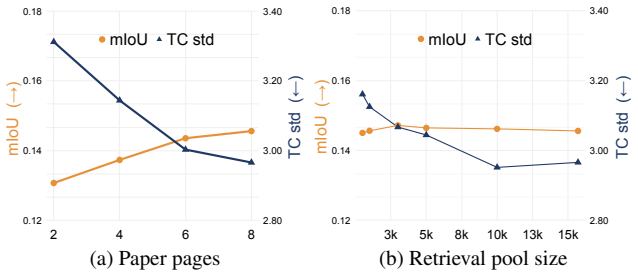
Failure Cases of the Retriever. Figure 5 shows failure cases of the retriever. Such cases often involve layouts that span multiple columns or adopt unique element arrangements. Handling these layouts requires combining retrieval with the generator rather than relying on the retriever alone.

Approach	Overall	Alignment†	mIoU ↑	TC_{std} ↓
Unconditional Setting w/ GPT-5				
Retrieved Top-3 (Avg.)	0.027	0.077	0.145	2.965
↪ Paper Structures	0.065	0.001	0.159	3.128
Conditional Setting w/ GPT-5				
Retrieved Top-3 (Avg.)	0.026	0.075	0.196	2.873
↪ Paper Structures	0.037	0.011	0.238	2.941
Retriever Upper Bound	—	—	0.347	0.390
Gold Layout	0.003	0.065	—	—

Table 6. Comparison of retrieved layouts, retriever upper bound, and gold layouts: "Overall" and "Alignment" metrics are better when their values are closer to those of the gold layouts. Values in the columns marked with † are scaled by a factor of 100.

Poster Input	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Layout	0.145	0.651	-0.057	2.965
Poster	0.137	0.646	-0.705	3.456

Table 7. Results of retriever training with different poster inputs



(a) Paper pages (b) Retrieval pool size

Figure 4. Ablation results for the retriever: (a) Effect of changing the number of input paper pages n_p by training the retriever. (b) Effect of varying the retrieval pool size during inference ($n_p = 8$)

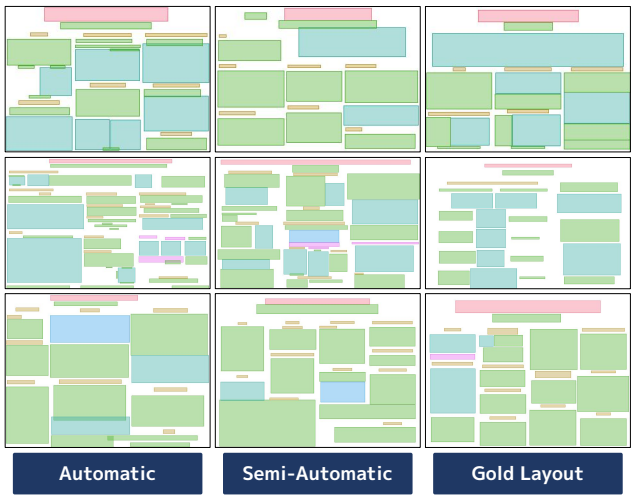


Figure 5. Failure cases of the retriever: we show the retrieved top-1 layouts under the automatic and semi-automatic settings

Comparison Across LLM Generators. Tables 8 and 9 compare different open-source LLM generators: GPT-OSS-120b[5], Qwen3-32B[6], and Qwen3-Coder-30B[7], in the automatic and semi-automatic settings [3, 6]. We observe that performance varies across models in the automatic setting, indicating that generation quality depends on the underlying LLM. In the semi-automatic setting, the best-performing open-source LLMs (Qwen3-32B and Qwen3-Coder-30B) achieve mIoU values close to GPT-5-mini. This suggests that the evaluated open-source LLMs can effectively follow user-specified layout constraints.

Ablation Study on Paper Structures. Table 10 shows the effect of different paper structures in the automatic setting, where GPT-5 is used as the generator. We observe that using all structures together yields higher mIoU than using any single structure, suggesting that these structures provide complementary information.

D. Limitations

We acknowledge that SciPostGen is limited to computer science papers and primarily landscape-format posters, which may restrict the diversity of represented domains and layout styles. Future extensions to encompass broader research domains and poster orientations could mitigate this limitation. Our framework focuses on layout generation and does not address multimodal summarization. We also do not include comparisons with existing layout generation models, since generating layouts from lengthy and structured scientific papers remains an exploratory problem, and existing models are not directly applicable without substantial adaptation.

[5] https://huggingface.co/openai/gpt-oss-120b
[6] https://huggingface.co/Qwen/Qwen3-32B
[7] https://huggingface.co/Qwen/Qwen3-Coder-30B-A3B-Instruct

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.145	0.651	-0.057	2.965
↪ w/o Paper Structures	0.133	0.638	1.484	2.816
↪ w/ Paper Structures	0.139	0.630	0.444	2.905

(a) **Qwen3-32b**

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.145	0.651	-0.057	2.965
↪ w/o Paper Structures	0.138	0.646	0.946	2.715
↪ w/ Paper Structures	0.131	0.645	0.095	2.826

(b) **Qwen3-Coder-30b**

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.145	0.651	-0.057	2.965
↪ w/o Paper Structures	0.120	0.617	1.596	3.272
↪ w/ Paper Structures	0.127	0.612	-0.175	3.311

(c) **GPT-OSS-120b**

Table 8. Performance comparison across different LLM generators in automatic generation setting

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.196	0.662	0.194	2.873
↪ w/o Paper Structures	0.214	0.654	1.185	2.598
↪ w/ Paper Structures	0.206	0.650	0.362	2.823

(a) **Qwen3-32b**

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.196	0.662	0.194	2.873
↪ w/o Paper Structures	0.214	0.660	0.979	2.650
↪ w/ Paper Structures	0.197	0.658	0.178	2.732

(b) **Qwen3-Coder-30b**

Approach	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.196	0.662	0.194	2.873
↪ w/o Paper Structures	0.199	0.631	2.119	3.050
↪ w/ Paper Structures	0.196	0.610	-0.674	3.816

(c) **GPT-OSS-120b**

Table 9. Performance comparison across different LLM generators in semi-automatic generation setting

Paper Input	mIoU ↑	LTSim ↑	TC_{mean} →	TC_{std} ↓
Retrieved Top-3 (Avg.)	0.145	0.651	-0.057	2.965
↪ Paper Structure	0.159	0.642	0.228	3.128
↪ Element Counts	0.152	0.643	-0.082	3.298
↪ Characters	0.156	0.656	0.785	2.694
↪ Aspect Ratios	0.148	0.631	0.638	3.527

Table 10. Results of framework with different paper inputs in automatic setting using GPT-5 as the generator.

References

[1] Paul Gavrikov and Janis Keuper. Can biases in imagenet models explain generalization? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22184–22194, 2024. 2

[2] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the 7th International Conference on Learning Representations*, 2019. 3

[3] OpenAI. gpt-oss-120b & gpt-oss-20b model card. arXiv:2508.10925, 2025. 6

[4] Wei Pang, Kevin Qinghong Lin, Xiangru Jian, Xi He, and Philip Torr. Paper2Poster: benchmarking multimodal poster generation from long-context papers. In *Proceedings of the 39th Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2025. 1

[5] 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

[6] Qwen Team. Qwen3 technical report. arXiv:2505.09388, 2025. 6

[7] Tao Sun, Enhao Pan, Zhengkai Yang, Kaixin Sui, Jiajun Shi, Xianfu Cheng, Tongliang Li, Wenhao Huang, Ge Zhang, Jian Yang, and Zhoujun Li. P2P: Automated paper-to-poster generation and fine-grained benchmark. arXiv:2505.17104, 2025. 1

[8] Shohei Tanaka, Hao Wang, and Yoshitaka Ushiku. SciPostLayout: a dataset for layout analysis and layout generation of scientific posters. In *Proceedings of the 35th British Machine Vision Conference*, 2024. 1