How Panel Layouts Define Manga: Insights from Visual Ablation Experiments

Siyuan Feng (9445233883@g.ecc.u-tokyo.ac.jp)[1], Teruya Yoshinaga (t-yoshinaga-3k6@eagle.sophia.ac.jp)[2], Katsuhiko Hayashi (katsuhiko-hayashi@g.ecc.u-tokyo.ac.jp)[1], Koki Washio (kkwashio3333@gmail.com)[3], Hidetaka Kamigaito (kamigaito.h@is.naist.jp)[4]

[1]The University of Tokyo, Department of Language and Information Sciences, Tokyo, Japan
[2]Sophia University, Graduate Degree Program of Applied Data Sciences, Tokyo, Japan
[3]Freelance Researcher, Tokyo, Japan
[4]Nara Institute of Science and Technology, Division of Information Science, Ikoma, Nara, Japan

Abstract

Manga has gained global popularity, yet how its visual elements, such as characters, text, and panel layouts, reflect the uniqueness of individual works remains underexplored. This study investigates the contribution of panel layouts to manga identity through both quantitative and qualitative analysis. We trained a deep learning model to classify manga titles based solely on facing page images, and performed ablation experiments by removing characters and text, retaining only panel frame structures. Using 10,122 images from 104 works in 12 genres in the Manga109 dataset, we demonstrate that panel layouts alone enable high-accuracy classification. Grad-CAM visualizations further reveal that the models focus on layout features such as size, spacing, and alignment. These findings suggest that panel layouts encode work-specific stylistic patterns and support visual narrative comprehension, highlighting their role as a key component of manga's visual identity.

Keywords: Artwork layout analysis; Digital comics; Comic classification; Deep learning; Visualization analysis

Introduction

Comic is a globally popular medium, encompassing a wide range of themes and genres. With the increasing adoption of e-books, the digital comic market has seen significant growth, and many comic works are now available in electronic formats. This shift has drawn considerable attention to multimedia processing research for digital comics, particularly Japanese comics (referred to simply as "manga" in this paper).

Manga, a distinctive form of Japanese comics, is characterized by its unique panel arrangements, which, along with creative character designs, guide the reader's eye through the story in a carefully orchestrated manner. Panel layouts play a crucial role in conveying the narrative, as emphasized by Shogakukan's "Newcomer Manga Award Manga Creator Training Course"[1], which outlines five fundamental rules for panel layout design. Manga artists can further enhance originality by creatively adhering to or intentionally breaking these rules to suit their storytelling needs. While readers may perceive subtle layout differences, it is often difficult to determine how these reflect the style of a work. This raises an important question: To what extent do panel layouts reflect the uniqueness of a manga work, and how do they contribute to its overall identity?

[1]https://shincomi.shogakukan.co.jp/training/004.html

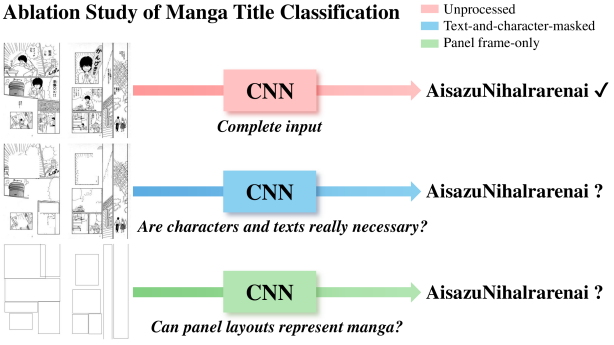


Figure 1: Brief description of the multi-class manga title classification experiment of the titles of the comics. The images shown in the figure are from AisazuNihalrarenai ©Masako Yoshi.

The significance of panel layouts has been acknowledged in multiple foundational works. For instance, Cohn (2013) emphasized the impact of page layout on reading order and compared the technical usage of panel arrangements between Japanese and Western comics. McCloud & Martin (1993) highlighted the role of panel-to-panel transitions and particularly explored how the spacing between panels (gutter) influences the flow of storytelling. Additionally, Natsume et al. (2020) delved deeply into the techniques and effects of panel layout in manga, offering insights into how layout decisions enhance the reader's experience. While these works underscore the importance of panel layouts, they do not delve into the extent to which panel layouts contribute to defining the unique identity of a manga work.

Despite these foundational studies, existing research on manga has predominantly focused on characters, text, and other visual elements, with limited exploration of panel layouts. While some studies leverage engineering approaches, such as image processing or machine learning, these often focus on practical applications rather than delving into the artistic or structural characteristics of manga. This imbalance leaves the role of panel layouts in reflecting the creative uniqueness of manga largely unexplored.

To address this gap, in this paper, we conducted visual ablation experiments. The experimental process is shown in Figure 1. We prepared three types of images: original unprocessed facing page images, text-and-character-masked

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

facing page images, and panel frame-only facing page images. These images were used as input for a deep learning model to perform a multi-class classification task on manga titles. The objective of this study was to analyze the changes in model performance by removing text and characters or retaining only panel layouts. If the model performs comparably well on processed images as it does on original unprocessed images, we believe this would validate the uniqueness of features such as manga panel layouts. By using panel frame-only facing page images, we aimed to evaluate the contribution of panel layout features from each work to the classification task. Subsequently, we employed Grad-CAM (Selvaraju et al., 2019) to visualize the key features in the input images and qualitatively interpret the results. Finally, we conducted additional experiments, including classification of manga publishers and genres, to rule out other factors influencing the uniqueness of panel layouts in manga works. We also performed classification experiments on noisy panel frames, further demonstrating that panel layouts are unique features of each manga work.

These results indicate that panel layouts possess inherent uniqueness, even in the absence of content such as characters or text. This suggests that layout structures carry stylistic information that may reflect the unique visual identity of each work. Although our approach is based on a classification task, the broader aim is to understand how layout alone encodes work-specific style and supports visual narrative interpretation. This line of inquiry is closely related to cognitive theories of attention and sequential image comprehension.

Related Work

Visual Analysis in Creative Works

Numerous studies have used images of creative works as input for predicting associated attributes such as titles or genres. Chu & Guo (2017) fine-tuned a pretrained CNN to classify movie genres based on poster images. Wi et al. (2020) applied a similar approach and visualized the model's focus using Grad-CAM to verify whether it extracted expected features. Godwin et al. (2018); Takahashi & Kondo (2023); Tüchler et al. (2021) examined how layout affects comprehension or user experience, including children's book design and web interfaces.

Several works also explored deep learning approaches in comic research. Laubrock & Dunst (2020); Augereau et al. (2018); Sharma & Kukreja (2024) surveyed computer vision applications in comics. Laubrock & Dubray (2019) conducted classification of comic authors and titles using CNNs on datasets like GNC (Dunst et al., 2017) and Manga109 (Fujimoto et al., 2016; Matsui et al., 2017). Dubray & Laubrock (2019) performed speech balloon detection, while Rigaud et al. (2021) worked on text segmentation inside speech balloons. Young-Min (2019) used CNN-based classification and feature visualization. Kukreja et al. (2023) applied CNN and Random Forest models to classify manga pages.

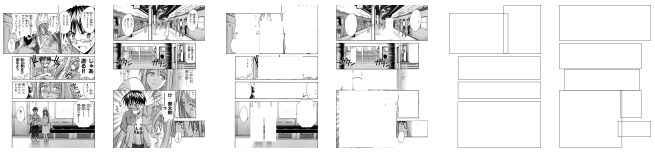
Existing Issues in Manga Research

While a growing number of studies have used manga page images, most focus on engineering applications, with limited attention to the underlying visual structure of manga. Laubrock & Dubray (2019) attempted to analyze manga features through classification, but their method treated pages as undivided wholes, without isolating individual components like panel layout. Young-Min (2019) proposed a CNN-based model for author classification and visualized learned features, but their analysis focused on stylistic textures and drawing patterns, rather than structural layout.

Although previous studies have explored page-level classification, none have specifically isolated panel layouts through ablation and visualization to analyze their contribution to manga identity. Although Kukreja et al. (2023) trained a CNN to classify genres, they did not consider whether each manga has unique stylistic traits. Their dataset was small and did not evaluate features like panel layout independently. In contrast, our work focuses on isolating panel layout features through ablation and visual interpretation, to investigate their role in distinguishing manga works.

Experiment

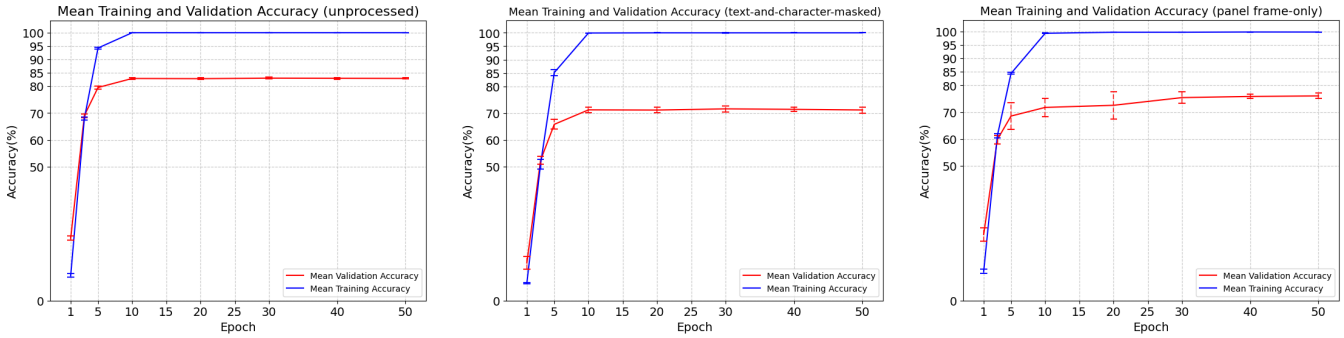
In the experiment, we analyze the characteristics of manga works by training multi-class classification models to predict the title of the work using manga facing page images as input. Specifically, we extracted 10,122 images containing panel information out of 10,602 facing page images obtained from Manga109 (Fujimoto et al., 2016; Matsui et al., 2017), and classified each image into 104 works based on information from `http://www.manga109.org/en/index.html`. The breakdown of genres includes 12 categories: 4-panel, animal, battle, fantasy, history, horror, humor, love, romantic comedy, SF, sport, and suspense. Table 1 provides an overview of the dataset.



(a) Unprocessed **(b)** Masked **(c)** Panel frame-only

Figure 2: Examples of images used in the experiment. The images shown in the figure are from LoveHina ©Ken Akamatsu.

In the experiments, we prepared unprocessed facing page images, masked facing page images, and facing page images with only the panel frames drawn. Figure 2a shows the original unprocessed images. Using annotation data from Manga109, we applied masking to both characters and text, as shown in Figure 2b. Additionally, we processed the images to retain only the panel frames, as shown in Figure 2c. By using panel frame-only facing page images, we aimed to evaluate how strongly the panel layout features of each work contribute to the classification task.



Unprocessed | Text-and-character-masked | Panel frame-only

Figure 3: Mean training and validation accuracy curves during model training, with standard deviation represented by error bars.

Table 1: Overview of the classification dataset consisting of 104 works constructed from Manga109: * indicates that different volumes of the same work are included, resulting in a total of 104 works.

Genre	Number of works	Number of pages with frames
4-panel	5	274
Animal	5	479
Battle	9	823
Fantasy	12	1,154
History	6	699
Horror	2	178
Humor	*15	1,391
Love	*13	1,198
Romantic comedy	*13	1,206
SF	14	1,294
Sport	*10	968
Suspense	5	458
Total	*109	10,122

Multi-class Classification for 104 Works

For data splitting, all page images were divided into a training set (80%), development set (10%), and test set (10%), resulting in 8,053, 1,011, and 1,058 images, respectively. To enhance generalization, 5-fold cross-validation was applied to the training set, with one part used as validation and the rest for training. The final model performance was evaluated on the test set after fine-tuning parameters using the development set.

We utilized ResNet101, a convolutional neural network (CNN) (He et al., 2015, 2016) pretrained on ImageNet (Russakovsky et al., 2015). Training used a mini-batch size of 32, cross-entropy loss, and Stochastic Gradient Descent (SGD) with a momentum factor of 90%. The initial learning rate was 0.001, reduced by 90% every 30 epochs. Training ran for up to 100 epochs, with classification experiments performed every epoch for the first 30 and every 10 epochs thereafter.

Figure 3 illustrates the classification accuracy of three models using different input types during the training process. Since the curves showed no significant changes after 50 epochs, only the first 50 epochs are displayed. This task is highly challenging, with a baseline classification accuracy of approximately 0.96% when randomly selecting one class from 104 categories.

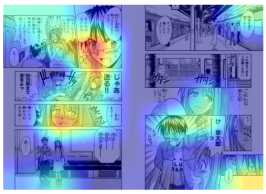
Testing on the test set was conducted using an ensemble voting approach with five models trained through 5-fold cross-validation. The final classification accuracies were 87.5% for unprocessed images, 79.7% for text-and-character-masked images, and 84.3% for panel frame-only images. Notably, the ResNet101 model, pretrained on the ImageNet dataset of natural images, achieved relatively high classification accuracy on manga data. This suggests that manga's visual features, particularly panel layouts, are sufficiently distinctive to be captured and leveraged by deep learning models, even those trained on vastly different domains.

Interestingly, the panel frame-only images yielded a relatively high accuracy of 84.3%, even though they lack explicit visual elements such as characters and text. This demonstrates that panel layouts themselves contain inherent distinguishing characteristics unique to each manga work. These findings quantitatively confirm that spatial relationships, such as panel size, spacing, and alignment, play a critical role in distinguishing manga works.

However, classification accuracy does not decrease consistently with increasing restrictions on input features. For instance, the higher accuracy of panel frame-only images compared to text-and-character-masked images suggests that the masking process inadvertently obscures some panel layout information. As shown in Figure 2, the Manga109 dataset annotates text and character locations as rectangles. Characters often appear near panel edges, and masking these characters frequently obscures panel lines. This phenomenon highlights the importance of preserving complete panel information for accurate classification, further validating the strong distinguishing characteristics of panel layout designs in manga.

Qualitative Analysis of Classification Results

To better understand the classification process, we conducted a qualitative analysis using Grad-CAM (Selvaraju et al., 2019), a visualization technique that highlights image regions important for the model's decision-making as heatmaps. This



Page 049 of LoveHina
©Ken Akamatsu



Page 016 of Arisa
©Ken Yagami

Figure 4: Grad-CAM heatmaps with unprocessed images. The model correctly classified input images and exhibited strong feature focus on areas where characters were drawn.

method reveals how the model interprets input types and identifies key features for classification.

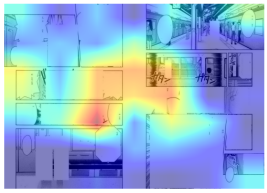
Figure 4 shows Grad-CAM heatmaps for unprocessed images of two correctly classified manga works, where the model focuses primarily on characters, reaffirming their importance as distinguishing features in manga. Figure 5 highlights heatmaps for page 49 of LoveHina_vol14, focusing on text-and-character-masked images and panel frame-only images. For masked images, the model shifts attention to panel layout details, such as border spacing and alignment, adapting to the absence of characters. In panel frame-only images, the model highlights two small, distinct frames at the bottom-right corner, commonly seen in LoveHina.

However, in text-and-character-masked images, these unique frames are frequently obscured due to the masking process in the Manga109 dataset, which uses rectangular annotations for text and character locations. As shown in Figure 5, the masking unintentionally removes the borders of these smaller frames, preventing the model from utilizing this distinctive layout information. This limitation provides a plausible explanation for the lower classification accuracy of text-and-character-masked images compared to panel frame-only images.

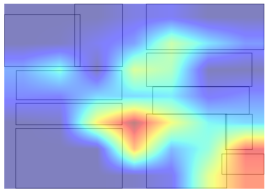
These observations underline the importance of panel layouts as a unique visual feature. The model's ability to accurately classify panel frame-only images demonstrates the strong discriminative power of layout-specific features, such as panel size, spacing, and alignment.

Building on this understanding, we further investigated whether the subtle differences in panel layouts could reflect an author's unique creative style, even within genres where such differences are difficult for human readers to discern.

To demonstrate that the panel layouts of each manga possess unique characteristics, we classified all five 4-panel manga works in the Manga109 dataset. The average classification accuracy was 93.1%, with only one misclassification within the 4-panel manga category. This high accuracy suggests that even subtle differences in panel layouts, which are challenging for human judgment, can be effectively captured by the model. These differences may reflect the unique preferences of authors in arranging panel content, such as the amount of space allocated for characters or dialogue.

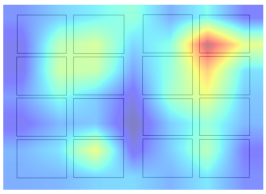


Text-and-character-masked

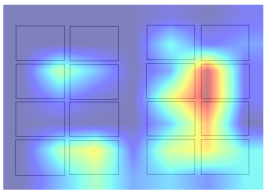


Panel frame-only

Figure 5: Grad-CAM heatmaps for page 49 of LoveHina_vol14 ©Ken Akamatsu, with text-and-character-masked facing page images and panel frame-only facing page images.



Akuhamu
©Satoshi Arai



KoukouNoHitotachi
©Ani Kuzuhara

Figure 6: Grad-CAM heatmaps of the input images for the classification model (using panel frame-only facing page images). Both input images belong to 4-panel manga works, and the model classified them correctly.

Next, the Grad-CAM heatmaps for classifying images from two 4-panel manga works are shown in Figure 6. The visualizations indicate that the model focuses on subtle layout features such as spacing between panels and the distribution of panels across the page. These features, while difficult for the human eye to distinguish, may be influenced by the author's preference for how much content (e.g. characters or speech balloons) they wish to include in each panel.

In conclusion, for panel frame-only facing page images, the Grad-CAM visualizations show that the model focuses solely on the spatial relationships and configurations of the panels. In the absence of any content, such as characters or text, the model relies entirely on panel layouts as distinguishing features for classification.

For 4-panel manga, which is particularly challenging for human judgment, the high classification accuracy achieved by the model highlights the subtle but critical differences in panel layouts. These differences, such as panel size, aspect ratio, spacing, and alignment, may reflect each author's unique approach to arranging content, such as characters or dialogue, within the panels. These layout decisions are not merely technical but also serve as expressions of the author's creative intent and storytelling style.

The high accuracy achieved across all manga works using panel frame-only facing page images supports the hypothesis that panel layouts possess inherent uniqueness across different manga works. Furthermore, it demonstrates that panel layouts play a critical role in distinguishing manga, even within genres like 4-panel manga, where differences are not easily perceived by the human eye. This finding underscores

Table 2: The results of the publisher classification experiment.

Publisher	Precision	Recall	F1-score
ENIX	0.00	0.00	0.00
Hakusensha	0.12	0.07	0.09
Asahi Sonorama	1.00	0.07	0.13
Tokuma Shoten	0.34	0.37	0.35
TokyoSanseisya	0.09	0.42	0.15
Shueisha	0.14	0.29	0.19
Kodansha	0.07	0.27	0.12
Kadokawa Shoten	0.09	0.02	0.03
Akita Publishing	0.05	0.01	0.02
Shogakukan	0.15	0.12	0.13
Gakken Educational	0.00	0.00	0.00
TAKESHOBO	0.00	0.00	0.00

Table 3: The results of the genre classification experiment.

Genre	Precision	Recall	F1-score
4-panel	0.89	0.70	0.78
Animal	0.00	0.00	0.00
Battle	0.02	0.01	0.01
Fantasy	0.00	0.00	0.00
History	0.06	0.01	0.02
Horror	0.00	0.00	0.00
Humor	0.34	0.67	0.46
Love	0.38	0.74	0.50
Romantic comedy	0.01	0.03	0.02
SF	0.30	0.49	0.38
Sport	0.01	0.02	0.02
Suspense	0.00	0.00	0.00

the strong discriminative power of layout features in manga classification tasks and their potential as a window into understanding the creative diversity of manga authors.

Publisher Classification and Genre Classification

In previous experiments, we observed that models focus on features like the spacing between panel frames and their distance to page edges. To determine whether these subtle features reflect the unique style of individual manga works, are influenced by publishers, or dictated by manga genres, we trained separate CNN models to classify manga by publisher and by genre.

From the Manga109 dataset, we selected 12 publishers (excluding two that contributed only one manga each). For the publisher classification task, one manga from each publisher was randomly chosen as the test set, with the remaining works as the training set. This setup provided sufficient training data while preventing the model from learning stylistic features specific to the test set. We used panel frame-only facing page images, with training methods and parameters identical to previous experiments.

As shown in Table 2, the publisher classification experiment achieved an average accuracy of 12.6%, compared to 8.3% for random guessing. These results suggest that publishers have minimal influence on panel layout design, indicating that panel layouts are not dictated by editorial policies but are instead unique to each manga work.

In Table 1, we have already seen the distribution of different manga genres in the Manga109 dataset. The dataset was split in the same manner as in the publisher classification experiment, ensuring that the model did not learn unique features specific to the test set during training. Training methods and parameters were again consistent with those used in previous experiments.

The results of the genre classification experiment are presented in Table 3. The overall average classification accuracy was 20.8%. Notably, the model demonstrated a relatively higher classification ability for 4-panel manga, which adheres to strict panel layout rules, as compared to other genres. For some genres, such as Humor and Love, the model showed moderate classification ability, likely due to the long histories of these genres, which may have led to the development of more established panel layout conventions. However, for other genres, the model struggled to classify the works effectively. When compared to the significantly higher average classification accuracy of 84.3% achieved in predicting specific manga titles based on panel layouts, these results indicate that manga genres contribute relatively less to panel layout features. This reinforces the conclusion that panel layouts are distinct and unique characteristics of individual manga works.

In conclusion, the publisher and genre classification experiments revealed that neither publishers nor genres significantly influence the panel layout of manga. The details of panel layouts, such as panel size, spacing between borders, and the spacing between borders and page edges, are likely determined by the creative choices of individual manga authors. These findings further emphasize the individuality of panel layouts as a defining feature of manga works.

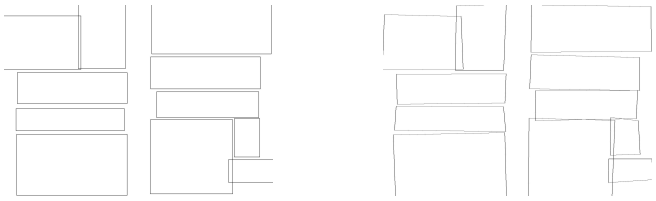
Classification Experiments with Noisy Panel Frame-only Facing Page Images

In the previous section, we confirmed that panel layout features are unique to each manga and are not determined by the publisher or genre. To further investigate the role of panel layout features in manga classification, we conducted additional experiments using noisy panel frame-only facing page images as input. In these experiments, random noise was introduced to disrupt the spatial relationships between panels while retaining the overall layout structure to some extent. This approach allowed us to test whether the classification accuracy would significantly decrease, thereby validating the importance of precise panel layout information. We employed two types of noise, as illustrated in Figure 7:

Rectangular noisy panels (Figure 7a): The first type retained the rectangular characteristics of the panels but randomly shifted their vertices by a certain value. Two experiments were conducted by adding noise of (-10, 10) and (-20,

20) to the vertices based on their original positions.

Quadrilateral noisy panels (Figure 7b): The second type only retained the quadrilateral characteristics of the panels, with vertices randomly shifted by a certain value. Similarly, two experiments were conducted with noise of (-10, 10) and (-20, 20).



(a) Rectangle noisy panels. **(b)** Quadrilateral noisy panels.

Figure 7: Noisy image examples.

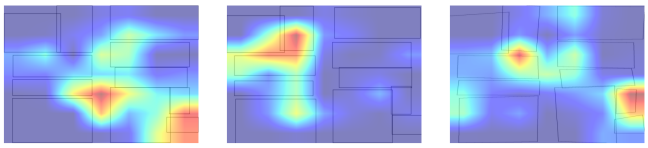
Table 4: Classification accuracy on noisy panel frame-only images.

Noise Type	Noise Range	Accuracy
Rectangular	(-10, 10)	0.67
Rectangular	(-20, 20)	0.53
Quadrilateral	(-10, 10)	0.71
Quadrilateral	(-20, 20)	0.59

We further added noise to panel frame positions to test the model's sensitivity to layout distortion. As shown in Table 4, the classification accuracy decreased as the noise level increased, indicating that precise panel positioning plays a meaningful role in layout-based classification. Although classification accuracy for noisy panel frame images was lower than that for original panel frame images, the model still demonstrated a certain level of accuracy. These results indicate that while noise reduces the model's ability to utilize panel layout features, the distinctive characteristics of panel designs across works remain sufficient for classification, even under noisy conditions.

Interestingly, the classification accuracy for quadrilateral noisy panels was consistently higher than for rectangular noisy panels. We hypothesize that this is because adding noise to rectangular panels disrupts unique characteristics of the original work more significantly, such as the spacing between panels and the margins between panels and page edges. In contrast, quadrilateral noisy panels preserve more of the original layout features, supporting higher classification accuracy.

For instance, consider Figure 8, which shows three panel frame-only images based on the original page 49 of LoveHina_vol14, with Grad-CAM heatmaps overlaid. The model trained on the rectangle noisy panel frame image failed to classify correctly. In both Figure 8a and Figure 8c, the model strongly focuses on the spacing between the small bottom-right panel and the page edge. This slightly outward-protruding bottom-right corner panel, relative to the row



(a) Unprocessed **(b)** Rectangle noise **(c)** Quadrilateral noise

Figure 8: Three types of panel frame-only facing page image.

above, appears to be a distinctive layout feature of LoveHina. When noise is added while retaining rectangular characteristics, such layout features are more likely to be lost compared to when only quadrilateral characteristics are preserved.

These findings provide further evidence that panel layouts in manga, even in subtle details that are challenging for human perception, play a decisive role in distinguishing works. This experiment highlights that the spatial relationships and configurations of panels are deeply tied to the unique characteristics of each manga, reinforcing the importance of panel layouts as a defining feature in manga classification tasks.

Conclusion

This study used deep learning to explore whether panel page designs in manga vary by work. Our experiments showed that even without characters and text, panel layouts exhibit work-specific uniqueness, enabling reliable classification and highlighting their role as a distinguishing visual structure in manga. This was validated through classification tasks and supported by Grad-CAM visualizations.

Additional experiments confirmed that panel layouts are minimally influenced by publishers or genres, underscoring their individuality at the work level. Even with added noise, panel layouts retained sufficient distinctiveness for classification, emphasizing their intricate role in manga creation.

While our study isolates panel layouts, we acknowledge that other visual elements—such as character design, gaze direction, and speech balloon positioning—also influence manga style and storytelling. Future research could investigate the interplay between these features and layout.

These findings highlight the significance of panel layouts as a unique visual feature in manga, reflecting the stylistic diversity of individual works and their contribution to storytelling structure. Future research directions include conducting experiments with a larger dataset to verify whether classification accuracy can remain high as the number of categories increases. Additionally, due to current data limitations, the differences in page design between works by the same author have not been thoroughly explored. Further investigation into this question is planned for future studies.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP24K02993.

References

Augereau, O., Iwata, M., & Kise, K. (2018). A survey of comics research in computer science. *Journal of imaging*, *4*(7), 87.

Chu, W.-T., & Guo, H.-J. (2017). Movie genre classification based on poster images with deep neural networks. In *Proceedings of the workshop on multimodal understanding of social, affective and subjective attributes* (p. 39-45). Association for Computing Machinery.

Cohn, N. (2013). *The visual language of comics: Introduction to the structure and cognition of sequential images*. Bloomsbury.

Dubray, D., & Laubrock, J. (2019). Deep cnn-based speech balloon detection and segmentation for comic books. *CoRR*, *abs/1902.08137*.

Dunst, A., Hartel, R., & Laubrock, J. (2017). The graphic narrative corpus (gnc): Design, annotation, and analysis for the digital humanities. In *2017 14th iapr international conference on document analysis and recognition (icdar)* (Vol. 03, p. 15-20).

Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., & Aizawa, K. (2016). Manga109 dataset and creation of metadata. In *Proceedings of the 1st international workshop on comics analysis, processing and understanding* (pp. 1–5).

Godwin, K. E., Eng, C. M., Todaro, R., Murray, G., & Fisher, A. V. (2018). Examination of the role of book layout, executive function, and processing speed on children's decoding and reading comprehension. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 40).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, *abs/1512.03385*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *Proceedings of the european conference on computer vision (eccv)* (pp. 630–645).

Kukreja, V., et al. (2023). Unmasking comic narratives: Cnn-rf approach to genre classification. In *2023 4th international conference on data analytics for business and industry (icdabi)* (pp. 555–560).

Laubrock, J., & Dubray, D. (2019, 01). Cnn-based classification of illustrator style in graphic novels: Which features contribute most?: 25th international conference, mmm 2019, thessaloniki, greece, january 8–11, 2019, proceedings, part ii. In (p. 684-695).

Laubrock, J., & Dunst, A. (2020). Computational approaches to comics analysis. *Topics in cognitive science*, *12*(1), 274–310.

Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., & Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, *76*(20), 21811–21838.

McCloud, S., & Martin, M. (1993). *Understanding comics: The invisible art* (Vol. 106). Kitchen sink press Northampton, MA.

Natsume, F., Holt, J., & Fukuda, T. (2020). Panel configurations in shōjo manga. *US-Japan Women's Journal*, *58*(58), 58–74.

Rigaud, C., Nguyen, N.-V., & Burie, J.-C. (2021). Text block segmentation in comic speech bubbles. In *Pattern recognition. icpr international workshops and challenges: Virtual event, january 10–15, 2021, proceedings, part vi* (p. 250-261). Springer-Verlag.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, *115*(3), 211–252.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019, oct). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359.

Sharma, R., & Kukreja, V. (2024). Image segmentation, classification and recognition methods for comics: A decade systematic literature review. *Engineering Applications of Artificial Intelligence*, *131*, 107715.

Takahashi, M., & Kondo, T. (2023). What is a textbook layout that everyone understands?: Towards a comprehensive understanding of text and supplementary information. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Tüchler, A. F., Zarina, L., & Skilters, J. (2021). The impact of interface alignment structure on aesthetic appreciation and usability rating. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Wi, J. A., Jang, S., & Kim, Y. (2020). Poster-based multiple movie genre classification using inter-channel features. *IEEE Access*, *8*, 66615-66624.

Young-Min, K. (2019). Feature visualization in comic artist classification using deep neural networks. *Journal of Big Data*, *6*(1), 56.