RGC Ref. No.:
UGC/FDS11/E01/21
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR
THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines: *1. Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
*2. Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

Deep Comic Screening via Tone-aware Semantic Layer Analysis

2. Investigator(s) and Academic Department(s) / Unit(s) Involved

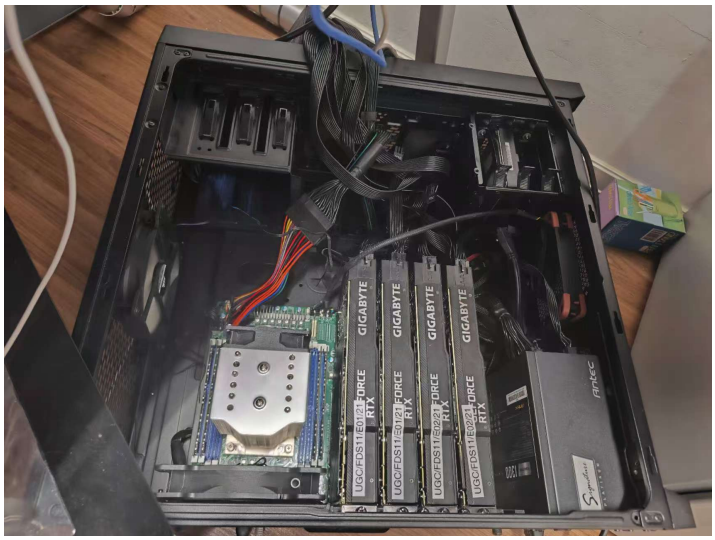
Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Dr. LIU Xueting / Assistant Professor	School of Computing and Information Sciences / Saint Francis University
Co-Investigator(s)	Prof. WONG Tien Tsin / Professor	Department of Computer Science and Engineering / The Chinese University of Hong Kong
Others		

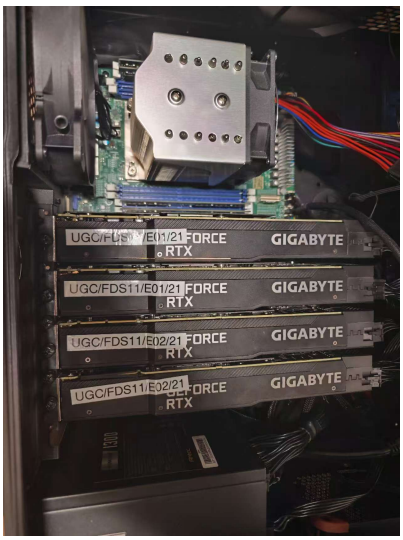
3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	01/01/2022		
Project Completion Date	31/12/2023	30/06/2024	Approved by UGC: 16/06/2023 Approved by Institute:

			16/06/2023
Duration *(in month)*	24	30	Approved by UGC: 16/06/2023 Approved by Institute: 16/06/2023
Deadline for Submission of Completion Report	31/12/2024	30/06/2025	Approved by UGC: 16/06/2023 Approved by Institute: 16/06/2023

4.4 Please attach photo(s) of acknowledgement of RGC-funded facilities / equipment.





Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. Develop a new deep learning network to predict the tones of an inked line drawing.

2. Design a new algorithm to extract the semantic layers based on a tone map.

3. Develop a new deep learning network to generate a screened comic from a stack of semantic layers.

4. Prepare a large supervised training dataset of line drawings, screened comics, and comics. Train the networks on the prepared dataset.

5. Evaluate the system via quantitative experiments, user study, and ablation study.

5.2 Revised objectives

Date of approval from the RGC: N/A

Reasons for the change: N/A

5.3 Realisation of the objectives
(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

The objectives were fully achieved.

1. Develop a new deep learning network to predict the tones of an inked line drawing.

Towards this objective, we proposed a diffusion model to generate a color illustration from the input sketch, which is fine-tuned on color illustrations from the community. Specifically, we used MeinaPastel, a diffusion model finetuned on colored illustration, as our colorization model, and control_v11p_sd15_lineart2 to condition the denoising process on line-art sketches. Subsequently, we proposed another diffusion model to obtain intensity values, i.e., the tones derived from the grayscale version of the color illustration to guide the comic generation.

2. Design a new algorithm to extract the semantic layers based on a tone map.

We attempted to adopt semantic segmentation methods to extract the semantic layers, but it worked unsatisfactorily as the existing methods commonly recognized a character as a single region, which is unable to acquire the semantic differences of the body parts. Therefore, we proposed to apply K-means clustering to the generated color illustration to obtain a set of regions where pixels in each region are connected and with similar colors. These regions, though low-level, worked better to guide the generation of screentones than the semantic regions extracted by the semantic segmentation methods.

3. Develop a new deep learning network to generate a screened comic from a stack of semantic layers.

To create comic images with screentones from a sketch, we obtained a large number of high-resolution comic training samples, as in the next objective. With the prepared training data, we finetuned both denoise diffusion U-Net and the VAE decoder components in stable diffusion. We further finetuned the decoder by following the paradigm of sd-vae-ft-mse that used a weighted sum of adversarial loss, MSE loss, and LPIP loss as the objective function. We found that LPIP loss degraded screentone reconstruction, so we further proposed to remove the LPIP loss from the objective function, which significantly enhanced the quality of the generated screentones.

4. Prepare a large supervised training dataset of line drawings, screened comics, and comics. Train the networks on the prepared dataset.

We scraped 186k high-resolution comic images from the Internet and further restored the high-resolution versions of the Manga109 dataset using a manga restoration method to obtain 103k more high-quality manga images. Altogether 289k high-resolution comic and manga images are collected as the training dataset.

5. Evaluate the system via quantitative experiments, user study, and ablation study.

We conducted visual comparisons, user studies, and ablation studies. We did not perform quantitative evaluations since there are no established metrics suitable for this purpose. Metrics like LPIP and FID, which rely on DNNs trained on labeled natural images, are not applicable to comics due to the absence of comparable datasets. For user studies, we randomly picked 20 unseen sketches for sketch-to-comic and 20 unseen illustrations for illustration-to-comic from Danbooru. 17 participants are invited. We also conducted ablation study on the importance of using tone maps as intermediary between line drawing and screened comics. All evaluations validated the outperformance of our method compared to existing methods.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. Develop a new deep learning network to predict the tones of an inked line drawing.	✓	100%
2. Design a new algorithm to extract the semantic layers based on a tone map.	✓	100%
3. Develop a new deep learning network to generate a screened comic from a stack of semantic layers.	✓	100%
4. Prepare a large supervised training dataset of line drawings, screened comics, and comics. Train the networks on the prepared dataset.	✓	100%
5. Evaluate the system via quantitative experiments, user study, and ablation study.	✓	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

In this project, we present a novel diffusion-based framework for the automated generation of screened comics from sketches. The key innovation of our approach lies in the introduction of tone maps as an intermediary step between sketch-based input and screened comic output. By leveraging tone maps as a proxy, we decompose the traditionally complex task of comic screening into two simpler subtasks. First, we generate a tone map from the sketch using a text-to-image diffusion model conditioned on line art. Second, we transform the tone map into a high-frequency screentone comic image. This intermediary transition is both intuitive and effective, as it capitalizes on the alignment between color shading and screentone patterns.

To the best of our knowledge, no fully automated method currently exists for this task. Existing approaches are limited to photo-to-comic conversion, which suffers from significant domain gaps, or reference-based screening techniques that require additional user-provided inputs. Although recent advancements in diffusion-based image synthesis have demonstrated impressive capabilities for producing high-quality, controllable outputs, these models still struggle to generate high-quality screentones. This limitation stems primarily from the absence of robust, high-resolution comic datasets for training both the variational autoencoder (VAE) and U-Net components of diffusion models.

Our framework addresses these challenges by enhancing the stable diffusion model through fine-tuning with a newly curated, high-quality comic dataset. Specifically, we reformulate the training objectives of the VAE to better capture the features of high-frequency screentones. For the U-Net denoising model, we introduce an additional conditioning factor: the tones. The tones, combined with text prompts, ensure that the generated screentones align seamlessly with the shading of the intermediary tone map. To further refine the output, we tackle issues such as minor structural distortions and unnatural shading artifacts that may arise during the denoising diffusion process. We propose an adaptive scaling method that integrates the generated screentones into the tone map, producing comic outputs that are both structurally accurate and visually appealing. Additionally, we introduce tone-aware semantic layer analysis, a layer-wise approach that decouples shading regions based on semantic context. This enables precise screentone application with greater consistency and fidelity.

We validate the effectiveness of our framework through comparative evaluations against existing sketch-to-comic and illustration-to-comic methods. Empirical results demonstrate that our approach significantly outperforms prior techniques in both visual quality and adherence to desired comic aesthetics.

Our research outputs include the following:

- A novel approach for generating high-quality screened comics from sketches, utilizing tone maps as an intermediary step to bridge the gap between line art and finalized comics.
- A fine-tuned diffusion model with a tailored objective and intensity conditioning to produce high-frequency screentones that align with intermediary shading.
- A layer-wise method for precise screentone application based on the understanding of different semantic regions.
- An adaptive scaling method to seamlessly integrate generated screentones into tone maps, ensuring structural coherence and aesthetic appeal in the comic output.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

We propose some potential for further development of the research and the proposed course of action as follows.

Firstly, while the current framework demonstrates strong performance, generating high-resolution screentones can be computationally intensive, which may limit its scalability for large-scale applications, application to a full comic page or a line cartoon animation. For the proposed course of action, it may be possible to explore distributed or parallel processing frameworks to enable faster generation of comics.

Secondly, although the adaptive scaling method and tone-aware semantic layer analysis mitigate many issues, minor structural distortions or unnatural shading artifacts may still arise in complex cases. For the proposed course of action, it is possible to investigate the use of post-processing techniques, such as neural style transfer or error correction algorithms, to refine outputs further. We may also train the model on more challenging examples that include intricate shading or fine details to improve its robustness or incorporate feedback loops where the model evaluates and corrects its own outputs, potentially using a discriminator network in a GAN-like setup.

7. Layman's Summary

(Describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Comics are a widely enjoyed form of entertainment, yet the process of adding screentones to inked line drawings can be labor-intensive. Existing methods for automatic comic screening often relied on user input for screentone examples or reference images, which hindered efficiency. Recent advancements in deep learning showed promise for automating this process, but existing models struggled to produce high-quality shaded screentones, primarily due to a lack of specialized frameworks and high-quality training datasets. This project proposed a novel framework for comic creation that generated high-quality shaded screentones based on tone-aware semantic layer analysis. Our approach first predicted tones from inked drawings and extracted semantic layers based on a tone map. Subsequently, we employed a deep learning network to generate a screened comic from these layers. By leveraging a large supervised dataset comprising line drawings, screened comics, and original comics, our method significantly outperformed current techniques in producing high-quality comics with high-frequency screentones. Our project showed significant importance in streamlining the comic creation process, enhancing both the efficiency and quality of comic production.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project

(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				**Author(s)** *(denote the corresponding author with an asterisk*)*	**Title and Journal / Book** *(with the volume, pages and other necessary publishing details specified)*	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
2022				Zibo Zhang, Xueting Liu*, Chengze Li, Huisi Wu*, Zhenkun Wen	Vectorizing Line Drawings of Arbitrary Width via Contour-based Topology Reconstruction. Computer Graphics Forum, 41(2), 433-445.	2022	Yes	Yes	Yes
2022				Cheng Xu, Wei Qu, Xuemiao Xu*, Xueting Liu	Multi-scale Flow-based Occluding Effect and Content Separation for Cartoon Animations. IEEE Transactions on Visualization and Computer Graphics, 29(9), 4001-4014.	2022	Yes	Yes	No
2023				Chenshu Xu, Xuemiao Xu*, Nanxuan Zhao, Weiwei Cai, Huaidong Zhang*, Chengze Li, Xueting	Panel-Page-Aware Comic Genre Understanding. IEEE Transactions on Image Processing, 32, 2636-2648.		Yes	Yes	No

				Liu					

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered
(Please attach a copy of each conference abstract)

Month / Year / Place	**Title**	**Conference Name**	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
10 / 2024 / Abu Dhabi, UAE	Sketch2manga: Shaded manga screening from sketch with diffusion models	IEEE International Conference on Image Processing (ICIP 2024)		Yes	Yes	Yes

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning
(Please elaborate)

This work contributed to teaching and learning in both the Bachelor of Science in Digital Entertainment Technology and the Bachelor of Science in Artificial Intelligence programmes provided by the PI's institution. For Digital Entertainment Technology students, this project introduces advanced techniques in digital art creation, emphasizing how artificial intelligence can streamline and enhance the production of high-quality comic images. Students gain knowledge of how computational methods can transform traditional artistic workflows in animation, game design, and digital storytelling.

For Artificial Intelligence students, the project provides insights into the application of AI in creative industries, expanding their understanding of deep learning beyond conventional domains like predictive modeling or data analysis. It highlights key concepts such as image processing, neural network design, and training models for artistic tasks. Students learn to integrate deep learning algorithms into niche areas, fostering interdisciplinary thinking and problem-solving skills. It also introduces students to challenges in AI model optimization and evaluation metrics in creative contexts.

Overall, this research bridges the gap between technology and creativity, enriching both programmes with practical knowledge and real-world applications of AI in digital media and entertainment. It fosters innovation and equips students with versatile skills for emerging career opportunities at the intersection of technology and art.

11. Student(s) Trained
(Please attach a copy of the title page of the thesis)

Name	**Degree Registered for**	**Date of Registration**	**Date of Thesis Submission / Graduation**
	Bachelor of Science (Honours) in Digital	September 2019	June 2023

	Entertainment Technology		
	Bachelor of Science (Honours) in Digital Entertainment Technology	September 2020	June 2024

* Remark: These two students are bachelor's degree candidates and are not required to write a thesis to complete the programme. They participated in this project for five months each, primarily assisting with data collection and conducting user studies.

12. Other Impact

(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

We have collaborated with researchers from Shenzhen University and South China University of Technology, combining the strengths of both institutions in artificial intelligence and creative design. By pooling resources and expertise, the collaboration aims to push the boundaries of deep learning applications in art and entertainment industries. In addition to advancing technology, the results of this research will be shared with industry partners to foster commercialization opportunities, empowering creators with cutting-edge tools that streamline their workflows and enhance artistic expression. Furthermore, this project contributes to teaching enhancement by delivering the research outcomes via seminars and invited talks in related courses. Students could gain experience with the deep learning methods developed, enriching their understanding of AI applications in creative fields.

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	**Conference Papers**	**Scholarly Books, Monographs and Chapters**	**Patents Awarded**	**Other Research Outputs (please specify)**	
No. of outputs arising directly from this research project	3	1	N/A	N/A	Type	No.
					N/A	N/A

14. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	**Reasons**
N/A	