RGC Ref. No.:
UGC/FDS11/E02/21
(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

Instance-aware Cartoon Stylization of Photo and Videos

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

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Dr. LI Chengze / Assistant Professor	School of Computing and Information Sciences / Saint Francis University
Co-Investigator(s)	N/A	N/A
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: 29/12/2022 Approved by Institute: 29/12/2022

Duration *(in month)*	24	30	Approved by UGC: 29/12/2022 Approved by Institute: 29/12/2022
Deadline for Submission of Completion Report	31/12/2024	30/06/2025	Approved by UGC: 29/12/2022 Approved by Institute: 29/12/2022

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. *Preparation for the semantical cartoon instance dataset*
2. *Network design and implementation of the instance cartoon stylization*
3. *Canvas refinement and motion-specific optimization*
4. *System evaluation and integration*

5.2 Revised objectives

Date of approval from the RGC: _____

Reasons for the change: _____

_____

1.

2.

3. ....

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)

1. **Preparation of a Semantical Cartoon Instance Dataset.** We successfully created a large-scale dataset containing over 100,000 image samples of cartoon and anime instances. To address the challenge of extracting high-quality instances from noisy, unlabeled internet data, we developed a motion-based approach to analyse the salient objects (which later forms the key idea of [J4]). The dataset was further extended to include abstract forms like storyboards and sketches, for which we proposed a framework [C1] to parse and integrate this information into our data pipelines. To handle diverse sources such as vintage animations, we developed a remastering technique [J2] that improves video frame quality and performs color correction. A journal paper [J5] details the practical applications of this dataset, including its labeling, and use in character-level stylization. The dataset has been open-sourced to benefit this project and facilitate future research.

2. **Network Design and Implementation of Instance Cartoon Stylization.** For the instance and contextual-level translation of cartoon and anime styles from still photographs, we have designed and implemented several targeted approaches:
   - *Fine-grained Portrait Stylization:* For the stylization of frontal faces and portraits, we propose a method [J1] to convert human portraits into an anime style. This method utilizes a two-stage design with an intermediate domain of "coser" portraits (humans costuming as anime characters) to bridge the significant domain gap. This approach ensures high-quality translation while preserving the key facial features of the original portrait.
   - *Full-body Scene Stylization:* For scenes containing full-body characters, we propose a semantic-aware framework [J5]. This approach processes tagged body parts (e.g., hair, arms, legs) to convert them into an anime style while maintaining their salient features and original poses.
   - *Reducing Translation Complexity:* The complex shading in anime data can increase training difficulty. To mitigate this, we propose an approach [J3] to separate shading from cartoon scenes, creating a plain-color intermediate representation that reduces the domain gap with natural photos. The shading can be subsequently reapplied using a corresponding model also presented in [J3].

3. **Canvas Refinement and Motion-Specific Optimization.** While the previous objectives address still image stylization, ensuring temporal consistency in video translation remains a challenge. Our work in [J5] discusses animation production by translating body poses on a frame-by-frame basis, providing a foundation for video stylization. To further improve temporal stability, we propose a stabilization method in [J4], a stabilization method originally designed for aligning motions across long-range anime frames using attention mechanisms. Critically, we discovered this pipeline can be reversed to stabilize temporal consistency when translating natural videos to anime. By reusing the optical flow from natural videos through this reversed pipeline, we achieve significantly improved temporal coherence in the translated anime sequences. This approach also addresses the outpainting problem for frame extension, to convert square or vertical framed videos to 16:9 formats.

4. **System Evaluation and Integration.** For all developed methods, we conducted comprehensive quantitative and qualitative evaluations. For the interactive methods [J1, J2, J3, J5], we also performed subjective user studies to assess visual quality and usability. The source code for [C1, J1, J2, J3, J5] has been open-sourced to inspire further work and ensure the methods can be encapsulated for various applications. To achieve the goal of system integration for production use, we have adapted our core algorithms into custom nodes for ComfyUI, a de facto standard node-based framework for generative art. This modular integration allows users to freely incorporate our stylization components into their production workflows, effectively providing a flexible middleware for software and services as outlined in the original proposal.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. Preparation for the semantical cartoon instance dataset	√	100%
2. Network design and implementation of the instance cartoon stylization	√	100%
3. Canvas refinement and motion-specific optimization	√	100%
4. System evaluation and integration	√	100%

6. Research Outcome

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

This project addressed the long-standing and complex task of photo-to-cartoon stylization, making valuable contributions to the fields of computer graphics and generative AI. Our work was designed with a dual focus: to advance the scientific understanding of cross-domain image translation and to develop practical tools for real-world use. The project's duration coincided with the recent boom in generative art, allowing our research to both benefit from and contribute to this rapidly evolving landscape, ultimately providing a solid foundation for broader impacts across academia and the creative industries.

A primary outcome of this project is the development of foundational infrastructure to support research in this domain. We identified that a critical bottleneck hindering progress was the profound scarcity of structured, high-quality cartoon data. Unlike natural images, cartoon styles are incredibly diverse and abstract, making large-scale, consistent annotation a formidable challenge. To address this gap, we developed and released an open-source dataset containing over 100,000 annotated instances. This resource, complemented by new frameworks for parsing abstract sketch and drawing elements [C1], remastering legacy animations [J2] and pose and body parts estimation [J5], provides the community with essential data for training and evaluating modern generative models, and we are pleased to see it inspiring concurrent work in the field.

Our core scientific contributions focus on developing novel strategies to bridge the wide and challenging domain gap between photorealistic images and artistic cartoons. This gap manifests in several ways, including differences in the drawing styles of the structures, non-photorealistic shading, and exaggerated character features in terms of the character and scene appearances, which often cause standard end-to-end translation models to fail. We found that a more effective strategy is to decompose this complex problem into more manageable, targeted stages. Our solutions include proposing an "intermediate domain" to better preserve facial identity during portrait stylization [J1], developing a learning-based method to separate shading from base colors to simplify the domain adaptation task [J3], and designing a specialized pose-based segmentation framework tailored for anime characters, as general-purpose models perform poorly on this distinct data distribution [J5].

For dynamic content understanding and generation, we addressed the critical issue of temporal consistency, which is essential for producing high-quality stylized videos. To this end, we developed a framework that enhances motion stability and content alignment across frames by using region-guided motion inference [J4]. While this technique has practical applications such as converting narrow frame-size animations to modern formats, its core contribution to this project is its ability to serve as a foundation for producing temporally stable and consistent anime and cartoon video sequences.

The merit and novelty of this research are underscored by four publications in top-tier venues, including CVPR and IEEE TVCG. The project's lifespan fortuitously met the explosion of generative arts, providing an invaluable educational opportunity for students to gain hands-on experience with cutting-edge models through their coursework and final-year projects. By open-sourcing our code and integrating key algorithms into popular creative platforms like ComfyUI, we have ensured our findings empower not only researchers but also artists and independent creators, contributing to the vibrant and rapidly evolving field of generative art.

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

First, the project successfully established a curated dataset that is highly effective for our retrieval-based stylization approach. Building on this success, a promising future direction is to scale this process. The foundation from this project enables the development of an automated pipeline, using models like the Segment Anything Model (SAM), to extract character and other salient instances from vast public databases. This would further enhance the model's generalization capabilities and extend its application far beyond the grant's original scope.

Second, our stylization pipeline was intentionally designed as a non-end-to-end system to achieve superior artistic quality and control. With this high-quality benchmark established, the pipeline could be further optimized to improve processing efficiency for high-throughput or real-time applications. Similarly, the current background generation is highly effective for static shots and provides a solid base that can be extended to handle more dynamic scenes with complex camera motion, representing a clear avenue for future enhancement.

Finally, the project's technology was successfully implemented and validated on open-source platforms like ComfyUI, ensuring wide reach and reproducibility within the research community. This successful proof-of-concept paves the way for integration into professional production environments. The natural next step is to build on this stable foundation by developing plugins for industry-standard software, such as Adobe After Effects and Toon Boom Studio, translating our research directly into practical tools for animators and studios.

7. Layman's Summary

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

Creating cartoon animation is traditionally a slow, labor-intensive process requiring artists to draw frame by frame. This project tackles that challenge by developing an innovative AI system that automatically transforms photos or live-action video into cartoon styles. Our proposed solution goes beyond a simple filter; it intelligently analyzes the input content, learning from a curated database of artistic examples to understand and redraw the material. This process ensures the main character's appearance remains consistent and provides a high degree of artistic control over the final look.

The significance of this work is twofold. First, it can greatly reduce the time and cost of making animations, opening the door for independent creators and professional studios. Second, the project makes key contributions to both research and education. The developed methods achieved state-of-the-art in some of the stylization tasks. To encourage further innovation, we have publicly shared our computer code and the data of artistic examples our AI learned from. This allows other researchers to build on our work and gives students at our institution practical tools to learn about this technology and create their own artworks or graphics-based projects.

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				Wenpeng Xiao; Cheng Xu; Jiajie Mai; Xuemiao Xu*; Yue Li; **Chengze Li**; Xueting Liu; Shengfeng He	[J1] Appearance-preserved Portrait-to-anime Translation via Proxy-guided Domain Adaptation **IEEE Transactions on Visualization and Computer Graphics** Volume: 30, Issue: 7	**Yes**	**Yes**	**Yes**	**Yes**
2023				Yinghua Liu, **Chengze Li*,** Xueting Liu, Huisi Wu; Zhenkun Wen	[J2] AddCR: a data-driven cartoon remastering **The Visual Computer** Volume: 39, pp. 3741-3753	**No**	**Yes**	**Yes**	**Yes**
				Ziheng Ma; **Chengze Li**; Xueting Liu; Huisi Wu*; Zhenkun Wen	[J3] Separating Shading and Reflectance From Cartoon Illustrations **IEEE Transactions on Visualizati**	**Yes**	**Yes**	**Yes**	**Yes**

					on and Computer Graphics Volume: 30, Issue: 7				
2024				Huisi Wu*; Hao Meng; **Chengze Li**; Xueting Liu; Zhenkun Wen; Tong-Yee Lee	[J4] Cartoon Animation Outpainting With Region-Guided Motion Inference **IEEE Transactions on Visualization and Computer Graphics** Volume: 31, Issue: 4	**No**	**Yes**	**Yes**	**Yes**
				Zhenhua Ou, Xueting Liu, Chengze Li, Zhenkun Wen, Ping Li, Zhijian Gao*; Huisi Wu	[J5] Body Part Segmentation of Anime Characters **Computer Animation and Virtual Worlds** Volume: 35, Issue: 6	**No**	**Yes**	**Yes**	**Yes**

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)*
06/22 New Orleans, Louisiana	Neural Recognition of Dashed Curves With Gestalt Law of Continuity	The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)	Yes	Yes	Yes	Yes

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

(Please elaborate)

Yes, the research experience and new knowledge from this project have been transferred to our teaching and learning activities. A primary contribution was the significant enhancement of the "Computer Vision and Image Processing" course. We updated the curriculum to include new lecture modules covering core concepts from this project, such as diffusion models and controllable image synthesis. This integration ensures that our students' education remains current, bridging classical theories of computer vision and image understanding with the latest advancements in generative AI and its creative applications.

The theoretical knowledge was complemented by practical, hands-on experience. We designed new lab exercises centered around the ComfyUI workflows developed during this research, allowing students to perform tasks like anime-style stylization and gain an intuitive understanding of complex generative models. Furthermore, the project's concepts guided a final year student project on human pose estimation, providing the student with practical experience in applying research to solve real-world problems.

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	2/9/21	July 2025
	Bachelor of Science	7/9/20	July 2024

12. Other Impact

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

We established collaborations with researchers from Shenzhen University and South China University of Technology, combining the strength of the institutes and personnels in contributing to the field of Creative Media, Computer Graphics and Creative Design.
Parts of the project's findings were also disseminated to industry and academic communities through invited keynote presentations. These included a briefing for the "AI-Assisted Animation Pilot Scheme" hosted by the Hong Kong Digital Entertainment Association (9 October 2023) and a talk at the IEEE Workshop on Deep Learning (3-4 April 2023).
Furthermore, the research directly enhanced our curriculum. The "Computer Vision and Image Processing" course was updated to include project methodologies and tools, featuring hands-on labs where students used our ComfyUI workflows for anime-style image and video synthesis, from a set of input conditions. The project's concepts also guided a final year student project on pose estimation, providing students with practical experience applying research to solve real-world problems.

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs	Patents Awarded	Other Research Outputs (please specify)

			and Chapters			
No. of outputs arising directly from this research project	5	1			Type	No.

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**