Generative Models for Visual Content Editing and Creation

ZHENG WEI, The Hong Kong University of Science and Technology, Hong Kong
XIAN XU, Lingnan University, Hong Kong
YUQING LIU, The University of Hong Kong, Hong Kong
GRACE HAN, Stanford University, United States of America
ANYI RAO, The Hong Kong University of Science and Technology, Hong Kong

Fig. 1. SIGGRAPH Asia 2025 Teaser.

Generative AI now drives storyboarding, previs, and look-development, yet two gaps slow adoption: artists struggle with opaque tools, while ML engineers lack cinematic grammar. This half-day master class closes both gaps by pairing concise theory with hands-on, human-in-the-loop practice and built-in ethics. Through an Explain → Show → Do rhythm, each concept moves from a crisp technical snapshot to a live demo and a guided task. Team exercises turn peer critique into a rapid feedback loop, while questions of authorship, bias, and legal clearance surface at every step—embedding responsible practice into real production workflows. Live demos built on the CineVision pipeline transform a log-line into reference frames, shot lists, and colour-graded contact sheets, showcasing diffusion, LoRA, ControlNet, AnimateDiff, and IP-Adapter in action. Participants leave able to (i) explain how modern diffusion and multimodal generators work, (ii) customise toolchains without ceding creative control, (iii) integrate AI assets into coherent, ethically sound sequences, and (iv) assess—and build—production-ready pipelines that enhance director–cinematographer collaboration.

CCS Concepts: • **Computing methodologies → Graphics systems and interfaces**.

Additional Key Words and Phrases: Generative AI, Diffusion Models, Previsualization, Storyboarding, Cinematic VR, Human–AI Collaboration, Responsible AI

ACM Reference Format:
Zheng Wei, Xian Xu, Yuqing Liu, Grace Han, and Anyi Rao. 2025. Generative Models for Visual Content Editing and Creation. In *SIGGRAPH Asia 2025 Courses (SA Courses '25), December 15-18, 2025*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3757371.3763257

Authors' addresses: Zheng Wei, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, zwei302@connect.ust.hk; Xian Xu, Lingnan University, Hong Kong, Hong Kong, xianxu@LN.edu.hk; Yuqing Liu, The University of Hong Kong, Hong Kong, Hong Kong, liuyuqing990831@connect.hku.hk; Grace Han, Stanford University, Palo Alto, United States of America, ghahahan@stanford.edu; Anyi Rao, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, anyirao@ust.hk.

SA Courses '25, December 15-18, 2025, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2131-1/2025/12
https://doi.org/10.1145/3757371.3763257

1 INTRODUCTION

Generative models have moved rapidly from research labs into preproduction rooms across film, animation, and immersive media [Chen et al. 2024; Jiang et al. 2024; Rao et al. 2024a, 2023, 2022; Wei et al. 2025b; Zhang et al. 2025b; Zhou et al. 2025]. Text-to-image diffusion, lightweight personalization (e.g., LoRA [Hu et al. 2022]), structure guidance (e.g., ControlNet [Zhang et al. 2023]), relighting (e.g., IC-Light [Zhang et al. 2025a]) and emerging text-to-video extensions [Guo et al. 2024, 2025] now accelerate concept art,

look development, and storyboarding. Yet adoption in professional pipelines remains uneven because two gaps persist: (i) creative control and continuity—artists confront opaque systems that may destabilize style, identity, and shot-to-shot coherence; and (ii) cinematic integration—engineers often lack the grammar of lenses, blocking, lighting, and coverage needed to embed models into director–cinematographer workflows rather than one-off image trials.

This course addresses those gaps with a compact and practice-centered curriculum that couples how it works with how to use it well. We build a clear model of latent diffusion and multimodal conditioning, then translate that model into mixed-initiative tool-chains that preserve authorship and support reproducible pre-visualization. Our reference implementation CineVision demonstrates how prompt design, reference stacks, structure guidance, and small adapters for character and costume can be composed into a sequence-aware pipeline. Our pedagogy follows an "Explain, Show and Do" cadence. Each concept is introduced briefly, demonstrated live with annotated settings, and reinforced through a guided micro-exercise. Participants collectively build a six-shot sequence from a log-line, producing a contact sheet and beat-by-beat shot list. Throughout, we surface common failure modes—prop and pose drift, perspective inconsistency, temporal flicker—and provide checklists that map symptoms to remedies. Responsible practice is integrated rather than appended. We log prompts and adapter choices for attribution; and we prefer models and materials with unambiguous rights to support downstream clearance. The goal is not merely to generate striking frames but to assemble coherent, ethically sound sequences that can enter a production track with minimal rework.

This course offers: (1) a concise, production-oriented explanation of diffusion and multimodal control for artists and developers; (2) a pattern language for composing control signals (prompt, reference, and structure) into reliable shot design; (3) a hands-on lab that yields a reproducible six-shot mini sequence; and (4) practical guardrails—rubrics and checklists—for stabilizing identity, palette, and motion while maintaining creative agency. Together, these elements enable directors, cinematographers, and HCI practitioners to adopt generative tools without ceding authorship or coherence to the model.

1.1 From Log-line to Contact Sheet

As shown in Figure 2, our live demo transforms a one-sentence log-line into a six-shot sequence [Wei et al. 2025b]: (1) draft multiple visual references per shot; (2) converge on character/costume via reference-guided diffusion; (3) lock composition and camera; (4) iterate palette/grading; (5) export a contact sheet and a beat-by-beat shot list.

2 PEDAGOGY, UNITS, AND LEARNING OUTCOMES

We follow an *Explain → Show → Do* cadence with progressive disclosure: brief theory, an annotated live demo, then a guided micro-exercise. Active learning (polls, pair programming, peer critique) surfaces misconceptions in real time.

Units (Half-Day).

- **Unit 1: Diffusion & Multimodal Foundations** — latent diffusion, cross-attention, LoRA/ControlNet, image–text alignment; rapid Colab demo.
- **Unit 2A: Prompt Craft & Style Control** — prompt taxonomy (subject, camera, lighting, mood), negative prompts, classifier-free guidance; hands-on: multiple distinct shot styles.
- **Unit 2B: Mixed-Initiative Tool-Chains & Lab** — Group task: build a six-shot sequence from a log-line; export contact sheet.
- **Unit 3A: The Aesthetics of Noise in Generative AI** — metastability, indexicality, multiplanarity; temporality in predictive generation; illustrated talk with case study.
- **Unit 3B: VR Storytelling & AI Collaboration** — cinematic VR virtual reality education [Wei et al. 2024a,b, 2025a, 2023; Xu et al. 2023, 2025], production workflow, emerging AI tools; comparative case discussion.

By the end, participants can (i) explain diffusion/multimodal basics, (ii) craft prompts and reference stacks for continuity, (iii) assemble reproducible tool-chains that preserve creative control, and (iv) integrate assets ethically into coherent sequences.

3 TEACHING TEAM AND ADVISORY COMMITTEE

Lecturers. Zheng Wei (HKUST), Xian Xu (Lingnan University), Yuqing Liu (HKU), Grace Han (Stanford), Anyi Rao (HKUST).
Advisory. Maneesh Agrawala (Stanford), Huamin Qu (HKUST), James A Evans (UChicago), Shane Denson (Stanford), Pan Hui (HKUST(GZ)) Tim Gruenewald (HKU), Bárbara Fernández-Melleda (HKU).

ACKNOWLEDGMENTS

We thank the CVEU workshop community (ICCV 2021, ECCV 2022, ICCV 2023, CVPR 2024, CVPR 2025, SIGGRAPH 2024 [Rao et al. 2024b], SIGGRAPH 2025 [Patashnik et al. 2025]), the 2025 HKUST AI Film Festival, and colleagues in HKUST VisLab and Multimedia Creativity Lab (MMLab).

REFERENCES

Yiran Chen, Anyi Rao, Xuekun Jiang, Shishi Xiao, Ruiqing Ma, Zeyu Wang, Hui Xiong, and Bo Dai. 2024. Cinepregen: Camera controllable video previsualization via engine-powered diffusion. *arXiv preprint arXiv:2408.17424* (2024).

Yuwei Guo, Ceyuan Yang, Anyi Rao, Zhengyang Liang, Yaohui Wang, Yu Qiao, Maneesh Agrawala, Dahua Lin, and Bo Dai. 2024. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. *International Conference on Learning Representations* (2024).

Yuwei Guo, Ceyuan Yang, Anyi Rao, Chenlin Meng, Omer Bar-Tal, Shuangrui Ding, Maneesh Agrawala, Dahua Lin, and Bo Dai. 2025. Keyframe-Guided Creative Video Inpainting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 13009–13020.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR* 1, 2 (2022), 3.

Xuekun Jiang, Anyi Rao, Jingbo Wang, Dahua Lin, and Bo Dai. 2024. Cinematic behavior transfer via nerf-based differentiable filming. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6723–6732.

Or Patashnik, Gaurav Parmar, Anyi Rao, Ozgur Kara, Fabian Caba Heilbron, Daniel Cohen-Or, James Matthew Rehg, and Jun-Yan Zhu. 2025. AI for Creative Visual Content Generation, Editing and Understanding. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Frontiers*. 1–2.

Anyi Rao, Jean-Peïc Chou, and Maneesh Agrawala. 2024a. ScriptViz: A Visualization Tool to Aid Scriptwriting based on a Large Movie Database. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*.

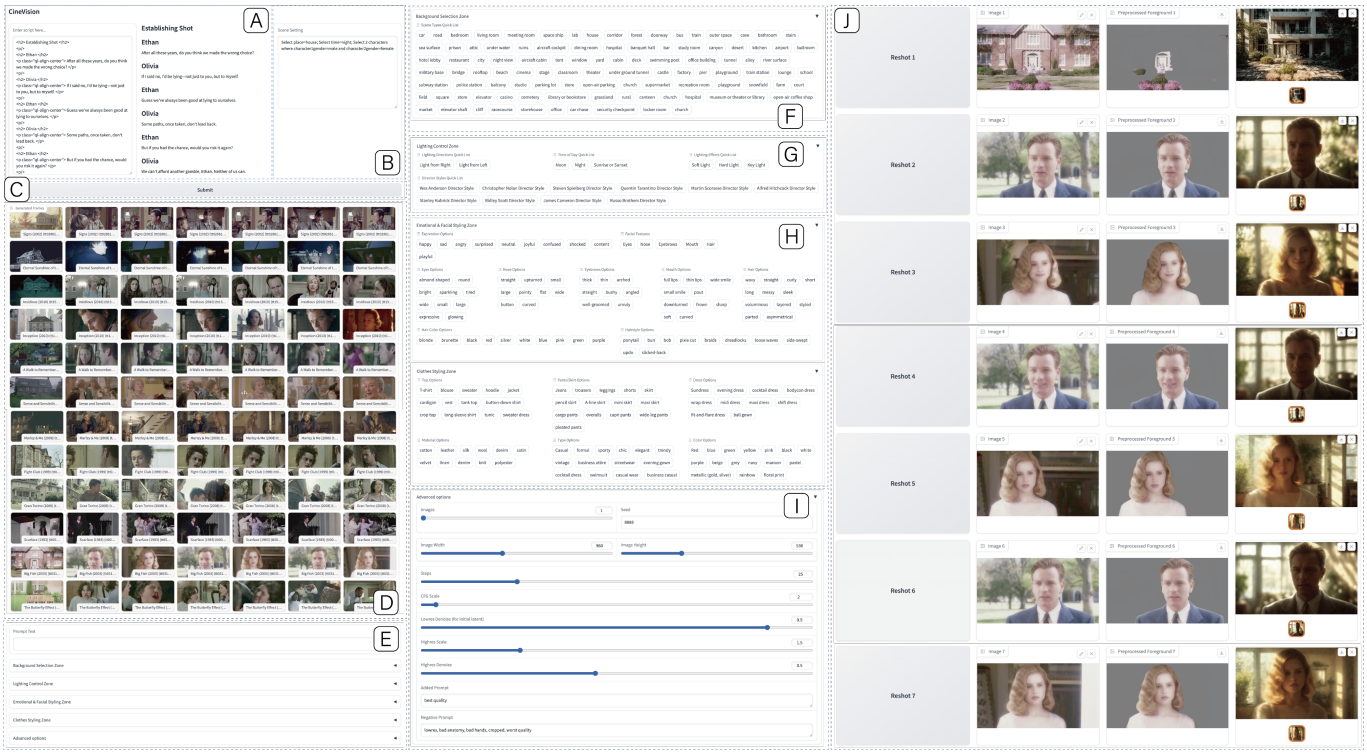

Fig. 2. CineVision pipeline: composing GUIs and scripts into a reproducible pre-production workflow for directors and cinematographers.

Anyi Rao, Xuekun Jiang, Yuwei Guo, Linning Xu, Lei Yang, Libiao Jin, Dahua Lin, and Bo Dai. 2023. Dynamic storyboard generation in an engine-based virtual environment for video production. In *ACM SIGGRAPH 2023 Posters*. 1–2.

Anyi Rao, Yuanbo Xiangli, Yuwei Guo, Mia Tang, Chenlin Meng, and Maneesh Agrawala. 2024b. Generative Models for Visual Content Editing and Creation. In *ACM SIGGRAPH 2024 Courses*. 1–6.

Anyi Rao, Linning Xu, and Dahua Lin. 2022. Shoot360: Normal view video creation from city panorama footage. In *ACM SIGGRAPH 2022 conference proceedings*. 1–9.

Zheng Wei, Yuzheng Chen, Wai Tong, Xuan Zong, Huamin Qu, Xian Xu, and Lik-Hang Lee. 2024a. Hearing the moment with metaecho! from physical to virtual in synchronized sound recording. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 6520–6529.

Zheng Wei, Shan Jin, Wai Tong, David Kei Man Yip, Pan Hui, and Xian Xu. 2024b. Multi-role vr training system for film production: Enhancing collaboration with metacrew. In *ACM SIGGRAPH 2024 Posters*. 1–2.

Zheng Wei, Jia Sun, Junxiang Liao, Lik-Hang Lee, Chan In Sio, Pan Hui, Huamin Qu, Wai Tong, and Xian Xu. 2025a. Illuminating the scene: How virtual environments and learning modes shape film lighting mastery in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* (2025).

Zheng Wei, Hongtao Wu, Xian Xu, Yefeng Zheng, Pan Hui, Maneesh Agrawala, Huamin Qu, Anyi Rao, et al. 2025b. CineVision: An Interactive Pre-visualization Storyboard System for Director-Cinematographer Collaboration. *arXiv preprint arXiv:2507.20355* (2025).

Zheng Wei, Xian Xu, Lik-Hang Lee, Wai Tong, Huamin Qu, and Pan Hui. 2023. Feeling present! from physical to virtual cinematography lighting education with metashadow. In *Proceedings of the 31st ACM International Conference on Multimedia*. 1127–1136.

Xian Xu, Wai Tong, Zheng Wei, Meng Xia, Lik-Hang Lee, and Huamin Qu. 2023. Cinematography in the metaverse: Exploring the lighting education on a soundstage. In *2023 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW)*. IEEE, 571–572.

Xian Xu, Wai Tong, Zheng Wei, Meng Xia, Lik-Hang Lee, and Huamin Qu. 2025. Transforming cinematography lighting education in the metaverse. *Visual Informatics* 9, 1 (2025), 1–17.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2025a. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *The Thirteenth International Conference on Learning Representations*.

Ruihan Zhang, Borou Yu, Jiajian Min, Yetong Xin, Zheng Wei, Juncheng Nemo Shi, Mingzhen Huang, Xianghao Kong, Nix Liu Xin, Shanshan Jiang, et al. 2025b. Generative AI for Film Creation: A Survey of Recent Advances. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 6267–6279.

Yujie Zhou, Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, Qidong Huang, Jinsong Li, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, et al. 2025. Light-a-video: Training-free video relighting via progressive light fusion. *arXiv preprint arXiv:2502.08590* (2025).