

Research Statement

Shengfeng He
Associate Professor of Computer Science
School of Computing and Information Systems, Singapore Management University
Email: shengfenghe@smu.edu.sg; Homepage: <https://shengfenghe.github.io/>
December 2, 2025

1. RESEARCH MOTIVATION AND CENTRAL QUESTIONS

Humans are natural consumers of visual content, engaging seamlessly with images, videos, and interactive media in their daily lives. Yet there remains a profound asymmetry between consumption and creation. While anyone can enjoy visual content, only a small fraction with professional training can create or manipulate it convincingly. For most people, attempts at producing visual media fall outside the natural manifold of images, resulting in unrealistic, incoherent, or stylized outputs that fail to capture human expressiveness. This gap between effortless consumption and challenging creation motivates my research: to develop computational models that not only understand human perception of the visual world, but also enable its faithful and controllable generation.

The challenge is multifaceted. From a *scientific perspective*, we need models that capture human-centric visual perception, including attention, semantics, and robustness to noise or missing data. From a *technological perspective*, generative models must move beyond black-box design toward interpretable, controllable, and reusable systems that allow both experts and non-experts to guide creation. From a *societal perspective*, the rapid growth of generative AI brings urgent questions of responsibility, intellectual property protection, and trustworthy deployment. Finally, from a *future-oriented perspective*, generative AI must evolve toward multimodal and embodied intelligence, capable of adapting to open environments and interacting with the physical world.

These perspectives converge into four central research questions:

- (1) **Understanding Human-Centric Perception:** How can we design algorithms that replicate human-like perception, focusing on saliency, semantics, and contextual reasoning, and that remain robust when trained on imperfect or limited data?
- (2) **Democratizing Visual Creation:** How can generative models lower the barrier of entry for creative tasks, enabling non-experts to produce photorealistic, stylistically rich, and controllable content at scale?
- (3) **Interpretability and Responsibility:** How can generative systems be made transparent and accountable, allowing their latent semantics to be understood, audited, and aligned with ethical and legal frameworks, especially in domains such as digital security and IP protection?
- (4) **Real-World Grounding:** How can we extend generative AI from synthetic benchmarks into multimodal and embodied contexts, where models must integrate visual, auditory, and physical cues, and operate reliably in real-world, open-ended environments such as robotics and interactive art?

Together, these questions articulate the long-term vision of my research program. They connect my foundational work on visual understanding, my contributions to generative modeling and interpretability, and my ongoing efforts in responsible and reality-centric generative AI.

As illustrated in Figure 1, the evolution of my research follows a coherent trajectory. The upper half of the figure captures my foundational work in **Visual Understanding** and **Visual Generation**, which established methods for perception, segmentation, synthesis, and manipulation of visual data. The lower half highlights my more recent directions in **Generative Understanding** and **Reality-Centric Generation**, where the emphasis shifts toward interpretability, responsibility, and embodied intelligence. This progression provides the context for the next sections: first, I present my *previous research accomplishments and the foundations they created*, and then I describe my *ongoing research directions and vision*, which extend these foundations toward responsible and reality-grounded generative AI.

2. PREVIOUS RESEARCH AND FOUNDATIONS

My early research laid the groundwork for machine perception, content creation, and model interpretability. This phase established the technical foundations upon which my current and future research directions are built. It can be divided into two interrelated strands: **Visual Understanding** and **Visual Generation**. Together, these contributions provided the basis for moving from perception to creation and toward controllable and responsible generative AI.

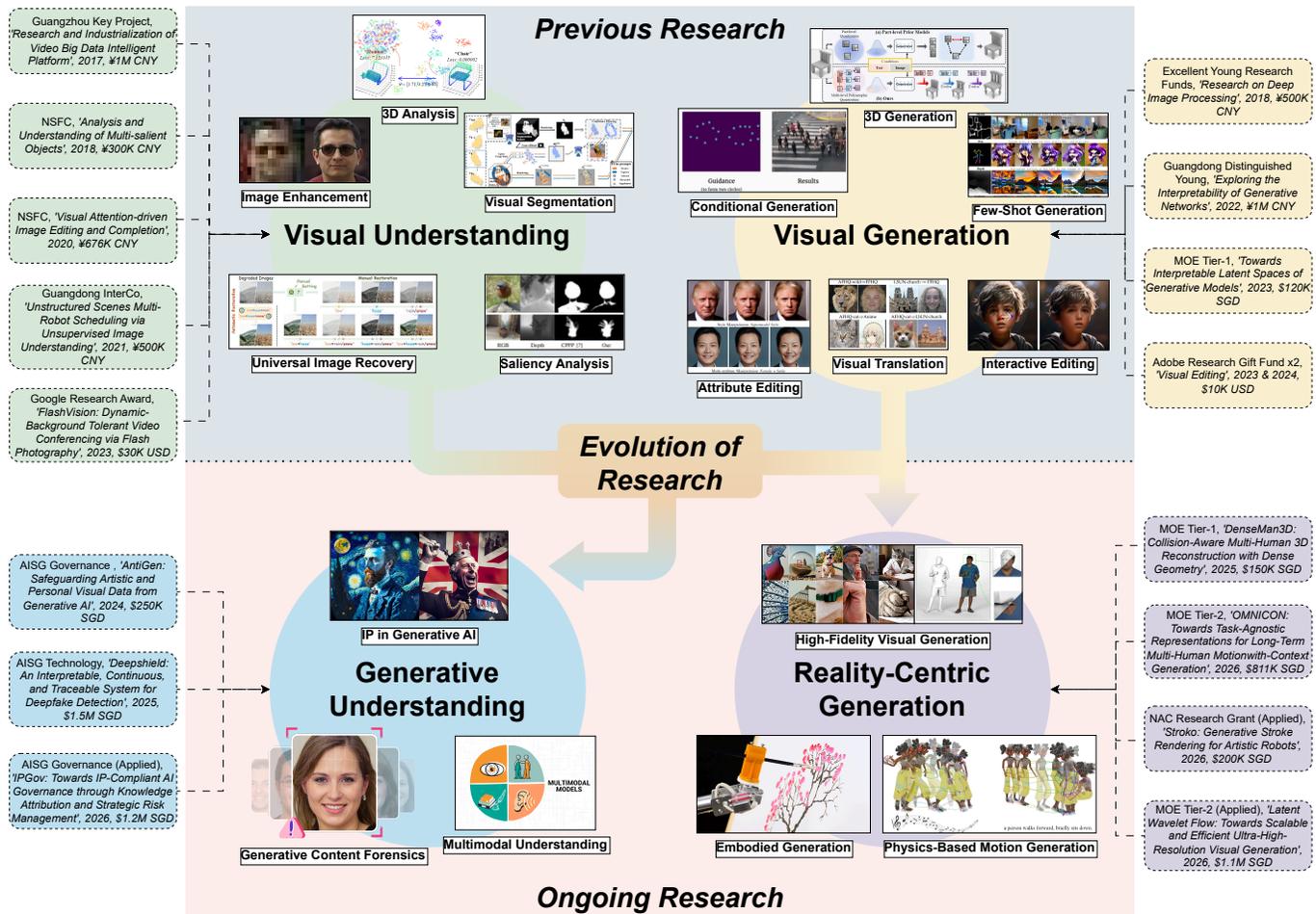


FIGURE 1. The evolution of my research from *Visual Understanding* and *Visual Generation* to ongoing directions in *Generative Understanding* and *Reality-Centric Generation*. The diagram illustrates the progression from perception, to creation, and now toward interpretability and real-world embodiment.

2.1. Visual Understanding. My earliest work investigated how humans perceive and prioritize information in visual scenes. I studied saliency detection by introducing alternative flash/no-flash stimuli to better model human attention [8], and developed one of the first deep learning frameworks for saliency detection [10]. To address efficiency, I proposed a real-time saliency detector capable of running at 30 FPS on a CPU [31]. Recognizing that perception extends beyond static imagery, I extended saliency to RGB-D [28, 26] and video domains [29, 32], enabling applications such as bullet chatting and visual question answering [7].

In parallel, I pursued object and scene analysis, beginning with visual tracking [11] and deformable tracking modules [25], then advancing orientation-aware proposals [9], stereo proposals [13], and egocentric temporal proposals [14]. Together, these contributions emphasized the importance of *multi-cue integration* (e.g., appearance, geometry, and motion) and *context-awareness* in achieving robust perception under diverse and challenging conditions.

This research agenda has since evolved toward richer representations of dynamic and multimodal scenes. Recent works advance saliency and detection into the multi-sensor domain through RGB-T fusion [30], extend depth reasoning to group-level and occlusion-heavy video scenes [19], and enable 3D visual grounding by decomposing structured multi-view information [12]. Viewed together, these studies form a coherent progression: from modeling isolated attention and objects, to understanding complex interactions and spatial structures in multimodal, dynamic, and three-dimensional environments.

This line of work was supported by several national and provincial projects in China during my early career, which enabled systematic investigations into perception under complex environments. In addition, after moving to Singapore, my flash/no-flash research on video conferencing received a *Google Research Award*, highlighting the international recognition and applied value of these early contributions.

2.2. Visual Generation. Building on the foundations of visual understanding, my research advanced into generative modeling, exploring both the creation of new content and the interpretability of generative models. This work established a bridge from perception to controllable generation and has been strengthened by both industry and competitive funding.

Creative Content Generation. A key motivation of my research has been to make visual content creation more accessible, controllable, and expressive. I developed methods for synthesizing and manipulating visual data, such as reflection removal [33], multi-view face synthesis [41, 36], and interactive crowd video synthesis with minimal user input (CrowdGAN) [1]. I also designed algorithms for artistic transformations, including deep pixelization [6, 34], cartoonization [18], and robust makeup transfer [3]. Collaborations with Tencent translated these ideas into practical applications such as snapshot video recovery from a single image [51], demonstrating the value of generative techniques in entertainment, social media, and beyond.

More recently, this line of work has evolved toward broader and training-free solutions that emphasize universality and artistic creativity. *Stroke2Sketch* leverages stroke attributes for training-free sketch generation [44], capturing stylistic essence without requiring paired datasets. *One-for-All* extends StyleGAN into a universal translation framework [4], achieving domain-agnostic image transformation with a single model. Viewed together, these contributions illustrate a coherent trajectory: from application-specific pipelines to general-purpose frameworks that democratize creative content generation.

This strand of work was further supported by three *Tencent Gift Funds* during my time in China and, more recently, two *Adobe Research Gift Funds* (2023, 2024), which enabled my group to expand its efforts in visual editing and dynamic video generation.

Interpretable Generative Models. As generative models grew in scale, a critical challenge emerged: lack of interpretability. I addressed this by uncovering latent semantics within pre-trained GANs, moving beyond binary attributes to discover richer and more nuanced editing directions [43]. To enable faithful editing of real data, I introduced the first video-based inversion method [37], preserving temporal continuity while allowing user-guided manipulation. I further explored generative priors, including progressive upscaling for up to $64\times$ enhancement [46], invertible rescaling frameworks [49], and disentangled identity/style spaces for cartoon and anime synthesis [35]. This research line was consolidated under competitive funding, including grants from China and the *MOE AcRF Tier 1 grant* on interpretable latent spaces, which supported advances in disentangled generative priors and informed my more recent works on diffusion inversion [40] and recurrent GAN inversion for video editing [38].

3. ONGOING RESEARCH

Building on the foundations of visual understanding, generation, and interpretability, my current research addresses pressing challenges at the frontier of generative AI. These efforts are organized into two complementary directions: **Generative Understanding**, which focuses on governance, interpretability, and responsibility, and **Reality-Centric Generation**, which pushes generative modeling into embodied, high-fidelity, and real-world applications. Both directions are supported by awarded projects under AI Singapore and MOE AcRF, as well as applied projects that expand my research trajectory into more in-depth exploration of interpretability and embodiment.

3.1. Generative Understanding. We approach generative understanding from three perspectives: (i) safeguarding artistic and personal visual data, (ii) continual and traceable detection of manipulated or malicious content, and (iii) robustness and interpretability as foundations for trustworthy generative pipelines.

This line of research is anchored by awarded and applied projects under AI Singapore. The *AI Singapore AntiGen project* focuses on protecting personal and artistic data against misuse by generative models. The *AI Singapore Deepshield project* develops continual and traceable detection of manipulated content, ensuring accountability and resilience against evolving deepfakes. The *AI Singapore IPGov project (applied)* extends these directions toward governance, investigating attribution and strategic risk management for IP-compliant generative systems.

Research outputs connect closely with these priorities. On privacy and protection, *G2Face* proposes high-fidelity reversible anonymization [42], enabling data sharing without sacrificing downstream utility, while *SITA* [15] reveals vulnerabilities in stylized image generation through structurally imperceptible adversarial attacks, motivating more defense-aware designs. On traceable and controllable generation, works such as *StyleGAN- ∞* [2] expand StyleGAN into a universal and modular translation framework, providing interpretable latent “StyleBooks” that improve attribution and control. On robustness and interpretability, [23] enhances cross-domain generalization via orientation learning, [21] exposes how latent structures influence one-shot rendering quality, and *Cross-Subject Mind Decoding* [39] demonstrates generative robustness by aligning brain signals with vision–language representations despite noisy inputs.

Finally, generative understanding also underpins **general visual recovery**. *Instruct2See* [17] develops instruction-based obstruction removal that generalizes across distributions, while *OneRestore* [5] unifies heterogeneous degradations into a single restoration pipeline. Together with modality-bridging approaches [16] and multimodal adaptation [24], these works demonstrate how interpretability and robustness can translate into practical pipelines for resilient visual recovery.

3.2. Reality-Centric Generation. We advance reality-centric generation from three perspectives: (i) high-fidelity 3D reconstruction, (ii) long-term context-aware motion generation, and (iii) embodied artistic generation in real-world environments.

This line of research is supported by a growing portfolio of awarded and applied grants. The *MOE AcRF Tier 1 DenseMan3D project* focuses on collision-aware reconstruction of dense multi-human motion. The *MOE AcRF Tier 2 OMNICON project* targets task-agnostic and long-term motion generation with contextual grounding. The *National Arts Council Stroko project (applied)* explores generative stroke rendering for artistic robots, bridging visual generation with embodied creativity. In parallel,

the *MOE AcRF Tier 2 Latent Wavelet Flow project (applied)* extends this trajectory toward scalable ultra-high-resolution image and video generation.

Recent works reflect progress across these perspectives. For 3D reconstruction, *NexusGS* [48] introduces sparse-view synthesis with epipolar depth priors, providing geometric consistency under limited observations. For motion generation, *Lagrangian Motion Fields* [45] develops a principled framework for stable long-horizon dynamics. For 2D-to-3D synthesis, *RecDreamer* [47] and *Stable Score Distillation* [50] extend diffusion-based models into three-dimensional domains, advancing consistency and fidelity in generative pipelines. For embodied artistic generation, *GenPoly* [22] and *Gaussian Prompter* [27] enable controllable, geometry- and prompt-aware outputs that serve as building blocks for interactive and robotic creativity. This direction also extends to the physical world, where studies on adversarial robustness [20] highlight the need for resilient generative perception when deployed in real environments.

These awarded and applied projects define a coherent trajectory in reality-centric generation: from reconstructing complex 3D environments, to generating long-term motion dynamics with contextual grounding, to embedding generative pipelines into embodied and artistic agents. This strand directly connects fundamental advances in generative modeling with national priorities in robotics, creative industries, and the safe deployment of real-world AI.

4. CONCLUSION

The evolution of my research reflects a consistent pursuit of expanding the boundaries of how machines perceive, create, and ultimately act in the world. From early explorations of visual perception to current advances in generative understanding and reality-centric generation, each phase has built on the last to address increasingly complex questions of interpretability, responsibility, and embodiment.

Looking forward, the significance of this trajectory lies not only in advancing technical capabilities, but in shaping generative AI as a trustworthy and creative partner in science, art, and society. By embedding governance into generative pipelines, grounding models in physical and multimodal realities, and preparing them for open-world deployment, my work aims to establish a foundation for a new generation of AI systems that are transparent, reliable, and impactful.

At the same time, my contributions extend beyond individual research outputs to leadership on the international stage. Through publications, editorial service, conference leadership, and collaborations, I seek to help define the global discourse on responsible generative AI and position Singapore as a visible hub in this rapidly evolving field.

Ultimately, my goal is to contribute to a future where generative intelligence enriches human creativity, enables resilient and responsible applications, and builds enduring bridges across disciplines and regions. This broader vision continues to guide my research and define its long-term impact.

REFERENCES

- [1] Liangyu Chai, Yongtuo Liu, Wenxi Liu, Guoqiang Han, and Shengfeng He. Crowdgan: Identity-free interactive crowd video generation and beyond. *IEEE TPAMI*, 2020. 3
- [2] Yihua Dai, Tianyi Xiang, Bailin Deng, Yong Du, Hongmin Cai, Jing Qin, and Shengfeng He. Stylegan- ∞ : Extending stylegan to arbitrary-ratio translation with stylebook. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 3
- [3] Han Deng, Chu Han, Hongmin Cai, Guoqiang Han, and Shengfeng He. Spatially-invariant style-codes controlled makeup transfer. In *CVPR*, pages 6549–6557, 2021. 3
- [4] Yong Du, Jiahui Zhan, Xinzhe Li, Junyu Dong, Chen Sheng, Ming-Hsuan Yang, and Shengfeng He. One-for-all: Towards universal domain translation with a single stylegan. *IEEE TPAMI*, 2025. 3
- [5] Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, 2024. 3
- [6] Chu Han, Qiang Wen, Shengfeng He, Qianshu Zhu, Yinjie Tan, Guoqiang Han, and Tien-Tsin Wong. Deep unsupervised pixelization. *ACM TOG*, 37(6):1–11, 2018. 3
- [7] Shengfeng He, Chu Han, Guoqiang Han, and Jing Qin. Exploring duality in visual question-driven top-down saliency. *IEEE TNNLS*, 31(7):2672–2679, 2019. 2
- [8] Shengfeng He and Rynson WH Lau. Saliency detection with flash and no-flash image pairs. In *ECCV*, pages 110–124, 2014. 2
- [9] Shengfeng He and Rynson WH Lau. Oriented object proposals. In *ICCV*, pages 280–288, 2015. 2
- [10] Shengfeng He, Rynson WH Lau, Wenxi Liu, Zhe Huang, and Qingxiong Yang. Supercnn: A superpixelwise convolutional neural network for salient object detection. *IJCV*, 115(3):330–344, 2015. 2
- [11] Shengfeng He, Qingxiong Yang, Rynson WH Lau, Jiang Wang, and Ming-Hsuan Yang. Visual tracking via locality sensitive histograms. In *CVPR*, pages 2427–2434, 2013. 2
- [12] Ronggang Huang, Haoxin Yang, Yan Cai, Xuemiao Xu, Huaidong Zhang, and Shengfeng He. Viewsr: 3d visual grounding via structured multi-view decomposition. In *ICCV*, 2025. 2
- [13] Shao Huang, Weiqiang Wang, Shengfeng He, and Rynson WH Lau. Stereo object proposals. *IEEE TIP*, 26(2):671–683, 2016. 2
- [14] Shao Huang, Weiqiang Wang, Shengfeng He, and Rynson WH Lau. Egocentric temporal action proposals. *IEEE TIP*, 27(2):764–777, 2017. 2
- [15] Jingdan Kang, Haoxin Yang, Yan Cai, Huaidong Zhang, Xuemiao Xu, Yong Du, and Shengfeng He. Sita: Structurally imperceptible and transferable adversarial attacks for stylized image generation. *IEEE TIFS*, 2025. 3
- [16] Guanzhou Ke, Shengfeng He, Xiao-Li Wang, Bo Wang, Guoqing Chao, Yuanyang Zhang, Yi Xie, and Hexing Su. Knowledge bridger: Towards training-free missing modality completion. In *CVPR*, 2025. 3
- [17] Junhang Li, Yu Guo, Chuhua Xian, and Shengfeng He. Instruct2see: Learning to remove any obstructions across distributions. In *ICML*, 2025. 3

- [18] Simin Li, Qiang Wen, Shuang Zhao, Zixun Sun, and Shengfeng He. Two-stage photograph cartoonization via line tracing. In *Computer Graphics Forum*, volume 39, pages 587–599, 2020. 3
- [19] Yuan Liang, Yang Zhou, Ziming Sun, Tianyi Xiang, Guiqing Li, and Shengfeng He. Instance-level video depth in groups beyond occlusions. In *ICCV*, 2025. 2
- [20] Guixu Lin, Muyao Niu, Qingtian Zhu, Zhengwei Yin, Zhuoxiao Li, Shengfeng He, and Yinqiang Zheng. Adversarial attacks on event-based pedestrian detectors: A physical approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 5227–5235, 2025. 4
- [21] Yihong Lin, Xuemiao Xu, Huaidong Zhang, Cheng Xu, Weijie Li, Yi Xie, Jing Qin, and Shengfeng He. Delving into invisible semantics for generalized one-shot neural human rendering. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 3
- [22] Bangzhen Liu, Yuyang Yu, Xuemiao Xu, Cheng Xu, Chenxi Zheng, and Shengfeng He. Genpoly: Learning generalized and tessellated shape priors via 3d polymorphic evolving. *IEEE TPAMI*, 2025. 4
- [23] Bangzhen Liu, Chenxi Zheng, Xuemiao Xu, Cheng Xu, Huaidong Zhang, and Shengfeng He. Rotation-adaptive point cloud domain generalization via intricate orientation learning. *IEEE TPAMI*, 2025. 3
- [24] Shanglin Liu, Jianming Lv, Jingdan Kang, Huaidong Zhang, Zequan Liang, and Shengfeng He. Modfinitly: Unsupervised domain adaptation with multi-modal information flow intertwining. In *CVPR*, 2025. 3
- [25] Wenxi Liu, Yibing Song, Dengsheng Chen, Shengfeng He, Yuanlong Yu, Tao Yan, Gehard P Hancke, and Rynson WH Lau. Deformable object tracking with gated fusion. *IEEE TIP*, 28(8):3766–3777, 2019. 2
- [26] Yuzhen Niu, Guanchao Long, Wenxi Liu, Wenzhong Guo, and Shengfeng He. Boundary-aware rgbd salient object detection with cross-modal feature sampling. *IEEE TIP*, 29:9496–9507, 2020. 2
- [27] Honghan Pan, Bangzhen Liu, Xuemiao Xu, Chenxi Zheng, Yongwei Nie, and Shengfeng He. Gaussian prompter: Linking 2d prompts for 3d gaussian segmentation. *IEEE TPAMI*, 2025. 4
- [28] Liangqiong Qu, Shengfeng He, Jiawei Zhang, Jiandong Tian, Yandong Tang, and Qingxiong Yang. Rgbd salient object detection via deep fusion. *IEEE TIP*, 26(5):2274–2285, 2017. 2
- [29] Sucheng Ren, Chu Han, Xin Yang, Guoqiang Han, and Shengfeng He. Tenet: Triple excitation network for video salient object detection. In *ECCV*, pages 212–228, 2020. 2
- [30] Hao Tang, Zechao Li, Dong Zhang, Shengfeng He, and Jinhui Tang. Divide-and-conquer: Confluent triple-flow network for rgb-t salient object detection. *IEEE TPAMI*, 2025. 2
- [31] Wei-Chih Tu, Shengfeng He, Qingxiong Yang, and Shao-Yi Chien. Real-time salient object detection with a minimum spanning tree. In *CVPR*, pages 2334–2342, 2016. 2
- [32] Bo Wang, Wenxi Liu, Guoqiang Han, and Shengfeng He. Learning long-term structural dependencies for video salient object detection. *IEEE TIP*, 29:9017–9031, 2020. 2
- [33] Qiang Wen, Yinjie Tan, Jing Qin, Wenxi Liu, Guoqiang Han, and Shengfeng He. Single image reflection removal beyond linearity. In *CVPR*, pages 3771–3779, 2019. 3
- [34] Zongwei Wu, Liangyu Chai, Nanxuan Zhao, Bailin Deng, Yongtuo Liu, Qiang Wen, Junle Wang, and Shengfeng He. Make your own sprites: Aliasing-aware and cell-controllable pixelization. *ACM Transactions on Graphics (TOG)*, 41(6):1–16, 2022. 3
- [35] Chenshu Xu, Yangyang Xu, Huaidong Zhang, Xuemiao Xu, and Shengfeng He. Dreamanime: Learning style-identity textual disentanglement for anime and beyond. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 3
- [36] Xuemiao Xu, Keke Li, Cheng Xu, and Shengfeng He. Gdface: Gated deformation for multi-view face image synthesis. In *AAAI*, volume 34, pages 12532–12540, 2020. 3
- [37] Yangyang Xu, Yong Du, Wenpeng Xiao, Xuemiao Xu, and Shengfeng He. From continuity to editability: Inverting gans with consecutive images. In *ICCV*, pages 13910–13918, 2021. 3
- [38] Yangyang Xu, Shengfeng He, Kwan-Yee Wong, and Ping Luo. Rigid: Recurrent gan inversion and editing of real face videos and beyond. *IJCV*, 2025. 3
- [39] Yangyang Xu, Bangzhen Liu, Wenqi Shao, Yong Du, Shengfeng He, and Tingting Zhu. Cross-subject mind decoding from inaccurate representations. In *ICCV*, 2025. 3
- [40] Yangyang Xu, Wenqi Shao, Yong Du, Haiming Zhu, Yang Zhou, Jiayuan Xie, Ping Luo, and Shengfeng He. Invert your prompt: Editing-aware diffusion inversion. *IJCV*, 2025. 3
- [41] Yangyang Xu, Xuemiao Xu, Jianbo Jiao, Keke Li, Cheng Xu, and Shengfeng He. Multi-view face synthesis via progressive face flow. *IEEE TIP*, 30:6024–6035, 2021. 3
- [42] Haoxin Yang, Xuemiao Xu, Cheng Xu, Huaidong Zhang, Jing Qin, Yi Wang, Pheng-Ann Heng, and Shengfeng He. G2face: High-fidelity reversible face anonymization via generative and geometric priors. *IEEE TIFS*, 2024. 3
- [43] Huiting Yang, Liangyu Chai, Qiang Wen, Shuang Zhao, Zixun Sun, and Shengfeng He. Discovering interpretable latent space directions of gans beyond binary attributes. In *CVPR*, pages 12177–12185, 2021. 3
- [44] Rui Yang, Huining Li, Yiyi Long, Xiaojun Wu, and Shengfeng He. Stroke2sketch: Harnessing stroke attributes for training-free sketch generation. In *ICCV*, 2025. 3
- [45] Yifei Yang, Zikai Huang, Chenshu Xu, and Shengfeng He. Lagrangian motion fields for long-term motion generation. *IEEE TPAMI*, 2025. 4
- [46] Zhou Yang, Yangyang Xu, Yong Du, Qiang Wen, , and Shengfeng He. Pro-pulse: Learning progressive encoders of latent semantics in gans for photo upsampling. *IEEE TIP*, 2022. 3
- [47] Chenxi Zheng, Yihong Lin, Bangzhen Liu, Xuemiao Xu, Yongwei Nie, and Shengfeng He. Recdreamer: Consistent text-to-3d generation via uniform score distillation. In *ICLR*, 2025. 4
- [48] Yulong Zheng, Zicheng Jiang, Shengfeng He, Yandu Sun, Junyu Dong, Huaidong Zhang, and Yong Du. Nexugs: Sparse view synthesis with epipolar depth priors in 3d gaussian splatting. In *CVPR*, 2025. 4
- [49] Zhixuan Zhong, Liangyu Chai, Yang Zhou, Bailin Deng, Jia Pan, and Shengfeng He. Faithful extreme rescaling via generative prior reciprocated invertible representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5708–5717, 2022. 3
- [50] Haiming Zhu, Yangyang Xu, Chenshu Xu, Tingrui Shen, Wenxi Liu, Yong Du, Jun Yu, and Shengfeng He. Stable score distillation. In *ICCV*, 2025. 4
- [51] Qianshu Zhu, Chu Han, Guoqiang Han, Tien-Tsin Wong, and Shengfeng He. Video snapshot: Single image motion expansion via invertible motion embedding. *IEEE TPAMI*, 2021. 3