SHIRONG YANG 「杨世荣」

yangshr2024@shanghaitech.edu.cn Shanghai, China Tel: (+86) 18183287493
Academic Homepage Sloan-Yang

EDUCATION

ShanghaiTech University Shanghai, China
Master of Science in Computer Science *2024 – Present*
GPA: 3.50/4.0; research in generative AI, computer vision, and graphic design generation.

Sichuan University Chengdu, China
Bachelor of Science in Artificial Intelligence *Sept. 2020 – Jul. 2024*
GPA: 3.67/4.0

HONORS & AWARDS

Academic Scholarship, ShanghaiTech University *2025–2026*
Outstanding Graduate, Sichuan University *2023–2024*
National Encouragement Scholarship (Top 3%), Sichuan University *2020–2021*
Tang Zhongying Moral Education Scholarship *Class of 2020*
Outstanding Student Leader, Sichuan University *2020–2021*

RESEARCH INTERESTS

Generative Modeling; Multimodal Understanding; Image Generation; Graphic Design Generation.

PUBLICATIONS

[1]. **Prior-Informed Design: Text-to-Design Generation with Multimodal Priors**
Shirong Yang, Ying Cao
*Under review at **ACM Transactions on Graphics (TOG)**, 2026.*

Data-efficient text-to-design generation using pretrained image and language models as multimodal priors.

[2]. **Learning Interaction between Image and Layout Priors for Joint Image-Layout Generation in Design Templates**
Shirong Yang, Bo Yang, Ying Cao
*Under review at **IEEE Transactions on Visualization and Computer Graphics (TVCG)**, 2026.* arXiv

Joint image-layout diffusion with learned backbone communication for faster, more faithful design template generation.

[3]. **Content-Edited Layout Generation for Graphic Design**
Shirong Yang, Zheyi Pan, Ying Cao
*Under review at **IEEE Transactions on Multimedia (TMM)**, 2026.* arXiv

Coupled flow matching for layout-aware image editing and layout generation from raw visual assets.

RESEARCH EXPERIENCE

ShanghaiTech University Shanghai, China
Graduate Researcher *2024 – Present*

Research on multimodal generative models for editable graphic design, structured visual representations, and coordinated image-layout synthesis.

[1]. **Prior-Informed Design: Text-to-Design Generation with Multimodal Priors**
Project Lead; text-to-design generation with pretrained multimodal priors

- Proposed **PriDe**, a data-efficient framework that uses frozen text-to-image models and large language models to generate structured, editable graphic designs.
- Developed residual generative modeling and prior-feature injection in a learned design latent space, reusing pretrained knowledge of appearance, color, layout, and typography.
- Achieved high-quality generation using only **10K layered designs** while retaining competitive quality in further reduced-data settings.

[2]. **Learning Interaction between Image and Layout Priors for Joint Image-Layout Generation in Design Templates**
Project Lead; joint image-layout diffusion and cross-modal interaction learning

- Proposed **InterIL**, a joint diffusion framework that simultaneously generates background images and foreground layouts instead of relying on a sequential design pipeline.
- Connected frozen image and layout diffusion backbones with a learnable **bidirectional communication module** that models cross-modal interaction while preserving strong single-modality priors.
- Developed training-free inference guidance for incorporating user preferences without retraining the generative model.

[3]. **Content-Edited Layout Generation for Graphic Design**
Project Lead; joint image editing and layout generation with flow matching

- Formulated **content-edited layout generation**, which jointly edits a background image and generates a foreground layout from a textual intention.
- Proposed **Search-Compare-Edit (SCE)**, an automatic pipeline for constructing paired image-editing data and textual editing instructions from large-scale image collections.
- Developed a two-flow multimodal **flow-matching** framework that couples background-image editing with foreground-layout generation.

TECHNICAL SKILLS

Programming	Python, C/C++, Rust
Generative Modeling	Diffusion Models, Flow Matching, VAE
Vision & Multimodal	PyTorch, Transformers, CLIP, DINO, Vision-Language Models
Graphics & Interactive	Unity, Unreal Engine, Bevy, Blender
Research Tools	Linux, Git, LaTeX, Hydra, Weights & Biases

LANGUAGES

Chinese: Native **English**: CET-6; Proficient in academic writing