

Yuanting Fan

(+86) 183-9112-6180 | fan0yuanting@gmail.com
AIGC Researcher · Tencent Youtu Lab
[retofan233.github.io](https://github.com/retofan233) | retofan-blog.pages.dev | [Google Scholar](https://scholar.google.com/citations?user=...)
Xi'an Jiaotong University M.S. in Software Engineering | Sep. 2022 - Jun. 2025 Chang'an University B.S. in Software Engineering | Sep. 2018 - Jun. 2022

Work Experience

Focus on 3D texture generation, industrial defect synthesis, and low-cost AI comic / short-drama production. Independently own the full stack from data and training to evaluation and launch. Awarded 2026H1 Outstanding.

 Tencent Technology (Shenzhen) · Youtu Lab

Researcher | Jun. 2025 - Present

Youtu3D (YT3D) | Single- and Multi-Reference Texture Generation

- Scope:** Within two months, independently drove the full Youtu3D pipeline for single- / multi-reference texture generation — data, training, evaluation, and release — targeting industry-leading fine-text rendering, multi-image reference following, and real-world physical fidelity.
- Data:** Built a distributed Blender PBR rendering pipeline producing 1.5M+ multi-view high-quality renders; introduced an OCR model to augment text-scene data, and a Blender pipeline for transparent / hollow asset synthesis to mitigate long-tail coverage.
- Model & training:** Multi-stage resolution-scaling training. Built the Klein 4B/9B multi-reference albedo framework with view-pose injection: view-dependent attention lets the target albedo interact only with same-orientation references, plus an illumination-consistency constraint to cut cross-view color drift and seams.
- Transparency / cutouts (MRT):** Led the industry's first texture model for transparent / hollow 3D assets. MRT jointly predicts metallic, roughness, and transparency on a multi-channel material map, separates cutout from transmission, and writes the result back to GLB.
- Scale:** Multi-round SFT / DPO / GRPO on a 300+ heterogeneous GPU cluster; 50+ GSB evaluation rounds with 200+ raters.
- Impact:** Leads the field in fine-text rendering, multi-image reference following, and real-world physical fidelity; objective metrics exceed the best in-industry models by over 30%; single- / multi-reference generation, including transparency / cutouts, is live on Tencent Cloud (China and international), with customers clearly preferring it over the latest production Hunyuan and Tripo models; awarded 2026H1 Outstanding.

UniDG | Universal Defect Synthesis Foundation Model [arXiv:2604.08915](https://arxiv.org/abs/2604.08915)

- Problem:** Industrial inspection lacks abnormal samples, and few-shot defect generators overfit to specific categories. Goal: a universal defect synthesis model that needs no per-category fine-tuning and supports both reference transfer and text-instruction editing.
- Data:** Built UDG with 300K normal-abnormal-mask-caption quadruplets from 50 public datasets and 269 raw defect types (mapped to 28 categories), ~23× the previous largest defect-generation set; a multi-agent pipeline (Inpainting / Captioner / Verifier) produced paired, verified samples.
- Model:** Based on FLUX.1-Fill-dev; proposed Defect-Context Editing (adaptive defect cropping + structured diptych input) and fused reference / target conditions via MM-DiT multimodal attention.
- Training:** Two-stage Diversity-SFT + Consistency-RFT (Flow-GRPO with dual reward models), balancing diversity and reference consistency for zero-shot cross-category transfer.
- Impact:** SOTA synthesis quality and downstream detection / localization on MVTec-AD and VisA; multi-class FG-mIoU +20.5% vs. AnoGen. Deployed on Tencent Cloud TIAOI for multiple B-side clients. Submitted to NeurIPS 2026.

AI Comic / Short-Drama Platform | Novel → Storyboard → Final Cut

- Scope:** Built the full platform-to-model stack for low-cost AI comic / short-drama generation, serving internal teams and B-side clients.
- Pipeline:** One-click generation (novel → auto storyboard → image / video generation → online FFmpeg editing). Python backend + Vue frontend + OpenAI-compatible model services, integrating multimodal understanding, image/video generation, video SR, and frame interpolation.
- Engineering:** Unified model gateway and task queue with storyboard-level retry, asset caching, and export; wrapped open-source understanding / generation models as pluggable services.

Joint Audio-Visual Video Generation | Multi-Speaker DPO

- Data & reward:** Built speaker-level preference pairs for multi-speaker scenes; used voiceprint-face embedding alignment as a reward to constrain lip sync, identity, and timbre.
- Training:** Multi-scale DPO on LTX-2.3-22B (global quality / speaker / segment), jointly optimizing overall quality and character-level audio-visual alignment.

HD Any2Any Face Swap | Wan2.2 + 4-Step Distillation

- Data:** Large-scale 4K face video slices (short side >768px) via Ray Data + LanceDB for efficient retrieval and reproducible training.
- Training:** Multi-stage progressive training (256 → 768px) + mixed sampling + LightX2V 4-step distillation on Stand-In Wan2.2-T2V-A14B; internal eval substantially leads prior SOTA on reference fidelity and facial dynamics.

Internship

 Tencent Technology (Shenzhen) · Youtu Lab

Research Intern | Apr. 2024 - Oct. 2024

Multimodal LLM for Industrial Inspection | Tool Use + Anomaly Detection

- Ownership:** Led the full pipeline of a multimodal LLM for industrial inspection: built instruction-tuning and tool-calling data, then ran SFT and human-feedback learning with LLaMA-Factory.
- Tools:** Trained backend tools on academic / industrial anomaly datasets, including MM-GroundingDINO (open-set detection) and WinCLIP (anomaly detection); served via vLLM with Gradio / Streamlit UIs.
- Follow-on:** Work on fine-grained VLMs, open-set detection, and interactive visual detection later became the first-author PR 2026 paper FineGrainedAD.

Selected Publications

NeurIPS 2026 (submitted) / First Author Large-Scale Universal Defect Generation [arXiv:2604.08915](#)

- **Problem:** Abnormal samples are scarce in industrial and medical settings; few-shot defect generators overfit to categories and transfer poorly across objects.
- **Contribution:** Built UDG (300K normal-abnormal-mask-caption quadruplets from 50 datasets and 269 defect types) and UniDG on FLUX.1-Fill-dev with Defect-Context Editing — reference transfer and text-instruction editing without per-category fine-tuning.
- **Results:** Leads synthesis quality and downstream detection / localization on MVTec-AD and VisA; multi-class FG-mIoU +20.5% vs. AnoGen. Deployed on Tencent Cloud TIAOI.

Pattern Recognition 2026 / First Author Towards Fine-Grained Vision-Language Alignment for Few-Shot Anomaly Detection

- **Problem:** Prior VLM few-shot detectors match image-level text to patch tokens, causing semantic misalignment and weak localization.
- **Contribution:** MFSC auto-builds multi-level fine-grained captions; FineGrainedAD uses learnable prompts (MLLP) and multi-level alignment (MLSA) so prompts lock onto visual regions.
- **Results:** Stronger overall localization than prior methods in few-shot settings on MVTec-AD and VisA.

ECCV 2024 / First Author AdaDiffSR: Adaptive Region-aware Dynamic Acceleration Diffusion Model for Real-World Image Super-Resolution

- **Problem:** Diffusion SR uses a uniform denoising budget over the whole image, wasting compute on invisible details.
- **Contribution:** MMLE estimates regional latent information gain to pick timesteps dynamically; PFJ injects original features for fidelity and realism.
- **Results:** Quality comparable to then-SOTA diffusion SR on synthetic and real data at about 60% of StableSR's compute.

IEEE TCSVT 2025 / First Author InstanceSR: Efficient Reconstructing Small Object with Differential Instance-level Super-Resolution

- **Problem:** Conventional SR optimizes whole-image appearance and under-allocates compute to small-object regions, hurting downstream detection.
- **Contribution:** Encodes the image into difficulty tokens by instance distribution; a multi-exit decoder spends more compute on foreground instances and less on background.
- **Results:** Outperforms prior SR methods on small-instance benchmarks and consistently boosts small-object detection.

AAAI 2024 / Student 2nd Author Decoupling Degradations with Recurrent Network for Video Restoration in Under-Display Camera

- **Problem:** UDC video mixes time-varying flare and haze while requiring temporal consistency.
- **Contribution:** D2RNet decouples flare / haze with dedicated attention; built a large UDC HDR video benchmark from measured PSFs.

Awards, Patents & Skills

- **Awards & patents:** First Prize, Northwest Region, China Graduate Electronic Design Competition (2023); National Third Prize, China Software Cup (2021); multiple patents filed.
- **Algorithms:** Diffusion / Flow Matching, SFT / DPO / Flow-GRPO, 3D texture generation, multimodal LLMs, anomaly detection and defect synthesis, super-resolution.
- **Engineering:** Python, PyTorch, LLaMA-Factory, vLLM, Ray, Blender PBR rendering, Docker, Linux, Git. Independently owns the full path from data to training, evaluation, and production serving (One Person Team).
- **Interests:** Agent applications, low-cost AI comic / short-drama generation, unified generation-understanding models, and world models. Happy to connect by email.