KEWEI ZHANG

(+86) 155-2770-6812 | **Homepage** | xiwen.kwzhang@gmail.com | zkwsdsg | github.com/xiwen1

A fourth-year CS undergraduate student at Wuhan University, expected to enroll at Peking University in Fall 2026, supervised by Prof. Daquan Zhou. My research interests lie in **AIGC**, **Efficient AI**, and **AI safety**, with a continuous passion and ambition to explore the cutting edge of AI.

BASIC INFORMATION

Wuhan University — Expected Graduation Date: June 2026

Bachelor of Software Engineering — *GPA: **3.92 / 4.0** | Rank: **1 / 235***

- **Academic Competence:** Fundamentals of Computer Graphics (98), Probability and Statistics (93), Linear Algebra (95), Discrete Mathematics (97), Object-Oriented Programming (94), Computer Architecture and Design (90), Database Systems (93), Software Requirements and Modeling (97), Systems Programming (93), Technical Writing (95), Digital Image Processing (94), Business Intelligence (94).
- **Programming Skills:** Proficient in **Python, Rust, Go, C#**, and **JavaScript**, along with relevant development frameworks and deep learning tools; Familiar with **GPU programming** with **CUDA** and **Triton**. I am skilled in multi-GPU parallel training methods such as **Accelerator** and **DeepSpeed**, with over **2000** hours of coding experience. [Programming Time Stats]
- **Engineering Experience:** I have extensive experience in building well-structured deep learning code and full-stack web development (Frontend: React.js, Next.js; Backend: Axum, Gin, Django). I have multiple open-source projects on GitHub [GitHub Profile], with over 100 stars.
- **Language: CET6 586**; Independently completed several academic papers in English. [Google Scholar]

SELECTED PUBLICATIONS

MHLA: Restoring Expressivity of Linear Attention via Token-Level Multi-Head

ICLR 2026 | ***First Author*** | *[Project Page]*

- **Background:** Existing linear attention methods suffer from "global context collapse", where compressing all tokens into a single summary limits rank and oversmooths attention, leading to degraded performance in long-sequence (high-resolution) tasks and large-scale models.
- **Method:** Proposed **Multi-Head Linear Attention (MHLA)**, which: 1) partitions sequences into token-level blocks with local KV summaries; 2) introduces a learnable **multi-head mixing matrix** to construct query-conditioned summaries and restore expressivity; 3) naturally supports chunkwise parallel training with linear complexity.
- **Results:** Achieved up to **+3.6%** Top-1 accuracy on ImageNet over self-attention, **+12.6%** FID improvement in image generation compared to DiT, and **+2.1%** Improvement over GLA in LLM, while maintaining throughput comparable to other linear attention baselines.

Revisiting Adversarial Patches for Designing Camera-Agnostic Attacks against Person Detection

NeurIPS 2024 | ***Co-first Author*** | *[Project Page]*

- **Background:** Existing patch-based pedestrian detection attacks face challenges in physical deployment due to camera ISP effects, causing up to 47.3% variance in attack success rates. We are the first to incorporate camera ISP modeling into the adversarial patch optimization framework to achieve camera-agnostic physical attacks.
- **Method:** 1) We propose a **differentiable ISP proxy network** to model nonlinear color space mappings across different cameras. 2) A **cooperative adversarial optimization framework** trains the attack and ISP modules to minimize the attack's effectiveness while maximizing detection failure. 3) We build a **multi-device evaluation benchmark** for comprehensive physical attack testing.
- Our method achieves a **15%** higher success rate than SOTA across multiple cameras. The adversarial optimization framework boosts physical attack success by **31.5%**, with results published in a patent.

FAB-Attack: Fabric-Aware Adversarial Attacks on Person Detectors under Motion Blur

ACMMM 2025 | ***Co-First Author***

- **Background:** Physical adversarial attacks expose vulnerabilities in vision systems used in safety-critical applications like autonomous driving and surveillance. While recent methods improve attack effectiveness, they overlook realistic garment deformations and motion blur, limiting real-world applicability. This work introduces FAB-Attack, a new adversarial attack method simulating realistic garment deformations and targeting both person detectors and image deblurring models.
- **Method:** 1) We propose a **Fabric-aware Texture Appliance (FTA)** module that applies adversarial textures to clothing regions, simulating realistic fabric dynamics with physics-inspired TPS; 2) We design a differentiable pipeline incorporating motion blur and deblurring processes to emulate real-world conditions, showing the stability of low-frequency information during motion blur; 3) We introduce a **frequency band separation mechanism** to suppress high-frequency components in adversarial patterns, enhancing robustness against motion blur.
- Experimental results show that FAB-Attack achieves state-of-the-art performance, reducing AP to 25.2% on the COCO dataset and achieving 94.4% attack success rate in the real world under severe motion blur.

Adversarial Watermarking under Diverse Manipulations: Benchmarking and Beyond.
IEEE TIP, Under Review | ***Co-first Author***

- **Background:** 1) Existing watermarking protection methods are vulnerable to composite attacks, with most defenses dropping over 30% effectiveness after common transformations (rotation, cropping, compression). 2) Current evaluation metrics like SSIM/PSNR are poorly correlated with human visual perception, limiting their effectiveness in assessing watermark defenses.
- **Method:** 1) We introduce a **transformation ensemble module** (TEM) that parameterizes 9 differentiable image operators, generating composite attack sequences for improved watermark robustness. 2) We design an attention-guided perturbation allocation strategy to enhance watermark concealment. 3) We also propose **watermark transparency consistency simulation** (ACE) to accurately assess watermark protection by fitting transparency data through the watermark removal network.
- Our method achieves 95.3% watermark retention on the CLWD dataset, a 23.7% improvement over SOTA. In tests on open-source models, we maintain 91.4% defense success in white-box and 84.6% in black-box scenarios, with 65.7% success against commercial APIs.

Moiré Backdoor Attack (MBA): A Novel Trigger for Pedestrian Detectors in the Physical World
ACM MM 2023 | ***Third Author***

- **Background:** 1) Current backdoor attacks on pedestrian detection systems lack targeted methods for dynamic human targets. 2) Traditional spatial-domain triggers have visible flaws, easily detectable in physical deployment. 3) Limited research on using natural phenomena as triggers, failing to leverage imaging system characteristics.
- **Method:** We propose using the Moire effect, a common phenomenon caused by camera sensors, as a backdoor trigger. The method models and simulates various real-world transformations to generate Moire-poisoned samples in pedestrian datasets, inducing backdoor effects across multiple detector models.
- Our method achieves a 77.3% attack success rate with only 10% of poisoned samples. A user survey shows only 8.3% detection, a 73.4% improvement over existing methods. Results published in a patent.

Subjective Camera: Training-Free Scene Sketch-Guided Generation and Spatial Reward Optimization
ICCV 2025 | ***Fifth Author***

- **Background:** Text-to-image models face challenges in spatial control: 1) Pure text inputs fail to describe complex scene geometries; 2) Sketch control methods require high-quality contours, limiting accessibility; 3) Multi-object generation suffers from semantic-geometric misalignment. This research develops a training-free bimodal control framework for joint sketch-text optimization.
- **Method:** 1) We propose a **hierarchical perceptual optimization framework** with CLIP/BLIP semantic alignment, PickScore quality loss, and aesthetic evaluation, iteratively optimizing in latent space; 2) We design an **incremental layout generation algorithm**, allowing users to input object contours and positions for gradual scene construction; 3) We introduce a **doodle realism module** using DDIM to map doodles to real images, enabling efficient style transfer with low iteration steps, addressing distortion from rough sketches.
- Our method addresses non-professional users' needs, offering an efficient solution for subjective image reconstruction. In multi-object scene creation, it outperforms other methods in user satisfaction.

SELECTED AWARDS

Lei Jun Computer Scholarship (20 of 1600) | *Wuhan University* — November 2024
National Scholarship (2 of 522) | *Ministry of Education* — October 2023
Honorable Mention | *COMAP's 2024 MCM/ICM.* — April 2024
National First Award | *The 18th China "Challenge Cup" Competition.* — September 2023