Wanzhou Lei

(917) 834-0120 | wanzhou_lei@berkeley.edu | Berkeley, CA

Education

University of California, Berkeley, Ph.D. in Applied Mathematics, GPA 3.94/4.0 — Sep 2024 – May 2028
Harvard University, M.S. in Data Science, GPA 4.00/4.0 — Sep 2022 – May 2024
New York University, B.S. in Mathematics and Computer Science, GPA 3.98/4.0 — Sep 2018 – May 2022

Internship Experience

Alibaba – Research Scientist Intern (Generative Vision), Beijing, China — May 2026 – Sep 2026

- Led RAESR, one-pass real-world super-resolution in a frozen vision-foundation latent; first-authored an under-review paper.
- Extended RAESR to video: a small 3.5M-parameter space-time transformer takes the per-frame RAESR outputs, aligns neighboring frames with frozen optical flow, and corrects flicker and inconsistent detail across frames, trained with losses that compare frames after motion alignment. The architecture is deployed in production in HappyHorse, Alibaba's flagship video generation model.

ASML – Source Mask Optimization (SMO) R&D Team, San Jose, CA — May 2025 – Aug 2025

- Led the replacement of the conventional lithography simulation workflow, which is non-smooth and hard to differentiate, with a residual U-Net that predicts lithography images from phi images, enabling end-to-end differentiable source-mask optimization.
- Partnered cross-functionally on data collection across diverse lithography conditions to build high-quality training data and improve generalization, and integrated the U-Net into ASML's high-performance C++ optimization framework for production use.
- Delivered substantial reductions in EPE (edge placement error) loss and faster convergence in gradient-based yield optimization.

Research Experience

RAESR: Real-World Super-Resolution in a Frozen Vision-Foundation Latent (Alibaba) — May 2026 – Oct 2026

- Led RAESR, a one-pass real-world super-resolution model that restores inside the fused latent of a frozen DINOv3-L encoder with a 415M-parameter adversarially trained decoder; it achieves state-of-the-art perceptual quality on four real-world benchmarks and the best perception–fidelity trade-off among all published methods, without any diffusion prior or sampling loop.
- Reduced inference to 719M parameters, 37 ms and 1.6 GiB per 512px image, 1.6× faster than the fastest one-step diffusion baseline and 59–514× faster than 20–50-step pipelines, while an 18B-parameter pipeline trails on all nine metrics; distilled the frozen encoder into a 3.5× smaller frontend without retraining the decoder (502M parameters, 30 ms, near-teacher quality).
- Rigorously analyzed the mechanism behind RAESR: the frozen encoder's shallow layers preserve pixel fidelity while its deep layers carry degradation-robust semantics, verified by layer ablations and a controlled comparison against a VAE latent.
- Designed the adversarial training recipe: a multi-scale critic on frozen DINOv3 features with per-layer convolutional heads, a relativistic loss with R1/R2 regularization, and diagnostics for critic saturation and patch-grid shortcuts that keep training stable.

SWIFT: Self-Guided Weak-to-Strong Diffusion Training via Target Extrapolation (Alibaba) — May 2026 – Oct 2026

- Co-developed SWIFT, which moves weak-to-strong guidance from sampling into training: each clean-sample regression target is extrapolated away from a frozen weak denoiser's prediction, so the correction is absorbed during training and the guide is discarded at inference, leaving one denoiser and the original sampler at about 9% extra training FLOPs.
- Proved, via a spectral analysis of finite-time population gradient flow (Gaussian-mixture data, RKHS dynamics), that a nonempty range of extrapolation strengths yields a denoiser strictly closer to the posterior mean than matched standard training: the correction partially cancels the finite-time shortfall on every learnable mode, whereas an overly large strength overshoots the posterior mean.
- One fixed recipe lowers unguided FID on ImageNet-256 in all 18 DiffusionBench configurations (16–49%, median 31%) across VAE, RAE, end-to-end latent and pixel diffusion; on RAEv2/DINOv3-L it reaches 1.15 FID unguided and 0.99 with retuned guidance, matching the 320-epoch baseline at 0.45× the training FLOPs, and remains complementary to autoguidance and CFG.

Text Encoder Transplantation for Frozen Image Generators — Jan 2026 – Aug 2026

- Co-developed TET, an image-free behavioral-distillation recipe that transplants an arbitrary text encoder into a frozen rectified-flow text-to-image generator: a resampler bridge plus LoRA adapters trained from captions alone by matching teacher velocities along the trajectory, preserving classifier-free-guidance geometry, and backpropagating through a differentiable 4-step rollout.
- Replaced the native encoder of FLUX.2-klein with a 5× smaller one at 96–98% of native alignment, shrank SD3.5's encoder stack by 72% at near-native quality, and made the frozen generator multilingual with zero-shot generation in 12 languages.

A Gradient Flow Analysis into Cross-Attention Dynamics, UC Berkeley — Dec 2025 – Present

- Analyzed cross-attention training under gradient descent and gradient flow on text-image classification (text queries, image-patch keys and values), giving a three-stage convergence framework for multimodal alignment informing CLIP- and LLaVA-style training.
- Proved the three-stage process: rapid convergence of the MLP head, nontrivial evolution of cross-attention scores that captures cross-modal correspondence, and final convergence guaranteeing successful classification; validated the predicted dynamics experimentally.

Reinforcement Learning for 3D Tetrahedral Mesh Refinement, UC Berkeley — Aug 2025 – Present

- Built an RL environment for topology-only tetrahedral mesh optimization using always-valid local bistellar flips, with an E(n)-equivariant GNN actor-critic that scores candidate edge and face operations to improve local mesh quality.

Physics-Informed Neural Networks (PINNs) for Nonlinear Differential Equations, Harvard University — May 2023 – Aug 2024

- Developed a one-shot transfer-learning method for nonlinear ODEs and PDEs with PINNs, improving efficiency by 50% over traditional numerical methods; presented at the NeurIPS 2023 workshop on deep learning and differential equations.

Publications

- Lei, W., Shen, C., He, Y. et al. Super-Resolution in The Right Latent Space: A Frozen Vision-Foundation Substrate. Under review, 2026.
- He, Y., Shen, C., Lei, W. et al. SWIFT: Self-guided Weak-to-strong Diffusion Models via Target Extrapolation. Under review, 2026.
- Yuan, H., Yuan, J., Lei, W. et al. Match the Behavior, Not the Features: Transplanting Text Encoders into Frozen Image Generators. Under review.
- Lei, W. et al. Why Transformers Can Successfully Learn Multimodal Data? A Theoretical Analysis into Cross-Attention Dynamics. In preparation.
- Lei, W., Nurbekyan, L. et al. Efficient Natural Gradient Descent Methods for Large-Scale PDE-Based Optimization Problems. *SISC*, 2023.
- Lei, W., Protopapas, P. et al. One-Shot Transfer Learning for Nonlinear ODEs. *NeurIPS 2023 DLDE Workshop*.