FG 2026 Doctoral Consortium Research Statement

Allen Tu
University of Maryland, College Park

Proposed Thesis Title: *Selective Computation for Reliable and Scalable Vision Systems*

1. Thesis Research Topic

Modern vision systems increasingly operate in unconstrained, mission-critical environments where failures are costly and observations are imperfect. In programs such as IARPA Biometric Recognition and Identification at Altitude and Range (BRIAR) and IARPA Walk-Through Rendering from Images at Varying Altitude (WRIVA), systems must function under degraded imagery, incomplete supervision, distribution shift, and finite computational budgets. In these regimes, scaling alone often fails to improve reliability and can destabilize system behavior, motivating principled mechanisms for determining what information to trust.

This dissertation advances the thesis that **reliable perception emerges when computation, data, and modeling capacity are selectively applied where they improve performance**. Rather than treating efficiency and robustness as competing objectives, my work frames selective computation as the mechanism that unifies them. Across 3D and 4D reconstruction, recognition, and generative enhancement, I develop vision systems that determine *what to compute, what to trust, and when additional modeling capacity should be invoked*.

2. Progress to Date

My completed work – including publications at CVPR and FG – establishes a unified research program on *selective computation mechanisms* for vision systems, demonstrating that principled selective computation enhances efficiency, robustness, and system reliability under real-world constraints.

Contribution-Aware Pruning. In *PUP 3D-GS* (CVPR 2025), I introduce a per-Gaussian pruning criterion derived from a second-order approximation of the ℓ_1 reconstruction loss to estimate each primitive's contribution to rendered images. While inspired by Hessian-based uncertainty estimation methods for 3D Gaussian Splatting (3DGS), parametric uncertainty correlates weakly with reconstruction error in practice. I therefore reinterpret the second-order signal as a *sensitivity score* that directly measures each Gaussian's influence on rendered images, enabling post-hoc removal of over 90% of primitives from pretrained 3DGS models. This substantially exceeds prior heuristic pruning strategies while preserving salient foreground structure and perceptual fidelity, revealing significant structural redundancy in explicit radiance field representations and establishing sensitivity-based pruning as a principled foundation for scalable 3D vision systems.

Scalable Systems Integration. *Speedy-Splat* (CVPR 2025) extends contribution-aware pruning from compression toward rendering throughput by addressing fundamental inefficiencies in the 3DGS pipeline. The standard renderer overestimates Gaussian-to-tile intersections, assigning primitives to tiles they do not influence and incurring substantial redundant computation. We introduce *AccuTile*, a drop-in rasterization algorithm that precisely localizes Gaussians and assigns them only to intersecting tiles, independent of the training procedure. We further enable pruning during training by reducing the storage requirement of the PUP 3D-GS sensitivity score by $36\times$, making second-order contribution estimation practical within the optimization loop. AccuTile alone improves rendering speed by $2\times$ while preserving visual fidelity; combined with training-time pruning, the system achieves over $6\times$ acceleration with more than $10\times$ fewer primitives.

Selective Temporal Modeling and Motion Structure. Dynamic 3DGS achieves high-fidelity reconstruction through per-Gaussian neural deformation, but its inference cost limits real-time scalability. In *SpeeDe3DGS* (CVPR 2026), I extend sensitivity-aware pruning to the temporal domain by aggregating contribution across time, preserving briefly visible details while compressing scenes by over $10\times$. Temporal perturbation during sensitivity estimation exposes unstable primitives that drift at unseen timestamps, suppressing floaters and improving coherence. *GroupFlow* further distills per-Gaussian neural motion into grouped SE(3) transformations by discovering shared trajectory structure, reducing deformation cost and stabilizing trajectories. Together, these mechanisms boost rendering speed by over $13\times$ while maintaining neural-field image quality.

Selective Evaluation for Recognition. In *TransFIRA* (FG 2026), I address the limitations of conventional Face Image Quality Assessment (FIQA) by defining recognizability directly in the decision geometry of the deployed face encoder, rather than through visual heuristics or curated annotations. I develop a transfer-learning framework that predicts recognizability using a lightweight regression head, enabling seamless integration into existing biometric pipelines with substantial component reuse. TransFIRA further introduces a recognizability-aware template aggregation strategy that filters unrecognizable probes and weights retained features, simultaneously improving accuracy while reducing unnecessary feature extraction and matching. On the proprietary IARPA BRIAR Protocol 3.1 surveillance benchmark, an ArcFace model trained solely on WebFace42M improves from 0.4269 to 0.8715 TAR at 10^{-3} FMR under TransFIRA aggregation, despite the pronounced distribution shift between public training data and operational imagery. This approach has been adopted within the operational BRIAR pipeline, where recognizability-based filtering replaced a computationally expensive keypoint-driven procedure, demonstrating that encoder-grounded selective computation improves both recognition reliability and system efficiency in real-world deployments.

Selective Generative Refinement. While pruning-based selection eliminates unnecessary computation, some reconstruction failures arise from insufficient supervision rather than redundancy. *SplatSuRe* (CVPR 2026) addresses this regime by governing where generative detail should be introduced to preserve multi-view consistency. Instead of uniformly applying super-resolution to all training images, SplatSuRe uses scene geometry and camera pose to identify regions where high-frequency ground-truth information is available and where it is fundamentally missing. This geometric scoring produces pixel-level maps that selectively apply super-resolution only in undersampled regions while relying on consistent low-resolution observations elsewhere. By suppressing generated detail in well-supervised areas, SplatSuRe avoids the view-dependent inconsistencies and blurring of uniform super-resolution, yielding sharper and more stable reconstructions.

Collectively, these works position selective computation as a unifying principle for reliable vision systems.

3. Working Plan and Remaining Dissertation Work

My dissertation investigates how diffusion priors can be integrated into 3D reconstruction pipelines in a controlled, uncertainty-aware manner, elevating generation from heuristic enhancement to a reliability-governed modeling tool. Experiments within the WRIVA program reveal a central failure mode of diffusion-in-the-loop reconstruction: unregulated synthetic views can overwhelm reliable real observations and degrade fidelity despite the expressive power of modern generative models. Generation must therefore operate not as indiscriminate augmentation, but as a selectively invoked component guided by explicit reliability measures.

I will develop an uncertainty-driven control framework for synthetic view generation and model densification. View-selection strategies will integrate diffusion priors with scene geometry and camera pose to identify when synthetic observations are likely to improve reconstruction rather than introduce hallucinations or redundancy. The same uncertainty signal will guide diffusion-aware densification, allocating additional primitives or resolution only where real imagery lacks sufficient geometric or photometric support. Collectively, this work elevates diffusion from heuristic augmentation to a governed component of the reconstruction process, unifying generation, densification, and pruning under a reliability-aware selective computation framework.

4. Career Goals and Community Engagement

I aim to pursue a research career advancing scalable and trustworthy vision systems across academia and industry. Engagement with the FG community will help refine my dissertation and shape high-impact directions at the intersection of biometrics, reconstruction, and generative modeling. I am also co-organizing the SPAR-3D Workshop on Security, Privacy, and Adversarial Robustness in 3D Generative Vision Models at CVPR 2026, reflecting my commitment to advancing reliable vision technologies for real-world deployment.