web: *https://lly00412.github.io/* email: lchen39@stevens.edu

Research Statement

Liyan Chen (CS PhD, Stevens Institute of Technology)

Artificial intelligence and machine learning models have become essential tools across a wide range of applications, such as autonomous navigation, immersive virtual environments, and robotics. However, in many real-world scenarios, obtaining sufficient supervision or ground-truth annotations for training remains a significant challenge. Developing generalizable, reliable, and efficient learning mechanisms — particularly those based on weak or self-supervision—is therefore critical. **My research in machine learning and computer vision focuses on building systems that can effectively learn from limited data and identify the uncertain or unreliable aspects within the current learning process.**

Research Progress

1. Data-Driven Structured Dropout for Convolutional Neural Networks

Convolutional neural networks (CNNs) have become a foundational tool in computer vision. However, CNNs are prone to overfit when neurons are likely to learn highly similar features due to the fact that visual inputs usually have strong spatial correlation among neighbors, especially when training data are limited.

Motivated by prior research [1–3], I introduce **a novel structured regularization method for convolutional layers**, called **DropCluster** [4]. The **contributions** include: (a) *a statistical measure to assess clustering tendency of activations* based on silhouette distances; (b) *a data-driven regularizer for CNNs* that randomly drop clustered activations together, aiming to balance between information preservation and spatial decorrelation of latent features. Empirical results show that DropCluster leverages the learned spatial structure to mitigate overfitting in CNNs for both classification and regression tasks under few-shot settings.

2. Precise Uncertainty Estimation by Learning the Distribution of Errors

In most real-world scenarios, even high-performing estimators are not error-free. Associating confidence or uncertainty with their estimates is of great importance, particularly in critical applications. Prior uncertainty predictors [5–8] can be used to rank the estimates approximately according to error, but they fail to match the *magnitude* of the actual errors. Of course, if we were able to predict the estimator's error at each pixel, otherwise, we could drive all errors down to zero and obtain a perfect estimator. **A more feasible objective** is to train an uncertainty estimator whose outputs follow the same distribution as the true errors.

I present an implementation of this concept via **a stereo matching network that jointly estimates disparity and its uncertainty from pairs of rectified images**, named **SEDNet** [9], for *Stereo Error Distribution Network*. To achieve the objective, I make several **contributions**: I first introduce *a novel uncertainty estimation subnetwork* that extracts information from the intermediate multi-resolution disparity maps generated by the disparity subnetwork. To train the network, we also need to formulate the distribution of errors and uncertainties in a differentiable manner. Hence, I introduce *a differentiable soft-histogramming technique used to approximate the distributions of disparity errors and estimated uncertainties*. Finally, I propose a matching error loss to to force the estimated uncertainties to match errors at the distribution level, i.e. to compute KL divergence between the differentiable histograms we obtained above. Experiments on both in-domain and cross-domain settings demonstrate that the proposed pipeline outperforms existing SOTA methods on both disparity and uncertainty estimation. I believe our method has the potential to achieve similar success on other pixel-wise regression tasks.

3. Efficient Uncertainty Quantification for Active 3D Reconstruction

Recent advancements in active mapping have demonstrated that effective exploration strategies can significantly enhance the completeness and fidelity of 3D reconstructions. These approaches typically incorporate efficient estimation of uncertainty and information gain for candidate views. On the other hand, NeRF [10] and 3DGS [11] significantly influenced computer vision and graphics due to their ability to synthesize high-quality images from novel viewpoints even under challenging imaging conditions or substantial geometric inaccuracies, while maintaining relatively low training costs. Following recent literature, I will use the term Radiance Fields (RF) to describe both. By bridging active mapping (AM) and novel view synthesis (NVS), I develop two research projects that leverage the strengths of both areas.

First, I propose **a novel uncertainty qualification approach** called *Virtual-Camera-based Uncertainty of Radiance Fields* [*that will be submmited soon*] designed for *measuring the inconsistencies among renderings by the RF model in virtual cameras sampled near the target viewpoint*. The key **contribution** is a *novel uncertainty quantification approach* that is *generally applicable*, as it treats the underlying RF models as black boxes *without requiring extra storage*. Our approach also shows superior performance compared to existing uncertainty estimators on standard NVS benchmarks [10,12–15], and can be applied to downstream tasks such as view selection and floater prunning.

Second, I am excited to introduce my work on **a carefully designed and comprehensive AM system** called **ActiveGAMER** [16], for *Active GAussian Mapping through Efficient Rendering*, that enables efficient exploration and high-fidelity 3D reconstruction. The core **contribution** of our system is *a rendering-based information gain module* that efficiently identifies the most informative viewpoints for next-best view planning *under both geometric and photometric reconstruction criteria*. *Coarse-to-fine exploration, post-refinement, and a global-local keyframe selection strategy* contribute to an effective trade-off between time efficiency and reconstruction quality in terms of completeness and fidelity. Experiments on Replica [17] and Matterport3D [18] validate the performance of the proposed system on both reconstruction quality (accuracy and completeness) and photometric quality (NVS metrics).

Most recently, I developed **ActiveSGM** [19], **the first dense *Active Semantic Gaussian Mapping* system** that integrates 3D Gaussian Splatting (3DGS) with semantics-aware exploration. ActiveSGM leverages both geometric and semantic uncertainty to guide next-best-view planning, and introduces a *sparse semantic representation* that retains only the top-k most likely categories per 3D primitive, reducing memory while preserving semantic richness. Unlike prior methods that rely on ground-truth labels, ActiveSGM uses *noisy predictions from a pre-trained model (OneFormer)* [20] and improves segmentation quality through exploration. Experiments on Replica and Matterport3D demonstrate superior performance of ActiveSGM in both reconstruction quality and semantic coverage, highlighting the system's ability to support efficient and adaptive scene exploration.

Ongoing Research and Future Directions

Additionally, I am highly interested in exploring active vision systems further in the future. I have recently started investigating this direction, which has also introduced new challenges.

- **Can we leverage higher-level signals, such as language feature, to enhance active mapping?** Semantic mapping has become an increasingly prominent topic in recent years, supported by powerful semantic segmentation networks and large language models (LLMs) [21–23]. However, the integration of semantic mapping with active vision remains largely unexplored. A key challenge lies in the nature of language features: it often involves high-dimensional representations that incur significant computational and memory costs. Moreover, compressing these high-dimensional features inevitably leads to information loss and semantic ambiguity, raising the question of how to preserve discriminative power while maintaining efficiency.

- **What metrics would be appropriate for measuring semantic uncertainty?** In real-world scenarios, dense semantic annotations are rarely available, and pseudo-supervision from pre-trained models is often noisy or uncertain. This makes accurate uncertainty estimation at the semantic level an ongoing research challenge. In closed-vocabulary settings, evaluation is relatively straightforward — metrics like accuracy or F1-score can be used to compare predictions against a fixed label set. However, in open-vocabulary settings, semantic ambiguity (e.g., desk vs. table) makes it difficult to define consistent ground truth, complicating both evaluation and supervision. This ambiguity directly impacts uncertainty estimation: since uncertainty reflects the model's likelihood of making a mistake, the inability to clearly define what constitutes an error undermines our ability to measure and interpret uncertainty effectively.

References

[1] S. Mostafa, D. Mondal, M. Beck, C. Bidinosti, C. Henry, and I. Stavness, "Visualizing feature maps for model selection in convolutional neural networks," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1362–1371.

[2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[3] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[4] L. Chen, P. Gautier, and S. Aydore, "Dropcluster: A structured dropout for convolutional networks," *arXiv preprint arXiv:2002.02997*, 2020.

[5] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *NeurIPS*, 2017, pp. 5574–5584.

[6] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *ICNN*. IEEE, 1994, pp. 55–60.

[7] E. Ilg, O. Cicek, S. Galesso, A. Klein, O. Makansi, F. Hutter, and T. Brox, "Uncertainty estimates and multi-hypotheses networks for optical flow," in *European Conference on Computer Vision*, 2018, pp. 652–667.

[8] M. Poggi, S. Kim, F. Tosi, S. Kim, F. Aleotti, D. Min, K. Sohn, and S. Mattoccia, "On the confidence of stereo matching in a deep-learning era: a quantitative evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5293–5313, 2021.

[9] L. Chen, W. Wang, and P. Mordohai, "Learning the distribution of errors in stereo matching for joint disparity and uncertainty estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 235–17 244.

[10] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision*. Springer, 2020, pp. 405–421.

[11] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering." *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 139–1, 2023.

[12] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–14, 2019.

[13] K. Yücer, A. Sorkine-Hornung, O. Wang, and O. Sorkine-Hornung, "Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction," *ACM Transactions on Graphics*, vol. 35, no. 3, pp. 1–15, 2016.

[14] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, "Tanks and Temples: Benchmarking large-scale scene reconstruction," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–13, 2017.

[15] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5855–5864.

[16] L. Chen, H. Zhan, K. Chen, X. Xu, Q. Yan, C. Cai, and Y. Xu, "ActiveGAMER: Active GAussian Mapping through Efficient Rendering," *arXiv preprint arXiv:2501.06897*, 2025.

[17] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma *et al.*, "The Replica dataset: A digital replica of indoor spaces," *arXiv preprint arXiv:1906.05797*, 2019.

[18] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matterport3D: Learning from RGB-D data in indoor environments," *arXiv preprint arXiv:1709.06158*, 2017.

[19] L. Chen, H. Zhan, H. Yin, Y. Xu, and P. Mordohai, "Understanding while exploring: Semantics-driven active mapping," *arXiv preprint arXiv:2506.00225*, 2025.

[20] J. Jain, J. Li, M. T. Chiu, A. Hassani, N. Orlov, and H. Shi, "Oneformer: One transformer to rule universal image segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 2989–2998.

[21] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. M. Ni, and H.-Y. Shum, "DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection," *arXiv preprint arXiv:2203.03605*, 2022.

[22] OpenAI, “ChatGPT: Optimizing Language Models for Dialogue,” https://openai.com/chatgpt, 2023, accessed: March 11, 2025.

[23] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan *et al.*, “DeepSeek-v3 technical report,” *arXiv preprint arXiv:2412.19437*, 2024.