Deep Learning for 3D World Representation

Resolution, Reliability and Remote Sensing

Luca Savant Aira, Torino, 2026

Long abstract

This dissertation studies volumetric representations of 3D scenes that can be reconstructed from real sensor data and rendered from arbitrary viewpoints. It concentrates on radiance-field-style models such as Neural Radiance Fields and 3D Gaussian Splatting, which encode geometry and appearance as continuous fields of density and color and are optimized end-to-end via differentiable volumetric rendering. A preliminary chapter revisits light transport and the volume rendering integral, and proves analytically how these classical concepts map to modern neural formulations and fixing notation for radiance, color, density, cameras and rendering operators.

On this foundation, the dissertation addresses four main challenges. The first is multi-image super-resolution (MISR) with large viewpoint changes, where traditional methods based on optical flow in the image plane break down. The proposed EpiMISR deep neural network replaces optical flow with explicit epipolar geometry, implicitly building 3D feature fields akin to Neural Radiance Fields but without per-scene optimization. This geometry-aware design handles arbitrary numbers of views, gracefully falls back to single-image SR, and yields clear gains over prior MISR approaches in settings with strong parallax.

The second challenge is reliability, in particular how to attach meaningful uncertainty estimates to reconstructions obtained with Gaussian Splatting. The dissertation introduces Stochastic Gaussian Splatting, which turns each Gaussian primitive into a Bayesian random variable. Rendering becomes stochastic, and Monte-Carlo evaluation produces both an expected image and a per-pixel predictive variance. A new loss term encourages these variances to correlate with true errors, leading to calibrated uncertainty maps, while preserving the speed and quality advantages of 3D Gaussian Splatting.

The third challenge is to cope with missing or extremely sparse viewpoints. To regularize volumetric models in such regimes, the thesis explores physics-based priors on motion via MotionCraft, a zero-shot video generator that operates in the latent space of a pretrained image diffusion model. The method is parameter-free, works from a single input image, and improves over existing zero-shot baselines on complex motions. Within the broader thesis, this line of work shows how physics-based latent warping can serve as a strong prior for imagining plausible dynamics and unseen viewpoints, complementing static volumetric reconstructions.

The fourth challenge is efficiency and scalability in real application, such as in satellite remote sensing, where multi-view high-resolution image collections must be processed under tight computational budgets. The Earth-Observation Gaussian Splatting (EOGS) framework adapts Gaussian Splatting to satellite photogrammetry by combining remote-sensing-specific ingredients such as radiometric corrections and physically motivated shadow modeling. Experiments show that EOGS attains reconstruction accuracy comparable to state-of-the-art NeRF-based Earth-observation methods while requiring orders of magnitude less training time, making volumetric radiance fields practical for high-throughput satellite pipelines.

Altogether, the dissertation shows that radiance-field-based scene representations can be adapted to address accuracy, reliability, data efficiency and scalability in a range of settings. Through four complementary but distinct methods, it advances our understanding of how physically grounded rendering, geometric constraints, Bayesian modeling and generative priors can be combined with volumetric models. The resulting contributions improve multi-image super-resolution, provide uncertainty estimates for Gaussian Splatting, enable plausible dynamics from sparse observations, and make radiance-field-style approaches more practical for large-scale remote-sensing pipelines.