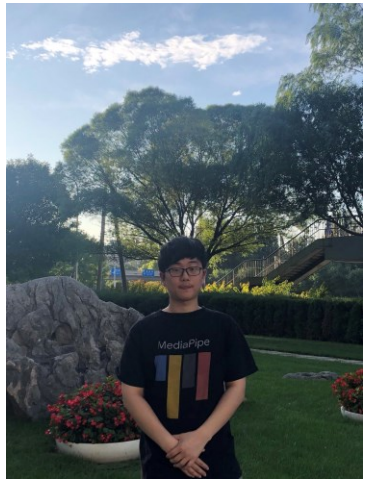
Department of Electrical Engineering
香港城市大學
City University of Hong Kong

**CityU Architecture Lab for Arithmetic and Security (CALAS) Seminar Series**

# VGGT: Visual Geometry Grounded Transformer

Jay Jianyuan WANG, Joint PhD Student
Meta AI Research and Visual Geometry Group (VGG), University of Oxford

CVPR 2025 Best Paper Award



## Abstract:

We present VGGT, a feed-forward neural network that directly infers all key 3D attributes of a scene, including camera parameters, point maps, depth maps, and 3D point tracks, from one, a few, or hundreds of its views. This approach is a step forward in 3D computer vision, where models have typically been constrained to and specialized for single tasks. It is also simple and efficient, reconstructing images in under one second, and still outperforming alternatives without their post-processing utilizing visual geometry optimization techniques. The network achieves state-of-the-art results in multiple 3D tasks, including camera parameter estimation, multi-view depth estimation, dense point cloud reconstruction, and point tracking. We also show that using pretrained VGGT as a feature backbone significantly enhances downstream tasks, such as non-rigid point tracking and feed-forward novel view synthesis.

## Biography:

Jianyuan Wang is a joint PhD student at Meta AI Research and the Visual Geometry Group (VGG), University of Oxford, currently in his third year. His research focuses on 3D understanding, particularly the reconstruction of 3D scenes from images, from PoseDiffusion, VGGSfM, to VGGT. His work has been recognized with several honors, including CVPR 2025 Best Paper Award.

31 July 2025 (Thur); 6pm - 7pm; G4302; https://cityu.zoom.us/j/96742093029