3D Gaussian Splats

Easily Attacked to Cause Harm

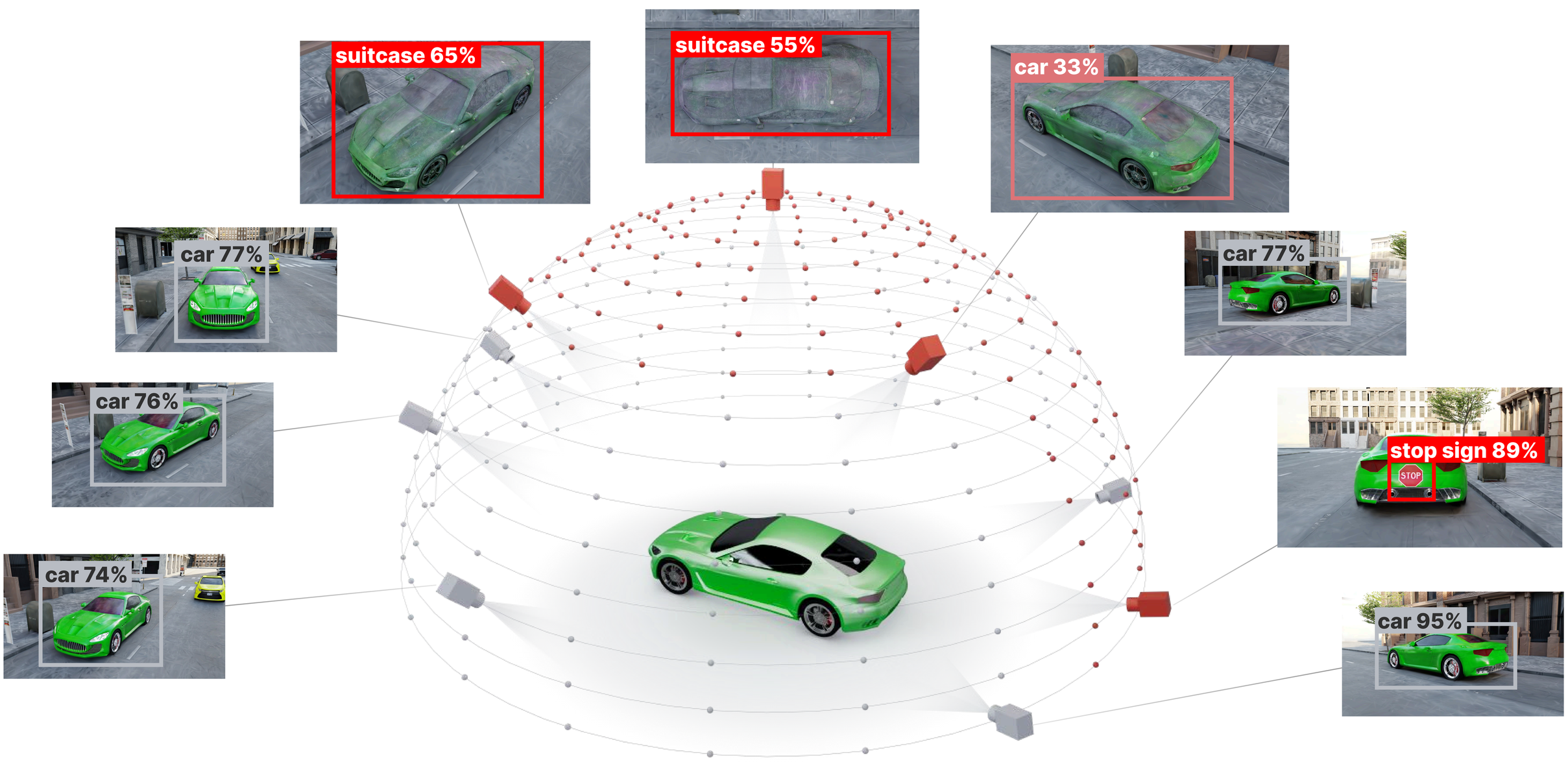
Matthew Hull, Haoyang Yang, Pratham Mehta, Mansi Phute, Aeree Cho, Haoran Wang, Matthew Lau, Wenke Lee, Willian Lunardi, Martin Andreoni, Polo Chau

Georgia Tech

Technology Innovation Institute

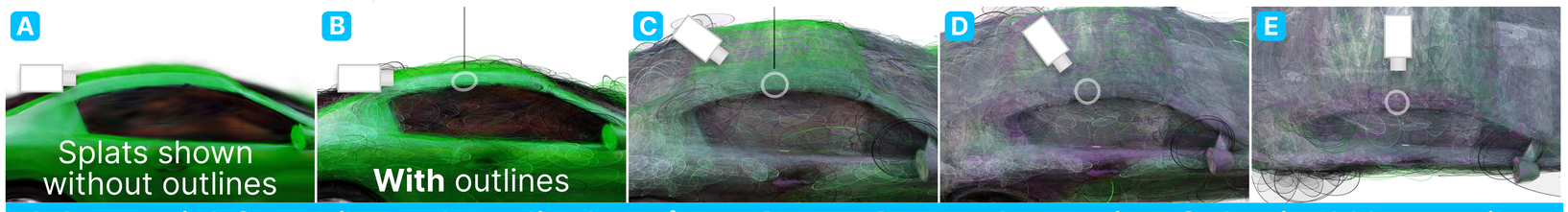
3DGS vulnerabilities are underexplored. We highlight new potential threats to robotic learning for autonomous navigation and other safety-critical 3DGS applications.

CLOAK attack conceals adversarial textures visible only from certain views



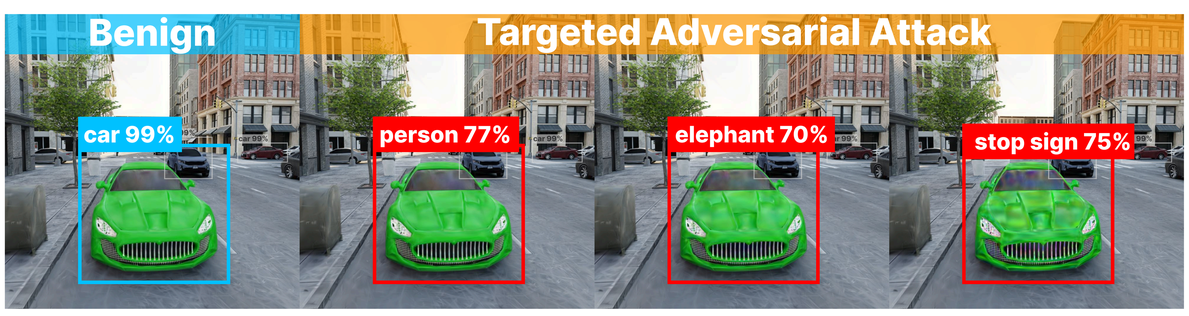
CLOAK is a poisoning attack method - benign images are replaced with adversarial images in the 3DGS training dataset for targeted camera viewpoints.

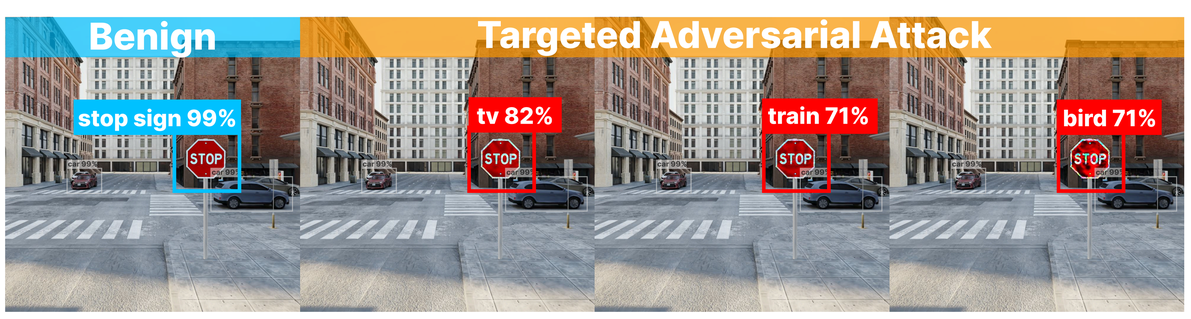
One splat's outline accentuated to visualize easier to across views.



Adversarial Gaussian splats display **view-dependent** colors using Spherical Harmonics.

DAGGER manipulates Gaussian attributes to induce misdetections on Faster R-CNN.





DAGGER generalizes projected gradient descent attack by exploiting the differentiability of the 3DGS scene representation to manipulate splat color, scaling, translation, rotation, or alpha attributes to fool object detectors.