SAGA: Spectral Adversarial Geometric Attack on 3D Meshes

Tomer Stolik*, Itai Lang*, Shai Avidan

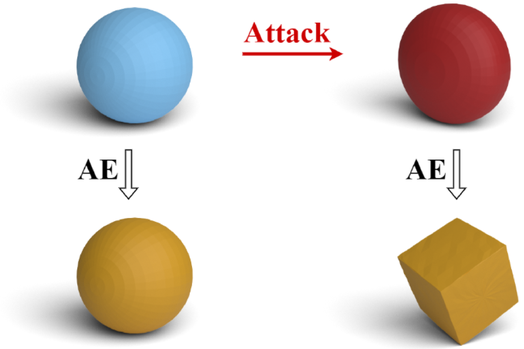
*Equal contribution

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

ICCV23 PARIS

Concept

Can we cubify a sphere?!



An original sphere mesh is accurately reconstructed by the autoencoder.

However, our *adversarial sphere* fools the autoencoder to reconstruct the output geometry of a *cube*!

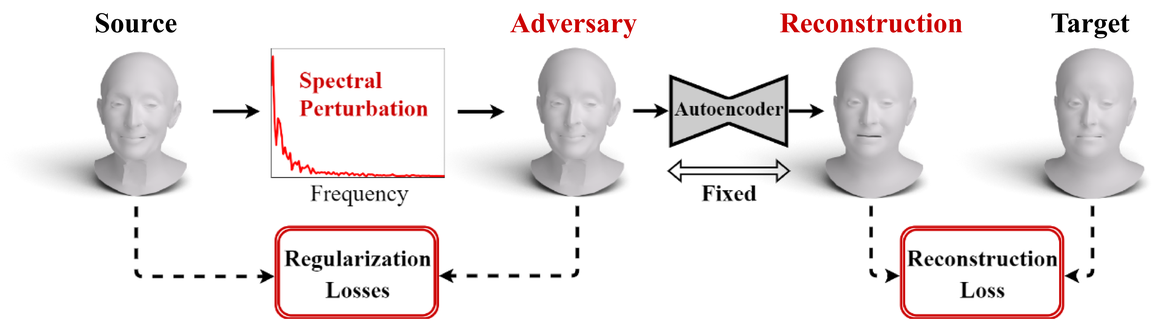
Contributions

- The first *geometric* adversarial attack on 3D meshes.
- The method is based on low-frequency spectral perturbations and regularizations of mesh attributes.
- SAGA crafts adversarial examples that change an AE's output into a different geometric shape.

Implementation

Our code is available!

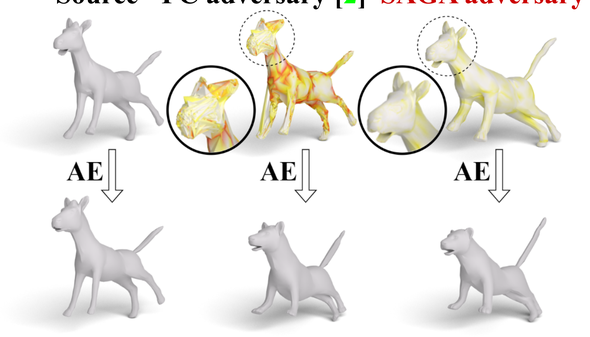
Method



- Perturbs the source shape spectral coefficients to craft an adversarial example.
- The malicious input mislead the AE to output the geometry of the target mesh.
- Optimized to reconstruct the target while preserving the source's properties.

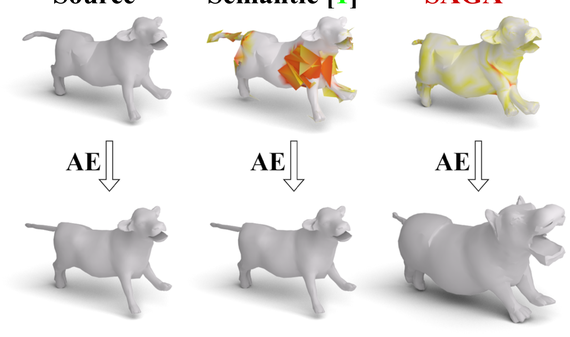
Attack

Comparison to a Geometric Attack



- **SAGA** changes the *horse*'s pose slightly while preserving its geometry and misleads the AE to reconstruct the target *leopard* shape.
- The **point cloud attack** causes apparent surface distortions to the source and its reconstruction lacks the fine-grained target mesh details.

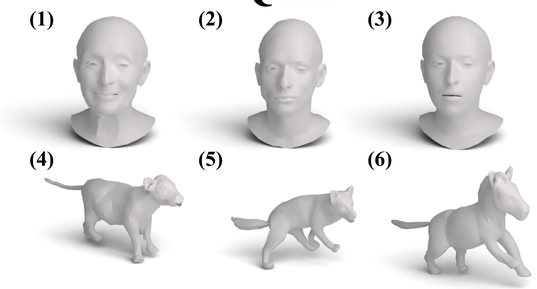
Comparison to a Semantic Attack



- **SAGA**'s adversarial example preserves the *cow* geometry and successfully tricks the AE to output the *hippo* mesh.
- The **semantic attack,** which is highly effective against classifiers, fails to mislead the AE and the reconstruction remains similar to the source.

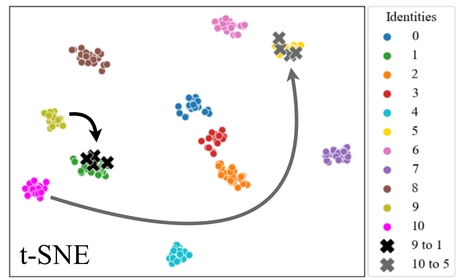
[1] Huang *et al.* Shape-invariant 3D Adversarial Point Clouds. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2022.*
[2] Lang *et al.* Geometric Adversarial Attacks and Defenses on 3D Point Clouds. *The International Conference on 3D Vision (3DV) 2021.*
[3] Ranjan *et al.* Generating 3D Faces Using Convolutional Mesh Autoencoders. *The European Conference on Computer Vision (ECCV) 2018.*

Quiz



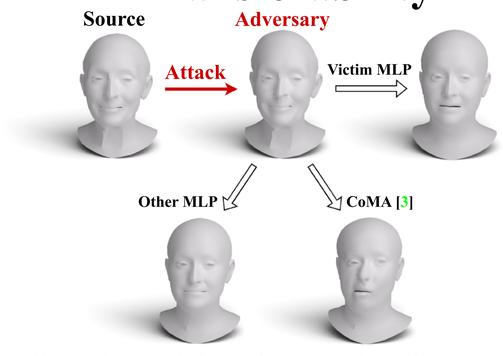
Which shape is an original mesh from the dataset and which is an adversarial example of SAGA?

Latent Space Analysis



Adversarial human faces from classes 9 and 10 are encoded to the typical latent region of the attack's target classes 1 and 5, respectively.

Transferability



Our adversarial meshes remain efficient even when applied to unseen autoencoders!