ϵ-Mesh Attack: A Surface-based Adversarial Point Cloud Attack for Facial Expression Recognition

Batuhan Cengiz, Mert Gülşen, Yusuf H. Şahin, Gözde Ünal
Istanbul Technical University

IEEE COMPUTER SOCIETY · IEEE Biometrics Council · 2024 27-31 May 2024 Istanbul, Turkey
The 18th IEEE International Conference on Automatic Face and Gesture Recognition

İTÜ VISION LAB

Overview

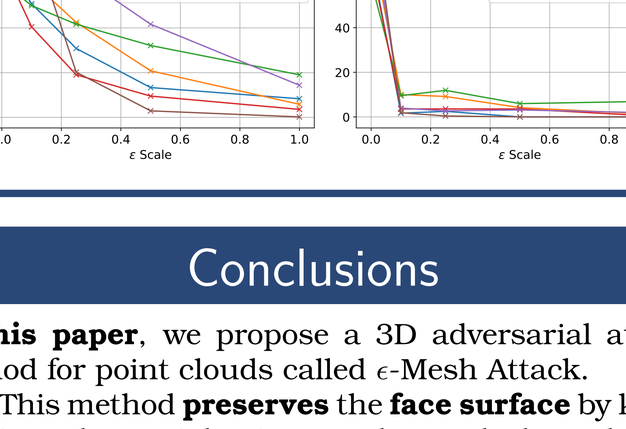
ϵ-Mesh Attack is an adversarial attack for 3D point clouds. Our framework is publicly available on **GitHub**.

Paper

Code

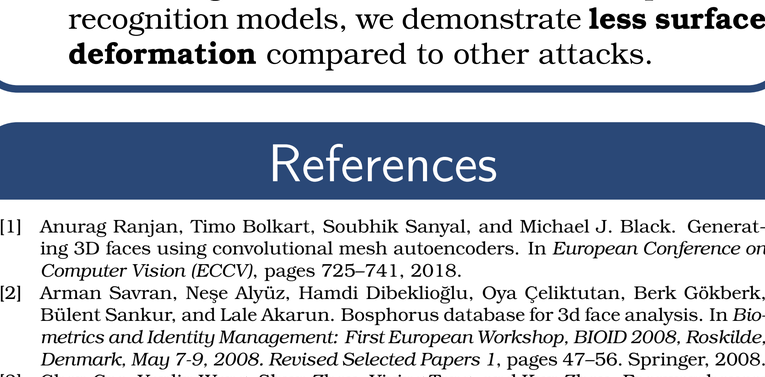
Motivation

As **3D point cloud data** becomes more available (accessible Lidar technology, mobile phones etc.), correct evaluation for **facial expression recognition** models become critical. However, current adversarial attacks **do not** consider **underlying surface** explicitly and **perturb facial surface** structure. Thus, we proposed ϵ-Mesh Attack which constrains the adversarial optimization by the underlying mesh.



Ablation Study

We showed that our proposed ϵ-Mesh Attack **scales** with parameter ϵ. **Perpendicular** method outperforms **central** method.



Conclusions

In this paper, we propose a 3D adversarial attack method for point clouds called ϵ-Mesh Attack.

- This method **preserves** the **face surface** by keeping adversarial points on the mesh through central and perpendicular projections onto mesh triangles.
- We parameterize our attack by ϵ to **scale attack boundaries** into similar triangles as shown in the figure below.
- Evaluating our method on 3D facial expression recognition models, we demonstrate **less surface deformation** compared to other attacks.

References

[1] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3D faces using convolutional mesh autoencoders. In *European Conference on Computer Vision (ECCV)*, pages 725–741, 2018.

[2] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3d face analysis. In *Biometrics and Identity Management: First European Workshop, BIOID 2008, Roskilde, Denmark, May 7-9, 2008. Revised Selected Papers 1*, pages 47–56. Springer, 2008.

[3] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, 2013.

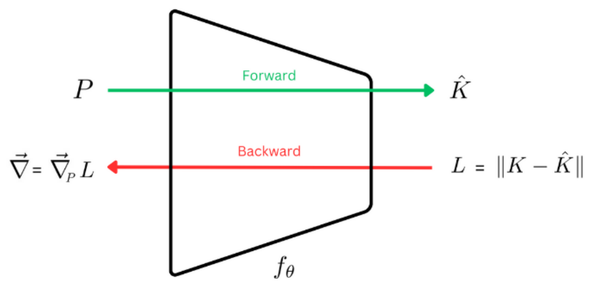
[4] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5):1–12, 2019.

[5] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017.

Proposed Method

For our ϵ-Mesh Attack, we proposed Central and Perpendicular projection methods. Our pipeline is as follows:

- Gradient with respect to input is calculated.
- Gradient is projected onto plane of mesh.
- If the projected point is out of the mesh:
  - **Central**: New point is projected to intersection of mesh with center pointing line.
  - **Perpendicular**: New point is projected to closest point of mesh.

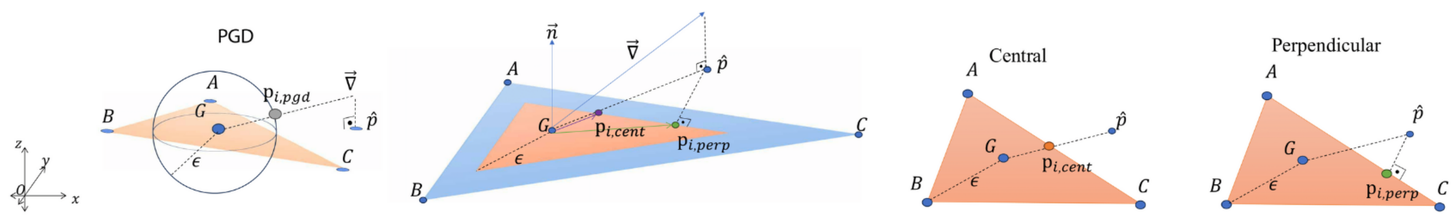


$$\vec{\nabla} = \frac{dL}{df_\theta}\frac{df_\theta}{dP}.$$

$$\vec{\nabla}_s = \vec{\nabla} - \frac{\vec{\nabla}\cdot\vec{n_i}}{\vec{n_i}\cdot\vec{n_i}}\vec{n_i}$$

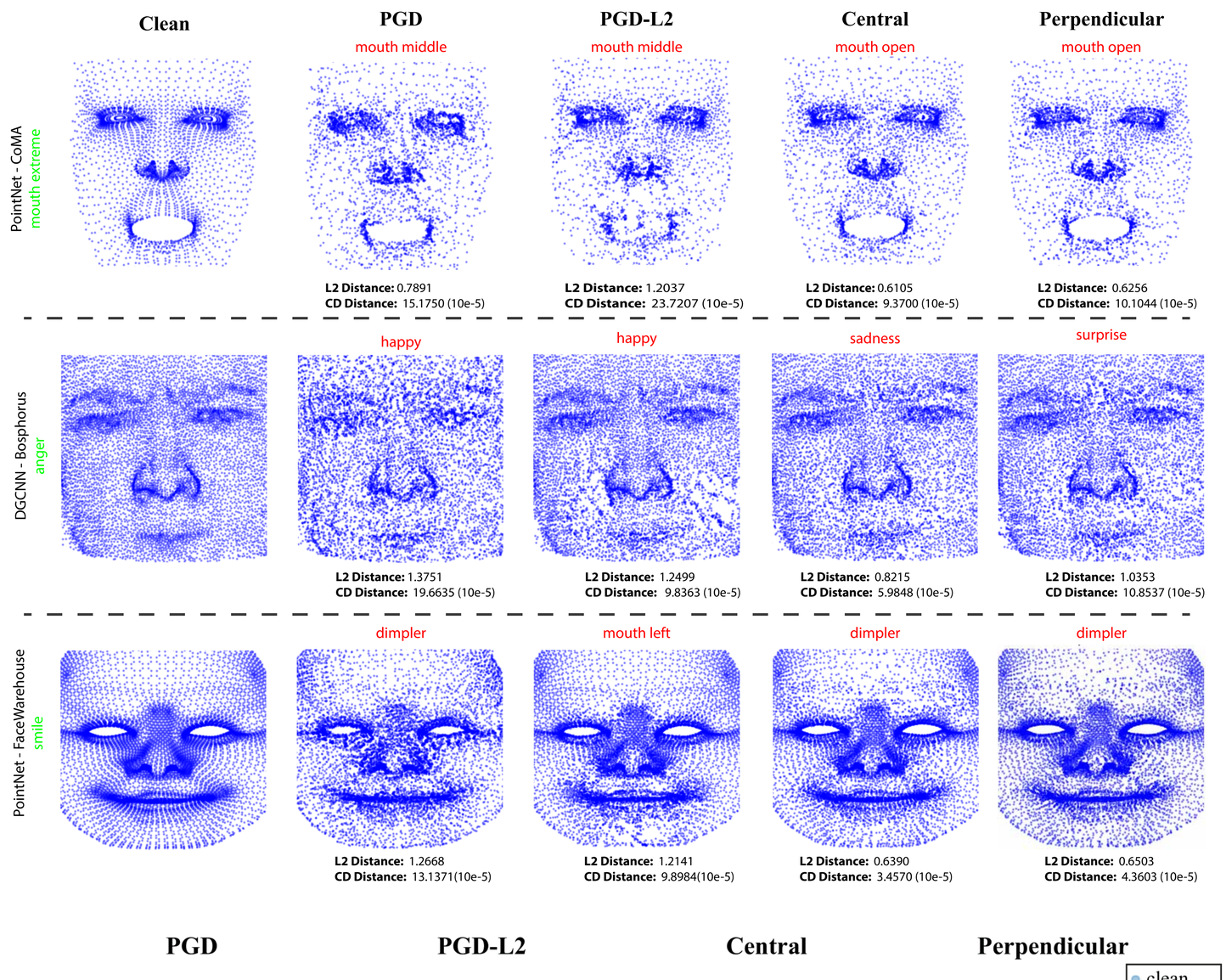
$$p_{i,cent} = \begin{cases} \hat{p}_i, & \hat{p}_i \in \triangle A_iB_iC_i \\ \triangle A_iB_iC_i \cap \overline{G_i\hat{p}_i}, & \text{otherwise} \end{cases}$$

$$p_{i,perp} = \begin{cases} \hat{p}_i, & \hat{p}_i \in \triangle A_iB_iC_i \\ \underset{x\in\triangle A_iB_iC_i}{\arg\min} \|x - \hat{p}_i\|, & \text{otherwise} \end{cases}$$



Qualitative Results

We have conducted our experiments on three well-known facial expression datasets: CoMA [1], Bosphorus [2] and FaceWarehouse [3].



Quantitative Results

Model	Attack	Eps	Alpha	Steps	CoMA Clean Acc (%)	CoMA Attacked Acc (%)	Bosphorus Clean Acc (%)	Bosphorus Attacked Acc (%)	FaceWarehouse Clean Acc (%)	FaceWarehouse Attacked Acc (%)
DGCNN[4]	PGD	0.01	0.0004	250	79.17	0.0	69.04	0.0	98.96	0.0
	PGD-L2	1.25	0.05			0.0		0.0		0.0
	(**Ours**) ϵ-mesh Central	1.00	0.10			5.83		3.57		0.21
	(**Ours**) ϵ-mesh Perpendicular	1.00	0.10			0.83		0.0		0.0
PointNet[5]	PGD	0.01	0.0004	250	71.67	0.0	60.71	0.0	88.96	0.0
	PGD-L2	1.25	0.05			0.0		0.0		0.0
	(**Ours**) ϵ-mesh Central	1.00	0.10			0.83		19.04		14.38
	(**Ours**) ϵ-mesh Perpendicular	1.00	0.10			0.0		7.14		1.67