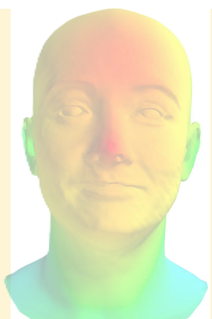
# 3DDGD: 3D Deepfake Generation and Detection Using 3D Face Meshes

Hichem Felouat, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen

## Motivation

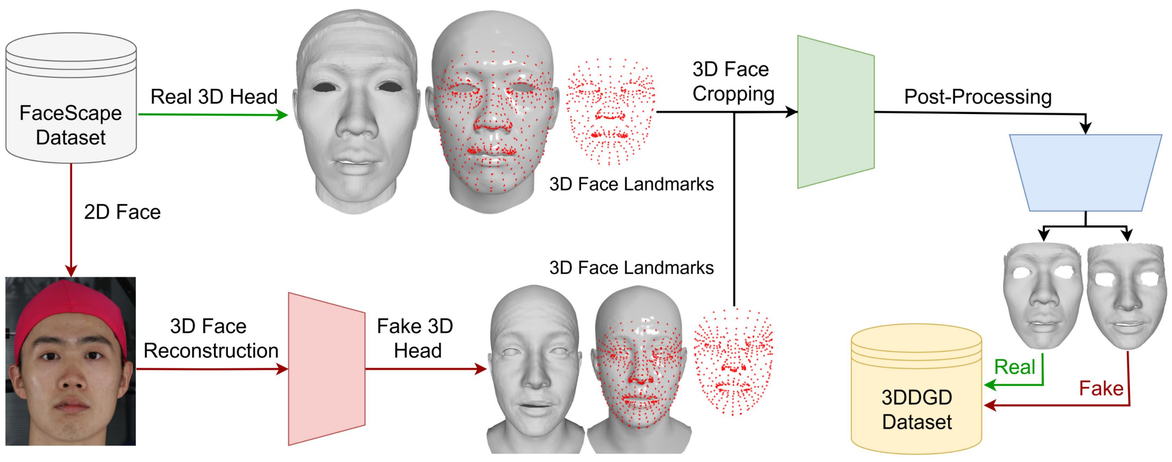
- 3D face recognition is widely used in security and virtual systems, including **biometric authentication** for smartphones (e.g., Apple's Face ID), and **remote identity verification** in eKYC.
- Rise of 3D deepfakes introduces new security risks.
- Strong need for improved detection and prevention methods.



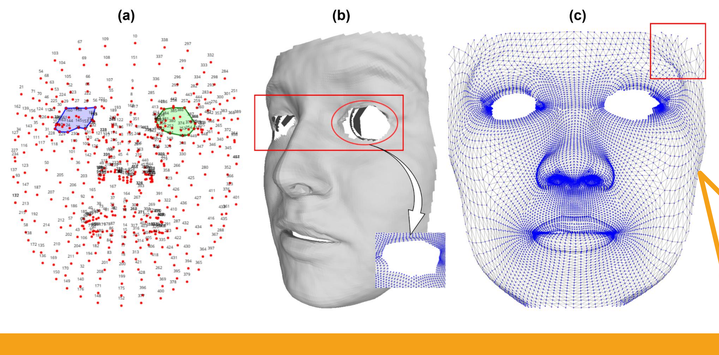
## Contribution

- New **dataset** includes real and fake 3D face models.
- Advanced **3D deepfake detection** models designed using 3D mesh learning.
- Models show strong generalization across multiple datasets.

## DDGD Dataset




- Real 3D head models sourced from the **FaceScape** and **BU-3DFE** datasets.
- Fake 3D models generated from 2D-to-3D face reconstruction.
- 3D facial landmarks used for face cropping.
- Post-processing applied to enhance model quality.
- **DECA** and **EMOCA** used for 3D face reconstruction.



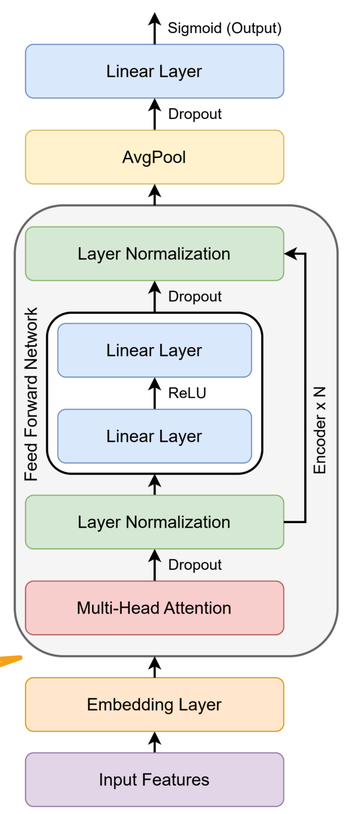

Postprocessing steps:
a) 3D facial landmarks.
b) Removing disconnected areas.
c) Removing outlier vertices near face and eyes.

## Dataset Statistics

| | | Source | No. of Expressions | No. of Identities | No. of Samples |
|---|---|---|---|---|---|
| Dataset 1 | Real | FaceScape | 20 | 846 | 16888 |
| | Fake | DECA | 20 | 359 | 7143 |
| | | EMOCA | 20 | 250 | 4958 |
| | | | | | **28989** |
| Dataset 2 | Real | BU-3DFE | 25 | 100 | 2500 |
| | Fake | DECA | 25 | 100 | 2500 |
| | | | | | **5000** |

Comprehensive statistics of datasets used in this study:
- **Two datasets** were created using the same pipeline.
- Differ in **source**, **resolution**, and **data collection** method.
- Used to evaluate models' **generalizability** on unseen data.

## Mesh TabTransformer




Mesh TabTransformer pipeline for 3D deepfake detection.

## Experiments

- Tested **mesh-MLP** model for 3D mesh classification.
- Developed two attention-based variants: **mesh-MLP-SA** and **mesh-MLP-MHA**.
- Introduced **Mesh-TabTransformer** model for 3D deepfake detection.

## Results

- 3D deepfake detection results reported as F1 scores (%). **Green** highlights highest scores; **red** indicates second-highest.

Trained on Dataset 1

| Models | Tested on Dataset 1 G0 | G1 | G2 | All | Tested on Dataset 2 G0 | G1 | G2 | All |
|---|---|---|---|---|---|---|---|---|
| Mesh_MLP | 95.39 | 95.66 | 95.61 | **96.30** | 63.36 | 60.72 | 62.13 | 61.78 |
| Mesh_MLP_SA | **97.61** | **98.30** | **98.38** | 95.57 | 71.59 | 71.79 | **72.50** | 60.62 |
| Mesh_MLP_MHA | 85.16 | 96.50 | 95.41 | **97.87** | **72.25** | **79.77** | **75.55** | **73.32** |
| TabTransformer | **98.17** | **98.01** | **97.98** | 93.87 | **71.66** | **72.26** | 71.71 | **83.18** |

Trained on Dataset 2

| Models | Tested on Dataset 2 G0 | G1 | G2 | All | Tested on Dataset 1 G0 | G1 | G2 | All |
|---|---|---|---|---|---|---|---|---|
| Mesh_MLP | **95.25** | **96.39** | **95.27** | **96.57** | 61.56 | **66.06** | 61.58 | **62.43** |
| Mesh_MLP_SA | **95.57** | **95.66** | **95.69** | **94.91** | 61.54 | 61.01 | **62.50** | 61.83 |
| Mesh_MLP_MHA | 93.48 | 92.38 | 92.51 | 94.79 | **61.77** | **64.17** | 59.45 | 62.08 |
| TabTransformer | 93.41 | 93.12 | 93.02 | 94.47 | **62.91** | 63.98 | **62.58** | **63.29** |

Trained on Dataset 1

| Model | Tested on Dataset 1 G0 | G1 | G2 | All | Tested on Dataset 2 G0 | G1 | G2 | All |
|---|---|---|---|---|---|---|---|---|
| Ensemble TabTransformer | 97.48 | 95.74 | 97.04 | 97.18 | 81.34 | 83.98 | 88.46 | 80.87 |

- Attention mechanisms improved performance and generalization.
- Mesh-TabTransformer performed well on unseen data.
- Ensemble TabTransformer remained stable across training sets.

Trained on Dataset 1 Tested on Dataset 1

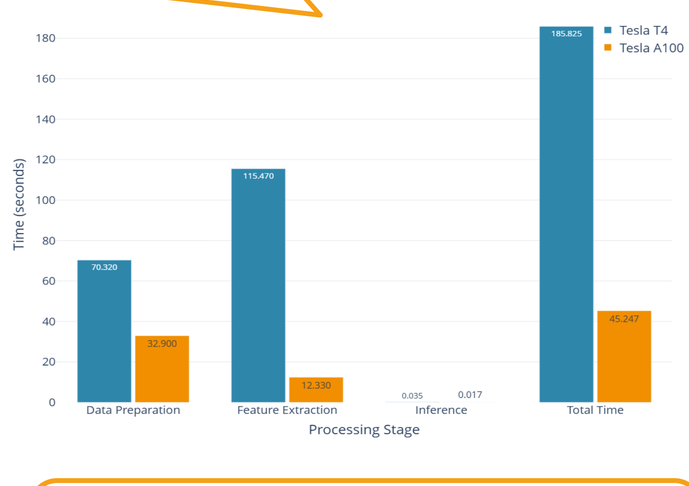
| Models | G0 | G1 | G2 | All |
|---|---|---|---|---|
| Mesh_MLP_MHA | 84.14 | 86.49 | 93.59 | 92.29 |
| TabTransformer | 87.88 | 85.75 | 92.98 | 89.76 |
| Ensemble TabTransformer | 95.97 | 94.78 | 95.83 | 96.73 |

Models demonstrated strong robustness to adversarial attacks. Results reported as F1 scores (%).

| Criterion | 2D Detection | 3D Detection |
|---|---|---|
| Data Availability | * | - |
| Feature Variety | + | * |
| Inference Time | * | + |
| Adversarial Robustness | - + | + * |

Compared with 2D deepfake detection. Ratings used: – = Weak, + = Good, * = Very Good.

Average processing time (seconds) per example.




## Conclusion

- 3D face data provides stronger protection against deepfakes than 2D images.
- Models showed strong performance across datasets and training conditions.
- Results highlight potential of 3D face technology for secure recognition and identity proofing.
- **Future Work:** Combine geometric features with facial textures to enhance detection accuracy and efficiency.

NII 越前研究室 国立情報学研究所

H. FELOUAT, H. H. NGUYEN, J. YAMAGISHI, and I. ECHIZEN, "3DDGD: 3D Deepfake Generation and Detection Using 3D Face Meshes, " IEEE Access (2025).

hichemfel@nii.ac.jp