3D-aware Facial Landmark Detection via Multi-view Consistent Training on Synthetic Data

Libing Zeng[1], Lele Chen[2], Wentao Bao[3], Zhong Li[2], Yi Xu[2], Junsong Yuan[4], Nima K. Kalantari[1]
[1]Texas A&M University, [2]OPPO US Research Center, InnoPeak Technology, Inc, [3]Michigan State University, [4]University at Buffalo

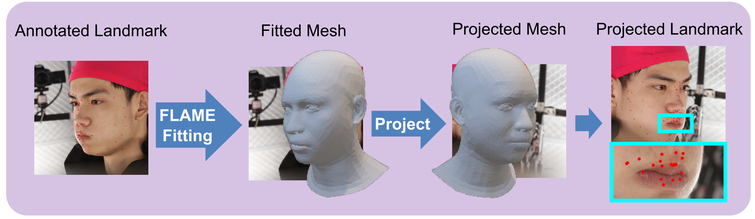
JUNE 18-22, 2023 CVPR

Motivation:

- Existing facial landmark detection approaches detect 3D/2D facial landmarks that are inconsistent with the 3D geometry.
- Current in-the-wild training data are manually annotated.
  - Such annotations are multiview inconsistent

Manual Annotation is Problematic:

- Invisible landmarks need to be guessed
- When FLAME fitted mesh using visible landmarks is projected to another view, multiview inconsistency appears

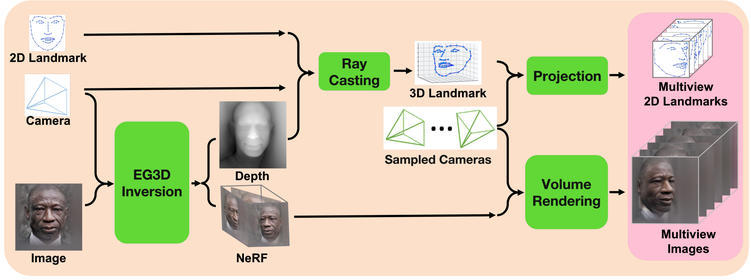


- The inaccuracy and multiview inconsistency results in the loss of the semantic meanings associated with landmarks
- Killing factor to learn an accurate landmark detector

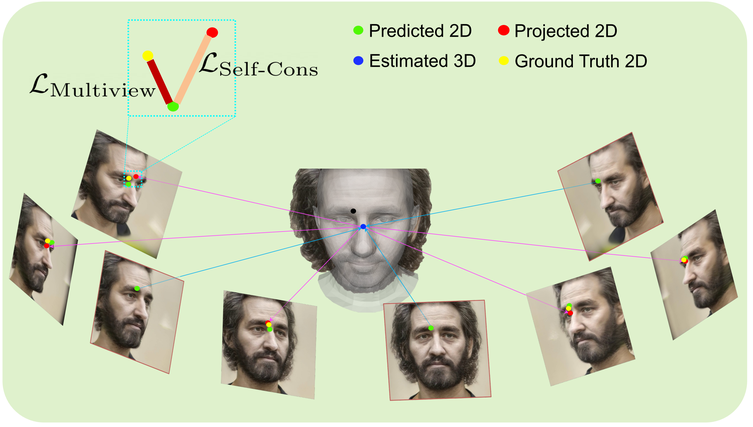
Contributions:

- We present a novel approach to producing a large-scale synthetic, but realistic, multi-view face dataset, titled DAD-3DHeads-Syn.
- We propose a novel 3D-aware optimization module, which can be plugged into any learning-based facial landmark detection method to improve its accuracy and multiview consistency.
- We demonstrate the performance improvements of our module built on top multiple baseline methods on simulated dataset, lab-captured datasets, and in-the-wild datasets.

Multiview Data Simulation:



Multiview Consistency Supervision:



Quantitative comparisons:

- Facial landmark detection result (NME↓)

	DAD-3DHeads	FaceScape	MultiFace
3DDFA	4.082	7.988	8.121
3DDFA+ (Ours)	**3.784**	**7.425**	**7.305**
DAD-3DNet	2.599	6.681	5.786
DAD-3DNet+ (Ours)	**2.503**	**6.050**	**5.480**

Visual comparisons:

- The visual comparisons between baseline methods and the refined methods on FOUR testing sets