Towards Fast, Accurate and Stable 3D Dense Face Alignment

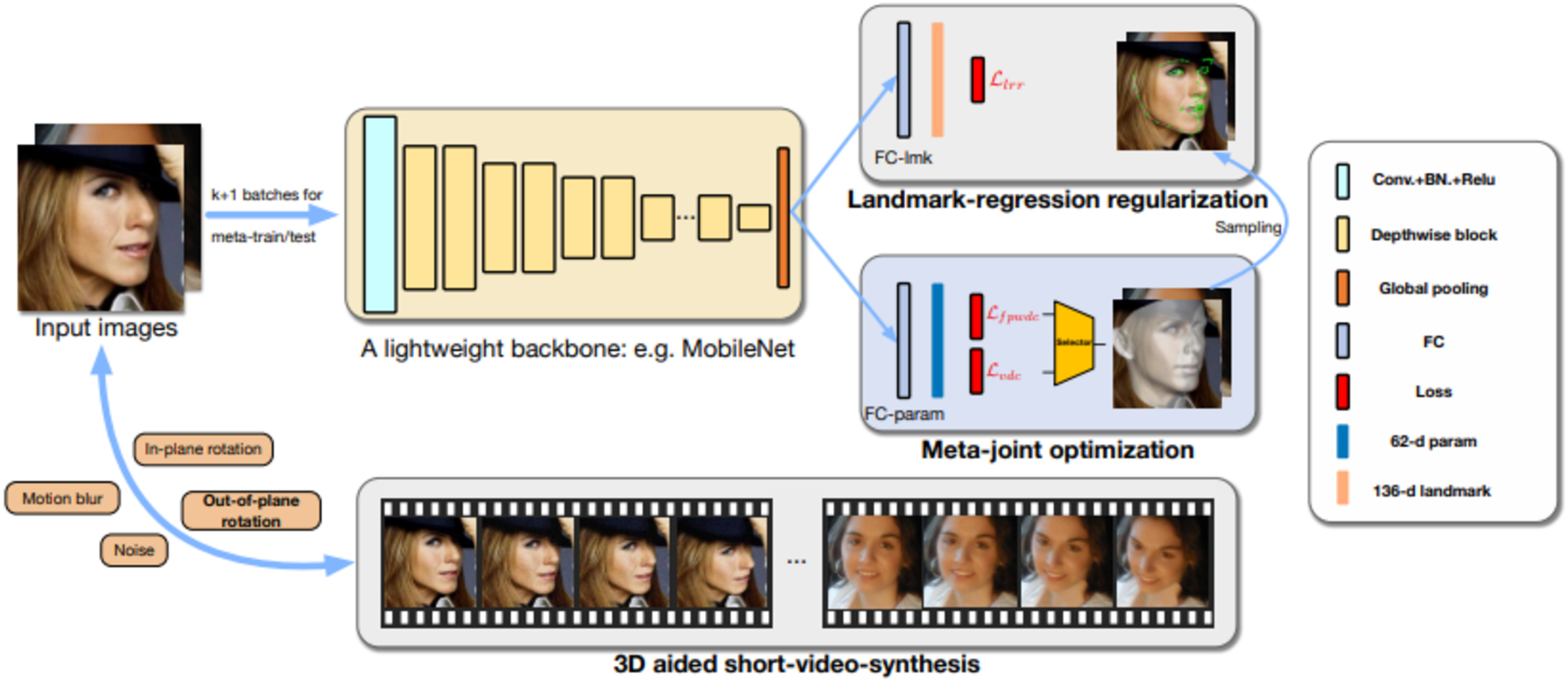
Team: Vision Impossible

Introduction to 3D Dense Face Alignment

- **Definition:** Aligning dense 3D face models to 2D images by mapping detailed 3D facial landmarks onto the image.
- **Applications:** Used in various real-time applications, including:
  - **Facial Recognition** – Enhances identity verification systems.
  - **AR/VR** – Provides realistic face tracking for immersive experiences.
  - **Animation** – Helps create accurate, lifelike facial animations in digital media.
- **Challenges:**
  - **Real-time performance** – Balancing computational load and speed.
  - **Accuracy** – Ensuring high precision in predicting 3D facial structure.
  - **Stability** – Minimizing jitter in video sequences.

NEXT →

3DDFA_V2 Model Overview



3D Morphable Model (3DMM) Parameters and Landmark Regression

3D face mesh:

$$\mathbf{S} = \overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}$$

- **Pose Parameters (12)** – Encodes face rotation and translation.
- **Shape Parameters (40)** – Encodes facial identity features.
- **Expression Parameters (10)** – Encodes variations in expressions.

2D Projection

$$V_{2d}(\mathbf{p}) = f * \mathbf{Pr} * \mathbf{R} * \left(\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}\right) + \mathbf{t}_{2d}$$

$$V_{2d}(\mathbf{p}) = \mathbf{Pr} * \mathbf{T} * \begin{bmatrix} \overline{\mathbf{S}} + \mathbf{A}\boldsymbol{\alpha} \\ \mathbf{1} \end{bmatrix}, \qquad T = [f \cdot R \quad f \cdot t_{3D}]$$

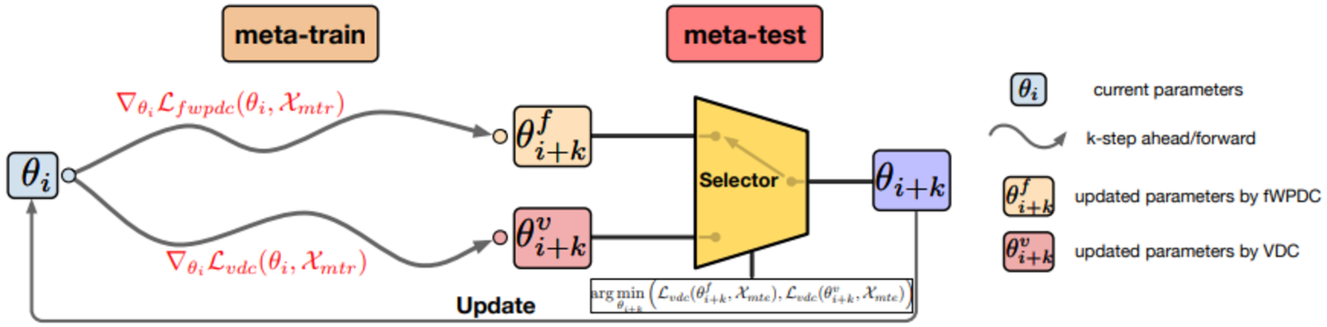
Meta-joint Optimization

- **Purpose:** Meta-joint optimization dynamically chooses between WPDC and VDC during training, which helps achieve optimal accuracy and faster convergence during training.
- **Loss Functions Used:**
  - **Weighted Parameter Distance Cost (WPDC):** Assigns weights to each parameter, focusing on important ones to improve initial convergence.

$$\mathcal{L}_{wpdc} = \left\| \mathbf{w} \cdot (\mathbf{p} - \mathbf{p}^g) \right\|^2$$

  - **Vertex Distance Cost (VDC):** Minimizes the distance between predicted and actual vertex positions, boosting final accuracy.

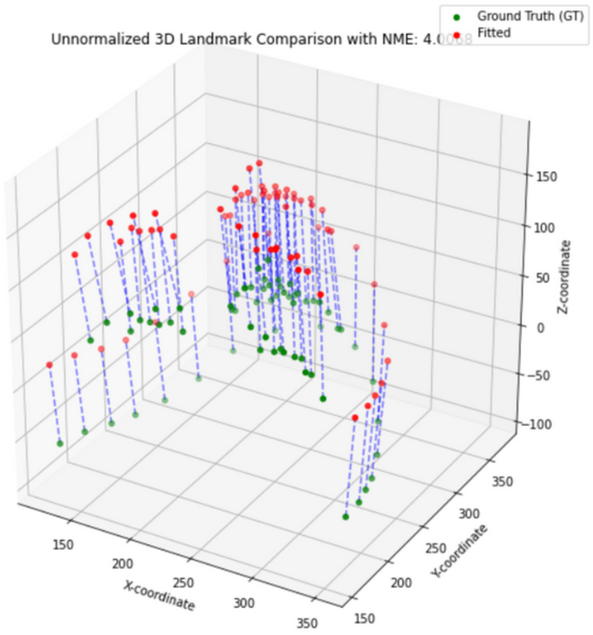
$$\mathcal{L}_{vdc} = \left\| V_{3d}(\mathbf{p}) - V_{3d}(\mathbf{p}^g) \right\|^2$$



Evaluation Metrics

- **Accuracy Metric: Normalized Mean Error (NME)**
  - Measures prediction accuracy by comparing ground-truth landmarks with predicted landmarks.
- **Stability Metric: Jitter Analysis**
  - Jitter is assessed on the Menpo-3D dataset to measure prediction consistency across sequential frames.
- This stability metric ensures the model produces smooth predictions across frames, reducing visual noise in video outputs. Our implementation achieved a NME of **4.0068**.

Method	AFLW2000-3D (68 pts)				AFLW (21 pts)			
	[0, 30]	[30, 60]	[60, 90]	Mean	[0, 30]	[30, 60]	[60, 90]	Mean
ESR [14]	4.60	6.70	12.67	7.99	5.66	7.12	11.94	8.24
SDM [50]	3.67	4.94	9.67	6.12	4.75	5.55	9.34	6.55
3DDFA [59]	3.78	4.54	7.93	5.42	5.00	5.06	6.74	5.60
3DDFA+SDM [59]	3.43	4.24	7.17	4.94	4.75	4.83	6.38	5.32
Yu et al. [53]	3.62	6.06	9.56	6.41	-	-	-	-
DeFA [36]	-	-	-	4.50	-	-	-	-
3DSTN [5]	3.15	4.33	5.98	4.49	**3.55**	**3.92**	5.21	**4.23**
3D-FAN [8]	3.15	3.53	4.60	3.76	4.40	4.52	5.17	4.69
3DDFA-TPAMI [60]	2.84	3.57	4.96	3.79	4.11	4.38	5.16	4.55
PRNet [18]	**2.75**	3.51	4.61	3.62	4.19	4.69	5.45	4.77
3DDFA-V2 (M+R)	**2.75**	**3.49**	**4.53**	**3.59**	4.06	4.41	**5.02**	4.50
3DDFA-V2 (M+R+S)	**2.63**	**3.42**	**4.48**	**3.51**	**3.98**	**4.31**	**4.99**	**4.43**



Proposed Extension:

- **Objective**: Leverage 3DDFA_V2 for enhanced expression manipulation and deepfake generation.
- **Dense 3D Landmark Mapping**: 3DDFA_V2 provides 35,000+ dense landmarks, capturing detailed facial structure.
- **Expression Parameter Adjustment**: Modify 3DMM expression parameters to control facial emotions.
- **Transfer Pipeline**:
  - **Landmark Generation**: Capture dense keypoints from source expression.
  - **Expression Manipulation**: Alter 3DMM parameters to create specific expressions (e.g., smile, frown).
  - **Target Mapping**: Apply modified landmarks onto a target face for realistic facial reenactment.

Team Vision Impossible

Uvindu
Kodikara

Ruchira
Abewardhane

Thank You!