**Contributions of HiFace & SD-DeTail**

- Model the static and dynamic details explicitly, demonstrate the benefits of synthetic data in detailed 3D face reconstruction.
- Novel loss functions to learn 3D representations of coarse shape and fine details simultaneously from both the synthetic and real-world images.
- SOTA reconstruction quality both quantitatively and qualitatively, with over 15% performance gains in the region-aware benchmark.
- Easy to plug-and-play into optimization-based methods and can transfer expressions and details from one to another for face animation.

**Overview of HiFace**

- **Reconstruction:** given a single image, HiFace faithfully reconstructs a coarse shape (2nd-row) with vivid details (3rd-row).
- **Animation:** given a source image, HiFace can animate the static (4th-row), dynamic (5th-row), or both (6th-row) details of the driving images.

**Model Architecture**

- Given (a), an image (top) to reconstruct its coarse shape (bottom), we formulate the detail as (b), a static factor and (c), a dynamic factor interpolated by polarized states w.r.t. compressed (top) and stretched (bottom). (d), the output displacement map is linearly combined by (b) and (c) to present vivid details.

**Train HiFace with Synthetic Images & Real-world Images**

Static and Dynamic Detail Losses: $\mathcal{L}_{\text{detail}} = \mathcal{L}_{\text{norm}} + \mathcal{L}_{\text{com}} + \mathcal{L}_{\text{ext}}$

Coarse Shape Losses: $\mathcal{L}_{\text{shape}} = \mathcal{L}_{\text{norm}} + \mathcal{L}_{\text{com}}$

Self-supervised Losses: $\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{norm}} + \mathcal{L}_{\text{com}} + \lambda_{\text{dis}} \mathcal{L}_{\text{dist}}$

Knowledge Distillation: $\mathcal{L}_{\text{kd}}$

Regularization: $\mathcal{L}_{\text{reg}}$

**3D Face Reconstruction Comparisons**

**Flexibility of SD-DeTail**

- Given a source image, we use the driving image to drive its expressions. DECA and EMOCA can animate the expression-driven details but lack realistic.

- As a comparison, HiFace is flexible to animate details from static (4th-row), dynamic (5th-row), or both (6th-row) factors, and presents vivid animation quality with realistic shapes.