Microsoft

CVPR

微软亚洲研究院创研论坛

CVPR 2020 论文分享会

伏羲实验室

Towards High-Fidelity 3D Face Reconstruction from In-the-Wild Images Using Graph Convolutional Networks

Jiangke Lin, Yi Yuan, Tianjia Shao, Kun Zhou

01 Background

01 Background

还有哪些属性可以编辑?

- 脸形
- 发型
- 妆容

如：眉型、眼妆、瞳色、睫毛、腮红、面纹等...

01 Background

- Face-to-Parameter Translation for Game Character Auto-Creation (ICCV, 2019)
- Fast and Robust Face-to-Parameter Translation for Game Character Auto-creation (AAAI, 2020)
- **Towards High-Fidelity 3D Face Reconstruction from In-the-Wild Images Using Graph Convolutional Networks (CVPR, 2020)**

01 Background

3DMMs

3D Morphable Model
Blanz, Volker, and Thomas Vetter. "A morphable model for the synthesis of 3D faces." 1999.

Facewarehouse
Cao, Chen, et al. "Facewarehouse: A 3d facial expression database for visual computing." 2013.

Surrey Face Model
Huber, Patrik, et al. "A multiresolution 3d morphable face model and fitting framework." 2016.

FLAME
Li, Tianye, et al. "Learning a model of facial shape and expression from 4D scans." 2017.

Basel Face Model 17
Gerig, Thomas, et al. "Morphable face models - an open framework." 2018.

Large Scale Facial Model
Booth, James, et al. "Large scale 3D morphable models." 2018.

01 Background

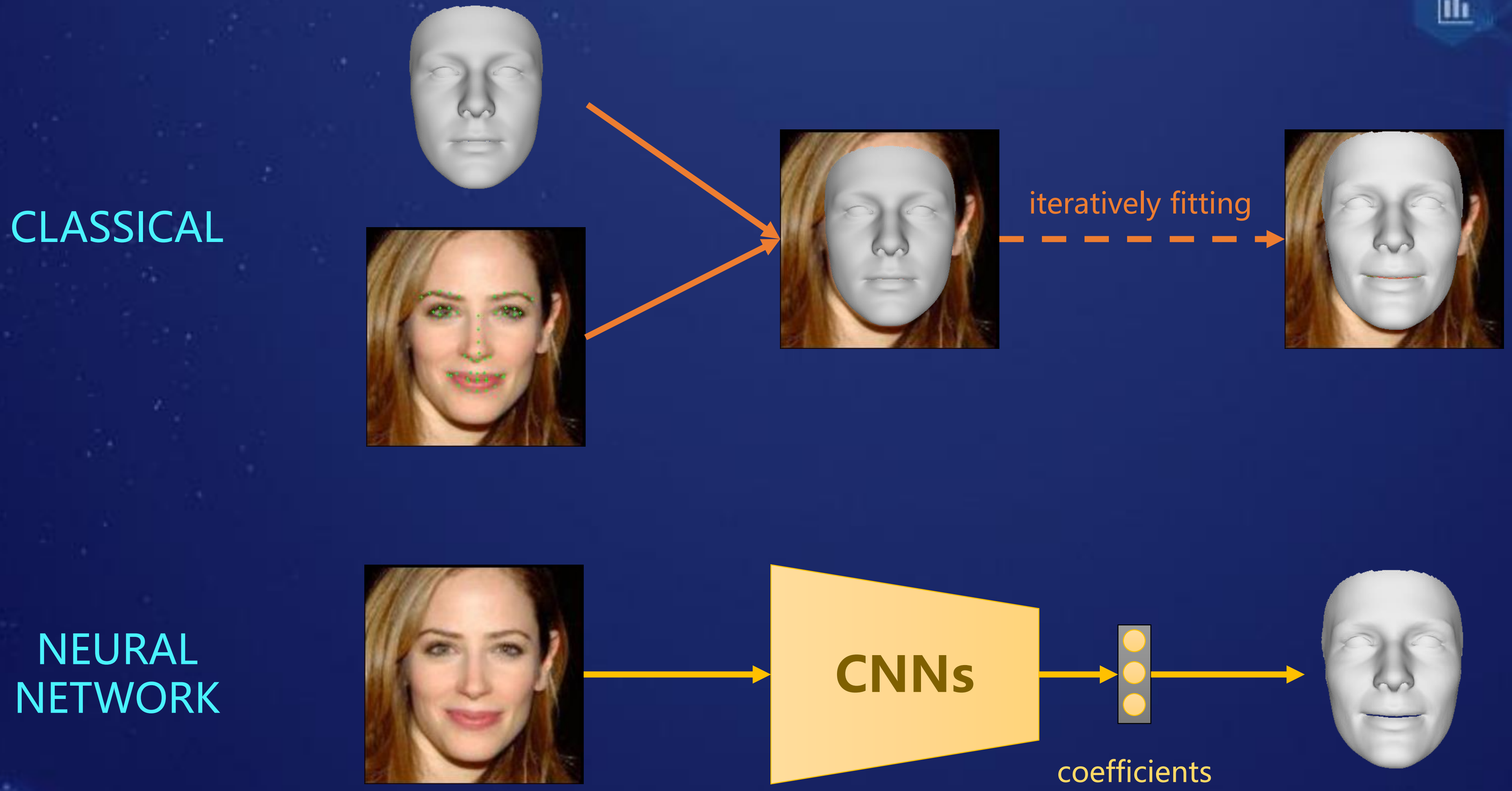
In a 3DMM, given the identity coefficients c_i, expression coefficients c_e and texture coefficients c_t, the face shape S and the texture T can be represented as:

$$S = S_{mean} + c_i I_{base} + c_e E_{base}$$

$$T = T_{mean} + c_t T_{base}$$

where S_{mean} and T_{mean} are the mean face shape and texture, and I_{base}, E_{base} and T_{base} are the PCA bases of identity, expression and texture respectively.

01 Background



02 The Main Idea

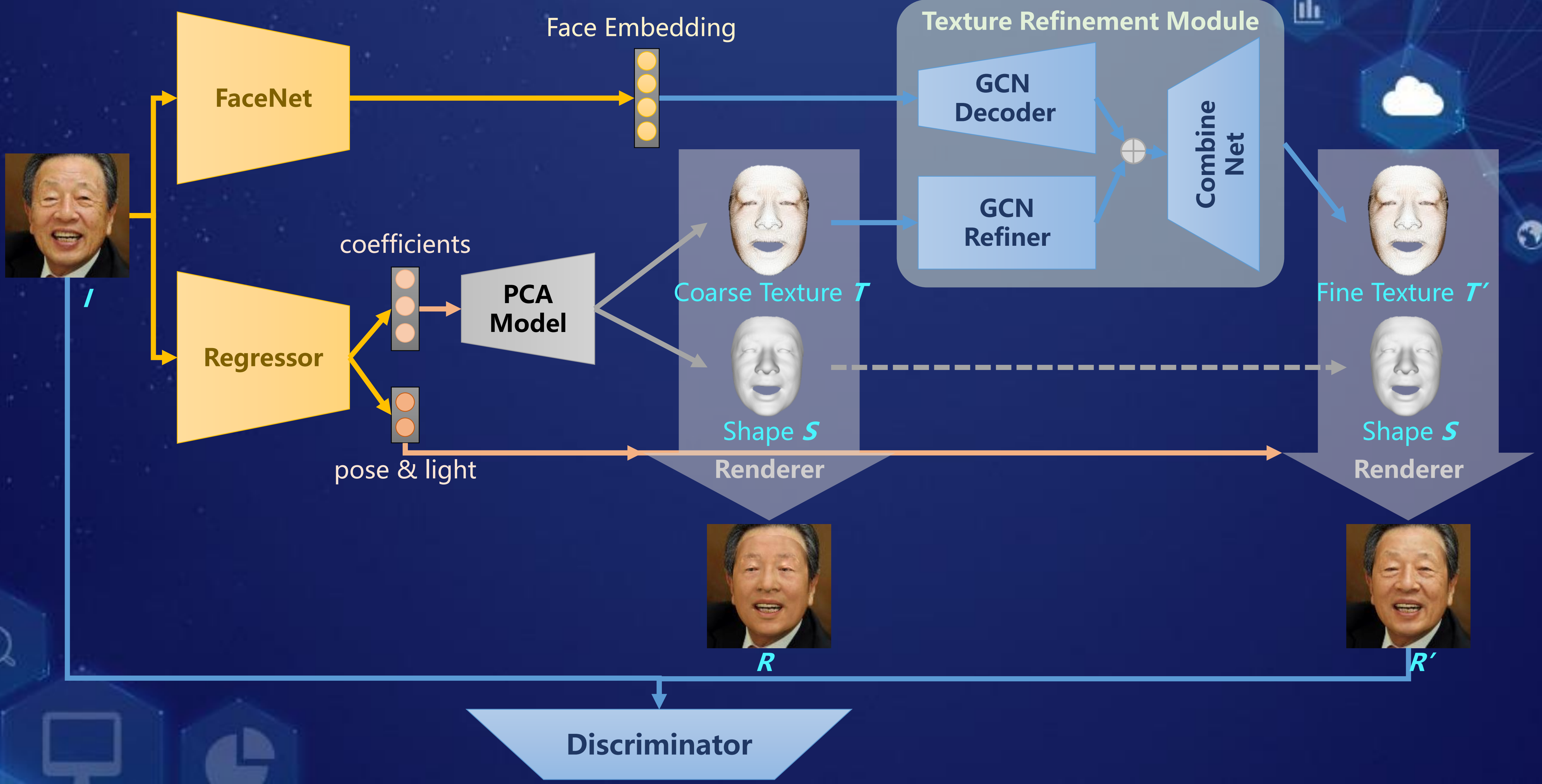
Input Image

Face Shape and Coarse Texture

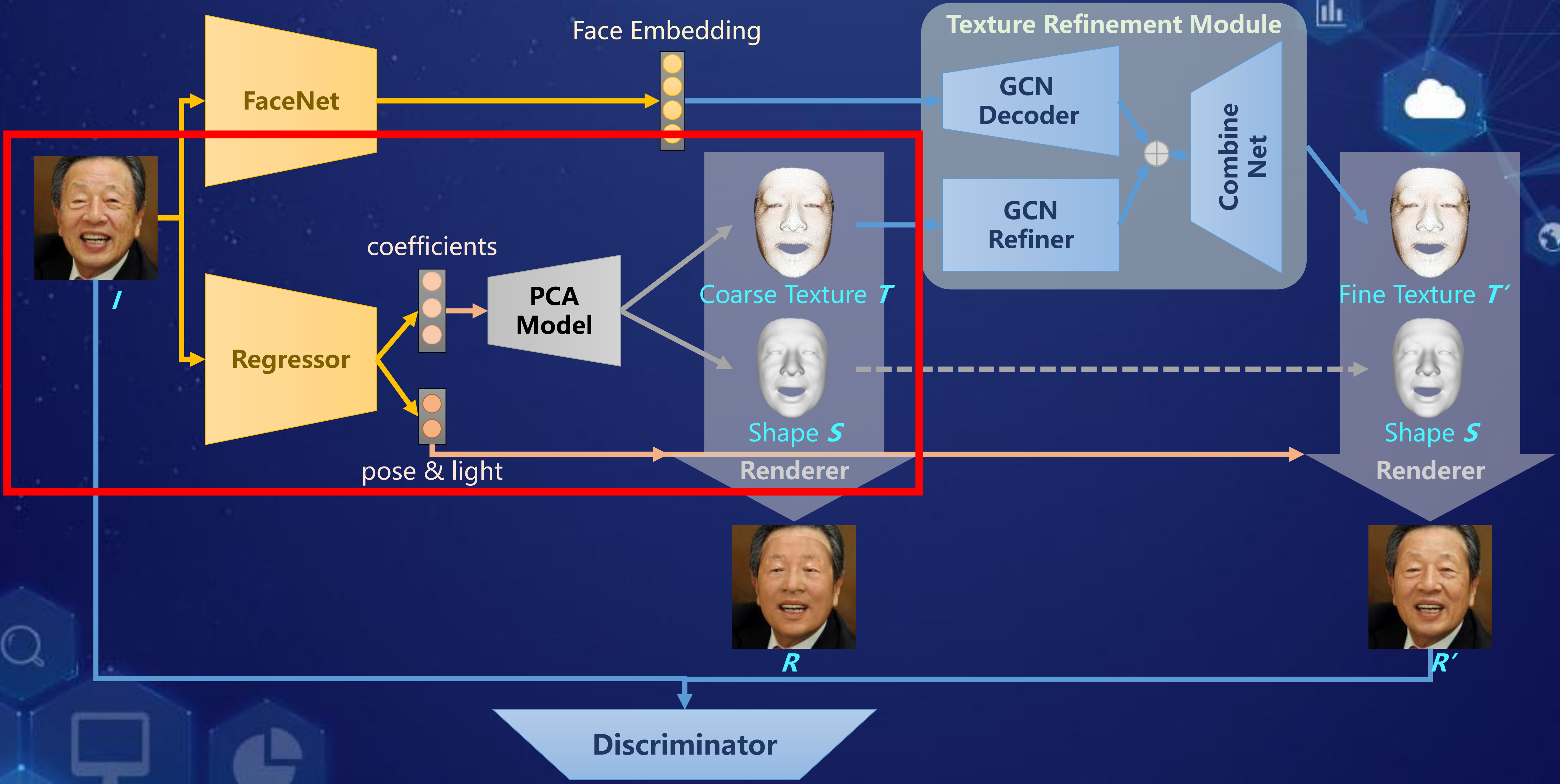
Fine Texture

Project to Image

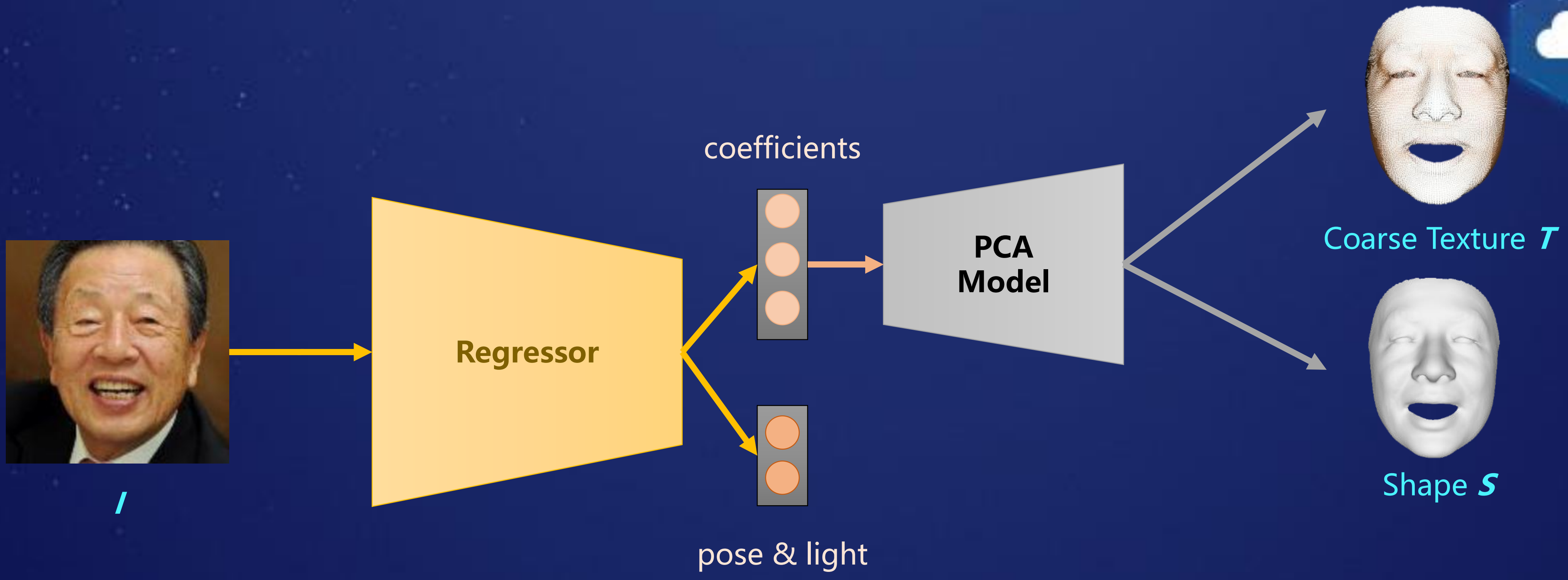
03 The Overall Framework



03 The Overall Framework

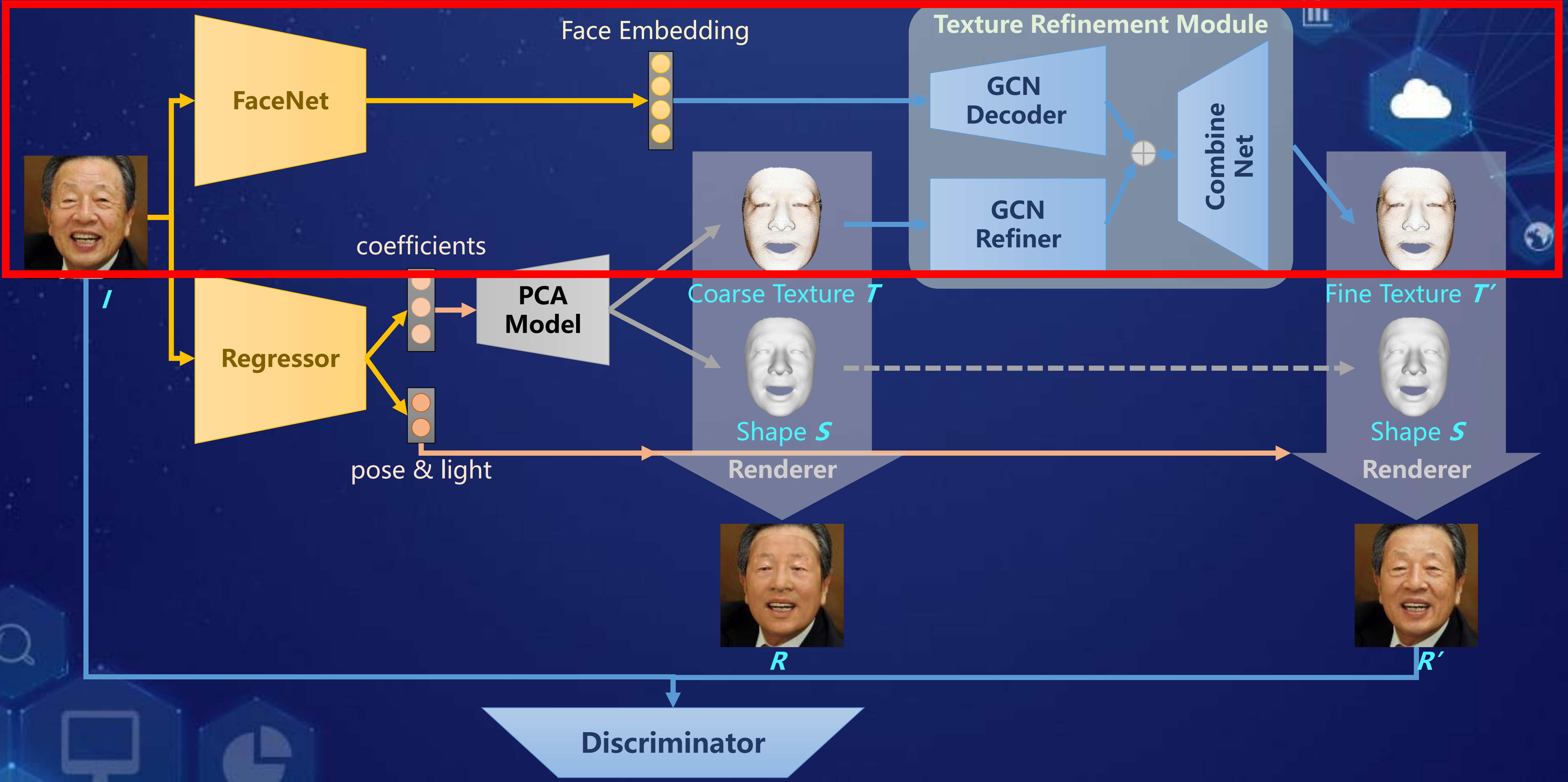


04 Approach

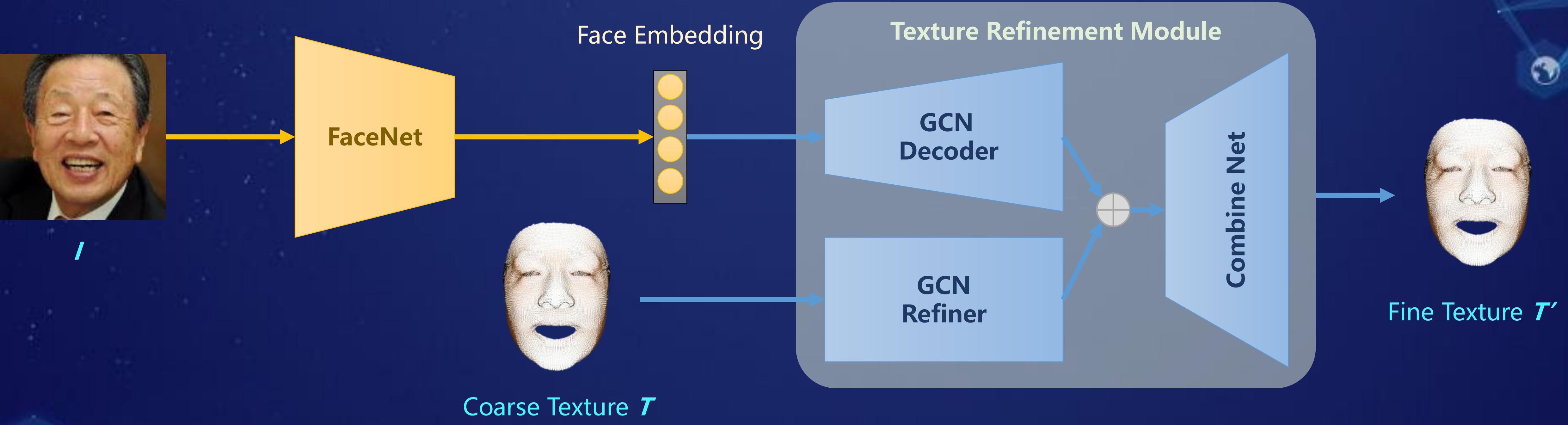


[1] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In IEEE Computer Vision and Pattern Recognition Workshops, 2019.

03 The Overall Framework

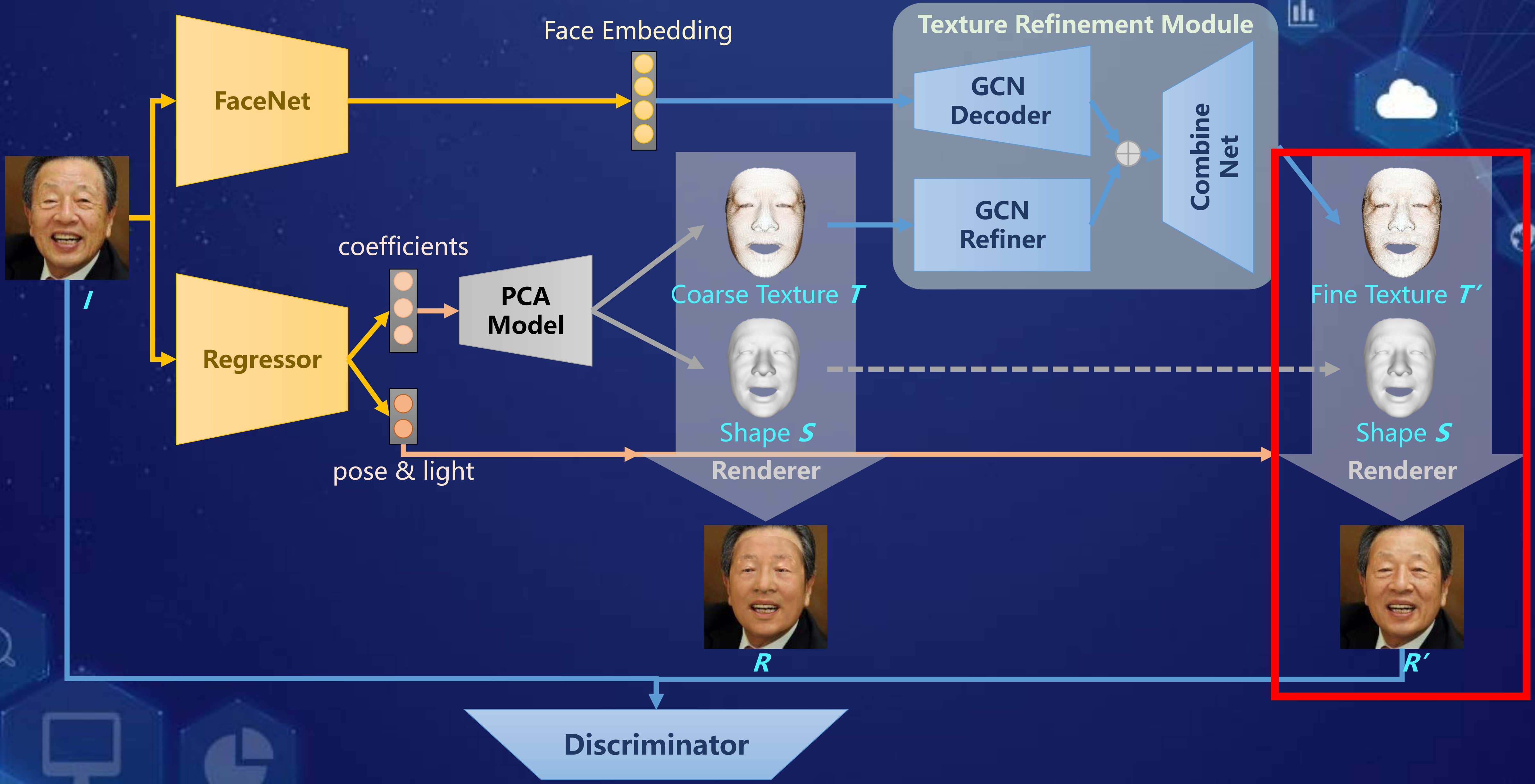


04 Approach

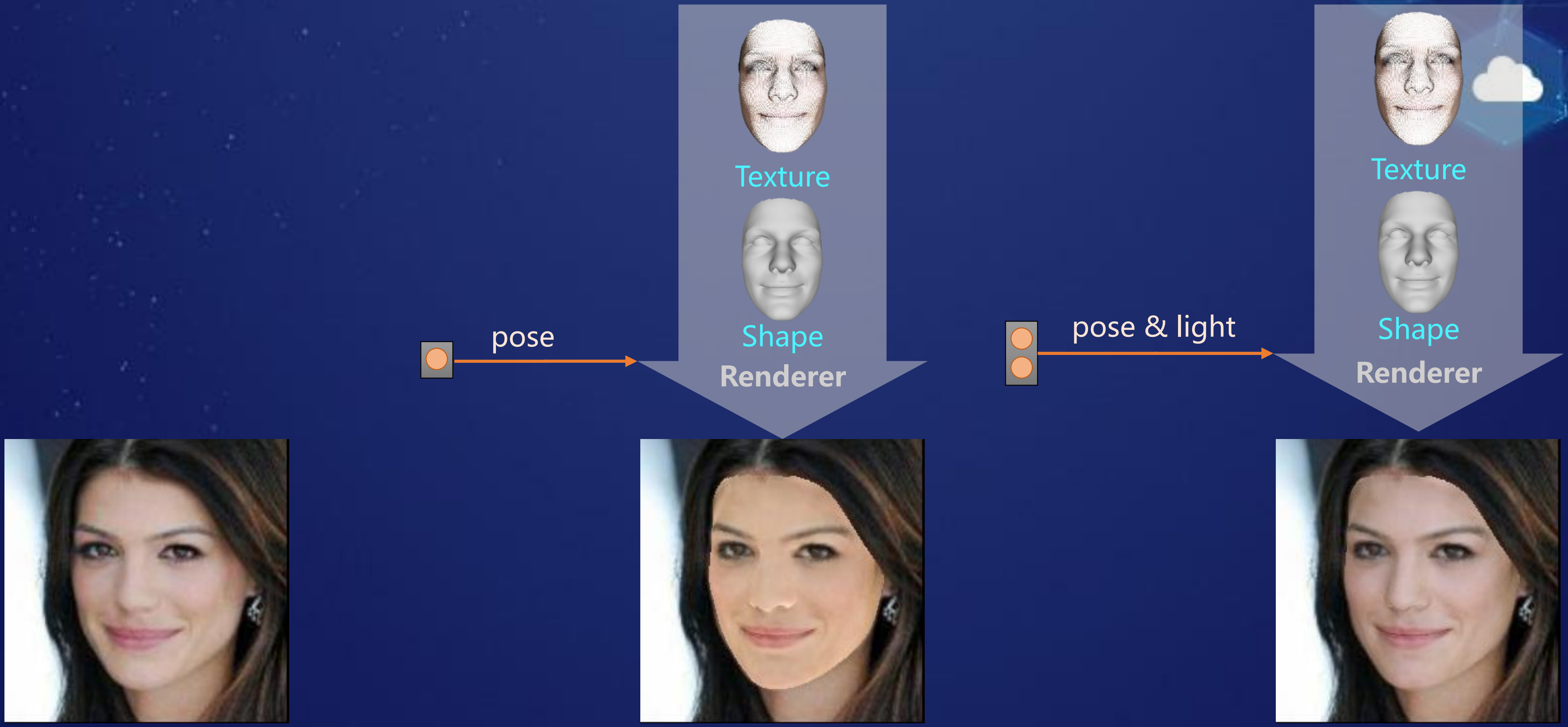


I
FaceNet
Face Embedding
Coarse Texture T
Texture Refinement Module
GCN Decoder
GCN Refiner
Combine Net
Fine Texture T′

03 The Overall Framework



04 Approach

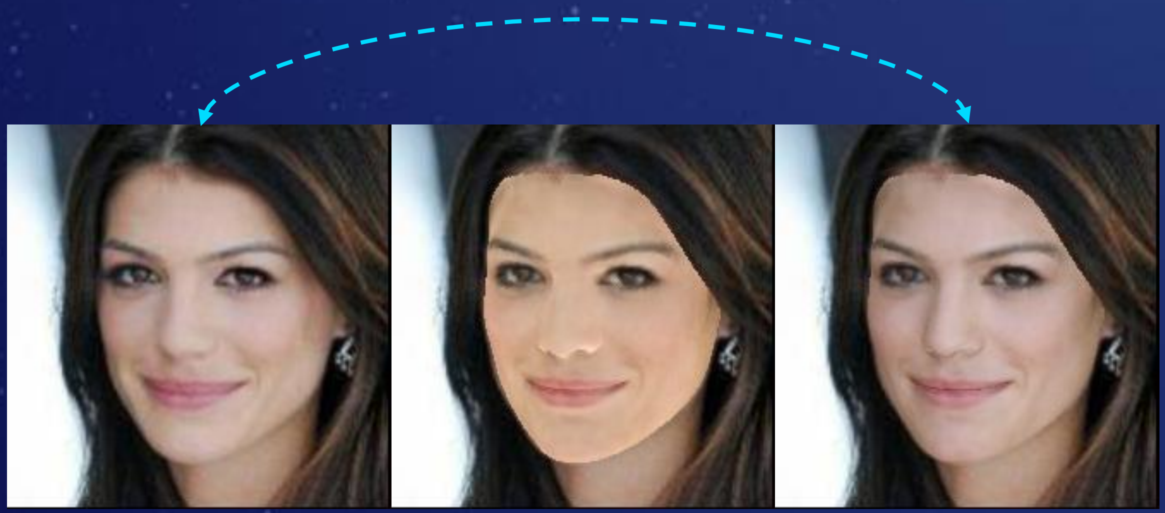


Input Image

Albedo
(generated by GCNs)

Rendered with
Illuminations

05 Losses



Input Image

Albedo (generated by GCNs)

Rendered with Illuminations

Pixel-wise Loss

$$L_{pix}(x, x') = \frac{\sum M_{proj} M_{face} \|x - x'\|_2}{\sum M_{proj} M_{face}}$$

Identity-Preserving Loss

$$L_{id}(x, x') = 1 - \frac{< F(x), F(x') >}{\|F(x)\| \cdot \|F(x')\|}$$

Adversarial Loss

$$L_{adv} = \mathop{\mathbb{E}}_{x' \sim \mathbb{P}_{R'}} [D(x')] - \mathop{\mathbb{E}}_{x \sim \mathbb{P}_I} [D(x)] + \lambda \mathop{\mathbb{E}}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

05 Losses

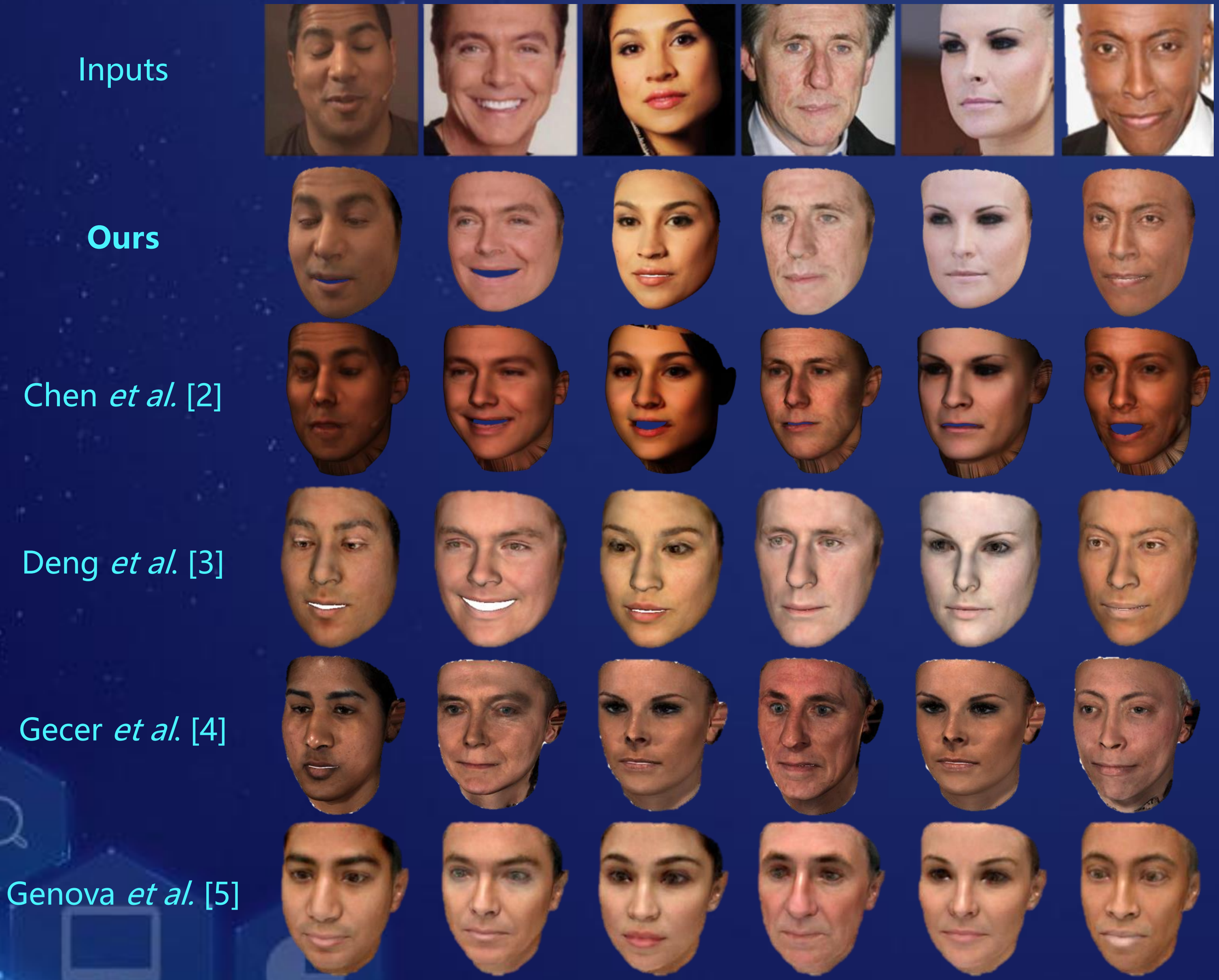
Coarse Texture ***T***   Fine Texture ***T′***

$$L_{vert}(x, x') = \frac{1}{N}\sum_{i=1}^{N} \|x_i - x_i'\|_2$$

05 Losses

$$L_{vert}(x, x') = \sigma_1 \left[L_{pix}\,(I, R') + \sigma_2\, L_{id}(I, R') + \sigma_3 L_{adv}(I, R')\right] + \sigma_4[L_{vert}(T, T') + L_{vert}(T_p, \widetilde{T'})]$$

06 Comparison



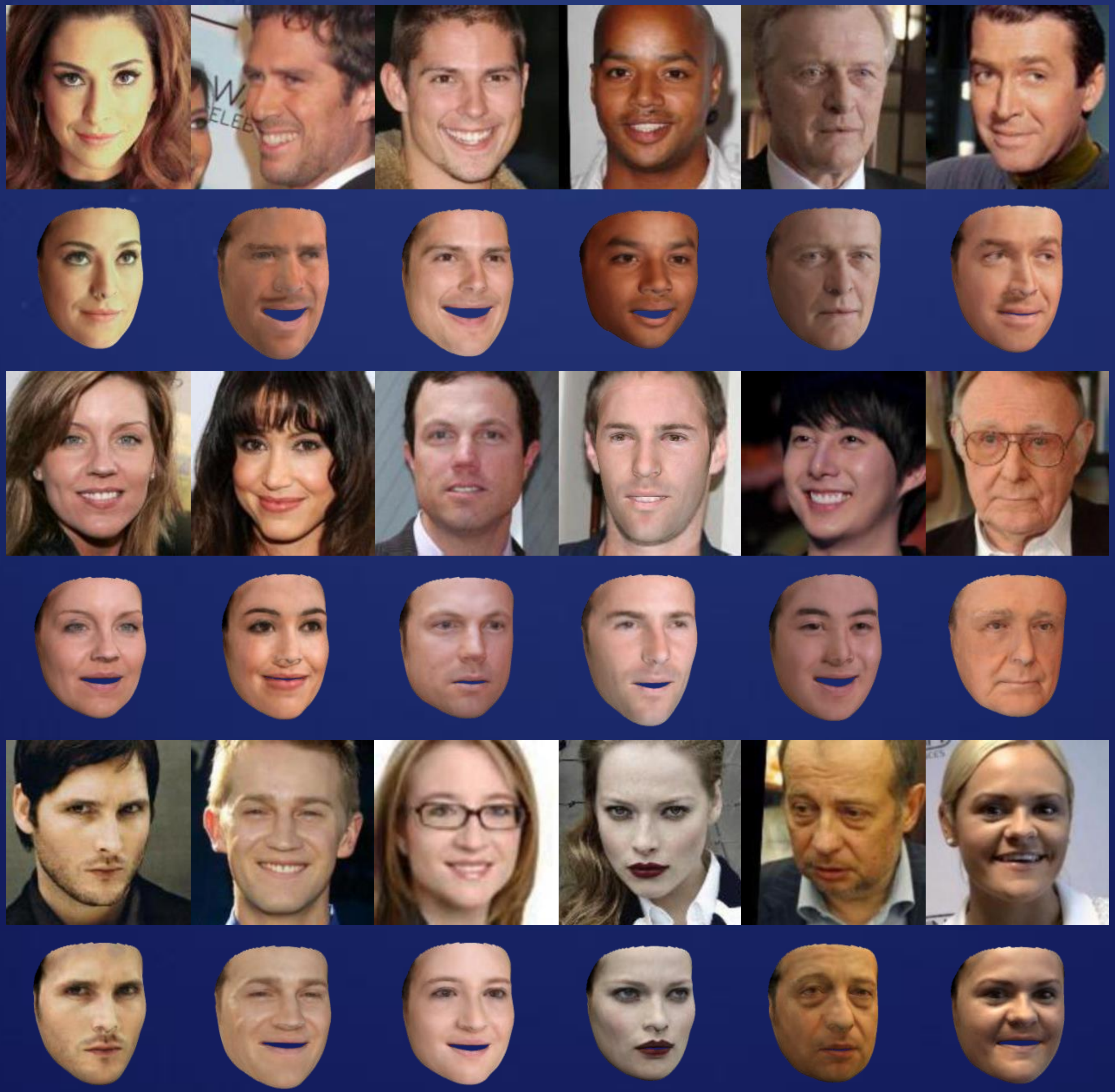
[2] Zhang Chen, Guli Zhang, Ziheng Zhang, Kenny Mitchell, Jingyi Yu, et al. Photo-realistic facial details synthesis from single image. arXiv preprint arXiv:1903.10873, 2019.
[3] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In IEEE Computer Vision and Pattern Recognition Workshops, 2019.
[4] Baris Gecer, Stylianos Ploumpis, Irene Kotsia, and Stefanos Zafeiriou. Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1155–1164, 2019.
[5] Kyle Genova, Forrester Cole, Aaron Maschinot, Aaron Sarna, Daniel Vlasic, and William T Freeman. Unsupervised training for 3d morphable model regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8377–8386, 2018.

07 Ablation Study

Losses				PSNR	SSIM	LightCNN	evoLVe
L_{pix}	L_{id}	L_{adv}	L_{vert}				
				26.58	0.826	0.724	0.641
√	√			28.57	0.863	0.828	0.738
√	√	√		29.30	0.872	0.840	0.755
√	√	√	√	**29.69**	**0.894**	**0.900**	**0.848**

08 More Results

08 More Results



谢谢观看

THANK YOU