Facial Microscopic Structures Synthesis from a Single Unconstrained Image

YOUYANG DU, School of Software, Shandong University, China
LU WANG*, School of Software, Shandong University, China
BEIBEI WANG*, School of Intelligence Science and Technology, Nanjing University, China

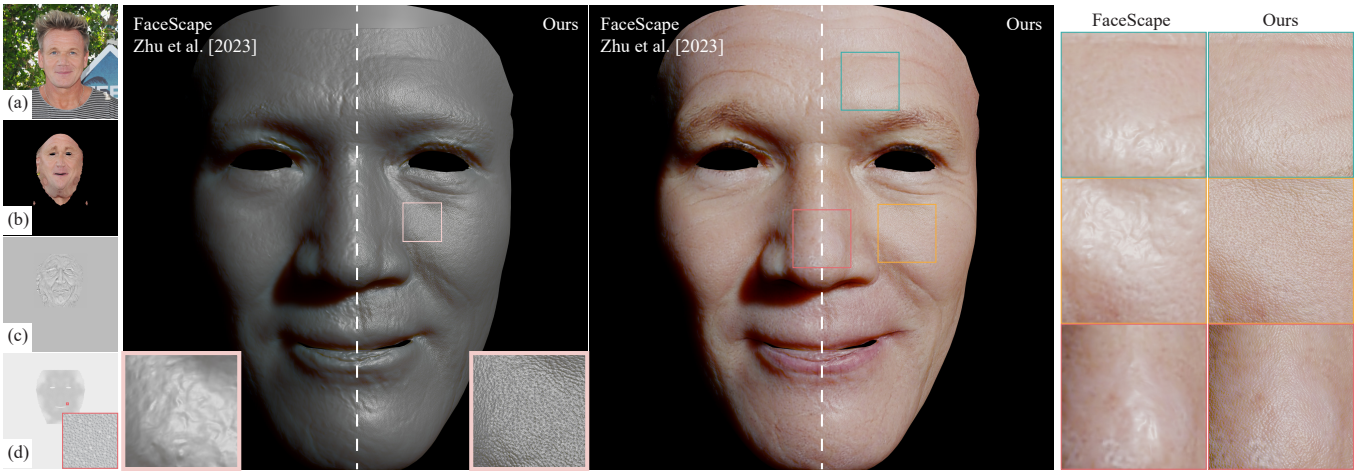

Fig. 1. We present a microscopic facial details synthesis framework (FLEX) from a single unconstrained image. We illustrate our results on the left, including (a) input image, (b) recovered facial texture, (c) macroscopic displacement map, and (d) microstructure displacement map synthesized by our method. The rendered results are shown on the right. Our method can obtain microscopic details, including wrinkles and pores, thereby enhancing the realism of rendering.

Obtaining 3D faces with microscopic structures from a single unconstrained image is challenging. The complexities of wrinkles and pores at a microscopic level, coupled with the blurriness of the input image, raise the difficulty. However, the distribution of wrinkles and pores tends to follow a specialized pattern, which can provide a strong prior for synthesizing them. Therefore, a key to microstructure synthesis is a parametric wrinkles and pore model with controllable semantic parameters. Additionally, ensuring differentiability is essential for enabling optimization through gradient descent methods. To this end, we propose a novel framework designed to reconstruct facial micro-wrinkles and pores from naturally captured images efficiently. At the core of our framework is a differentiable representation of wrinkles and pores via a graph neural network (GNN), which can simulate the complex interactions between adjacent wrinkles by multiple graph convolutions. Furthermore, to overcome the problem of inconsistency between the blurry input and clear wrinkles during optimization, we proposed a Direction Distribution Similarity that ensures that the wrinkle-directional features remain consistent. Consequently, our framework can synthesize facial microstructures from a blurry skin image patch, which is cropped from a natural-captured facial image, in around an average of 2 seconds. Our framework can seamlessly integrate with existing macroscopic facial detail reconstruction methods to enhance their detailed appearance. We showcase this capability on several works, including DECA, HRN, and FaceScape.

CCS Concepts: • **Computing methodologies** → *Rendering*.

Additional Key Words and Phrases: Facial Detail Synthesis, Microscopic Structures, Differentiable Simulation

ACM Reference Format:
Youyang Du, Lu Wang, and Beibei Wang. 2025. Facial Microscopic Structures Synthesis from a Single Unconstrained Image. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '25), August 10–14, 2025, Vancouver, BC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3721238.3730760

1 INTRODUCTION

Creating a 3D human face with rich details is essential for many applications in the modern digital era, including video games, virtual reality, and digital representations of humans. While existing methods have demonstrated impressive capabilities in reconstructing macroscale or even mesoscale structures from a single image, they often struggle to extend to microstructures, such as wrinkles and pores, which are crucial for achieving realism in human faces.

*Corresponding authors.

Authors' addresses: Youyang Du, School of Software, Shandong University, China, duyouyang957@gmail.com; Lu Wang, School of Software, Shandong University, China, luwang_hcivr@sdu.edu.cn; Beibei Wang, School of Intelligence Science and Technology, Nanjing University, China, beibei.wang@nju.edu.cn.

SIGGRAPH Conference Papers '25, August 10–14, 2025, Vancouver, BC, Canada

ACM ISBN 979-8-4007-1540-2/2025/08
https://doi.org/10.1145/3721238.3730760

This is particularly challenging when the single image is captured under natural conditions, as it may only show a blurry structure of the wrinkles and pores.

Extensive efforts have been made to recover high-quality facial geometry from professionally captured images, with either active lighting condition [Kampouris et al. 2018; Ma et al. 2007; Riviere et al. 2020], or passive lighting condition [Beeler et al. 2011; Gotardo et al. 2018]. While these methods successfully reconstruct facial geometries with microscopic details, they often depend on costly and specialized capture setups. On the other hand, following the seminal work of Blanz and Vetter [1999], many studies have focused on reconstructing facial geometries from a single unconstrained image, showing powerful capability at the macro scale [Feng et al. 2021; Lei et al. 2023] and the meso scale [Yamaguchi et al. 2018; Yang et al. 2020], yet they are not designed to handle the microscopic structures. One exception is the work by Weiss et al. [2023], which introduced a parametric facial microstructure model, successfully recovering microscopic structures from a close-up captured partial face image with apparent wrinkles and pores.

In this paper, we aim to synthesize realistic facial microstructure from a single unconstrained image. To achieve this, we introduce a novel framework called FLEX (Fine-Level facial detail EXtraction), which efficiently obtains the facial microscopic structures from a single image captured under unconstrained conditions. At the core of our framework is the differentiable simulation of the microstructure. Inspired by Weiss et al. [2023], the connection of wrinkles and pores follows a special pattern, which can be represented through a graph. Built upon this observation, we propose a novel parametric microstructure model via the graph neural network (GNN) controlled by semantic parameters. In our model, the structure among wrinkles can be modeled by the GNN, and the complex interactions between adjacent wrinkles are simulated through multiple graph convolutions. By leveraging the parametric microstructure model as a prior, the parameters of microstructure could be optimized, as long as the difference between the simulated image with clear microstructures and the blurry input image can be measured. To this end, we propose a novel objective function, Direction Distribution Similarity, which focuses on the consistency of the wrinkle's directional features. Consequently, our method can extract facial microstructure from a blurry skin patch in around 2 seconds. Experimental results show that our synthesized microstructure significantly enhances realism when combined with existing macroscopic facial detail recovery methods.

In summary, our main contributions are:

- A framework to synthesize realistic microstructure of wrinkles and tiny pores, from a single unconstrained facial image.
- A differentiable neural wrinkle simulation that is capable of synthesizing microstructures from semantic parameters by leveraging the graph neural network.
- A novel objective function, Direction Distribution Similarity, to bridge the gap between blurry and clear skin patches by focusing on directional features of wrinkles.

2 RELATED WORK

2.1 Facial Detail Recovery

High-Fidelity Geometry Capture. Capturing 3D detailed facial geometry has been a significant topic in computer graphics for decades. Early approaches primarily utilized active capture systems. Weyrich et al. [2006] corrected coarse facial geometry details by leveraging methods proposed by Nehab et al. [2005]. Ma et al. [2007] employed a polarized gradient-based illumination system to recover high-quality surface normals of human faces. Subsequent works [Fyffe et al. 2016; Kampouris et al. 2018; Riviere et al. 2020] aimed to accelerate capture time, simplify lighting setups, and reconstruct high-quality facial geometry along with reflectance materials. Graham et al. [2012] proposed a method for recovering microscopic details through an example-based synthesis approach [Hertzmann et al. 2001], yet it still relies on high-quality displacement maps and several specially scanned facial patches as inputs.

In contrast, another line of work focuses on reconstructed 3D facial geometry under passive lighting conditions. Beeler et al. [2010] proposed a passive stereo system capable of capturing high-quality 3D facial geometry with pore-level details under standard lighting conditions. Their subsequent study [Beeler et al. 2011] improved temporal consistency for dynamic facial performance capture. The following works [Gotardo et al. 2018; Lattas et al. 2022] shorten the capture time and leverage affordable equipment. Cooperating with deep learning, some works [Huynh et al. 2018; Li et al. 2021; Liu et al. 2022; Xiao et al. 2022] also utilize complicated deep neural networks to infer facial details from captured images. However, these approaches often rely on images from complex and structured capture setups.

Recovery from Unconstrained Images. To address the challenge of recovering facial details from unconstrained environments, including in-the-wild scenarios, many studies have extended the basic work of 3DMM [Blanz and Vetter 1999; Gerig et al. 2017; Jiang et al. 2021; Li et al. 2017]. Some methods [Chai et al. 2023; Chen et al. 2019; Lattas et al. 2023; Saito et al. 2016; Yamaguchi et al. 2018; Zhu et al. 2023] leverage image-to-image inference from textures extracted from monocular images. Others [Dib et al. 2023; Tewari et al. 2018] employ differentiable rendering techniques to recover facial details, while some works [Chai et al. 2023; Feng et al. 2021] enable networks to learn dynamic facial wrinkle reconstruction in a self-supervised manner. While these approaches achieve realistic appearance modeling at macroscopic or even mesoscopic levels, their designs, which focus on the entire face region, struggle to capture finer details such as micro-wrinkles and tiny pores.

2.2 Wrinkle Simulation

Microscopic wrinkle simulation methods have been empirical in the past, often providing a plausible appearance but not being realistic enough. Wu et al. [1996] proposed a method for simulating static and dynamic wrinkles by combining biomechanical modeling and bump mapping, which effectively balanced visual realism and computational efficiency. Building on such foundations, Bando et al. [2002] introduced a simpler method using intuitive parameters to model fine- and large-scale wrinkles based on Bézier curves. To

Table 1. A brief preview of the main parameters and their description defined by Weiss et al. [2023]. Here we only display the parameters that directly influence pores (top) and wrinkles (bottom).

Parameter	Description
ϕ	The average of pore distance
β	The blending strength from wrinkles to pores
θ	The main direction of wrinkles
μ	The degree of wrinkles' orientation uniformity
δ	The degree of influence of wrinkle length
σ	The degree of contiguity of wrinkle
γ	The tolerance of wrinkle crossing
τ	The tolerance of similar wrinkle
ρ	The deposit of wrinkle depth during iteration

make the static microstructure compatible with dynamic expression, Nagano et al. [2015] proposed to use convolution to simulate the deformation of the microstructure when the facial surface is stretching or shrinking.

Recently, Weiss et al. [2023] proposed a novel graph-based wrinkle simulation method, in which they consider the micro-wrinkles and pores as the edges and nodes in the graph, respectively. Their graph-based assumption is consistent with the natural relationship between pores and wrinkles, resulting in realistic simulation results. Furthermore, to help the artists model the microstructures, they also leveraged Particle Swarm Optimization (PSO) [Kennedy and Eberhart 2002] to obtain sets of parameters from clearly captured skin displacement patches. However, due to the non-differentiable simulation, their optimization is time-consuming. Their following work [Weiss et al. 2024] trained a StyleGAN [Karras et al. 2019] to generate micro-wrinkle patches. While the GAN-based synthesis yields impressive results due to the adversarial training mechanism, it is a data-driven method, which highly depends on training data and its resolution. In contrast, in our Neural Wrinkle Simulation, the simple GNN learns the complex interactions among wrinkles from a limited dataset, enabling us to synthesize microstructures with arbitrary density and number of layers.

3 PRELIMINARY

The graph-based parametric facial microstructure model was first introduced by Weiss et al. [2023]. Their model takes several semantic parameters as input and then simulates the facial micro-wrinkles and pores, which will be represented as a displacement map. We show the main semantic parameters that control the appearance of wrinkles and pores in Table 1. This graph-based microstructure model assumes that the wrinkles and pores correspond to the edges and nodes on the graph structure, respectively. In this graph, each wrinkle and pore is assigned a depth value. As wrinkles and pores are connected to each other, the depth change of one wrinkle will influence its adjacent wrinkles, and the depth updating is operated iteratively.

Starting from an empty graph with an initialized wrinkle depth of zero, each wrinkle will update its depth value in each iteration step with a probability. More precisely, for a wrinkle w_i, at the k^{th} step of iteration, the probability $p_i^{(k)}$ to increase its depth is determined by a function f:

$$p_i^{(k)} = f(\theta, \mu, \delta, \sigma, \gamma, \tau; w_\theta^i, w_l^i, W^{(k)}, G(\phi)), \tag{1}$$

where $G(\phi)$ is the graph structure created from the pore distance ϕ, w_θ^i is the wrinkle's direction, w_l^i is its length, and $W^{(k)}$ is the situation of other surrounding wrinkles. According to the probability $p_i^{(k)}$, the wrinkle's depth is iteratively updated from $d_i^{(k)}$ to $d_i^{(k+1)}$:

$$d_i^{(k+1)} = d_i^{(k)} + \Delta t(1 - d_i^{(k)})\rho, \tag{2}$$

in which the $\Delta t = 0.03$ is a fixed global time stride. The depth q of a pore can be calculated based on the depth of wrinkles connected to it:

$$q = \max(d_{max}, d_{max} + \beta(d_{sum} - d_{max})), \tag{3}$$

where d_{sum} is the sum depth of neighbouring wrinkles and d_{max} is the max depth among them.

In the model of Weiss et al. [2023], the parameters only participate in calculating the probability of the iteration instead of the final wrinkle depth itself, leading to the non-differentiability between the wrinkle depth and the parameters. Instead, we propose our graph-based parametric model to present microstructures in a neural way and give a differentiable solution to optimize them.

4 OUR METHOD

4.1 Overview

Given a single unconstrained facial image, our method aims at synthesizing the facial microstructures of micro-wrinkles and tiny pores from it. Unconstrained images refer to images captured under unconstrained photographic environments, where unconstrained environments denote conditions where lighting, camera setup, and hardware equipment are uncontrolled or unknown. We represent the microstructures by a differentiable parametric model, with several semantic parameters of wrinkles and pores. We treat our main microstructure recovery problem as an optimization problem. For the target facial microstructure $\mathcal{W}'$, it aims to obtain the microstructure parameters by optimizing:

$$\arg\min_x \mathcal{L}(\mathcal{W}(x), \mathcal{W}'), \tag{4}$$

where $\mathcal{W}$ is our parametric microstructure model and $\mathcal{L}$ is our objective function. We define $x = (\phi, \beta, \theta, \mu, \delta, \sigma, \tau, \rho)$ as our semantic parameter, in which ϕ and β are parameters of the pore, and the others are the parameters of the wrinkle. The description of each component can be found in Table 1.

We first present the pipeline of our framework, FLEX (Sec. 4.2), which takes an unconstrained facial image as input and synthesizes the microstructure of the whole face. Then we present the core of our framework, which is the differentiable parametric microstructure model $\mathcal{W}$ (Sec. 4.3), namely Neural Wrinkle Simulation. Finally, we show our proposed objective function (Sec. 4.4), Direction Distribution Similarity, to enable a reasonable optimization.

4.2 FLEX

FLEX is a framework focused on extracting the entire facial microstructure from a single unconstrained facial image, as shown in Figure 2. We first use current macroscopic facial detail recovery

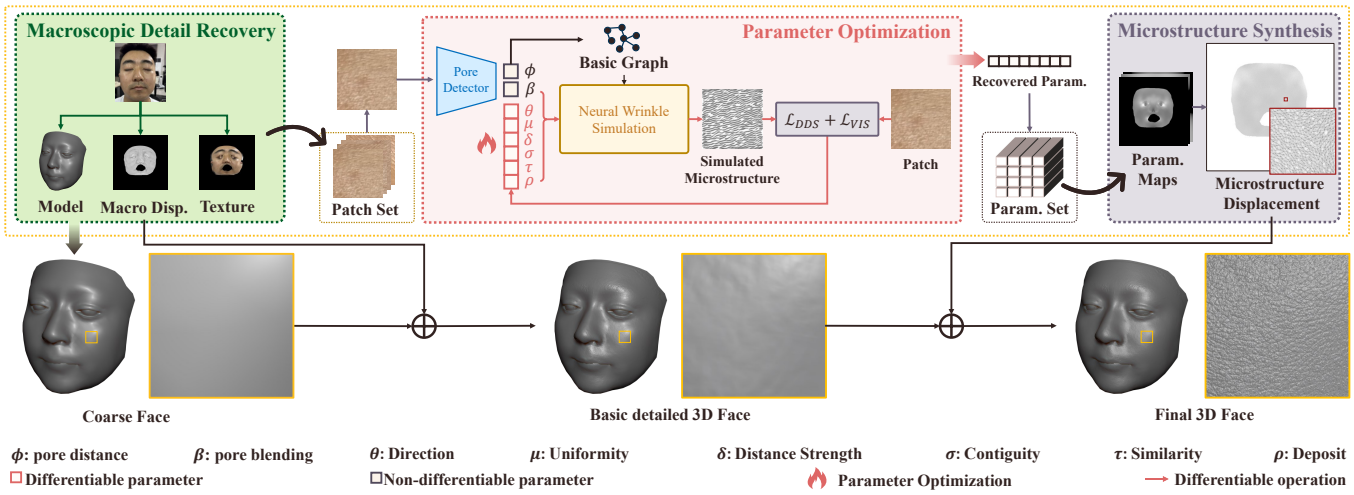

Fig. 2. An overview of the entire pipeline of FLEX. FLEX first extracts geometry, macroscopic details, and texture maps from the facial image. Then, a set of facial patches is sampled from the face, and each patch is passed to the parameter optimization module to obtain its parameters of wrinkles and pores. Finally, the optimized parameter set of all patches is used to synthesize the facial microstructure for the entire face.

methods to recover macro details for the whole face, then parameters are recovered from a skin patch to present its microstructures. Finally, the parameters of all patches are combined into a set to synthesize the microstructure displacement map of the entire face.

Macroscopic Detail Recovery. For a single unconstrained facial image, we use any off-the-shelf macroscopic detail recovery method (such as FaceScape [Zhu et al. 2023], DECA [Feng et al. 2021], and HRN [Lei et al. 2023]) to warp the input face image into texture space, and recover the 3D coarse model, face texture, and macroscopic displacement map for the whole face.

Parameter Optimization. We crop a skin patch from the texture space by uniformly random sampling, then recover pore and wrinkle parameters for this patch. The pore parameters, including pore distance ϕ and blending strength β, are easily predicted by a pore detector $\mathcal{P}$ from a skin patch I':

$$(\phi, \beta) = \mathcal{P}(\mathcal{Y}(I'), \mathcal{H}(I'), \mathcal{T}(I')), \tag{5}$$

where $\mathcal{Y}(I') \in \mathbb{R}^{H\times W}$ is the gray-scale of the facial patch I', $\mathcal{H}(I') \in \mathbb{R}^{H\times W}$ is the high-pass filtered image from I', and $\mathcal{T}(I') \in \{0,1\}^{H\times W}$ is the binary image thresholded from $\mathcal{H}(I')$. Details of patch sampling are presented in the supplementary material.

For the wrinkle parameters, we leverage our neural wrinkle simulation $\mathcal{W}$ for gradient descent optimization, as shown in Equation 4. A core component of our neural wrinkle simulation is a GNN, which is introduced in Section 4.3. To optimize the parameters of microstructure $\mathcal{W}'$ in the skin patch I', we define the objective function as:

$$\mathcal{L}(\mathcal{W}(x), \mathcal{W}') \triangleq \lambda_d \mathcal{L}_{DDS}(\mathcal{W}(x), I') + \lambda_v \mathcal{L}_{VIS}(\mathcal{W}(x), I'), \tag{6}$$

where $\mathcal{L}_{DDS}$ is the direction distribution similarity, and $\mathcal{L}_{VIS}$ is visual similarity. Details of these losses will be presented in Sec. 4.4 and Sec. 5.

Microstructure Synthesis. We obtain several discrete parameter points scattered across the texture space after optimization. To form the spatially-varying parameters map, we use the radial basis function interpolation in the texture space. Finally, we leverage a simplified version of Weiss23's forward model to synthesize the final displacement map using our recovered spatially-varying parameter maps.

4.3 Neural Wrinkle Simulation

Our goal is to find a differentiable solution that obtains the proper wrinkle depth based on given wrinkle parameters. An important characteristic of wrinkles is that adjacent wrinkles will affect the depth of each other. Previous work of Weiss et al. [2023] simulates the depth of wrinkles by probability-based iteration, which is an efficient way to consider the influence of neighboring wrinkles. However, it is difficult to establish a direct relationship between wrinkle parameters and wrinkle depth, which makes this method non-differentiable. Our key insight comes from the compatibility between graph neural networks and graph-based microstructure models. Specifically, given a graph structure of wrinkles and pores, the complex interactions between wrinkles can be simulated through multiple graph convolutions in the differentiable graph neural network. The design of our GNN-based neural wrinkle simulation is shown in Figure 3.

Graph-based Microstructure Model. Similar to Weiss et al. [2023], we represent the pattern of wrinkles and pores with a graph $G = (V, E)$, where a pore corresponds to a node, and a wrinkle corresponds to an edge. Each node or edge is assigned a depth value. The predicted pore distance ϕ (see Figure 4 a) provides the average distance between nodes, and each node is then connected to its six nearest neighboring nodes to form a basic graph structure G^{ϕ} (see Figure 4 b).

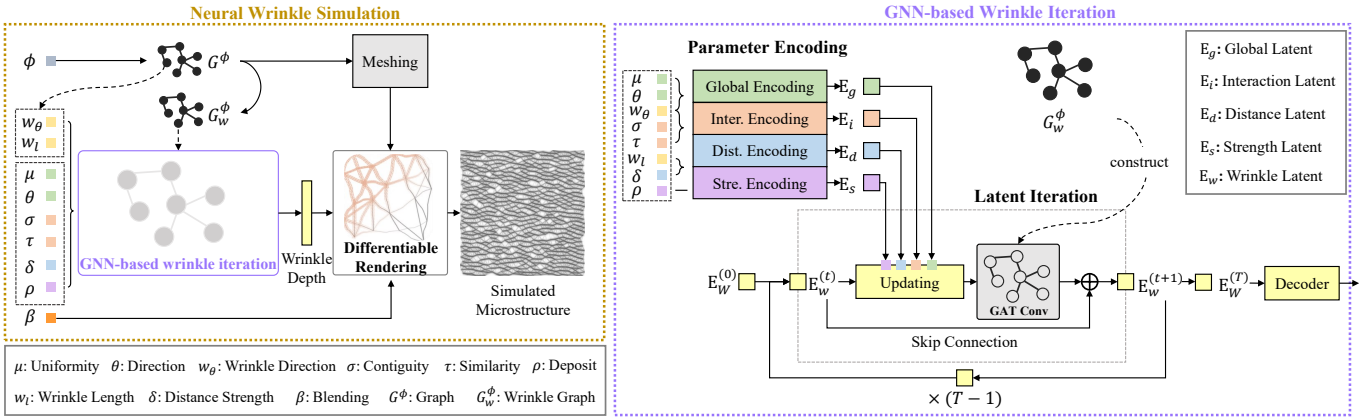

Fig. 3. The illustration of neural wrinkle simulation. We present the overall structure of our neural wrinkle simulation (left) and the details of GNN-based wrinkle iteration (right).

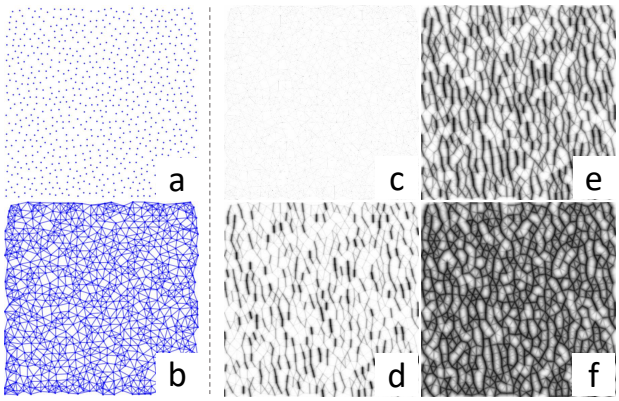

Fig. 4. The illustration of our graph-based microstructure model. We sampled the positions of pores based on pore distance (a) and connected them to graph (b). The iteration depth of wrinkles is shown from (c) to (f). Results are from our neural wrinkle simulation after the number of 0, 3, 5, and 7 iterations, respectively.

With the graph representation G^ϕ, we construct its dual graph G_w^ϕ, where wrinkles are represented as nodes. It indicates that if two wrinkles in graph G^ϕ share the same pore, their related nodes in G_w^ϕ are considered connected. This way, we can provide the GNN with G_w^ϕ, and perform wrinkle simulation by multiple graph convolutions.

Structure of Neural Wrinkle Simulation. Building upon the two key graph structures G^ϕ and G_w^ϕ, we propose a *neural wrinkle simulation* component designed to simulate facial microstructure based on semantic parameters. As shown in Figure 3 (left), it consists of two parts: GNN-based wrinkle iteration and Differentiable Rendering. Specifically, GNN-based wrinkle iteration uses G_w^ϕ to obtain the depth of each wrinkle based on the given parameters. Differentiable Rendering transforms G^ϕ into a microstructure base mesh and renders the wrinkles and pores in the UV space to simulate the microstructure (i.e., a displacement map).

GNN-based Wrinkle Iteration. Considering that the depth of adjacent wrinkles will affect each other, the wrinkle depth is computed iteratively through GNN (Figure 4 c-f). The specific iteration is shown in Figure 3 (right). At the beginning, we assign the template latent vector $\mathrm{E}_w^{(0)}$ onto each node in G_w^ϕ, representing the depth of the wrinkle in a higher dimension. Considering the different characteristics of the parameters, we pre-encode the parameter x, as well as wrinkle direction w_θ and wrinkle length w_l from G^ϕ, into four different latent vectors: global latent E_g, interaction latent E_i, distance latent E_d, and strength latent E_s. At t^{th} iteration, the updating rule from $\mathrm{E}_w^{(t-1)}$ to $\mathrm{E}_w^{(t)}$ is:

$$\mathrm{E}_w^{(t)} = \mathrm{E}_w^{(t-1)} + \mathcal{G}(\mathrm{E}'_g + \mathrm{E}'_i + \mathrm{E}'_d + \mathrm{E}'_s; G_w^\phi), \tag{7}$$

$$\mathrm{E}'_k = \Theta_k(\mathrm{E}_k \cdot \mathrm{E}_w^{(t)}), k \in \{g, i, d, s\}, \tag{8}$$

where the $\mathcal{G}(\cdot; G_w^\phi)$ is the GAT convolution [Velickovic et al. 2017] based on the wrinkle graph G_w^ϕ and Θ_k is a linear transformation for each latent of E_g, E_i, E_d and E_s. After T iterations, we obtain the final wrinkle-varying $E_w^{(T)}$, and then decode this wrinkle latent into the depth of wrinkle:

$$D_w = \Phi_w(\mathrm{E}_w^{(T)}), \tag{9}$$

where the decoder Φ_w is a 4-layer MLP and D_w is the list of wrinkle depth. In practice, we set $T = 7$. We obtain the list of pore depth D_p following the Equation 3, and the final displacement map of microstructure I_x can be obtained by:

$$I_x = \mathcal{F}(D_w, D_p; G^\phi), \tag{10}$$

where $\mathcal{F}$ is the differentiable rendering.

4.4 Direction Distribution Similarity

Skin patches from the unconstrained image are missing a certain degree of detail, leading to a blurry appearance. Thus, rather than accurately matching, we aim to recover the wrinkle parameters that

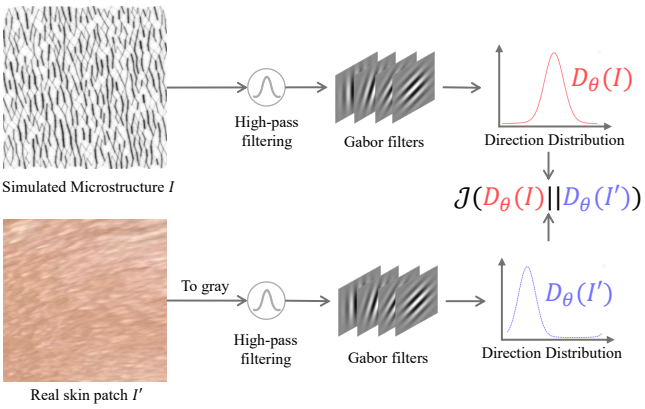

Fig. 5. Illustration of the direction distribution similarity. Given two patches, we first extract the high-pass filtered images from their grayscale images, then compute the Direction Distribution using several Gabor filters. Finally, we measure the similarity using Jensen-Shannon Divergence.

can synthesize a microstructure with the highest similarity to the target facial patch. We observe that the directional features of micro-wrinkles are the most important factor affecting the appearance of skin. However, common objective functions fail to capture such features and are prone to getting stuck in local minima during optimization. To overcome this, we propose a novel objective function, named Direction Distribution Similarity. Specifically, the intensity $p(I;\theta)$ of a given direction θ on a micro-wrinkle patch I is:

$$p(I;\theta) = \mathrm{Var}(f(\theta) \otimes I), \tag{11}$$

where $f(\theta)$ is the Gabor filter oriented to θ and the $\otimes$ is an image convolution operation. The variance of the filtered image indicates the strength of the image at direction θ. To make the directional feature obvious, we convert the patch into grayscale and then apply high-pass filtering to it before passing it to the Gabor filters. The direction intensities across the definition domain of $[0, \pi)$ form the direction distribution $\mathcal{D}_\theta(I)$, which is normalized:

$$\int_0^\pi p(I;\theta)\mathrm{d}\theta = 1. \tag{12}$$

To match up the directional features, the similarity between the simulated microstructure patch I to a given skin patch I' can be measured by:

$$\mathcal{L}_{DDS}(I, I') \triangleq \mathcal{J}(\mathcal{D}_\theta(I)||\mathcal{D}_\theta(I')), \tag{13}$$

where $\mathcal{J}(\mathcal{D}_\theta(I)||\mathcal{D}_\theta(I'))$ is the Jensen-Shannon divergence between two different distributions $\mathcal{D}_\theta(I)$ and $\mathcal{D}_\theta(I')$.

5 IMPLEMENTATION DETAILS

Dataset. We synthesize two different datasets for training the pore detector net and the neural wrinkle simulation net, respectively. Pore detector net is trained on 50k synthesized skin patches with the size of 256^2, in which micro-wrinkles and tiny pores can be seen. The neural micro-wrinkle simulation net is trained on 50k micro-wrinkle displacement maps with the size of 256^2, consisting of 20 different graph structures. The details of constructing those datasets are presented in the supplementary materials.

Training. The loss function of pore detector training is the MSE between the ground truth and predicted parameters. The loss function of neural wrinkle simulation network training consists of three terms with equivalent weights: the MSE between predicted wrinkle depth array and ground-truth depth array, image MSE, and LPIPS [Zhang et al. 2018] between drawn wrinkle patch and ground-truth wrinkle patch. The networks are implemented in the PyTorch framework. We use Adam [Kingma and Ba 2014] as our optimizer with a learning rate of $1e-4$ in our training. Both the pore detector and neural wrinkle simulation were trained with 10 epochs on each dataset, which took 5 hours and 18 hours on a single NVIDIA RTX 3090 GPU, respectively.

Optimization. We adopt the style loss proposed by Huang and Belongie [2017] as our visual similarity $\mathcal{L}_{VIS}$, which computes the MSE of the mean and var of the feature of some VGG layers. We select $\lambda_d = 10$ and $\lambda_v = 0.001$ as the corresponding weights, respectively, and the optimization takes 75 iterations on a single NVIDIA RTX 3090 GPU.

6 RESULTS

In this section, we demonstrate the results of our method by synthesizing microstructure displacement maps from single-view unconstrained facial images. Apart from our method **FLEX**, we additionally set up two other microstructure representation methods that are:

- **Uniform**: with pore distance recovered by our method and fixed parameters of wrinkles.
- **Noise**: generated by a noise model from von der Pahlen et al. [2014], which is commonly used in real-time face rendering.

The unconstrained images are from Pexels [Pexels 2025] and the open source repository of Zhu et al. [2023], in which the facial microstructures are visible and no cover is on the face.

6.1 Qualitative Evaluation

Compare with image patches. We integrate FLEX with FaceScape [Yang et al. 2020] and synthesize the microstructure displacement map from several unconstrained facial images, as shown in Figure 6. The results show the synthesized microstructure compared with the same view of the zoomed-in input image patch. The synthesized pores of our method have a similar density to the input, and the appearance of the synthesized wrinkles is also close to the input. We integrate FLEX with different existing macroscopic facial detail recovery methods, including FaceScape [Zhu et al. 2023], DECA [Feng et al. 2021], and HRN [Lei et al. 2023] to demonstrate our compatibility, as shown in Figure 7. Considering rendering with texture, we show our rendering result under a point light in Figure 1. We also show the comparison of rendering results with textures under different environment lighting in Figure 8, it is easy to see that our method greatly improved the realism of the face.

Compare with ground truth texture. To further validate the quality of our synthesized result, we compare our synthesized microstructure with the ground truth skin texture from Alexander et al. [2009], as shown in Figure 9. We replace the texture extracted from the

Table 2. A/B comparison results of the user study. Each cell indicates the number of times the method at the left beats the method at the top. We present the total number of wins and the page rank measurement on the right, with the optimal value highlighted with underline.

Winner	Loser			Total ↑	PageRank ↓
	FLEX	**Uniform**	**Noise**		
FLEX	0	216	246	462	0.2304
Uniform	162	0	162	324	0.2472
Noise	90	48	0	138	0.5224

Table 3. The quantitative evaluation result on the synthesized skin patch dataset. We show the average of LPIPS on the microstructure appearance and MSE on the optimized parameter, of five types of results under different optimization/synthesis strategies.

	Uniform	**Noise**	**DDS Only**	**VIS Only**	**Ours**
LPIPS ↓	0.3984	0.6864	0.3073	0.4282	0.3019
MSE ↓	-	-	0.0919	0.2603	0.0933

facial image with the ground truth skin texture to synthesize the final facial microstructure displacement map. To better compare the similarity, we rendered the skin texture in the texture space with either our synthesized microstructure displacement (top) or the ground truth displacement (bottom). Results show that our synthesized microstructure has a realistic spatially-varying appearance compared with ground truth skin texture.

6.2 Quantitative Evaluation

We conduct a quantitative evaluation of the microstructure patches synthesized from the skin patches in our synthesized skin patch dataset. We randomly select 1000 skin patches from the validation set as the optimization targets. Since we need to measure the visual similarity, we choose LPIPS to measure the similarity between the synthesized microstructure displacement and the ground-truth displacement map that was used to synthesize the skin patch. We establish multiple comparative approaches, including **DDS Only** (only DDS in objective function), **VIS Only** (only VIS in objective function), **Uniform**, and **Noise**. Apart from visual similarity, we also compare the parameter similarity via MSE between optimized parameters and ground-truth, as shown in Table 3. Results show that our method has the highest visual similarity on the synthesized microstructure, and the second highest parameter similarity on the optimized parameters, slightly lower than **DDS Only**.

6.3 User Study for Quantitative Evaluation

We conducted a user study to evaluate further whether FLEX enhances the visual similarity of reconstructed facial geometry compared to the methods mentioned at the beginning of Section 6.

Study Design. Our evaluation includes 42 participants. Each participant evaluated at least 22 groups of face patches, with each group showing a pair of patches by using two random methods. These two patches correspond to one facial region of a random individual. Participants were asked to do an A/B comparison to choose either the more visually realistic of the two patches.

Comparison Result. Table 2 shows the A/B comparison results. We also applied the PageRank algorithm [Page et al. 1999], following Pang and Ling [2013] and Weiss et al. [2024], to further analyze the comparison results. The evaluation results indicate that **FLEX** has the highest scores, followed by **Uniform**.

6.4 Performance

In all tests, for a single facial patch, our parameter recovery can be completed in an average of 2.35 seconds. In contrast, the parameter optimization method of Weiss et al. [2023] costs more than 140 seconds. Thanks to our differentiable microstructure simulation, we achieve a significant efficiency improvement compared to Weiss et al. [2023].

6.5 Ablation Study

In our ablation study, we aim to demonstrate the indispensability of both *Direction Distribution Similarity* (DDS) and *Visual Similarity* (VIS) during the optimization. The optimization results, as illustrated in Figure 10, we find that our proposed method, which combines DDS with VIS, achieves superior results. It not only keeps the directional features consistent with the input patches but also reproduces a visually similar micro-wrinkle appearance. In contrast, using DDS only may cause an unstable strength of wrinkle depth, while using VIS only will cause the directional features to be lost, leading to a problematic appearance.

6.6 Discussion and Limitations

Our method inevitably has several limitations. First, we assume facial microstructure comprises only concave features (e.g., micro-wrinkles and pores). However, certain facial regions (e.g., the nasal tip) exhibit convex structures caused by sebaceous filaments or fat deposits, which fall outside our model's representational scope and thus cannot be synthesized. Second, our method optimizes microstructure parameters based on visible facial textures in the input image. However, since the perceived microstructures are inherently affected by lighting conditions and viewing angles, we cannot guarantee microstructure consistency across varying illumination scenarios.

7 ETHIC CONCERN

While our work advances facial microscopic details synthesis from a single unconstrained facial image, we explicitly acknowledge its potential misuse implications. The proposed method's ability to synthesize high-fidelity microscopic facial microstructure could theoretically be exploited to create forgeries, particularly when combined with existing macroscopic facial synthesis techniques.

8 CONCLUSION

In this paper, we present FLEX, a framework for synthesizing facial microstructures, including micro wrinkles and tiny pores, from a single unconstrained facial image. Our key insight is the strong prior provided by the specialized pattern of the relationship between wrinkles and pores, which makes it possible to synthesize realistic facial microstructure from challenging inputs. Specifically, we propose a GNN-based differentiable parametric microstructure model,

using multiple graph convolutions to simulate the complex interactions between adjacent wrinkles. We treat the parameter recovery process as an optimization problem and propose Direction Distribution Similarity to resolve the inconsistency between the blurry input and the simulated microstructure, ensuring consistent directional features. Thanks to the differentiability, our optimization is approximately 60 times faster than previous methods. Experimental results demonstrate that the microstructures synthesized by FLEX seamlessly integrate with existing macroscopic geometry recovery methods, significantly enhancing the visual realism of the reconstructed faces. Promising future work includes considering dynamic expressions and a more comprehensive microstructure model.

ACKNOWLEDGMENTS

We thank the reviewers for the valuable comments and XiaoLong Wu for his help with our video. This work has been partially supported by the National Key R&D Program of China under grant No. 2022YFB3303203 and Natural Science Foundation of China under grant No. 62272275 and 62172220.

REFERENCES

Oleg Alexander, Mike Rogers, William Lambeth, Matt Jen-Yuan Chiang, and Paul E. Debevec. 2009. The Digital Emily project: photoreal facial modeling and animation. In *International Conference on Computer Graphics and Interactive Techniques*. https://api.semanticscholar.org/CorpusID:195721699

Y. Bando, T. Kuratate, and T. Nishita. 2002. A Simple Method for Modeling Wrinkles on Human Skin. In *10th Pacific Conference on Computer Graphics and Applications, 2002. Proceedings.* IEEE Comput. Soc, Beijing, China, 166–175. https://doi.org/10.1109/PCCGA.2002.1167852

Thabo Beeler, B. Bickel, Paul A. Beardsley, Bob Sumner, and Markus H. Gross. 2010. High-quality single-shot capture of facial geometry. *ACM SIGGRAPH 2010 papers* (2010). https://api.semanticscholar.org/CorpusID:17031083

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-Quality Passive Facial Performance Capture Using Anchor Frames. In *ACM SIGGRAPH 2011 Papers*. ACM, Vancouver British Columbia Canada, 1–10. https://doi.org/10.1145/1964921.1964970

Volker Blanz and Thomas Vetter. 1999. A Morphable Model For The Synthesis Of 3D Faces. *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (1999). https://api.semanticscholar.org/CorpusID:203705211

Zenghao Chai, Tianke Zhang, Tianyu He, Xu Tan, Tadas Baltrušaitis, HsiangTao Wu, Runnan Li, Sheng Zhao, Chun Yuan, and Jiang Bian. 2023. HiFace: High-Fidelity 3D Face Reconstruction by Learning Static and Dynamic Details. https://doi.org/10.48550/arXiv.2303.11225 arXiv:2303.11225 [cs]

Anpei Chen, Zhang Chen, Guli Zhang, Ziheng Zhang, Kenny Mitchell, and Jingyi Yu. 2019. Photo-Realistic Facial Details Synthesis from Single Image. https://doi.org/10.48550/arXiv.1903.10873 arXiv:1903.10873 [cs]

Abdallah Dib, Junghyun Ahn, Cédric Thébault, Philippe-Henri Gosselin, and Louis Chevallier. 2023. S2F2: Self-Supervised High Fidelity Face Reconstruction from Monocular Image. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, Waikoloa Beach, HI, USA, 1–8. https://doi.org/10.1109/FG57933.2023.10042713

Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. 2021. Learning an Animatable Detailed 3D Face Model from In-the-Wild Images. *ACM Transactions on Graphics* 40, 4 (Aug. 2021), 1–13. https://doi.org/10.1145/3450626.3459936

G Fyffe, P Graham, B Tunwattanapong, A Ghosh, and P Debevec. 2016. Near-Instant Capture of High-Resolution Facial Geometry and Reflectance. (2016).

Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. 2017. Morphable Face Models - An Open Framework. arXiv:1709.08398 [cs.CV] https://arxiv.org/abs/1709.08398

Paulo Gotardo, Jérémy Riviere, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2018. Practical Dynamic Facial Appearance Modeling and Acquisition. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 1–13. https://doi.org/10.1145/3272127.3275073

Paul Graham, Borom Tunwattanapong, Jay Busch, Xueming Yu, Andrew Jones, Paul E. Debevec, and Abhijeet Ghosh. 2012. Measurement-Based Synthesis of Facial Microgeometry. *Computer Graphics Forum* 32 (2012). https://api.semanticscholar.org/CorpusID:11732897

Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and D. Salesin. 2001. Image Analogies. *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (2001). https://api.semanticscholar.org/CorpusID:2201072

Xun Huang and Serge J. Belongie. 2017. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 1510–1519. https://api.semanticscholar.org/CorpusID:6576859

Loc Huynh, Weikai Chen, Shunsuke Saito, Jun Xing, Koki Nagano, Andrew Jones, Paul Debevec, and Hao Li. 2018. Mesoscopic Facial Geometry Inference Using Deep Neural Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Salt Lake City, UT, USA, 8407–8416. https://doi.org/10.1109/CVPR.2018.00877

Diqiong Jiang, Yiwei Jin, Fanglue Zhang, Zhe Zhu, Yun Zhang, Ruofeng Tong, and Ming Tang. 2021. Sphere Face Model: A 3D Morphable Model with Hypersphere Manifold Latent Space. *ArXiv* abs/2112.02238 (2021). https://api.semanticscholar.org/CorpusID:266687405

C Kampouris, S Zafeiriou, and A Ghosh. 2018. Diffuse-Specular Separation Using Binary Spherical Gradient Illumination. (2018).

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2019. Analyzing and Improving the Image Quality of StyleGAN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 8107–8116. https://api.semanticscholar.org/CorpusID:209202273

James Kennedy and Russell Eberhart. 2002. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks* 4 (2002), 1942–1948 vol.4. https://api.semanticscholar.org/CorpusID:3114196

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). https://api.semanticscholar.org/CorpusID:6628106

Alexandros Lattas, Yiming Lin, Jayanthi Kannan, Ekin Ozturk, Luca Filipi, Giuseppe Claudio Guarnera, Gaurav Chawla, and Abhijeet Ghosh. 2022. Practical and Scalable Desktop-Based High-Quality Facial Capture. In *European Conference on Computer Vision*. https://api.semanticscholar.org/CorpusID:253518903

Alexandros Lattas, Stylianos Moschoglou, Stylianos Ploumpis, Baris Gecer, Jiankang Deng, and Stefanos Zafeiriou. 2023. FitMe: Deep Photorealistic 3D Morphable Model Avatars. https://doi.org/10.48550/arXiv.2305.09641 arXiv:2305.09641 [cs]

Biwen Lei, Jianqiang Ren, Mengyang Feng, Miaomiao Cui, and Xuansong Xie. 2023. A Hierarchical Representation Network for Accurate and Detailed Face Reconstruction from In-The-Wild Images. *ArXiv* abs/2302.14434 (2023).

Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. 2017. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 36, 6 (2017), 194:1–194:17. https://doi.org/10.1145/3130800.3130813

Tianye Li, Shichen Liu, Timo Bolkart, Jiayi Liu, Hao Li, and Yajie Zhao. 2021. Topologically Consistent Multi-View Face Inference Using Volumetric Sampling. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Montreal, QC, Canada, 3804–3814. https://doi.org/10.1109/ICCV48922.2021.00380

Shichen Liu, Yunxuan Cai, Haiwei Chen, Yichao Zhou, and Yajie Zhao. 2022. Rapid Face Asset Acquisition with Recurrent Feature Alignment. *ACM Transactions on Graphics* 41, 6 (Dec. 2022), 1–17. https://doi.org/10.1145/3550454.3555509

Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Félix Chabert, Malte Weiss, and Paul E. Debevec. 2007. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. In *Rendering Techniques*.

Koki Nagano, Graham Fyffe, Oleg Alexander, Jernej Barbi, Hao Li, Abhijeet Ghosh, and Paul E. Debevec. 2015. Skin microstructure deformation with displacement map convolution. *ACM Transactions on Graphics (TOG)* 34 (2015), 1 – 10. https://api.semanticscholar.org/CorpusID:12082265

Diego F. Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM SIGGRAPH 2005 Papers* (2005). https://api.semanticscholar.org/CorpusID:2402326

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking : Bringing Order to the Web. In *The Web Conference*. https://api.semanticscholar.org/CorpusID:1508503

Yu Pang and Haibin Ling. 2013. Finding the Best from the Second Bests - Inhibiting Subjective Bias in Evaluation of Visual Tracking Algorithms. *2013 IEEE International Conference on Computer Vision* (2013), 2784–2791. https://api.semanticscholar.org/CorpusID:1155907

Pexels. 2025. Pexels: Free Stock Photos. https://www.pexels.com/ Accessed: 2025-01-17.

Jérémy Riviere, Paulo Gotardo, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2020. Single-Shot High-Quality Facial Geometry and Skin Appearance Capture. *ACM Transactions on Graphics* 39, 4 (Aug. 2020). https://doi.org/10.1145/3386569.3392464

Shunsuke Saito, Lingyu Wei, Liwen Hu, Koki Nagano, and Hao Li. 2016. Photorealistic Facial Texture Inference Using Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 2326–2335. https://api.semanticscholar.org/CorpusID:5464807

Ayush Tewari, Michael Zollhofer, Pablo Garrido, Florian Bernard, Hyeongwoo Kim, Patrick Perez, and Christian Theobalt. 2018. Self-Supervised Multi-level Face Model Learning for Monocular Reconstruction at Over 250 Hz. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Salt Lake City, UT, USA, 2549–2559. https://doi.org/10.1109/CVPR.2018.00270

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio', and Yoshua Bengio. 2017. Graph Attention Networks. *ArXiv* abs/1710.10903 (2017). https://api.semanticscholar.org/CorpusID:3292002

Javier von der Pahlen, Jorge Jimenez, Etienne Danvoye, Paul E. Debevec, Graham Fyffe, and Oleg Alexander. 2014. Digital ira and beyond: creating real-time photoreal digital actors. In *International Conference on Computer Graphics and Interactive Techniques*. https://api.semanticscholar.org/CorpusID:20679008

S. Weiss, J. Moulin, P. Chandran, G. Zoss, P. Gotardo, and D. Bradley. 2023. Graph-Based Synthesis for Skin Micro Wrinkles. *Computer Graphics Forum* 42, 5 (Aug. 2023), e14904. https://doi.org/10.1111/cgf.14904

S. Weiss, J. Stanhope, P. Chandran, G. Zoss, and D. Bradley. 2024. Stylize My Wrinkles: Bridging the Gap from Simulation to Reality. *Computer Graphics Forum* 43, 2 (May 2024), e15048. https://doi.org/10.1111/cgf.15048

Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, B. Bickel, Craig Donner, Chien-I Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus H. Gross. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM SIGGRAPH 2006 Papers* (2006). https://api.semanticscholar.org/CorpusID:455764

Yin Wu, Prem Kumar Kalra, and Nadia Magnenat-Thalmann. 1996. Simulation of static and dynamic wrinkles of skin. *Proceedings Computer Animation '96* (1996), 90–97. https://api.semanticscholar.org/CorpusID:14125528

Yunze Xiao, Hao Zhu, Haotian Yang, Zhengyu Diao, Xiangju Lu, and Xun Cao. 2022. Detailed Facial Geometry Recovery from Multi-View Images by Learning an Implicit Function. https://doi.org/10.48550/arXiv.2201.01016 arXiv:2201.01016 [cs]

Shugo Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olszewski, Shigeo Morishima, and Hao Li. 2018. High-Fidelity Facial Reflectance and Geometry Inference from an Unconstrained Image. *ACM Transactions on Graphics* 37, 4 (Aug. 2018), 1–14. https://doi.org/10.1145/3197517.3201364

Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. 2020. FaceScape: A Large-Scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 598–607. https://api.semanticscholar.org/CorpusID:214728393

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

Hao Zhu, Haotian Yang, Longwei Guo, Yidi Zhang, Yanru Wang, Mingkai Huang, Menghua Wu, Qiu Shen, Ruigang Yang, and Xun Cao. 2023. FaceScape: 3D Facial Dataset and Benchmark for Single-View 3D Face Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2023).

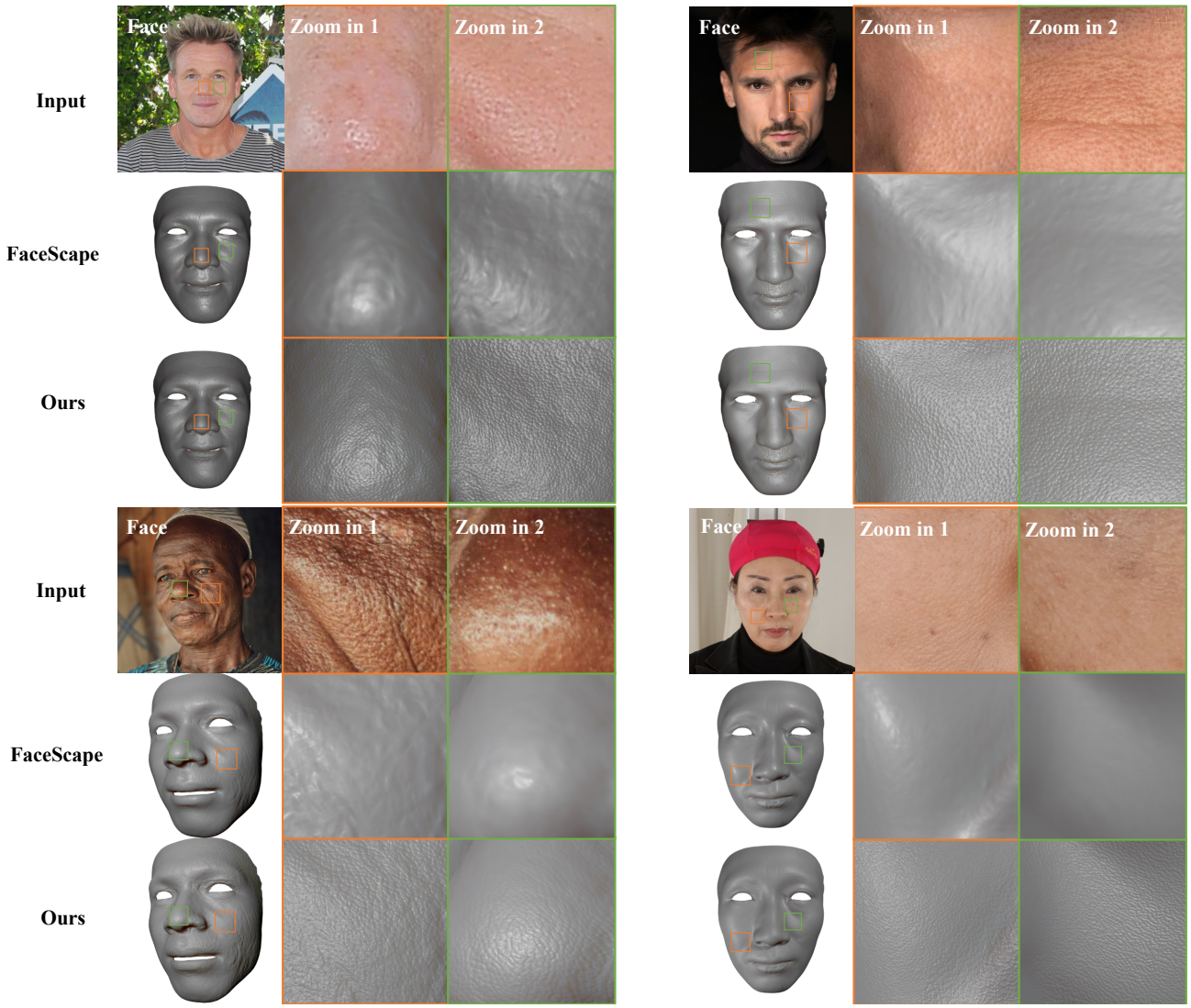

Fig. 6. The illustration of integrating FLEX with FaceScape [Zhu et al. 2023]. For each face, we show the whole face as well as two zoomed-in images from both the input facial image and our rendering results.

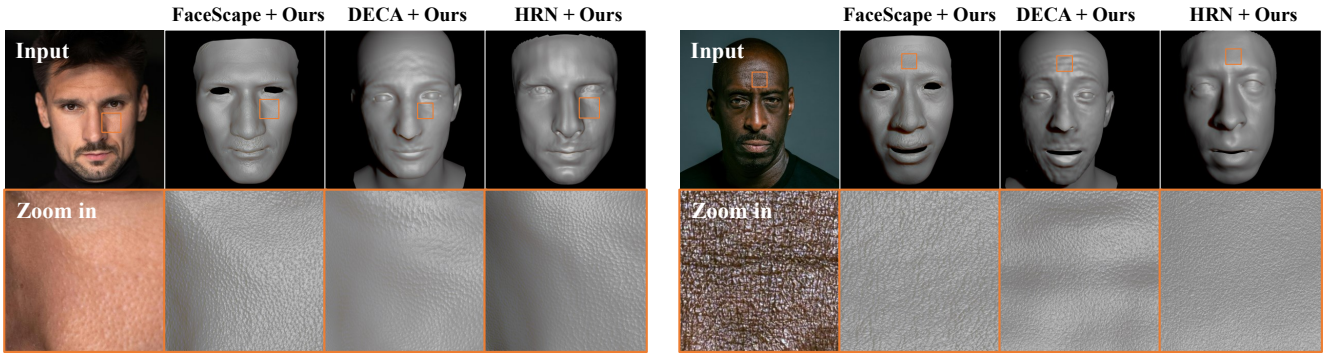

Fig. 7. The results of integrating with different macroscopic methods. We compare the same zoomed-in region across different integration setups. The tested facial images are all from Pexels [Pexels 2025].

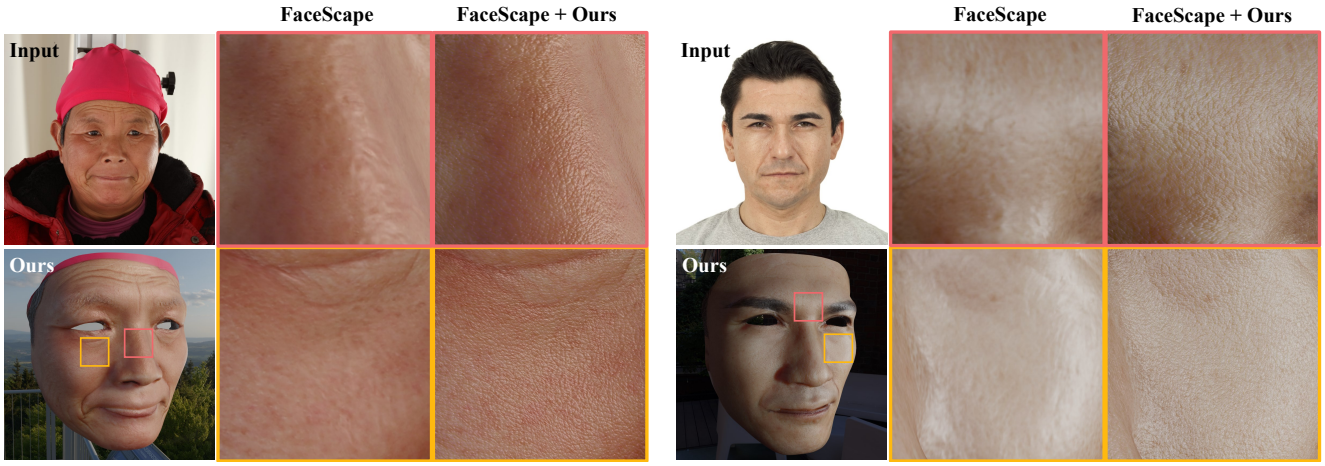

Fig. 8. Comparison of rendering results. We obtain the macroscopic details and microscopic details from FaceScape [Zhu et al. 2023] and our method, respectively, and then render them under environmental light. We select two zoomed-in rendered results for each individual. The tested facial images are from the open source repository of Zhu et al. [2023].

Fig. 9. The comparison of the synthesizing result from ground truth texture from Alexander et al. [2009]. We show the zoomed-in crops from several face regions.

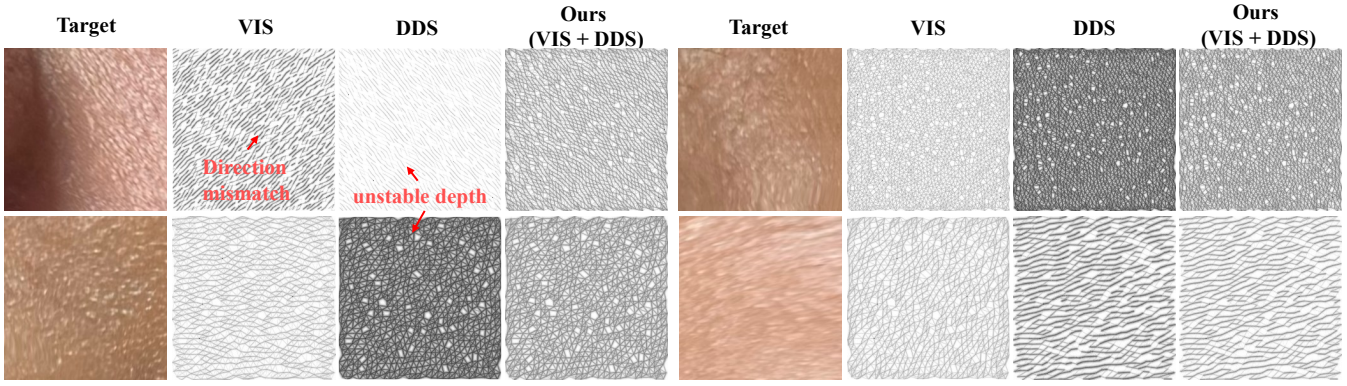

Fig. 10. The optimization result of different objective function setups. The highlighted results reveal the limitation of direction mismatch when using VIS only, and unstable depth when using DDS only.