EUROGRAPHICS 2024 / A. Bermano and E. Kalogerakis
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 2

Stylize My Wrinkles: Bridging the Gap from Simulation to Reality

S. Weiss*[1] and J. Stanhope*[2] and P. Chandran[1] and G. Zoss[1] and D. Bradley[1]

[1] DisneyResearch|Studios, Switzerland [2] ETH Zürich
Authors marked with * contributed equally

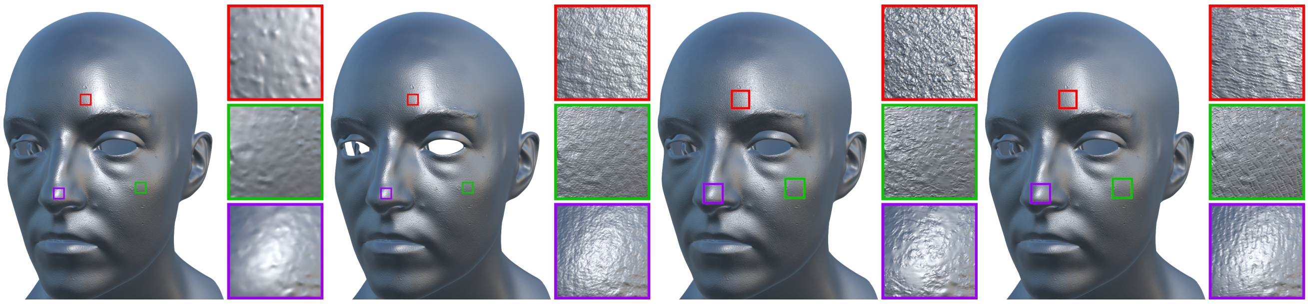

Figure 1: *Current full-face facial scanning techniques (a) can capture geometry sufficient for most shots, but start to lack details needed for close-up shots. Prior work either used wrinkle simulation (b) to fill in the missing details, or texture synthesis from sparse small patch scans (c). We obtain the best of both worlds by first simulating the wrinkles with spatial-varying parameters estimated from real skin and then stylizing the output given high-resolution patch scans (d).*

Abstract

Modeling realistic human skin with pores and wrinkles down to the milli- and micrometer resolution is a challenging task. Prior work showed that such micro geometry can be efficiently generated through simulation methods, or in specialized cases via 3D scanning of real skin. Simulation methods allow to highly customize the wrinkles on the face, but can lead to a synthetic look. Scanning methods can lead to a more organic look for the micro details, however these methods are only applicable to small skin patches due to the required image resolution. In this work we aim to overcome the gap between synthetic simulation and real skin scanning, by proposing a method that can be applied to large skin regions (e.g. an entire face) with the controllability of simulation and the organic look of real micro details. Our method is based on style transfer at its core, where we use scanned displacement maps of real skin patches as style images and displacement maps from an artist-friendly simulation method as content images. We build a library of displacement maps as style images by employing a simplified scanning setup that can capture high-resolution patches of real skin. To create the content component for the style transfer and to facilitate parameter-tuning for the simulation, we design a library of preset parameter values depicting different skin types, and present a new method to fit the simulation parameters to scanned skin patches. This allows fully-automatic parameter generation, interpolation and stylization across entire faces. We evaluate our method by generating realistic skin micro details for various subjects of different ages and genders, and demonstrate that our approach achieves a more organic and natural look than simulation alone.

1. Introduction

A crucial step in creating and rendering realistic-looking digital avatars is the accurate representation of the facial micro details, such as pores and fine scale wrinkles. This micro geometry helps to break up the specular reflections when rendering, and is one of the key components in achieving photorealism of digital characters. Although simple approximations like adding noise can be used for real-time applications [vdPJD*14], more sophisticated techniques are typically used to achieve high quality results when rendering skin offline [GTB*13, WMC*23]. The need for realistic micro details is exacerbated by the ever-expanding use of digital characters in high-end visual effects productions, where extreme close-up shots are common.

The most recent solution to achieve skin micro wrinkles has been through a simulation approach. Weiss *et al.* [WMC*23] proposed a graph-based simulation method to synthetically generate micro geometry for static faces, baked into a displacement map. While the simulation allows for fine details with full artist controllability, this solution has one major drawback; since the micro geometry is created through a computer simulation, it can sometimes appear too synthetic and fail to represent the organic chaotic variations in pore distribution and wrinkle shape that are present in real skin (refer to Fig. 1, b).

On the other hand, Graham *et al.* [GTB*13] proposed a scanning setup that can capture patches of real skin in an area of 2x2cm with very high quality. Then, texture synthesis methods like image analogies [HJO*01] or patch nearest neighbors [GFS*22] (refer to Fig. 1, c) are used to extend the scanned patches over the entire face. While this approach can create organic looking skin, it has several limitations. First, it requires a very intricate and custom-built hardware setup. Second, image analogies or patch nearest neighbors as the method to fill in the entire face are dependent on the quality of the UV map and do not consider stretching and compression in the parametrization. Third, artists cannot influence the result of the reconstruction, e.g. controlling which features are shown.

We present a pipeline that combines the best features of the two approaches. The main idea is to have an artist-controllable wrinkle simulation, where the output is further stylized by skin features from real scanned patches. At its core, the problem becomes displacement map stylization, where the style image comes from real data and the content image comes from the simulation. To obtain style images, we built a simplified patch scanning setup inspired by Chen *et al.* [CGS06], which does not require a custom-built light setup. This patch scanner allows us to quickly capture a library of displacement maps of small high-resolution patches of real skin, which can be selected as the style images for our method. To obtain content images, we use the recent wrinkle simulator of Weiss *et al.* [WMC*23], which can generate a multitude of different pore and micro-wrinkle formations. To facilitate usability, we introduce a library of preset simulation parameters that target different wrinkle structures, which can be spatially interpolated over the facial surface. We further present a new method to optimize simulation parameters to match our scanned patches, using a combination of an image classification network and image generation network. This fitting method provides an automatic way of expanding the preset parameter library. Ultimately, an artist interactively selects simulation presets for different parts of the face, interpolates them spatially, runs the forward simulation, and then stylizes the simulation result to re-introduce the organic nature of real skin, again in a controllable way by selecting from the library of scanned patch displacement maps. The final output is a full face displacement map of micro geometry with the organic look of real skin obtained in an artist-controllable manner (see to Fig. 1, d).

To summarize, we propose

- a simplified patch capturing setup requiring only three cameras and a hand-held flashlight,
- a simulation preset library representing a variety of micro-wrinkle types to help artists in designing full-face simulated wrinkles,
- a dual discriminator-generator reconstruction pipeline that can fit simulation parameters to match the captured displacement map of real skin, and
- an application of style transfer to introduce the organic nature of real skin into the simulated micro wrinkles.

2. Related Work

We now describe related work in the areas of scanning micro wrinkles from real skin and synthesizing micro geometry with simulation. As our method is based on high-resolution image style transfer, we also discuss relevant works in that field.

Scanning Micro Wrinkles. Capturing the full face in 3D is not a new problem, with several methods that utilize different 3D reconstruction techniques and can achieve impressive results [NHRD90, BBA*07, BBB*10, GFT*11, RGB*20b]. Many of these methods rely on expensive and complicated capture hardware including multiple cameras and lights. More lightweight, in-the-wild methods have also been proposed [WBGB16, IBP15, LLK*22], some of which utilize a pre-trained neural prior to fit a 3D face mesh to one or several images [TZK*17, GPKo19, DYX*19, GZY*20, FFBB21, DTA*21, WBH*22, CZGB23]. In any case, full face reconstruction methods are currently not capable of recovering the fine-scale micro details that we address in this work.

Scanning micro wrinkles often requires dedicated setups that focus on capturing high resolution scans of small skin patches. Graham *et al.* [GTB*13] present a hardware setup that can acquire skin patch geometry at an impressive resolution of $7\mu m$ per pixel. Their scanning method was derived from that of Ma *et al.* [MHP*07] and requires a complex lighting rig with multiple polarized, controllable lights. Later works such as Nagano *et al.* [NFA*15] used a similar setup to measure the skin while being stretched and compressed, and used this captured data to simulate deformations in reconstructed skin. Further works sought to improve the resolution of skin patch scans using microscopes [BMH*16].

These patch scanning methods are able to show high levels of fidelity and quality, but their reliance on complex hardware set-ups makes them difficult to operate and replicate. With this in mind, we are motivated to devise a patch scanning method that can be economically built and easily operated with minimal training.

Synthesizing Micro Geometry. As an alternative to micro-detail scanning techniques, simulation methods can be used to generate synthetic skin wrinkles and pores. Early works by Ishii *et al.* [IYYT93] and Wu *et al.* [WKT96] use Voronoi cells or Delaunay triangulation to create polygonal patterns on the skin. The edges of these polygons are then carved into the displacement map to form the micro wrinkles. These models assume isotropic stress: One hypothesis of why micro wrinkles are formed is as a skin reservoir for stretching [LC93]. In regions of isotropic muscle movements, wrinkles also form equally oriented in every direction. In regions of anisotropic muscle movements, e.g. the forehead, micro wrinkles tend to form in parallel lines. This anisotropic nature is explored by Bando *et al.* [BKN02] in a greedy simulation method where lines are drawn sequentially following a flow field of wrinkle orientations. Later work extended upon this method with improved

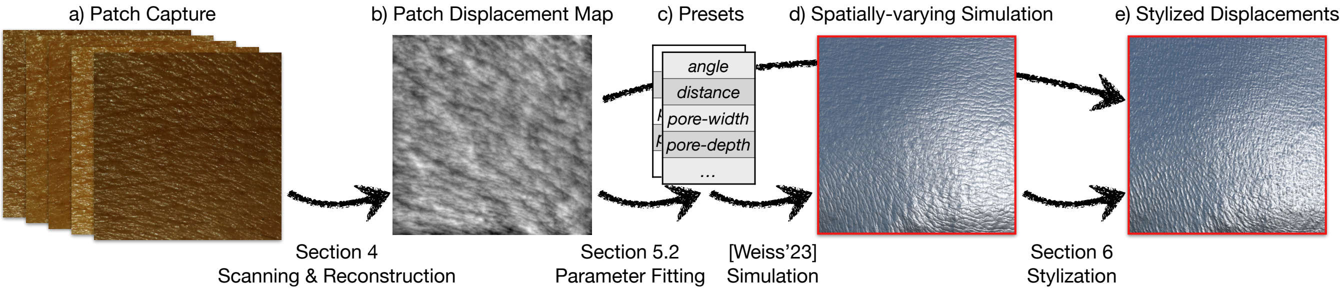

Figure 2: *We present a scanning and stylization pipeline to enhance the realism of synthetic facial wrinkles. We propose a new patch scanning setup, Section 4, to acquire highly detailed scans (a) and reconstruct the displacement map (b) per patch. A novel parameter fitting process, Section 5.2, then fits simulation parameters (c) for that preset, so that a customized wrinkle simulation [WMC*23] can create a spatially-varying synthetic wrinkle displacement map (d). Finally, a modified stylization method, Section 6, then stylizes this map to enhance the realism (e) using the reconstructed maps of real skin as the style.*

control parameters and drawing methods [VB18, LXZ07, LLLC11]. Recently, Weiss *et al.* [WMC*23] presented a graph-based method for skin micro details where the wrinkles are treated as the edges that connect the nodes of the graph representing the pores on the skin. This simulation allows for fine control over the appearance of the skin and consistent wrinkle generation over the entire face.

One major drawback of simulation methods is that they tend to look too synthetic, often lacking the natural organic structure of real skin. Our work aims to bridge the gap between simulation and reality. We build upon the simulation method of Weiss *et al.* [WMC*23] and extend it with spatially-varying parameter presets, better parameter fitting to scanned data, and a final stylization method that leverages the organic structure of real skin patches.

High-Resolution Style Transfer. Our approach to add a data-driven layer to simulated micro wrinkles is based on image style transfer. Early works in style transfer like image analogies [HJO*01] require matched pairs of input and stylized images or are limited to high-frequency details [Ash03, LSRY10, LSY11]. With the seminal work by Gatys *et al.* [GEB16] on neural style transfer and the introduction of a neural style loss, texture transfer over large ranges of frequencies became possible. Since then, many works have investigated the quality and controllability of style transfer [GEBo17, LWLH17] or improving the speed of style transfer by training neural networks to directly perform stylization in an inference pass [JAFF16, HB17, LZY*17, SYZ18, SKLO18, LLKY19, LCLB21, CZG*21]. We refer the reader to the recent survey of Jing *et al.* [JYF*19] for more a more in depth summary of the neural style transfer methods.

Many stylization algorithms mentioned above struggle with scalability to high-resolution images. For example, in our situation the displacement maps for skin micro wrinkles on a face may have a resolution of 16k or more. Performing stylization on large images at once is infeasible due to memory limitations, even with a reduced network size [WLW*20], and running it patch-by-patch leads to border artifacts. Chen *et al.* [CWX*21] show how to circumvent this issue by carefully controlling the normalization between border patches. Abdellatif and Elsheikh [AE23] store the inner feature maps along the border of the image to avoid artifacts due to padding. Alternatively, Lötzsch *et al.* [LRB*22] use differentiable image filters, a collection of manually designed, parameterized filters, to stylize the input image. As this method does not rely on neural network inference, it is also applicable to high-resolution input images. In our work, we employ the stylization method of Li *et al.* [LCLB21] which is fast and offers stylization even from only one (or a few) style images, which allows us to style the entire face (patch by patch) given only a small number of scanned real patches as style images.

3. Overcoming the Synthetic Nature of Simulated Wrinkles

To tackle the problem of overcoming the synthetic nature of simulated micro wrinkles, we propose to stylize the output of the simulation using real scan data, see Fig. 2 for an overview. To this end, we create a library of style images (high-resolution displacement maps) by scanning real patches on a number of actor's faces using a new simplified patch scanner (Section 4). For the "content" part of the image stylization we rely on a recent micro-wrinkle simulator that allows artistic control, and we propose to generate a library of preset simulation parameters depicting a variety of skin micro details. We further propose an approach to fit the simulation parameters to real scanned patches, allowing to automatically extend the preset library. With these simulation presets, the wrinkle simulation can create consistent micro wrinkles for the whole face by smoothly interpolating between the different styles of skin at different locations (Section 5). Finally, given the simulated content result and the scanned patch library for stylization, we propose to use a fast style transfer algorithm to generate the final micro-detail displacement map in a patch-by-patch manner (Section 6). The result will be a micro-detail displacement map of the entire face, artistically generated with real organic features. In this work, we create displacement maps at a resolution of 16k x 16k pixels.

4. Style Library via Real Patch Scanning

We present a simple hardware setup and reconstruction method for scanning real skin patches and use it to collect a diverse set of displacement maps of various facial regions from different subjects at a micro-meter resolution. The goal is to capture a vast amount of natural variation, thus providing us a rich collection of patches for style transfer.

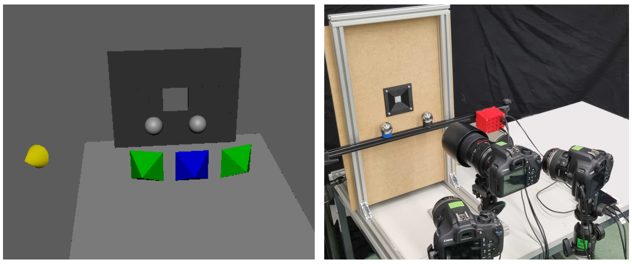

Figure 3: *Schematic overview (left) and realization (right) of the scanning setup featuring one main camera (blue) and two side cameras (green) for light source estimation from two mirror balls. A handheld flashlight (yellow) illuminates the skin which is pressed against a small cutout window from the back.*

While our patch scanner takes inspiration from several previous approaches, it has a few unique features. We aim to achieve a similar (or greater) level of detail as Graham *et al.* [GTB*13] but without the complex light dome hardware. Thus we adopt a method inspired by Chen *et al.* [CGS06] but adapted for skin patches.

4.1. Hardware Setup

Our simple hardware setup consists of three DSLR cameras, two mirror balls, a flashlight and a board with a small cutout window (see Fig. 3). In the center of the scanning setup is a board with a window (2cm x 2cm) for subjects to press their face against from the back. The window was modeled and 3D printed, such that the shape of the window is easy for subjects to press all parts of their face (cheeks, forehead, temple, etc) into the scanner. During capture, subjects are given a face rest to use to help them remain still. A capture session lasts about 30 seconds for a single patch. For the light source, a handheld flashlight is used. At each frame during capture, the light source is manually moved, illuminating the skin patch from different angles. No fixed set of light positions are required, and in practice a set of approximately 30 images are recorded in fast succession (around 1 image per second), covering 30 different light angles. Two mirror spheres are placed on either side of the face window, which are used to estimate the position of the light sources. The scanner uses three synchronized cameras, one main camera with 180mm macro lens framed and focused on the skin patch only, and two additional side cameras each with 60mm lenses framed to see both mirror balls for light position estimation. All three cameras are calibrated in a single world space and the positions and sizes of the mirror balls are also reconstructed in 3D by annotating their projections in the two side cameras and using their known diameter to triangulate their 3D positions. The main camera has a resolution of 18MP, and so with this setup the scanning resolution is approximately 5μm per pixel.

4.2. Light Source Estimation

To estimate the 3D position of the light source in each capture frame we threshold the side camera images and apply blob detection to find the two pixels per camera that most likely contain the reflection of the flashlight in the mirror balls. Projecting these four highlight pixels into 3D space gives us four intersection points on the mirror spheres, which are reflected to obtain four rays that should converge at the 3D light position. Due to inaccuracies in blob detection the rays do not perfectly converge, and thus we search for the point that minimizes the distance to the four light rays. A detailed explanation of this procedure can be found in the supplemental document. The optimized light source position is found for each captured frame, which gives us a light direction for each captured image of a skin patch, which will be used in the displacement map reconstruction method described next.

4.3. Displacement Map Reconstruction

Since small movements due to breathing or even blood flow cannot be avoided, we first align all captured images from the main camera using an image registration technique that works under varied illumination [PQE21]. Human skin is covered with a thin oily film, called the skin surface lipid film [INN07]. This film leads to a very specular reflection of the skin, especially at the micro scale and encodes most of the wrinkle structure. Therefore, to estimate the normal of the skin, we again follow the approach by Chen *et al.* [CGS06] using shape-from-specularity, instead of a more traditional reconstruction method based on a Lambertian assumption. Specular reflections are identified as pixels that exceed a certain threshold. Then the normal vector n_j at the current pixel j is given as the half vector between the direction to the camera and the direction to the light source. This gives rise to a sparse collection of normal samples over all captured images $N = \{n_1, n_2, \ldots, n_K\}$.

Since we capture only a small set of around 30 images to keep the stress for the subject as low as possible, the normal samples are very sparse. To obtain a dense displacement map D, we construct the following linear system,

$$\frac{\partial D}{\partial x} = n_x, \text{ normal constraint} \tag{1}$$

$$\frac{\partial D}{\partial y} = n_y, \text{ normal constraint} \tag{2}$$

$$\lambda \Delta D = 0, \text{ smoothness constraint} \tag{3}$$

$$D(\mathbf{c}) = 0, \text{ fix height of the center pixel,} \tag{4}$$

where $\mathbf{c}$ is the center pixel of the image, n_x and n_y are the observed normals in the x- and y-directions, and λ is a regularization weight. For normal constraints, $\frac{\partial D}{\partial x}, \frac{\partial D}{\partial y}$ is computed using a first order central differences filter. ΔD is computed with a Laplacian filter with a weighting of $\lambda = 1\mathrm{e}^{-3}$. These filters are constructed as sparse matrices that operate on the flattened image, which allows this system to be solved as a linear system with an out-of-the-box solver. An example of the sparse normal and displacement map result can be found for one skin patch in Fig. 4. Additionally, we apply a high-pass filter to the displacement map to filter out low-frequency variations introduced by the bulging of the skin when it is pressed against the observation window, as well as to filter out all the frequencies that may already be captured in the baseline full-face 3D reconstruction. The result is a displacement map containing only the micro details missing from the macro-scale full-face reconstruction.

The displacement maps, together with metadata containing general skin properties (eg. age) and the location of the patches on the

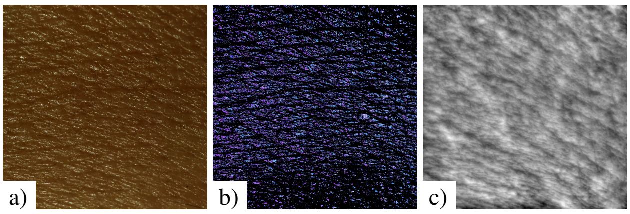

Figure 4: *From a set of input images (a) (only one frame is shown), we extract a sparse set of normal vectors corresponding to the specular reflections (b). These normal map samples are then used to reconstruct a dense displacement map (c).*

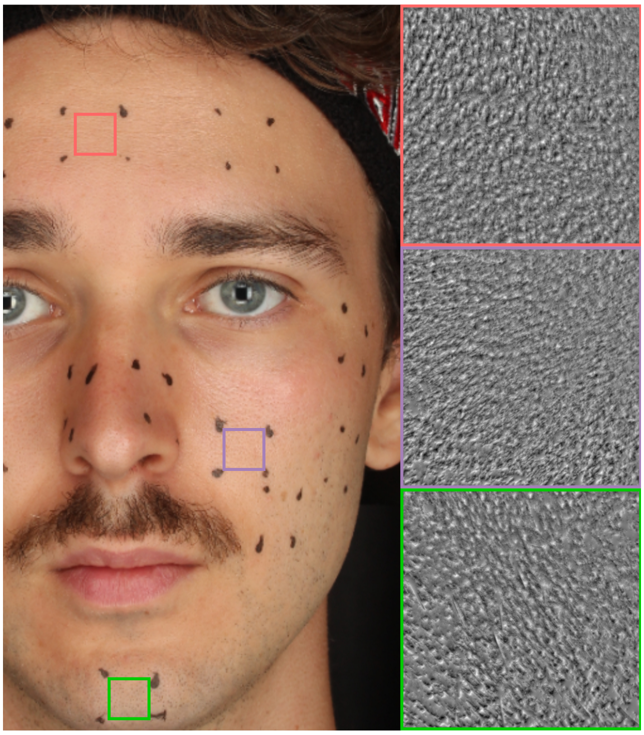

Figure 5: *An actor with markers drawn on the face to align scanned patches to a full face scan. 3 displacement map patches are shown for visualization, from the forehead, cheek, and chin regions.*

face (eg. forehead, cheek, etc.) are collected in a *Displacement Style Library*. An example of the captured locations for one subject together with three recovered displacement maps are shown in Figure 5. In order to locate the scanned patches on the face after scanning we place 4 dots at the patch corners. The displacement maps recovered have a resolution of 5μm per pixel, exceeding the resolution of Graham *et al.* [GTB*13] who report a spatial resolution of 7μm per pixel.

5. Wrinkle Simulation with a Preset Library

The scanned patches from the previous section form the style images for our displacement map style transfer application. We now discuss the creation of the content images. The main idea is to use the micro-wrinkle simulator of Weiss *et al.* [WMC*23], with a few key extensions. First, we prepare a number of preset parameter values to represent a variety of skin types. We also propose a method to fit parameter values to the real scanned patches from Section 4, extending the preset library. Then, given the preset parameter library, an artist can place different skin parameters at different places on the face and interpolate the parameter values spatially. Running the final simulation generates the base content for stylization over the entire face. In this section we elaborate on the preset library, fitting the simulation to real data and interpolating for simulation.

5.1. Artistically Designed Parameter Presets

Using the simulator of Weiss *et al.* [WMC*23], we empirically created different skin structure types by varying the simulation parameters. These preset values form the basis for a *Simulation Preset Library*. Pre-defining these preset values helps to speed up the interactive spatially-varying wrinkle design process for the whole face (described in Section 5.3). A collection of eight such user-defined presets representing various skin types are shown in Fig. 6.

5.2. Fitting the Simulation to Real Skin

We can additionally create simulation parameter presets by fitting the simulation to real data, i.e. the scanned patches from Section 4. This provides two benefits. First, the fitting process we propose is automatic, and thus one can avoid tedious manual parameter tweaking to achieve a certain wrinkle style in the preset library. Second, the scanned patches provide a lot of micro-detail variability, and converting them to simulation parameters is a great way to boost the variability of the preset library as well.

Fitting the parameters of the wrinkle simulator to real data is not a new idea. In fact, Weiss *et al.* [WMC*23] proposed a method themselves to fit to their own simulator. Their approach uses Particle Swarm Optimization (PSO) [KE95] for this task, however, we found that this optimization sometimes gets stuck in suboptimal local minima (as illustrated in Fig. 7c). We therefore present a more robust fitting pipeline using neural networks.

Our approach is to synthetically generate 50,000 simulated skin patches (at a resolution of 512×512) by randomly sampling the parameter space of the simulator, and then learn a mapping from the simulated patches back to the parameters. One method to accomplish this task is through training a classification network. To this end, we adapt the VGG-11 architecture by Simonyan *et al.* [SZ14], but change the input resolution from 224 to 512 and add a global pooling layer between the last convolutional layer and the first fully connected layer to make the network more agnostic to global translations [LMW*22, SK22]. After training, we can feed real scanned patches to the classifier and directly predict the simulation parameters. Unfortunately, when exposed to real-world displacement maps that were never seen during training (eg. Fig. 7a), the classification network fails to capture the parameters like wrinkle orientation, size, depth or pore rendering parameters, which create the "feel" of the texture. It does, however, accurately predict the density of the pores (Fig. 7d).

An alternative to classification is to train a class-conditioned generative network, where the classes are the simulation parameters,

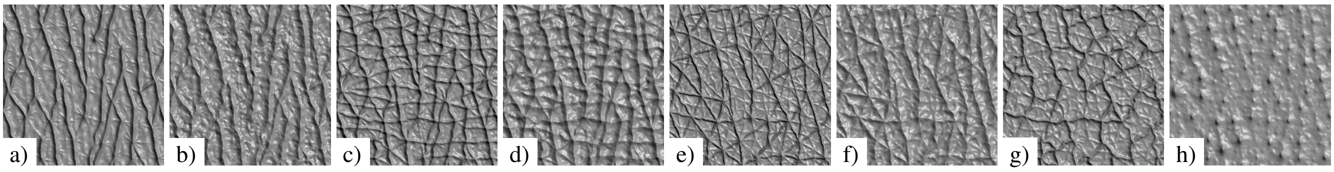

Figure 6: *User-defined simulation presets to represent a variety of different skin types: (a,b) strong directional wrinkles, (c,d) horizontal and vertical wrinkles, (e,f) crisscrossing wrinkles, (g) irregular cells, (h) mainly pores. The patches are shown as displaced geometry shaded with a light source coming slightly from the left.*

and then attempt to optimize for the class that produces a generated displacement map that best matches the scanned target. We evaluate this approach by training a generative network based on the StyleGAN2 architecture [KLA*20] with the last layer changed from RGB to grayscale and the simulation parameters as continuous class labels. One way to think of the generative network is like a differentiable proxy of the (non-differentiable) wrinkle simulator. This generator can be used to fit simulation parameters by embedding the target displacement map in the C-space, the space spanned by the input class labels, i.e. the simulation parameters. To that end, we freeze the network weights and optimize the class labels in order to minimize a style loss between the generator output and the target displacement map using Adam [KB14]. We use the same style loss as Huang and Belongie [HB17] as presented by Weiss *et al.* [WMC*23]. As shown in Fig. 7e, the StyleGAN-based optimization manages to fit the parameters responsible for the "feel" well, but fails in optimizing the pore distance.

Therefore, we propose a dual discriminator-generator pipeline, combining both networks: first, the classification network is used to infer an initial set of simulation parameters. Then, the StyleGAN embedding starts with those parameters and refines them. This combination leads to the best fit as shown in Fig. 7b. Additional comparisons and evaluations are presented in the supplemental document, together with some examples of our synthetic training data. Our proposed patch fitting process can be used to automatically obtain additional simulation parameters to enhance the Simulation Preset Library. Detailed hyperparameter choices are listed in the supplemental document (Section B.1).

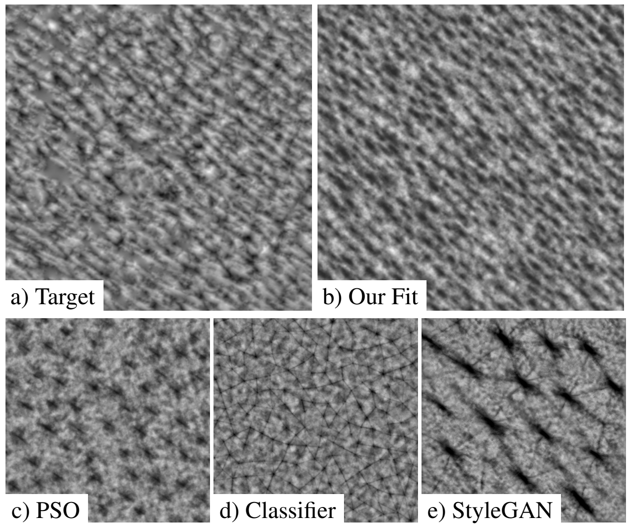

Figure 7: *Fitting simulation parameters to a target real displacement map (a). Using our approach of StyleGAN-embedding with initialization from the classifier (b) leads to the best fit. The PSO approach [WMC*23] (c) gets stuck in a suboptimal minima, the pure classifier (d) can capture the pore density well but not the other parameters, and the pure StyleGAN embedding (e) manages to capture the feel of the texture but not the pore density.*

5.3. Parameter Placement, Interpolation and Simulation

The final micro-detail displacement map that will be used as the content for stylization is obtained in a controllable way, by choosing simulation parameters from the Simulation Preset Library, placing them at various locations on the face, interpolating the parameters spatially across the entire face, and then running the simulator of Weiss *et al.* [WMC*23]. This can be accomplished in an interactive session as illustrated in Fig. 8.

6. Style Transfer for Final Micro Details

As stated from the beginning, the simulated displacement map created in Section 5 lacks the small-scale organic variations of real skin. To overcome this limitation, we apply style transfer to merge the full-face simulated wrinkles with the patch-based scanned displacement maps of real skin.

Due to the fine resolution of the displacement map, optimization-based style transfer approaches as, e.g., initially proposed by Gatys *et al.* [GEB16] become intractable. We therefore turn to fast stylization methods that train a neural network to perform the stylization at inference time. As a further limitation, only relatively few style images are available, since scanning thousands of skin patches is impractical, and so we focus on few-shot stylization algorithms. With these limitations in mind, we propose to employ the style transfer algorithm of Li *et al.* [LCLB21] (SITTA). Here, a convolutional neural network similar to a GAN-like architecture is trained to stylize *content images* taken from the database of synthetic patches (see Section 5.2) using a *style image* from the Displacement Style Library. The SITTA architecture preserves both the structure from the *content image* and texture from the *style image*, making it particularly useful for stylizing the simulated skin. We follow the training scheme by Li *et al.*, leading to a training time of around 30 minutes for a single style image.

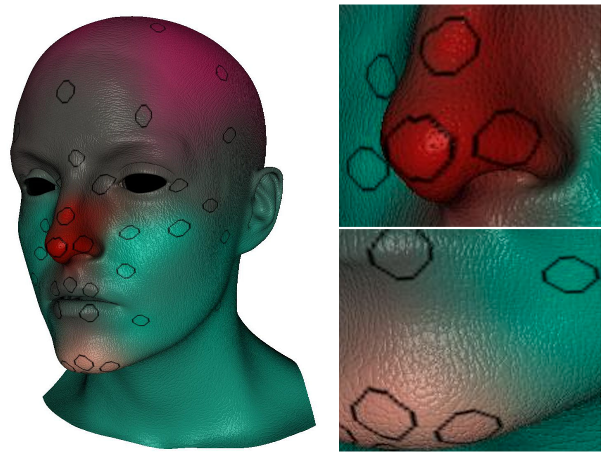

Figure 8: *Example of an interactive session where an artist applies the parameter presets from the Simulation Preset Library on the face (black circles) and the simulation creates micro wrinkles spanning the entire face with smoothly interpolated looks.*

After training, network inference is used to stylize the full 16k displacement map from the simulation in a patch-by-patch manner. To hide seams between the patches, we use an overlap of 50 pixels and feather the edges [Sze96]. Furthermore, to maintain the differing micro-detail skin structure from different regions of the face, additional SITTA models can be quickly trained from one or more additional scanned patches in the Displacement Style Library. Stylizing the entire displacement map takes around 35 seconds on a single RTX4070 GPU. This stylization of the simulation output can effectively re-introduce the high-frequency details and organic variations lacking in the simulation as shown in Fig. 9. To control the strength of the stylization in the final output, we blend the original displacement map and the stylized version with a user-controllable blending factor. Finally, users are free to choose different style images in order to achieve different fine-detail organic effects, offering another level of artistic control as illustrated in Fig. 10 and the supplemental video.

7. Results

We now demonstrate the results of our method by creating micro-wrinkle geometry for four different subjects (one elderly female and three younger subjects, two female and one male). In all cases the base geometry was obtained by a full-face 3D scan [RGB*20a], and we compare the proposed method to several alternatives as follows:

PNN - Texture synthesis to repeat the scanned displacement maps and fill the entire face using Patch Nearest Neighbors [GFS*22], an improved method over Image Analogies [HJO*01] as used by Graham [GTB*13].

Sim-Uniform - A pure simulation of the micro wrinkles with per-face constant simulation settings using the method by Weiss *et al.* [WMC*23].

Sim-Presets - The simulation extended by our spatially-varying Simulation Preset Library to obtain vastly different wrinkle types in different regions (see Section 5).

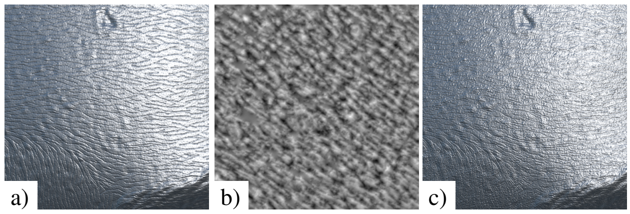

Figure 9: *The wrinkle simulation (a) creates globally consistent micro wrinkles and pores spanning the entire face. Combining this as the content image with a displacement map captured from real skin (b) as style image leads to the final output (c). Images best viewed zoomed-in.*

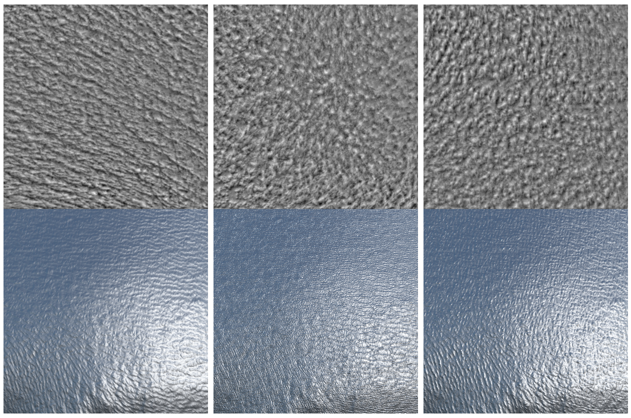

Figure 10: *Comparison of different styles on a forehead patch. The first row shows three different style images used to stylize the same simulated wrinkles, with the results in the bottom row. Images best viewed zoomed-in.*

Stylized - Our proposed method of stylizing the above spatially-varying simulation with displacement maps obtained from the Displacement Style Library using style transfer (see Section 6).

Qualitative comparisons are presented in Section 7.1. For a quantitative comparison between the four methods above we conducted a user study, which is summarized in Section 7.2.

7.1. Qualitative Comparison

Visual results for the four subjects can be found in Fig. 1, Fig. 11, Fig. 12, and Fig. 13. The geometry obtained from the full-face 3D reconstruction already contains macro-scale displacements, i.e. coarse wrinkles and features like moles (*Base*). It can be easily seen in the aforementioned figures that our methods add a substantial level of micro details.

For comparison, we have included the Patch Nearest Neighbors method (*PNN*), inspired by Graham *et al.* [GTB*13]. This method certainly improves over the base geometry, but the method does not preserve longer connected wrinkles that are commonly found in human skin. Furthermore, certain areas of the face appear stretched and deformed due to UV parametrization and flow lines are not respected.

Table 1: *Raw results of the A/B comparisons of the user study. 58 participants were tasks with ranking 56 pairs of images on which one looks more realistic to them. Each pair of images shows two methods for displacement maps from the same part of the face and subject. The table below shows how often a method won over another method in the survey.*

Winner \ Loser	PNN	Sim-Uniform	Sim-Presets	Stylized
PNN	0	84	71	73
Sim-Uniform	264	0	289	300
Sim-Presets	219	465	0	490
Stylized	275	338	380	0

The method of Weiss *et al.* [WMC*23] with uniform settings for the entire face (*Sim-Uniform*) shows relatively high quality results, however, due to the fact that the simulated wrinkles are tuned for a specific region of the face (we chose the forehead), that region looks convincing but the rest of the face lacks realism. The simulated wrinkles with spatially varying parameters provided by the preset library (*Sim-Presets*) overcomes this problem. Compare the nose in Fig. 11, as well as nose and cheek in Fig. 12 to see this clearly.

The stylization of the simulated wrinkles adds additional detail (*Stylized*). Specifically, we observe that it breaks up sharp lines found in the simulation and adds organic noise. For example, compare the nose in Fig. 11 and forehead in Fig. 13. The effect of stylization can be best seen in the accompanying video when swapping between different style images.

7.2. User Study for a Quantitative Ranking

We conducted a user study to evaluate the different wrinkle generation methods perceptually. At a high level, users were shown side-by-side results from two different methods on the same small skin patch and were asked to *"rank which image looks more realistic"* based on which one *"feels more natural to you"* as human skin. To focus only on the geometric detail, the skin patches were rendered without a color texture, similar to the qualitative figures presented in Section 7.1.

More specifically, we used four zoom-in regions (forehead, cheek, nose, crow's feet at the side of the eye) on each of our four subjects, for a total of 16 patches. For each patch, we rendered the displacement maps with the four methods *PNN*, *Sim-Uniform*, *Sim-Presets*, *Stylized*. Following the two-alternative forced choice approach, we showed the participants random pairs of two different methods for the same patch. To keep the survey short, we did not show all possible combinations, but a random selection of 56 samples.

58 participants took part in the study from different backgrounds, including some people with and some without a technical background as well as some people from the medical community. The raw results of the study can be found in Table 1 and the images used for the survey can be found in the supplementary material.

To evaluate this incomplete tournament, we use the page rank algorithm [PBMW98] as presented by Pang and Ling [PL13]. The rankings are shown in Table 2. Patch Nearest Neighbors *PNN* is

Table 2: *Page rank results given the comparisons from Table 1. The best method has the lowest ranking score, the best one the highest.*

1. (best)	Sim-Presets	Score = 0.168
2.	Stylized	Score = 0.187
3.	Sim-Uniform	Score = 0.209
4. (worst)	PNN	Score = 0.435

outperformed by the simulation-based methods by a large margin based on the user study. The winning method is *Sim-Presets*, the simulation with the parameters changing spatially based on our presets. A close second is *Stylized*, the Sim-Presets result stylized by our method. Overall our extension of the wrinkle simulator from Weiss *et al.* [WMC*23] to *Sim-Presets* has greatly increased its "realistic feel", according to the study. And while not always chosen as the preferred favorite, our *Stylized* method still outperforms previous work (e.g. PNN, Sim-Uniform) and offers a layer of artistic control on top of *Sim-Presets*.

8. Conclusion

We present an approach for generating realistic and convincing micro details for human skin, which combines the customizable benefits of simulation with the organic look of real skin patches. Our method is based on the recent simulator of Weiss *et al.* [WMC*23], which allowed for artistic and intuitive simulation of wrinkles, and we bridge the gap between simulation and realism by utilizing 3D scans of skin patches captured from real actors and creating simulation presets to vary the simulation across the face.

First, we use a novel skin patch scanning set up to recover geometry of the skin in various regions of the face. Next, a simulation parameter preset library is built from artistically designed simulated patches combined with fitted simulation parameters that match the geometry recovered from the skin patches, made possible by the dual discriminator-generator fitting method that we propose. This collection of presets can then be easily used by artists to quickly and intuitively create a 16k displacement map that contains bespoke, simulated micro geometry. This displacement map is then stylized using our captured Displacement Style Library, using a fast style transfer method to add further organic detail and enhance realism.

Our method is not without some limitations. As discussed in Section 7.2 it is clear that audiences do not always prefer the stylized wrinkle results over pure simulation with spatially-varying parameters (although they do quite often). However, we would note that the user study was performed with just one particular artistic choice of style images, and so we expect results to be a little subjective. Our method allows full artistic freedom to choose different styles for different skin regions, some of which may result in a more organic and realistic look than others. Additionally, the proposed capture setup is susceptible to vibrations and movements caused by the captured subject, due to breathing or shifting during capture. This motion may disrupt reconstruction when viewed at the scale we capture at. We largely alleviate the problem by aligning the captured images, but it is difficult to remove the motion completely. A possible future work would be to revisit the image registration step and improve the image alignment further.

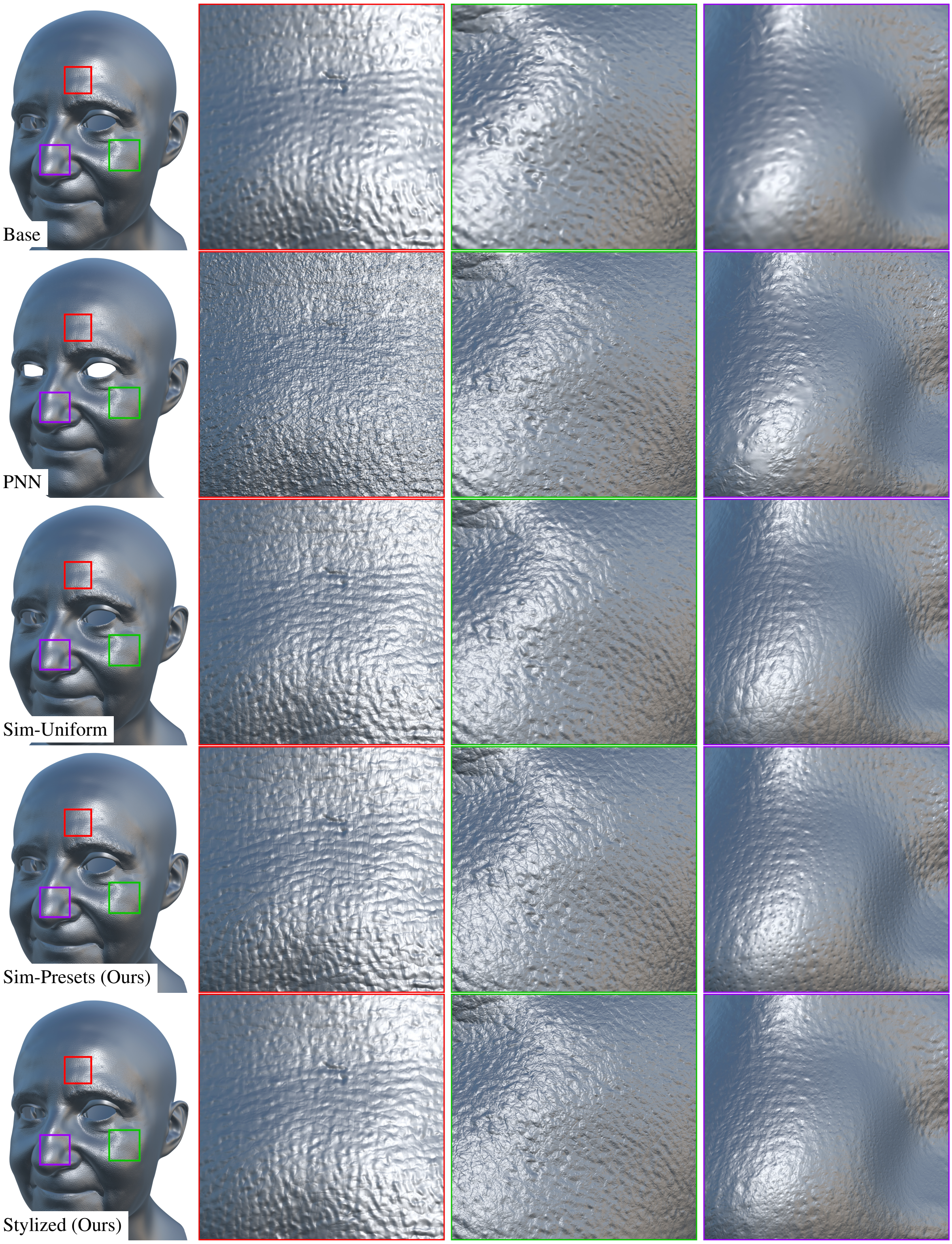

Figure 11: *Detailed comparison of Subject 1. The first row shows the base mesh as obtained from a full head 3D reconstruction. The second row shows micro geometry created by texture synthesis (PNN) similar to Graham* et al. *[GTB*13]. The third row shows the refinement using the simulation method by Weiss* et al. *[WMC*23] using constant uniform simulation parameters across the face. In the fourth row we improve this by spatially varying the simulation parameters using the proposed Simulation Preset Library. In the last row, this simulation is then stylized using the Displacement Style Library.*

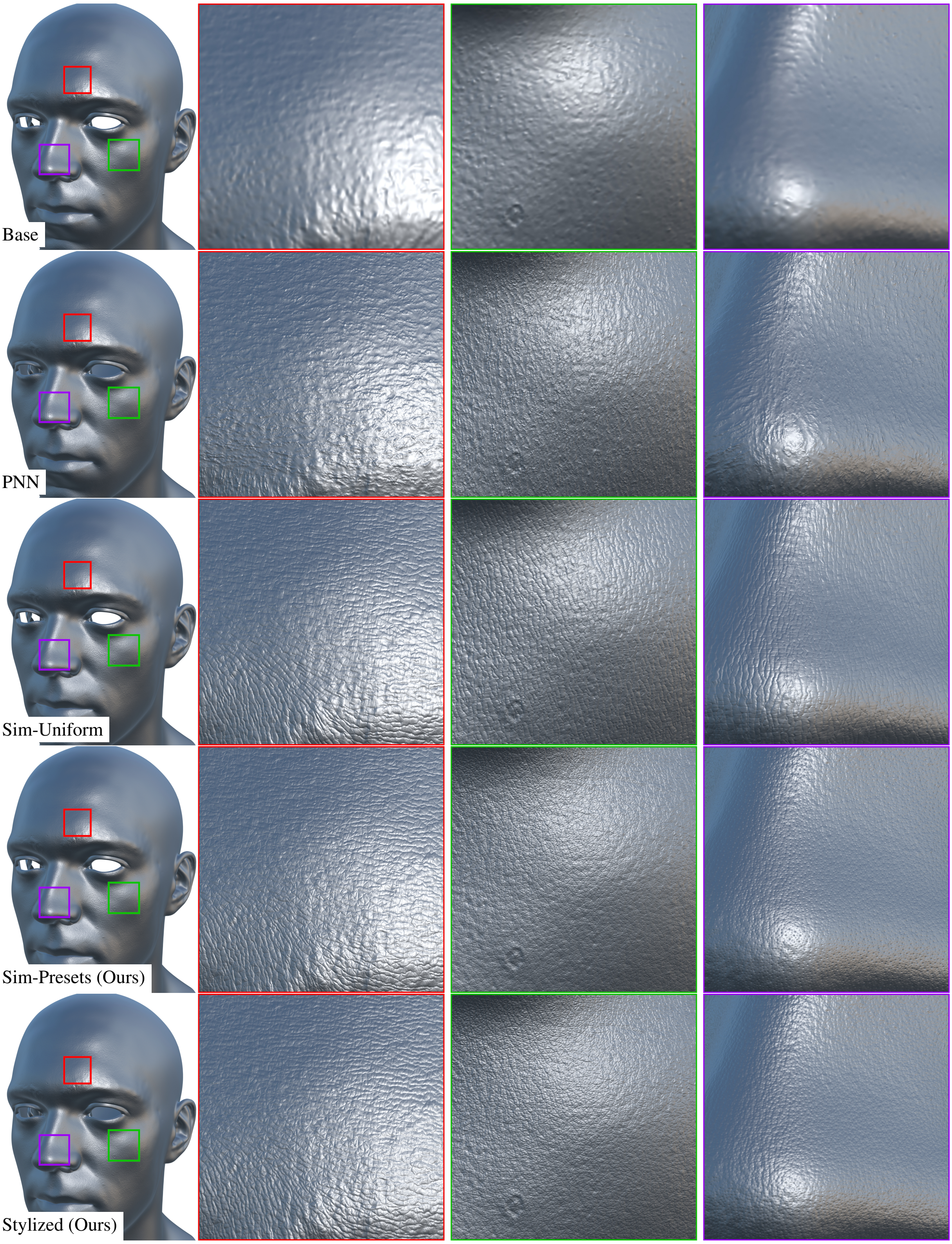

Figure 12: *Detailed comparison of Subject 2, with the same rows described in Fig. 11. Note that when estimating the micro geometry using texture synthesis (PNN) wrinkles that conform to the flow lines are lost and the structure of regions that cannot be captured, e.g. the nose, do not match expectations.*

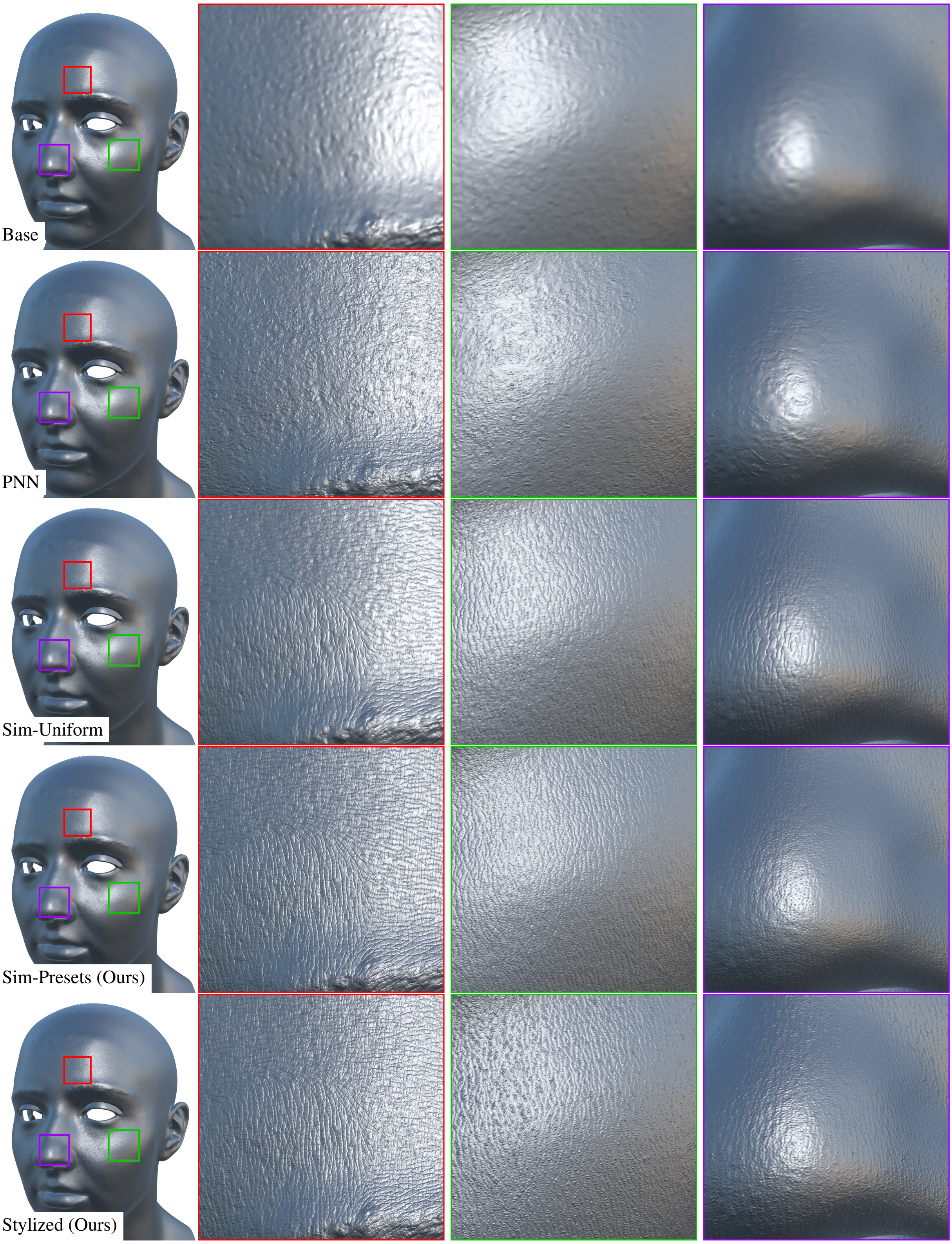

Figure 13: *Detailed comparison of Subject 3. Here, the differences between a simulation with uniform parameters and a simulation with spatially varying parameters drawn from the preset library are especially noticeable. A strong stylization was chosen for the cheek and nose region to highlight the features that can be introduced with this method.*

References

[AE23] ABDELLATIF A., ELSHEIKH A. H.: Generating Infinite-Resolution texture using GANs with Patch-by-Patch paradigm. arXiv:2309.02340. 3

[Ash03] ASHIKHMIN M.: Fast texture transfer. *IEEE Comput. Graph. Appl. 23*, 4 (July 2003), 38–43. 3

[BBA*07] BICKEL B., BOTSCH M., ANGST R., MATUSIK W., OTADUY M., PFISTER H., GROSS M.: Multi-scale capture of facial geometry and motion. *ACM Transactions on Graphics 26*, 3 (2007). 2

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M.: High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics 29*, 4 (2010). 2

[BKN02] BANDO Y., KURATATE T., NISHITA T.: A simple method for modeling wrinkles on human skin. In *10th Pacific Conference on Computer Graphics and Applications* (2002), pp. 166–175. 2

[BMH*16] BALU M., MIKAMI H., HOU J., POTMA E. O., TROMBERG B. J.: Rapid mesoscale multiphoton microscopy of human skin. *Biomed. Opt. Express 7*, 11 (2016), 4375–4387. 2

[CGS06] CHEN T., GOESELE M., SEIDEL H.-P.: Mesostructure from specularity. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (June 2006), vol. 2, pp. 1825–1832. 2, 4

[CWX*21] CHEN Z., WANG W., XIE E., LU T., LUO P.: Towards Ultra-Resolution neural style transfer via thumbnail instance normalization. arXiv:2103.11784. 3

[CZG*21] CHANDRAN P., ZOSS G., GOTARDO P., GROSS M., BRADLEY D.: Adaptive convolutions for structure-aware style transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 7972–7981. 3

[CZGB23] CHANDRAN P., ZOSS G., GOTARDO P., BRADLEY D.: Continuous landmark detection with 3d queries. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2023), pp. 16858–16867. 2

[DTA*21] DIB A., THEBAULT C., AHN J., GOSSELIN P., THEOBALT C., CHEVALLIER L.: Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2021). 2

[DYX*19] DENG Y., YANG J., XU S., CHEN D., JIA Y., TONG X.: Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *IEEE Computer Vision and Pattern Recognition Workshops* (2019). 2

[FFBB21] FENG Y., FENG H., BLACK M. J., BOLKART T.: Learning an animatable detailed 3D face model from in-the-wild images. *ACM Transactions on Graphics, (Proc. SIGGRAPH) 40*, 8 (2021). 2

[GEB16] GATYS L. A., ECKER A. S., BETHGE M.: Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). 3, 6

[GEBo17] GATYS L. A., ECKER A. S., BETHGE M., OTHERS: Controlling perceptual factors in neural style transfer. *Proceedings of the* (2017). 3

[GFS*22] GRANOT N., FEINSTEIN B., SHOCHER A., BAGON S., IRANI M.: Drop the GAN: In defense of patches nearest neighbors as single image generative models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), IEEE, pp. 13460–13469. 2, 7

[GFT*11] GHOSH A., FYFFE G., TUNWATTANAPONG B., BUSCH J., YU X., DEBEVEC P.: Multiview face capture using polarized spherical gradient illumination. In *Proceedings of the 2011 SIGGRAPH Asia Conference* (2011), no. 129, Association for Computing Machinery, pp. 1–10. 2

[GPKo19] GECER, PLOUMPIS, KOTSIA, OTHERS: Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). 2

[GTB*13] GRAHAM P., TUNWATTANAPONG B., BUSCH J., YU X., JONES A., DEBEVEC P., GHOSH A.: Measurement-based synthesis of facial microgeometry. In *Computer Graphics Forum* (2013), vol. 32, pp. 335–344. 1, 2, 4, 5, 7, 9

[GZY*20] GUO J., ZHU X., YANG Y., YANG F., LEI Z., LI S. Z.: Towards fast, accurate and stable 3d dense face alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2020). 2

[HB17] HUANG X., BELONGIE S.: Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2017). 3, 6

[HJO*01] HERTZMANN A., JACOBS C. E., OLIVER N., CURLESS B., SALESIN D. H.: Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), SIGGRAPH '01, pp. 327–340. 2, 3, 7

[IBP15] ICHIM A. E., BOUAZIZ S., PAULY M.: Dynamic 3d avatar creation from hand-held video input. *ACM Transactions on Graphics (ToG) 34*, 4 (2015), 1–14. 2

[INN07] IGARASHI T., NISHINO K., NAYAR S. K.: The appearance of human skin: A survey. *Foundations and Trends® in Computer Graphics and Vision 3*, 1 (2007), 1–95. 4

[IYYT93] ISHII T., YASUDA T., YOKOI S., TORIWAKI J.-I.: A generation model for human skin texture. In *Communicating with Virtual Worlds* (1993), pp. 139–150. 2

[JAFF16] JOHNSON J., ALAHI A., FEI-FEI L.: Perceptual losses for Real-Time style transfer and Super-Resolution. In *Computer Vision – ECCV 2016* (2016), Springer International Publishing, pp. 694–711. 3

[JYF*19] JING Y., YANG Y., FENG Z., YE J., YU Y., SONG M.: Neural style transfer: A review. *IEEE transactions on visualization and computer graphics 26*, 11 (2019), 3365–3385. 3

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. arXiv:1412.6980. 6

[KE95] KENNEDY J., EBERHART R.: Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks* (1995), vol. 4, pp. 1942–1948 vol.4. 5

[KLA*20] KARRAS T., LAINE S., AITTALA M., HELLSTEN J., LEHTINEN J., AILA T.: Analyzing and improving the image quality of stylegan. 8110–8119. 6

[LC93] LEVEQUE J. L., CORCUFF P.: The surface of the skin — the microrelief. In *Noninvasive Methods for the Quantification of Skin Functions: An Update on Methodology and Clinical Applications*. 1993, pp. 3–24. 2

[LCLB21] LI B., CUI Y., LIN T.-Y., BELONGIE S.: SITTA: Single image texture translation for data augmentation. arXiv:2106.13804. 3, 6

[LLK*22] LATTAS A., LIN Y., KANNAN J., OZTURK E., FILIPI L., GUARNERA G. C., CHAWLA G., GHOSH A.: Desktop-based High-Quality facial capture for everyone. In *ACM SIGGRAPH 2022 Talks* (2022), no. 15, pp. 1–2. 2

[LLKY19] LI X., LIU S., KAUTZ J., YANG M.-H.: Learning linear transformations for fast image and video style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 3809–3817. 3

[LLLC11] LI L., LIU F., LI C., CHEN G.: Realistic wrinkle generation for 3D face modeling based on automatically extracted curves and improved shape control functions. *Computuers & Graphics 35*, 1 (2011), 175–184. 3

[LMW*22] LIU Z., MAO H., WU C.-Y., FEICHTENHOFER C., DARRELL T., XIE S.: A ConvNet for the 2020s. 11976–11986. arXiv:2201.03545. 5

[LRB*22] LÖTZSCH W., REIMANN M., BÜSSEMEYER M., SEMMO A., DÖLLNER J., TRAPP M.: WISE: Whitebox image stylization by Example-Based learning. In *Computer Vision – ECCV 2022* (2022), Springer Nature Switzerland, pp. 135–152. 3

[LSRY10] LEE H., SEO S., RYOO S., YOON K.: Directional texture transfer. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering* (New York, NY, USA, June 2010), NPAR '10, Association for Computing Machinery, pp. 43–48. 3

[LSY11] LEE H., SEO S., YOON K.: Directional texture transfer with edge enhancement. *Comput. Graph. 35*, 1 (Feb. 2011), 81–91. 3

[LWLH17] LI Y., WANG N., LIU J., HOU X.: Demystifying neural style transfer. arXiv:1701.01036. 3

[LXZ07] LI Y.-B., XIAO H., ZHANG S.-Y.: The wrinkle generation method for facial reconstruction based on extraction of partition wrinkle line features and fractal interpolation. In *Fourth International Conference on Image and Graphics (ICIG 2007)* (2007), pp. 933–937. 3

[LZY*17] LU M., ZHAO H., YAO A., XU F., CHEN Y., ZHANG L.: Decoder network over lightweight reconstructed feature for fast semantic style transfer. In *2017 IEEE International Conference on Computer Vision (ICCV)* (Oct. 2017), IEEE, pp. 2469–2477. 3

[MHP*07] MA W.-C., HAWKINS T., PEERS P., CHABERT C.-F., WEISS M., DEBEVEC P. E., ET AL.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques* (2007), 183–194. 2

[NFA*15] NAGANO K., FYFFE G., ALEXANDER O., BARBIČ J., LI H., GHOSH A., DEBEVEC P.: Skin microstructure deformation with displacement map convolution. *ACM Transactions on Graphics 34*, 4 (2015), 1–10. 2

[NHRD90] NAHAS M., HUITRIC H., RIOUX M., DOMEY J.: Facial image synthesis using skin texture recording. *The Visual Computer 6*, 6 (1990), 337–343. 2

[PBMW98] PAGE L., BRIN S., MOTWANI R., WINOGRAD T.: *The pagerank citation ranking: Bring order to the web.* Tech. rep., Technical report, stanford University, 1998. 8

[PL13] PANG Y., LING H.: Finding the best from the second bests - inhibiting subjective bias in evaluation of visual tracking algorithms. In *2013 IEEE International Conference on Computer Vision* (Dec. 2013), IEEE. 8

[PQE21] PIZENBERG M., QUÉAU Y., ELMOATAZ A.: Low-Rank registration of images captured under unknown, varying lighting. In *Scale Space and Variational Methods in Computer Vision* (2021), Springer International Publishing, pp. 153–164. 4

[RGB*20a] RIVIERE J., GOTARDO P., BRADLEY D., GHOSH A., BEELER T.: Single-shot high-quality facial geometry and skin appearance capture. *ACM Transactions on Graphics 39*, 4 (2020). 7

[RGB*20b] RIVIERE J., GOTARDO P., BRADLEY D., GHOSH A., BEELER T.: Single-shot high-quality facial geometry and skin appearance capture.(2020). 2

[SK22] SADEGHZADEH H., KOOHI S.: Translation-invariant optical neural network for image classification. *Scientific Reports 12*, 1 (Oct. 2022), 17232. 5

[SKLO18] SANAKOYEU A., KOTOVENKO D., LANG S., OMMER B.: A style-aware content loss for real-time HD style transfer. 698–714. arXiv:1807.10201. 3

[SYZ18] SHEN F., YAN S., ZENG G.: Neural style transfer via meta networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (June 2018), IEEE, pp. 8061–8069. 3

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for Large-Scale image recognition. arXiv:1409.1556. 5

[Sze96] SZELISKI R.: Video mosaics for virtual environments. *IEEE Computer Graphics and Applications 16*, 2 (Mar. 1996), 22–30. 7

[TZK*17] TEWARI A., ZOLLÖFER M., KIM H., GARRIDO P., BERNARD F., PEREZ P., CHRISTIAN T.: MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction. In *The IEEE International Conference on Computer Vision (ICCV)* (2017). 2

[VB18] VANDERFEESTEN R., BIKKER J.: Example-Based skin wrinkle displacement maps. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (2018), pp. 212–219. 3

[vdPJD*14] VON DER PAHLEN J., JIMENEZ J., DANVOYE E., DEBEVEC P., FYFFE G., ALEXANDER O.: Digital ira and beyond: creating real-time photoreal digital actors. In *ACM SIGGRAPH 2014 Courses* (New York, NY, USA, 2014), no. Article 1 in SIGGRAPH '14, Association for Computing Machinery, pp. 1–384. 1

[WBGB16] WU C., BRADLEY D., GROSS M., BEELER T.: An anatomically-constrained local deformation model for monocular face capture. *ACM transactions on graphics (TOG) 35*, 4 (2016), 1–12. 2

[WBH*22] WOOD E., BALTRUSAITIS T., HEWITT C., JOHNSON M., SHEN J., MILOSAVLJEVIC N., WILDE D., GARBIN S., RAMAN C., SHOTTON J., SHARP T., STOJILJKOVIC I., CASHMAN T., VALENTIN J.: 3d face reconstruction with dense landmarks. *European Conference on Computer Vision* (2022). 2

[WKT96] WU Y., KALRA P., THALMANN N. M.: Simulation of static and dynamic wrinkles of skin. In *Proceedings Computer Animation* (1996), pp. 90–97. 2

[WLW*20] WANG H., LI Y., WANG Y., HU H., YANG M.-H.: Collaborative distillation for ultra-resolution universal style transfer. 1860–1869. arXiv:2003.08436. 3

[WMC*23] WEISS S., MOULIN J., CHANDRAN P., ZOSS G., GOTARDO P., BRADLEY D.: Graph-Based synthesis for skin micro wrinkles. In *COMPUTER GRAPHICS forum* (2023), vol. 42. 1, 2, 3, 5, 6, 7, 8, 9