Generative Texture Super-Resolution via Differential Rendering

Milena Bagdasarian[1], Peter Eisert[1,2] and Anna Hilsmann[1]
[1]*Fraunhofer Heinrich-Hertz-Institute, Berlin, Germany*
[2]*Humboldt University of Berlin, Germany*

Keywords: Texture Super-Resolution, Differentiable Rendering, GAN.

Abstract: Image super-resolution is a well-studied field that aims at generating high-resolution images from low-resolution inputs while preserving fine details and realistic features. Despite significant progress on regular images, inferring high-resolution textures of 3D models poses unique challenges. Due to the non-contiguous arrangement of texture patches, intended for wrapping around 3D meshes, applying conventional image super-resolution techniques to texture maps often results in artifacts and seams at texture discontinuities on the mesh. Additionally, obtaining ground truth data for texture super-resolution becomes highly complex due to the labor intensive process of hand-crafting ground truth textures for each mesh. We propose a generative deep learning network for texture map super-resolution using a differentiable renderer and calibrated reference images. Combining a super-resolution generative adversarial network (GAN) with differentiable rendering, we guide our network towards learning realistic details and seamless texture map super-resolution without a high-resolution ground truth of the texture. Instead, we use high-resolution reference images. Through the differentiable rendering approach, we include model knowledge such as 3D meshes, projection matrices, and calibrated images to bridge the domain gap between 2D image super-resolution and texture map super-resolution. Our results show textures with fine structures and improved detail, which is especially of interest in virtual and augmented reality environments depicting humans.

1 INTRODUCTION

Image super-resolution (SR) is an extensively researched topic where a high-resolution image is estimated from a low-resolution input. It aims not only at increasing the pixel count but also at recovering fine details and realistic features. Research on this problem has produced many statistical, model, and learning-based approaches (Yang et al., 2014). However, due to the ill-posed nature of the task, inferring a high-resolution image from its low-resolution version remains an active area of research.

In recent years, the widespread adoption of virtual and augmented reality (VR/AR) has sparked an increasing interest in creating realistic human 3D models. One particular area of focus is the creation of highly detailed and lifelike personal avatars (Eisert et al., 2022). Alongside the development of 3D models for virtual humans, there is a growing need for super-resolving and improving their textures. While texture maps can be represented as 2D images, super-resolving them poses new challenges. Since the textures are intended to be wrapped around 3D meshes, they are often arranged in non-contiguous patches, which can introduce distortions when viewed in image space. Therefore, applying a regular image SR approach to a texture map would inevitably lead to the creation of artifacts and seams at the texture discontinuities when applied to a mesh. Consequently, texture super-resolution algorithms must address these specific challenges.

Learning-based super-resolution networks often rely on supervision and, consequently, high-resolution ground truth images (Wang et al., 2021a). However, in the context of textures, obtaining ground truth data becomes considerably more complex. For every mesh, a hand-crafted ground truth texture would be required, necessitating an impractical amount of manual effort. Although captured images of the object or subject under study can potentially provide the necessary details for super-resolution, they do not directly present these details in texture space, rendering them unsuitable as direct ground truth substitutes.

To address these challenges, we contribute to the advancement of texture super-resolution techniques,

Bagdasarian, M., Eisert, P. and Hilsmann, A.
Generative Texture Super-Resolution via Differential Rendering.
DOI: 10.5220/0012303300003660
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 3: VISAPP*, pages 282-289
ISBN: 978-989-758-679-8; ISSN: 2184-4321
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

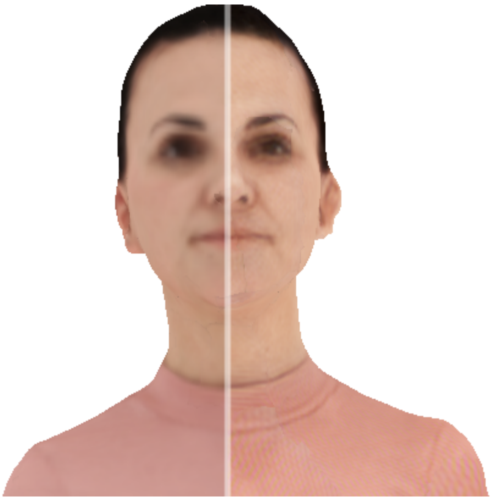

Figure 1: Rendered 3D model with a low resolution versus a super-resolved texture.

specifically addressing the challenges that arise in the context of 3D mesh textures. For this, we propose a generative texture super-resolution architecture that leverages a differentiable renderer to bridge the domain gap between texture space and images. In our proposed architecture, a generator network is employed to super-resolve a low-resolution texture, while a discriminator network distinguishes between a mesh rendering with the super-resolved texture and the real-world images of the object. By utilizing this approach, we eliminate the need for a ground-truth texture during the training of our model. To validate the effectiveness of our approach, we evaluate our model on volumetric sequences of humans, where we demonstrate its ability to enhance the visual fidelity of the textures. Our experimental results demonstrate the potential of our approach to significantly improve the visual quality and realism of textures, paving the way for the creation of highly detailed and lifelike personal avatars in virtual and augmented reality applications, generated from volumetric capturings of humans (Paier et al., 2023; Hilsmann et al., 2020).

2 RELATED WORK

2.1 Super-Resolution

Texture super-resolution aims at enhancing the resolution and details of texture maps, which plays a crucial role in creating realistic and highly detailed 3D models, particularly in virtual and augmented reality (VR/AR) environments. The field of texture map super-resolution has evolved from traditional image super-resolution problems but poses specific challenges as the texture space is defined on the mesh surface rather than on a 2-dimensional image grid. Early image super-resolution methods predominantly relied on interpolation-based techniques which compute color values based on the local neighborhood of pixels. However, this cannot guarantee the recovery of high-frequency details (Glasner et al., 2009). Model-based approaches interpret the low-resolution image as a downsampled version of the high resolution image and integrate a-priori knowledge on smoothness or similarity in order to solve for the high-resolution image (Ancuti et al., 2018; Wang et al., 2021b). In recent years, learning-based approaches have emerged as a dominant paradigm for image super-resolution. Deep learning models, such as using Transformer (Lu et al., 2022; Cao et al., 2022), Convolutional Neural Networks (CNN) (Dong et al., 2014), employing sub-pixel convolutions (Shi et al., 2016), Laplacian pyramid networks (Lai et al., 2017) or Generative Adversarial Networks (GAN) (Ledig et al., 2017; Wang et al., 2018), have demonstrated remarkable success in capturing intricate details and realistic features. These models learn the mapping between low-resolution and high-resolution texture maps by leveraging large-scale datasets of paired texture examples.

However, image super-resolution networks are usually limited to a regular 2D image space. While texture maps can be represented through 2D images, neighboring pixels in the texture space do not automatically correspond to neighboring pixels after projection. Therefore, classical image SR approaches are not suitable for texture map super-resolution as they may produce unwanted distortions and seams between patches when applied to texture maps.

Only a handful of papers have addressed these challenges of texture map or appearance super-resolution. Some approaches overcome the domain gap between texture and image space by leveraging multi-view datasets (Cheng et al., 2022; Richard et al., 2019; Tsiminaki et al., 2019; Goldluecke and Cremers, 2009), exploiting temporal redundancy (Tsiminaki et al., 2019), and jointly optimizing geometry and texture (Maier et al., 2017). Li et al. incorporated geometric information by introducing normal maps in order to guide the super-resolution of texture images by interpreting them as high-level features (Li et al., 2019). Recently, Ranade et al. introduced a rendering loss using differentiable rendering and high-resolution ground-truth textures as well as the geometry for training (Ranade et al., 2022).

However, often, these works assume ground truth high resolution textures for training. Especially for objects that have been reconstructed from real-world image data, such as capturings of human performances, high-resolution versions of the textures are not available as the textures have been created by fusing and merging multi-view image data. Pesavento et al. addressed this problem especially for the case of 4D captured human performances by transferring a super-resolution texture captured with high-resolution static cameras to a dynamic video performance capture (Pesavento et al., 2021). However, while this approach transfers high resolution texture details onto a dynamic performance, all the dynamic dependent details (wrinkling of clothes etc.) get lost with this approach.

We propose a new appearance super-resolution approach in which we rely on a learning-based network architecture and include structural information in the form of meshes through differentiable rendering. In contrast to existing approaches that require high-resolution ground-truth textures, we only use high-resolution reference images.

2.2 Differentiable Rendering

In Computer Graphics, rendering plays a crucial role in simulating the forward image formation process, which involves generating visually realistic images from 3D scenes by considering various factors such as material properties, lighting conditions, camera settings, and geometric information. However, a common limitation faced by renderers is their discretization step during rasterization, which leads to a loss of differentiability and hampers their integration with machine learning models.

The demand for differentiable renderers has emerged to address this limitation and enables the seamless incorporation of the rendering process into gradient-based optimization techniques that leverage backpropagation. While some existing approaches, such as OpenDR (Loper and Black, 2014) and the neural 3D mesh renderer (Kato et al., 2018), provide more traditional rasterization methods and approximate the rendering gradient during backpropagation, other approaches offer fully differentiable rendering functions. Examples of such approaches include the PyTorch3D renderer (Johnson et al., 2020), SoftRas (Liu et al., 2019), and a study by Laine et al. (Laine et al., 2020). Additionally, differentiable rendering of point clouds has been demonstrated in a previous work by Insafutdinov et al. (Insafutdinov and Dosovitskiy, 2018). More realistic rendering results yet at a higher computing cost are provided by differentiable ray tracing methods, such as those proposed by Li et al. (Li et al., 2018) and Nimier-David et al. (Nimier-David et al., 2019).

We use the PyTorch3D renderer (Johnson et al., 2020) to include model knowledge into our learning-based super-resolution approach. We show that texture super-resolution can be learned from reference images instead of high-resolution ground truth textures.

3 OVERVIEW AND ARCHITECTURE

In our proposed GAN-based texture super-resolution architecture, we leverage the power of a differentiable renderer. The overall framework consists of three key components: a generator (G), a differentiable renderer (DR), and a discriminator (D). The generator is responsible for super-resolving a low-resolution texture, while the differentiable renderer is used to render the mesh with the generated texture. Finally, the discriminator plays a crucial role in distinguishing between the mesh rendering with the super-resolved texture and a real-world image of the object or subject.

To construct our texture generating network, we draw inspiration from the architecture of SRGAN (Ledig et al., 2017), which is a deep learning framework for achieving photo-realistic single image super-resolution. The generator in our network follows the design principles outlined in (Ledig et al., 2017) and is composed of 16 residual blocks (He et al., 2016) that are identical in structure. However, instead of employing the sub-pixel convolution layers proposed in SRGAN, we employ up-sampling layers to increase the resolution of the generated texture. On the other hand, the discriminator network, which plays a critical role in assessing the realism of the super-resolved textures, consists of eight convolution layers, two dense layers, and utilizes a sigmoid activation function for classifying the probability of the input being real or fake, following the approach in (Ledig et al., 2017).

While the SRGAN architecture is well-suited for 2D image super-resolution, our objective focuses on enhancing texture maps. To bridge this gap, we introduce a differentiable renderer (Johnson et al., 2020) between the generator and discriminator networks. This renderer facilitates the integration of the generated textures onto a mesh and enables the evaluation of the rendered results against real-world images. By leveraging the differentiable renderer, we can exploit the benefits of gradient-based optimization techniques, allowing us to train the network. more effectively and generate high-quality super-resolved texture maps. Figure 2 illustrates the architecture of our texture generating network, showcasing the interactions between the generator, differentiable renderer,

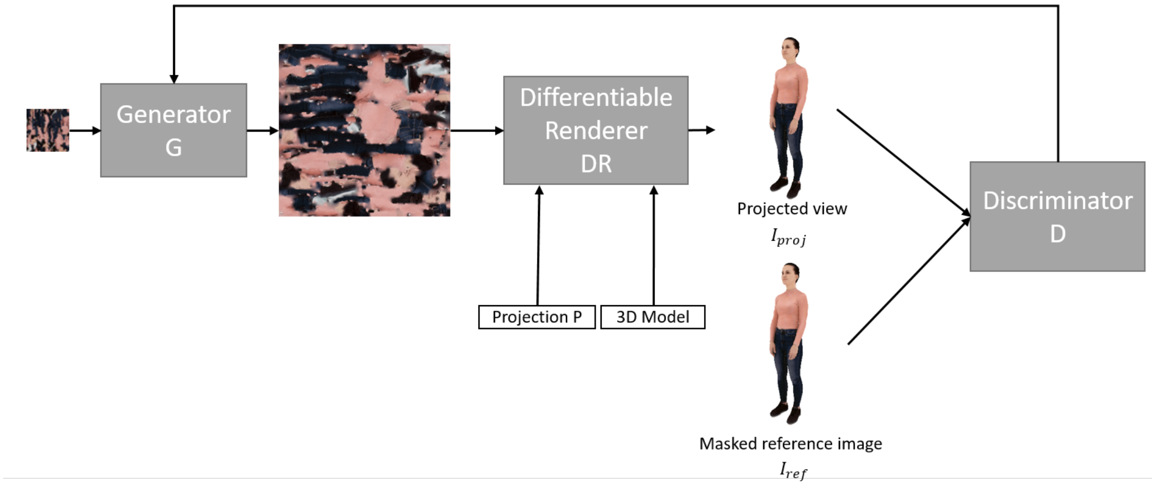

Figure 2: Architecture overview. Given a low-resolution texture (T_{LR}), a 3D model, a projection matrix, and a calibrated reference image I_{ref}, we synthesize a high resolution texture T_{HR} through the generator. We project the 3D model and its generated high-resolution texture (T_{HR}) and compare this projection to the corresponding masked reference image.

and discriminator components.

During the training phase, the generator module takes a low-resolution texture map denoted as T_{LR} as input. The primary objective of the generator is to produce a corresponding high-resolution texture map denoted as T_{HR}. This process enables the enhancement of the details and quality of the input texture map. To incorporate model knowledge into the training process, we provide the high-resolution texture map T_{HR} along with additional information, including a 3D mesh representation and a projection matrix denoted as P, to the differentiable renderer. The differentiable renderer takes these inputs and generates a textured and projected mesh, denoted as I_{proj}. To evaluate the realism and fidelity of the rendered output, we compare I_{proj} with a calibrated high-resolution reference image denoted as I_{ref}, which depicts the same person in the same pose. However, to account for unwanted backgrounds and focus solely on the person of interest, we apply a segmentation mask computed from the 3D model. This segmentation mask effectively isolates the person in the photograph, allowing us to concentrate on the quality of the texture super-resolution. We exclusively utilize low-resolution textures and high-resolution images from the volumetric dataset. This approach eliminates the need for high-resolution textures in the training phase, as we focus on learning the mapping between low-resolution textures and their corresponding high-resolution counterparts.

Once the training phase is completed, during inference or the application phase, we only utilize the generator component of our network. Given a specific low-resolution texture map, the generator is employed to super-resolve the input, producing a high-resolution texture map that exhibits enhanced details and improved quality.

3.1 Loss Functions

In the generator module of our network, we incorporate two essential pixel-wise loss functions to guide the training process: a content loss based on a pre-trained VGG19 network and an adversarial loss. These losses contribute to the overall network loss, denoted as L_{SR}, which is formed as a weighted sum of the individual losses. Here, λ_a and λ_b represent the respective weights for the losses.

The total network loss L_{SR} is defined as follows:

$$L_{SR} = \lambda_a L^{T}_{MSE} + L^{P}_{MSE} + L^{vgg_{ij}}_{MSE} + \lambda_b L^{SR}_{BCE} \qquad (1)$$

To enforce color coherence between the low-resolution input texture T_{LR} and the generated high-resolution output texture T_{HR}, we calculate a texture loss termed L^{T}_{MSE} (Eq. 2), expressed as a pixel-wise mean squared error (MSE) between the two textures. In particular, we employ a down-sampled version of the generated texture, denoted as T^{ds}_{HR} to avoid the need for the introduction of a ground truth texture into our network. This texture loss can be formulated as follows:

$$L^{T}_{MSE} = \frac{1}{wh}\sum_{x=1}^{w}\sum_{y=1}^{h}(T_{LR}(x,y) - T^{ds}_{HR}(x,y))^2 \qquad (2)$$

Furthermore, we introduce a projection loss denoted as L^{P}_{MSE}, which represents the mean squared error between the output of the differentiable renderer (I_{proj}) and the masked reference image (I_{ref}). This loss ensures that the rendered and projected 3D model aligns well with the corresponding reference image. The projection loss can be computed as follows:

$$L_{MSE}^{P} = \frac{1}{wh} \sum_{x=1}^{w} \sum_{y=1}^{h} (I_{proj}(x,y) - I_{ref}(x,y))^2 \qquad (3)$$

Additionally, we incorporate a content loss $L_{MSE}^{vgg_{ij}}$, which quantifies the discrepancy between the rendered output (I_{proj}) and the reference image (I_{ref}) in the feature-space. Inspired by the approach presented in (Ledig et al., 2017), we extract features from a pre-trained VGG19 model after the j-th convolution layer and before the i-th max-pooling layer. The content loss is defined as follows:

$$L_{MSE}^{vgg_{ij}} = \frac{1}{wh} \sum_{x=1}^{w} \sum_{y=1}^{h} (vgg_{ij}(I_{proj})(x,y) - vgg_{ij}(I_{ref})(x,y))^2 \qquad (4)$$

Lastly, the adversarial loss, denoted as L_{BCE}^{SR}, plays a central role in aligning the distribution of the generated high-resolution textures with that of the target textures. This loss is formulated as a binary cross-entropy loss, where $D(G(T_{LR}))$ represents the probability that the reconstructed image is a target high-resolution texture. The adversarial loss can be expressed as:

$$L_{BCE}^{SR} = \sum_{n=1}^{N} -logD(G(T_{LR})) \qquad (5)$$

By incorporating these diverse loss components in our generator module, we enable a comprehensive and multi-faceted training process that emphasizes color coherence, texture fidelity, and adversarial learning. These losses collectively contribute to the training objective and guide the generator in producing high-quality super-resolved textures.

3.2 Input Data

To conduct our experiments and evaluate the performance of our proposed method, we rely on a carefully curated dataset (see Fig. 3) consisting of volumetric video sequences of human performances. Each frame within the dataset possesses its own reconstructed 3D model, corresponding texture map, and a set of 32 calibrated high-resolution images. The acquisition of this dataset involved capturing volumetric video in a specialized studio setting that was specifically designed to capture human subjects. The studio setup includes a total of 32 RGB cameras arranged in 16 stereo pairs, effectively covering the subject from all angles.

The data pipeline employed in the acquisition process is detailed in (Schreer et al., 2019), providing a comprehensive overview of the steps involved in capturing and processing the volumetric video sequences. In our study, we specifically focus on the recorded volumetric sequences that have a minimum length of 800 frames. Consequently, this results in an equivalent number of 3D models available for each recorded subject within the dataset. It is worth noting that the cameras utilized in the setup are accurately calibrated to ensure consistency and alignment between different views and frames.

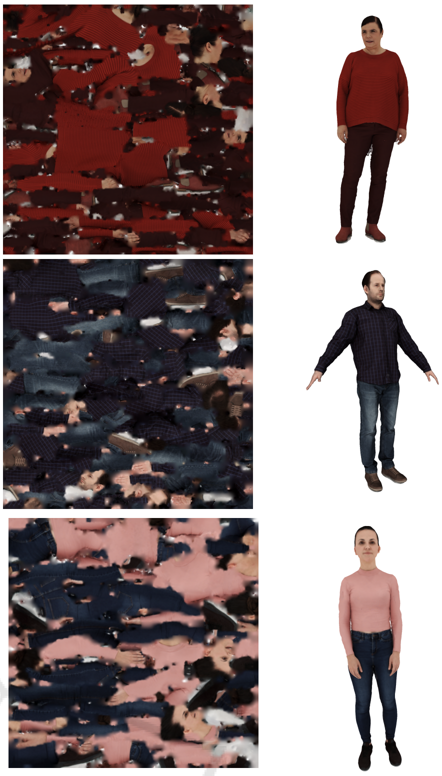

Figure 3: Texture and projected mesh, example frames from the dataset.

4 EXPERIMENTS AND RESULTS

In our training process, we adopt a systematic approach by dividing the available texture maps into three separate sets, each corresponding to a different volumetric sequence. We utilize all the available textures within each set and train our network architecture accordingly. The training process is stopped after 20 epochs to prevent overfitting and to ensure that the network has effectively learned the desired texture enhancement. It is important to note that we do not employ any pre-training for our network architecture. However, we incorporate a pre-trained VGG19-layer model in the content loss calculation to lever-

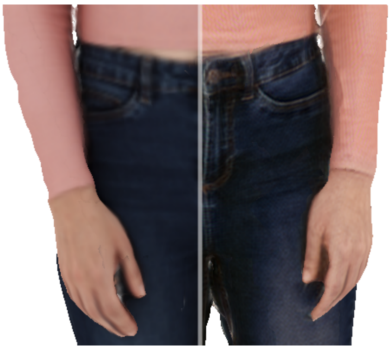

Figure 4: Rendered 3D model with a low resolution versus a super-resolved texture.

age its learned representations. During each training step, a random image is selected from the 32 available camera views to be compared with the projection generated by the network. We refrain from using high-resolution textures as ground truth, as our objective is to enhance the textures based solely on the information contained within the reference images.

To ensure that our network captures the individual characteristics and nuances of each human 3D model, we train a separate model for each volumetric sequence. By adopting this approach, we can tailor the training process to the specific characteristics and variations present in each sequence, ultimately leading to more accurate and personalized texture super-resolution.

In Figure 4, we present a visual comparison between the rendered 3D model with a low-resolution texture (left) and the same model with a super-resolved texture (right). The improvement in texture quality and detail is evident, highlighting the effectiveness of our approach.

Furthermore, in Figure 5, we showcase an example of a low-resolution input texture and the significant increase in resolution achieved in the generated texture. Additionally, we display the corresponding projected images, including the low-resolution projection and the projection with the generated high-resolution texture. Notably, our results demonstrate enhanced fine details, such as the seams of the clothing, such as the jeans' pocket, as well as improvements in the representation of eyes and hair. It is important to emphasize that the super-resolved details in our outputs are not directly derived from high-resolution ground-truth textures but rather extracted and introduced through the discrimination process and loss backpropagation, leveraging the information captured in the reference images I_{ref}. As a result, our approach minimizes visible seams when applied to the 3D mesh, ensuring a more coherent and visually appealing texture enhancement.

5 CONCLUSION AND FUTURE WORK

In this paper, we have proposed a GAN-based texture super-resolution approach that incorporates a differentiable renderer to super-resolve texture maps in volumetric video sequences. The differential renderer enables us to directly address the problem of super-resolution in texture space instead of image space. Our training process involves pixel-wise losses, including content loss based on a pre-trained VGG19 network and adversarial loss, to guide the generator in learning the mapping between low-resolution and high-resolution textures. Experimental results have demonstrated significant improvements in texture quality, capturing fine details and realistic representation of clothing, hair, and facial features. Our work contributes to advancements in texture super-resolution and holds promise for applications in virtual reality, gaming, and computer animation. Future work includes leveraging the similarity of consecutive volumetric sequence frames as is common in video super-resolution. Furthermore, different super-resolution base architectures will be explored, and the ability to generalize to new meshes will be studied.

ACKNOWLEDGEMENTS

This work has been funded by the German Federal Ministry of Education and Research (BMBF) through the Research Program MoDL under contract no. 01 IS 20044.

REFERENCES

Ancuti, C., Ancuti, C. O., and Timofte, R. (2018). Ntire 2018 challenge on image dehazing: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Salt Lake City, US.

Cao, J., Liang, J., Zhang, K., Li, Y., Zhang, Y., Wang, W., and Gool, L. V. (2022). Reference-based image super-resolution with deformable attention transformer. In *Proc. European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel.

Cheng, R., Sun, Y., Yan, B., Tan, W., and Ma, C. (2022). Geometry-aware reference synthesis for multi-view image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 6083–6093, New York, NY, USA. Association for Computing Machinery.

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). Learning a deep convolutional network for image super-

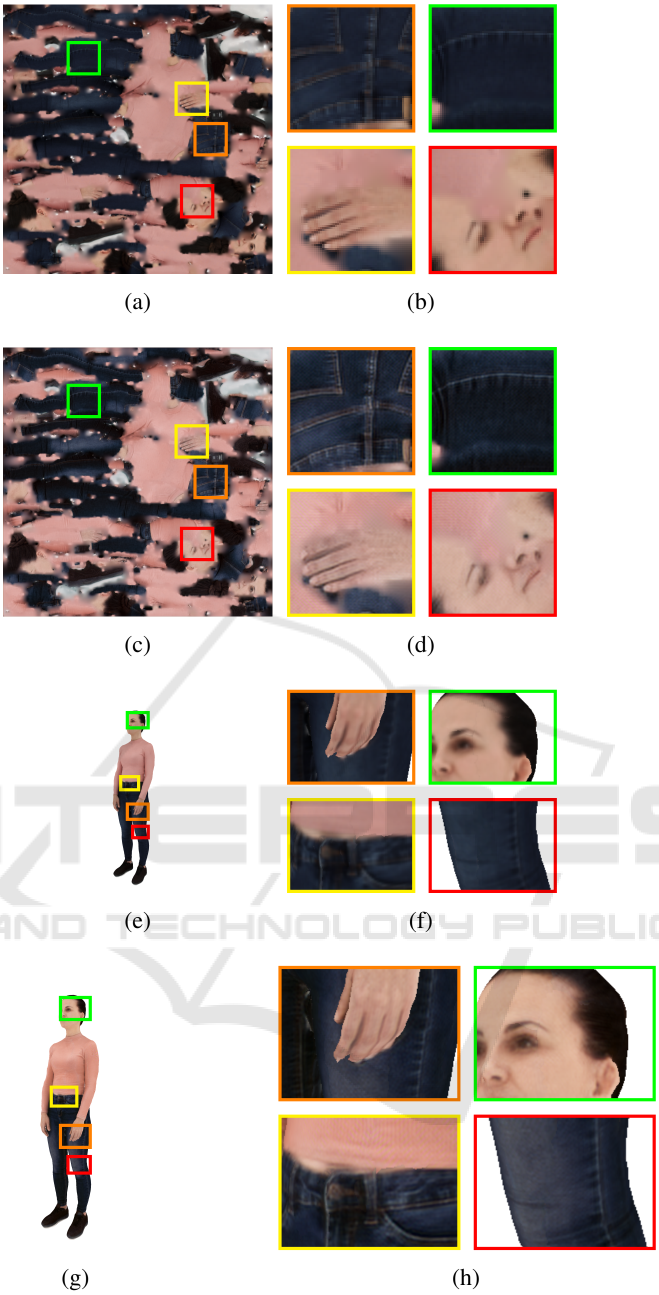

Figure 5: (a) Low-resolution texture, (c) generated high-resolution texture, (e) rendering with low-resolution texture, (g) rendering with generated high-resolution texture. (b), (d), (f), (h) depict color-coded zoomed patches.

resolution. In *European conference on computer vision*. Springer.

Eisert, P., Schreer, O., Feldmann, I., Hellge, C., and Hilsmann, A. (2022). Volumetric video - acquisition, interaction, streaming and rendering. In *Immersive Video Technologies*, pages 289–326. Academic Press.

Glasner, D., Bagon, S., and Irani, M. (2009). Super-resolution from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 349–356.

Goldluecke, B. and Cremers, D. (2009). Superresolution texture maps for multiview reconstruction. In *IEEE 12th Int. Conference on Computer Vision*. IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proc. IEEE Conference on computer vision and pattern recognition*.

Hilsmann, A., Fechteler, P., Morgenstern, W., Paier, W., Feldmann, I., Schreer, O., and Eisert, P. (2020). Going beyond free viewpoint: creating animatable volumetric video of human performances. *IET Computer Vision*, 14:350–358(8).

Insafutdinov, E. and Dosovitskiy, A. (2018). Unsupervised learning of shape and pose with differentiable point clouds. *Advances in neural information processing systems*, 31.

Johnson, J., Ravi, N., Reizenstein, J.and Novotny, D., Tulsiani, S., Lassner, C., and Branson, S. (2020). Accelerating 3D deep learning with pytorch3d. *SIGGRAPH Asia*.

Kato, H., Ushiku, Y., and Harada, T. (2018). Neural 3D mesh renderer. In *Proc. IEEE conference on computer vision and pattern recognition*.

Lai, W.-S., Huang, J.-B., Ahuja, N., and Yang, M.-H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proc. IEEE conference on computer vision and pattern recognition*.

Laine, S., Hellsten, J., Karras, T., Seol, Y., Lehtinen, J., and Aila, T. (2020). Modular primitives for high-performance differentiable rendering. *tog*, 39(6).

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Johannes, T., Wang, Z., and Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proc. IEEE conference on computer vision and pattern recognition*.

Li, T.-M., Aittala, M., Durand, F., and Lehtinen, J. (2018). Differentiable monte carlo ray tracing through edge sampling. *tog*, 37(6).

Li, Y., Tsiminaki, V., Timofte, R., Pollefeys, M., and Gool, L. V. (2019). 3D appearance super-resolution with deep learning. *Proc. IEEE/CVF conference on computer vision and pattern recognition*.

Liu, S., Li, T., Chen, W., and Li, H. (2019). Soft rasterizer: A differentiable renderer for image-based 3D reasoning. In *Proc. IEEE/CVF Int. Conference on Computer Vision*.

Loper, M. M. and Black, M. J. (2014). OpenDR: An approximate differentiable renderer. *European Conference on Computer Vision*.

Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., and Zeng, T. (2022). Transformer for single image super-resolution. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Maier, R., Kim, K., Cremers, D., Kautz, J., and Nießner, M. (2017). Intrinsic3D: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *Proc. IEEE Int. conference on computer vision*.

Nimier-David, M., Vicini, D., Zeltner, T., and Jakob, W. (2019). Mitsuba 2: A retargetable forward and inverse renderer. *tog*, 38(6).

Paier, W., Hilsmann, A., and Eisert, P. (2023). Unsupervised learning of style-aware facial animation from real acting performances. *Graphical Models*, 129:101199.

Pesavento, M., Volino, M., and Hilton, A. (2021). Super-resolution appearance transfer for 4d human performancesmore realistic rendering results yet at a higher computing cost are provided by differentiable ray tracing methods . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1791–1801.

Ranade, R., Liang, Y., Wang, S., Bai, D., and Lee, J. (2022). 3D texture super resolution via the rendering loss. In *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Richard, A., Cherabier, I., Oswald, M. R., Tsiminaki, V., Pollefeys, M., and Schindler, K. (2019). Learned multi-view texture super-resolution. *Proc. Int. Conference on 3D Vision (3DV)*.

Schreer, O., Feldmann, I., Renault, S., Zepp, M., Worchel, M., Eisert, P., and Kauff, P. (2019). Capture and 3D video processing of volumetric video. In *Proc. International Conference on Image Processing*.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proc. IEEE conference on computer vision and pattern recognition*.

Tsiminaki, V., Dong, W., Oswald, M. R., and Pollefeys, M. (2019). Joint multi-view texture super-resolution and intrinsic decomposition. In *30th British Machine Vision Conference (BMVC 2019)*.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proc. European conference on computer vision (ECCV) workshops*.

Wang, Z., Chen, J., and Hoi, S. C. H. (2021a). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10).

Wang, Z., Chen, J., and Hoi, S. C. H. (2021b). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10).

Yang, C.-Y., Ma, C., and Yang, M.-H. (2014). Single-image super-resolution: A benchmark. In *European conference on computer vision*. Springer.