frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 19 January 2022
doi: 10.3389/frvir.2021.785104

Study on Automatic 3D Facial Caricaturization: From Rules to Deep Learning

Nicolas Olivier[1,2]*, *Glenn Kerbiriou*[1,3]*, *Ferran Arguelaguet*[2], *Quentin Avril*[1], *Fabien Danieau*[1], *Philippe Guillotel*[1], *Ludovic Hoyet*[2] and *Franck Multon*[2,4]

[1]InterDigital, Cesson-Sévigné, France, [2]Inria, Univ Rennes, CNRS, IRISA, Rennes, France, [3]Institut national des sciences appliquées de Rennes, Rennes, France, [4]Laboratoire Mouvement, Sport, Santé (M2S), Bruz, France

Facial caricature is the art of drawing faces in an exaggerated way to convey emotions such as humor or sarcasm. Automatic caricaturization has been explored both in the 2D and 3D domain. In this paper, we propose two novel approaches to automatically caricaturize input facial scans, filling gaps in the literature in terms of user-control, caricature style transfer, and exploring the use of deep learning for 3D mesh caricaturization. The first approach is a gradient-based differential deformation approach with data driven stylization. It is a combination of two deformation processes: facial curvature and proportions exaggeration. The second approach is a GAN for unpaired face-scan-to-3D-caricature translation. We leverage existing facial and caricature datasets, along with recent domain-to-domain translation methods and 3D convolutional operators, to learn to caricaturize 3D facial scans in an unsupervised way. To evaluate and compare these two novel approaches with the state of the art, we conducted the first user study of facial mesh caricaturization techniques, with 49 participants. It highlights the subjectivity of the caricature perception and the complementarity of the methods. Finally, we provide insights for automatically generating caricaturized 3D facial mesh.

Keywords: caricature, style transfer, machine learning, geometry processing, 3D mesh, perceptual study

Edited by:
Ruofei Du,
Google, United States

Reviewed by:
Daisuke Iwai,
Osaka University, Japan
John Dingliana,
Trinity College Dublin, Ireland

***Correspondence:**
Nicolas Olivier
Nicolas.Olivier@InterDigital.com
Glenn Kerbiriou
Glenn.Kerbiriou@InterDigital.com

Specialty section:
This article was submitted to Technologies for VR, a section of the journal Frontiers in Virtual Reality

Received: 28 September 2021
Accepted: 06 December 2021
Published: 19 January 2022

Citation:
Olivier N, Kerbiriou G, Arguelaguet F, Avril Q, Danieau F, Guillotel P, Hoyet L and Multon F (2022) Study on Automatic 3D Facial Caricaturization: From Rules to Deep Learning. Front. Virtual Real. 2:785104. doi: 10.3389/frvir.2021.785104

1 INTRODUCTION

Caricatures have been used for centuries to convey humor or sarcasm. References can be found during the Antiquity with Aristotle referring to these artists as "grotesque," or in the works of Leonardo Da Vinci who was eagerly looking for people with deformities to use as models. Caricature can be defined as the art of drawing persons (usually faces) in a simplified or exaggerated way through sketching, pencil strokes, or other artistic drawings. Caricatures have been commonly used to entertain people, to laugh at politics or as a gift or souvenir sketched by street artists. These artists have the ability to capture distinct facial features, and then exaggerate those features (Redman, 1984). With the development of social VR networks or games, users may wish to use stylized avatars, including avatars preserving their identity (Olivier et al., 2020) but with such exaggerated features. Hence, automatically generating such caricatured avatars becomes a key issue, as having artists manually creating caricatured avatars would not be feasible for such applications involving large numbers of users. Let us consider a 3D mesh representing the user's face (either using 3D scanning or computer vision methods to build 3D shape from a minimum set of images). An automatic caricature system should maintain the

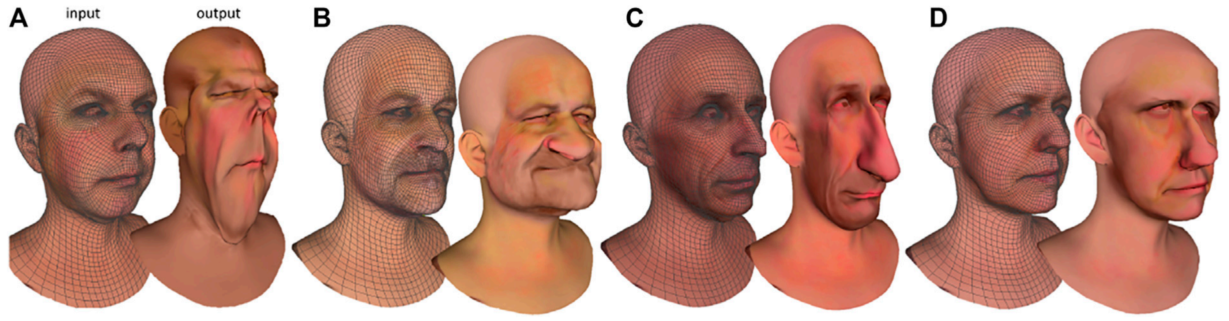

FIGURE 1 | Results of our novel user-controlled rule-based approach. Each pair **(A, B, C, and D)** presents the input facial scan (wired on the left) and its automatically generated caricature on the right.

relative geometric location of facial components, while emphasizing the subject's facial features distinct from others. While different caricature experts would generate different styles of faces (more or less cartoonish style for example), they would all be exaggerating facial traits of the individual (Brennan, 1985; Liang et al., 2002; Mo et al., 2004). The ability of creating a variety of plausible caricatures for each single face is therefore a key challenge when automatically generating caricatures, as different artists would create visually different caricatures, which should also be taken into account when evaluating the subjective quality of the results.

Previous works for the generation of 3D caricatures can be separated into two main families: interactive and automatic methods. Interactive methods offer tools to caricature experts to design the resulting caricature (Akleman, 1997; Akleman et al., 2000; Chen et al., 2002; Gooch et al., 2004), while fully automatic methods use hand-crafted rules (Brennan, 1985; Liang et al., 2002; Mo et al., 2004), often derived from the drawing procedure of artists. However, these approaches are typically restricted to a particular artistic style, e.g., sketch or a certain cartoon, and predefined templates of exaggeration. From the works in the literature in other domains, two different solutions could be envisioned to automatically generate caricatures. First, in the context of exaggerating distinct features, Sela et al. (2015) proposed a generic method to exaggerate the differences between the 3D scan of an object and an average template model of such type of object. However, this method has never been formally evaluated for human faces. Second, deep learning methods could be considered. As mentioned above, automatic methods mainly use hand-crafted rules that may fail to capture some complex choices made by caricature experts. In contrast, generative adversarial networks (GANs) are a promising mean to attempt to learn these choices based on a set of examples made by experts, without being limited to hand-crafted rules, but it has been never applied for the generation of 3D caricatures. The main goal of this paper is to propose and evaluate novel methods for the automatic generation of 3D caricatures from real 3D facial scans, first with a rule-based method, in order to keep tunable and interpretable parameters, and a deep learning method, to leverage real caricature data and hence generate caricatures closer to real ones. The main hypotheses we wish to address in this paper are:

H1: the specialization of generic exaggeration methods for human faces should allow to produce convincing caricatures. To this end, we adapted the generic method proposed by Sela et al. (2015) in order to generate caricatures by exaggerating facial features from a 3D face scan (see **Figure 1**). This method has two main stages, one based on a curvature EDFM (Exaggerating the Difference From the Mean), and another based on a nearest-neighbors search in a 3D caricature dataset, to apply the proportion exaggeration.

H2: deep learning should allow to overcome some of the limitations of rule-based methods by their ability to generalize based on a set of examples. Thus, we designed a method leveraging advances in the field of GAN-based style transfer, which has shown great success in the 2D domain, for instance on drawn caricatures (Cao et al., 2019).

H3: both methods should reach and overcome the state-of-the-art results when trying to automatically generate caricatures from a human face 3D scan. To assess the advantages and disadvantages of the proposed methods, we conducted a perceptual study considering the base method proposed by Sela et al. (2015) and an additional EDFM method (Akleman and Reisch, 2004).

The results of the study support hypotheses H1 and H2, as the perceptual study demonstrated no significant preference of the subjects for any of the tested methods, for the proposed human faces. Although this result shows that the two proposed methods reached state of the art performance (H3), the perceptual study did not show a clear winner, highlighting the difficulty to simulate and evaluate such artistic caricatures for which a large variety of styles and solutions exists. The remainder of the paper is structured as follows. First, **Section 2** reviews the state of the art, and identifies the gaps between existing techniques. **Section 3** and **Section 4** present the proposed rule-based and deep learning-based caricaturization methods respectively. Then, **Section 5** presents the perceptual evaluation of the proposed methods with state-of-the-art methods. Finally, we discuss the results and provide insights on the automatic caricature generation in **Section 6**.

2 RELATED WORK

Computer assisted caricature generation has been a topic of interest for researchers since the beginning of Computer Graphics (Brennan, 1985). Typically, techniques from drawing guides, such as Redman's practical guide 1984) on how to draw caricatures, are exploited. This guide sets the fundamental rules of caricatures and proposes some concepts that are massively used. Among them, the "mean face assumption" implies the existence of an average face, and the process of "Exaggerating the Difference From the Mean" (EDFM) consists in emphasizing the features that make a person unique, i.e., different from the average face. Existing methods for automatic caricature generation split into two main categories: rule-based and learning-based methods.

2.1 Rule-Based Methods

Rule-based methods use *a priori* known procedures to caricaturize a shape. They can be further divided into two branches depending if their domain of application is on human faces or other shapes.

Face rule-based methods follow caricature drawing guidelines (e.g., EDFM) to generate deformed faces with emphasized features. Brennan (1985) first proposed an implementation of EDFM in two dimensions. They built an interactive system where a user can select facial feature points which are matched against the average feature points, then the distance between them is exaggerated. This algorithm was later extended by Akleman *et al.* in 2D and 3D domains (Akleman, 1997; Akleman and Reisch, 2004). Their software relies on a low-level procedure which requires the user to decide whether the exaggeration of a feature increases likeness or not. In the same spirit, Fujiwara et al. (2002) developed a piece of software named PICASSO for automatic 3D caricature generation. They used a set of feature points to generate simplified 3D faces before performing EDFM. EDFM was also used by Blanz and Vetter (1999) in an application example of their morphable model. They learn a principal component analysis (PCA) space from 200 3D textured faces. Their system allows caricature generation by increasing the distance to the statistical mean in terms of geometry and texture. Statistical dispersion has been taken into account by Mo et al. (2004) who showed that features should be emphasized proportionally to their standard deviation to preserve likeness. Chen et al. (2006) created 3D caricatures by fusing 2D caricatures generated using EDFM from different views. Redman's guide (Redman, 1984) not only introduces EDFM but also high levels concepts such as the five head types (oval, triangular, squared, round and long) and the dissociation between local and global exaggeration. These concepts have been exploited by Liu et al. (2012) to perform photo to 3D caricature translation. They applied EDFM with respect to the shape of the head (global scale) and to the distance ratios of a set of feature points (local scale). Face rule-based methods can generate a caricature from an input photograph or a 3D model but fail at reproducing artistic styles. Different caricaturists would make different caricatures from the same person. To avoid this issue, they usually provide user control at a relatively low-level of comprehension, which often requires artistic knowledge.

Non face specific rule-based methods rely on intrinsic or extracted features of geometrical shapes. They generalize the concept of caricature beyond the domain of human faces. Eigensatz et al. (2008) developed a 3D shape editing technique based on principal curvatures manipulation. With no reference model, their method can enhance or reduce the sharpness of a 3D shape. The link between saliency and caricature has been explored by Cimen et al. (2012). They introduced a perceptual method for caricaturing 3D shapes based on their saliency using free form deformation technique. A computational approach for surface caricaturization has been presented by Sela et al. (2015). They locally scale the gradient field of a mesh by its absolute Gaussian curvature. A reference mesh can be provided to follow the EDFM rule, and the authors show that their method is invariant to isometries, i.e., invariant to poses. General shape rule-based methods can also caricature a 2D or 3D shape without any reference model. As they do not take into account any statistical information nor the concept of artistic style, they try to link low-level geometry information to high-level caricature concepts, e.g., the fact that the most salient area should be more exaggerated (Cimen et al., 2012). As a result, they do not take into account the semantic of faces nor the art of human face caricature.

Since this work only tackles human face caricaturization, we refer to "face rule-based methods" as simply "rule-based methods".

2.2 Learning Based Methods

Existing learning-based methods for caricature generation can use both paired and unpaired data as training material.

Supervised data-driven methods would automatically find rules by relying on pairs of exemplars to learn a mapping between the domain of normal faces and the domain of caricatures. Xie et al. (2009) proposed a framework that learns a PCA model over 3D caricatures and a Locally Linear Embedding (LLE) model over 2D caricatures, both made by artists. The user can manually create a deformation that is projected into the PCA subspace and refined using the LLE model. Li et al. (2008) and Liu et al. (2009) both focused on learning a mapping between the LLE representation of photographs and their corresponding LLE representation of 3D caricatures modeled by artists. In the same vein, but only in the 3D domain, Zhou et al. (2016) regressed a set of locally linear mappings from sparse exemplars of 3D faces and their corresponding 3D caricature. As far as we know, Clarke et al. (2011) are the only authors that proposed a physics-oriented caricature method. They capture the artistic style of 2D caricatures by learning a pseudo stress-strain model which describes physical properties of virtual materials. All these data-driven approaches are based on paired datasets which require the work of 2D or 3D artists. Such datasets are costly to produce, therefore techniques of this kind are hardly applicable.

Unsupervised learning based methods learn how to caricature from unpaired face and caricature exemplars. Chen et al. (2001) and Liang et al. (2002) generated 2D caricatures by learning a nonlinear mapping between photos and corresponding caricatures made by artists. Derived from the image synthesis literature, where they have been used for unpaired one-to-one translation (Liu et al., 2017; Taigman et al., 2017; Yi et al., 2017; Zhu et al., 2017), or unpaired many-to-many translation (Huang et al., 2018b; Liu et al., 2019; Choi et al., 2020), Generative Adversarial Networks (GANs) have also shown impressive results on mesh synthesis and mesh-to-mesh translation (Goodfellow et al., 2014). Other approaches achieve 2D stylization using 3D priors and a differentiable renderer (Wang et al. (2021).) Cao et al. (2019) proposed a photo to 2D caricature translation framework CariGANs based on a large dataset of over 6,000 labeled 2D caricatures (Huo et al., 2018), and two GANs, namely CariGeoGAN for geometry exaggeration using landmark warping, and CariStyGAN for stylization. CariStyGAN allows to use a reference graphic style, or else, it will generate a random style. This framework was first extended by Shi et al. (2019) with a feature point-based warping for geometric exaggeration, then by Gu et al. (2021) which provides a random set of deformation styles in addition to the random set of graphics styles, offering consequent user control. In the 3D domain, Wu et al. (2018) then Cai et al. (2021) proposed robust methods for 3D caricature reconstruction from meshes, enlarging the set of available in-the-wild 3D caricatures, when used in combination with WebCaricature (Huo et al., 2018). Guo et al. (2019) showed an approach for producing expressive 3D caricatures from photos using a VAE-CycleGAN. Ye et al. (2020) proposed an end-to-end 3D caricature generation from photos method, using a GAN-based architecture with two symmetrical generators and discriminators. A step of texture stylization is performed with CariStyGAN. The recent works for caricature generation in 3D domain allow to reproduce the style of artists but they do not feature much user control. Ye et al. (2020) introduced Facial Shape Vectors so the user can choose the facial proportions on the caricature, but this is a quite low-level interaction and thus should be done by an artist. These works also show a weakness from the use of CariStyGAN for texture stylization. CariStyGAN tends to emphasize the shadows and light spots of the photos in order to make the reliefs sharper. In the case of textured 3D models, the shadows and light spots should be induced by the geometry and the lighting conditions, not by the texture albedo. If lighting information is entangled within texture information, changing the lighting condition can make the 3D model appear to be enlightened by non-existent lights.

Adopting a 3D mesh representation requires application of mesh convolutions defined on non-Euclidean domains (i.e., geometric deep learning methodologies). Over the past few years, the field of geometric deep learning has received significant attention (Litany et al., 2017; Maron et al., 2017). Methods relevant to this paper are auto-encoder structures such as used by Ranjan et al. (2017) and Gong et al. (2019), that showcase the efficiency of recent 3D convolutional operators at capturing the distribution of 3D facial meshes. Several approaches resort to mapping 3D faces to a 2D domain, and using 2D convolution operators (Moschoglou et al., 2020). Projecting a 3D surface to a 2D plane for 2D convolutions requires locally deforming distances, which translates to higher computing and memory costs compared to recent 3D convolution approaches, and some high-frequency information loss (Gong et al., 2019).

Deep learning based approaches, leveraging recent advancements in the field, could produce caricatures more similar to the kind produced by professionals, and allow global style control using handmade caricatures as style examples. On the opposite side, a user-controlled rule-based approach enabling a local control of the facial mesh deformation would allow for fine-tuned local control. We develop both approaches in **Section 3** and **Section 4**. Finally, there is no overall perception user study of this specific field, limiting any qualitative comparison between approaches. We present the first study of this kind in **Section 5**, in order to evaluate the strengths and drawbacks of these two novel methods in comparison to two state-of-the-art approaches.

3 RULE-BASED USER-CONTROLLED CARICATURIZATION

We present a novel method featuring short computation time and providing meaningful user control over the generated caricatures. It is based on two main modules depicted in **Figure 2** (in green and in yellow). First, a curvature exaggeration module (in green) enhances the facial lines by applying EDFM technique to the main PCA scores of the mesh gradients of the input face. This emphasizes only the 3D surface details such as ridges, peaks, and folds, and does not affect the global shape of the face (such as eyes, nose, and mouth relative positions). Second, a proportion exaggeration module (in yellow) leverages compositions of real artists (see Section 3.1) to caricature the general shape of the face. It projects the input face into a 3D caricature shape space thanks to a *k*NN regressor. This process applies a smooth and large scale deformation to the input face while preserving its local features. The curvature exaggeration and proportion exaggeration modules are thus complementary. They are combined to provide the user with a bilateral control (small scale versus large scale) over the resulting caricature. Lastly, an optional texture blurring and contrast enhancement module (in pink) makes the resulting caricature less realistic and more graphic. The reason behind this step is to make the result more acceptable for human observers. As shown by Zell et al. (2015), we use texture blurring because it increases the appeal and lowers the eeriness of a virtual character. The increase in contrast is meant to make the caricatures less realistic, but one could have used another technique to this end. In addition to these modules, our user-controlled method features semantic mesh segmentation in four regions (see Section 3.2). In total, the method exposes ten knobs to the user.

3.1 Datasets

Realistic 3D faces were sampled from the LSFM dataset (Booth et al., 2016) which contains nearly 10k distinct 3D faces. In order to have textured meshes, we completed this set with 300 in-house 3D face scans. Their topologies are unified through automatic facial landmarking and geometry fitting (Danieau et al., 2019). To build our 3D caricatured mesh dataset, we run the 2D to 3D

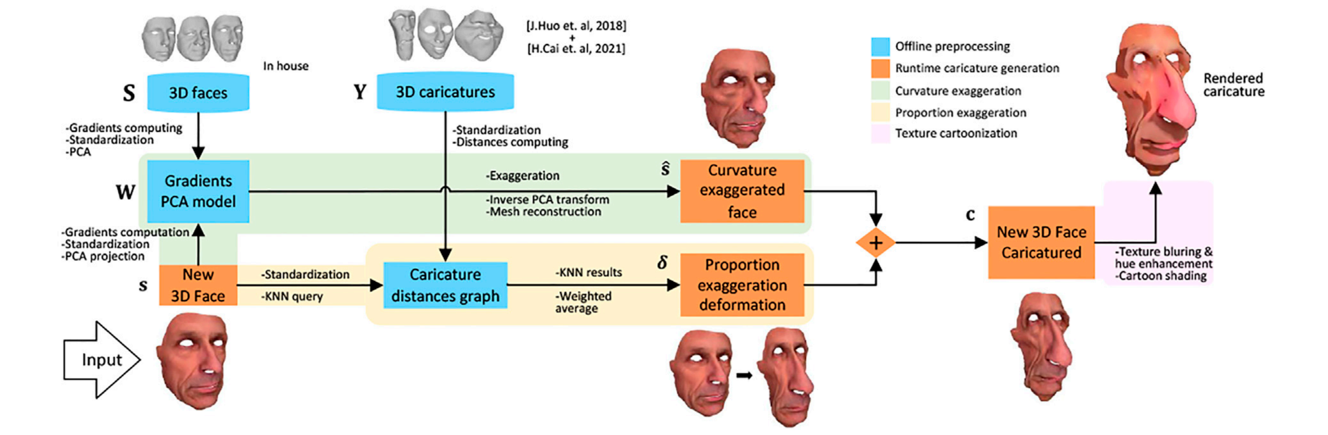

FIGURE 2 | Overview of our user-controlled method presented in **Section 3**. Arrows and diamond shapes represent algorithms while boxes represent data. Offline and online processing are represented by the blue and orange colors, respectively. Green, yellow, and pink highlights show the different modules which compose the core of the user-controlled caricature system. For simplification purposes, the face segmentation is not shown.

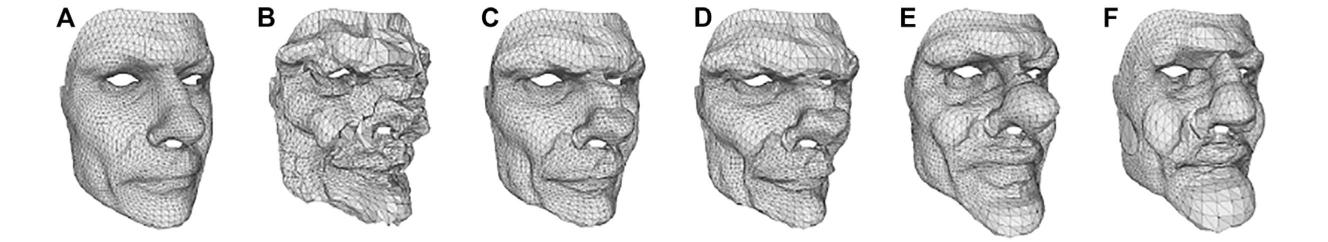

FIGURE 3 | Different curvature exaggeration techniques: **(A)** Original 3D Mesh. **(B)** Naive gradient EDFM without segmentation ($f_{grad} = 5$). **(C)** Gradient EDFM with PCA denoising, without segmentation and **(D)** with segmentation ($f_{grad} = 5$). **(E)** Sela et al. (2015)'s method, without reference model ($\gamma = 0.3$) and **(F)** with the mean face as reference model ($\beta = 4$). More examples in Supplementary Material, at Supplementary Figure 23.

caricature inference method of Cai et al. (2021) on the WebCaricature dataset (Huo et al., 2018), which enables to extract the 3D caricatured face mesh from each 2D image. The WebCaricature dataset contains over 6k 2D caricatures. When Cai's algorithm did not successfully estimate the faces, due to extreme drawing composition (quick sketch, incomplete drawings, drafts, cubism etc.) the generated output remains the same default caricature mesh. All faces were then registered, in order to have a fixed topology (Sumner and Popović, 2004).

3.2 Facial Segmentation

In face modeling, cartoonization and caricaturing, semantic segmentation is a popular technique for increasing expressivity and user interaction (Blanz and Vetter, 1999; Liu et al., 2009; Zhou et al., 2016). In the proposed system, the 3D faces are segmented using the scheme proposed by Blanz and Vetter (1999) *i.e.* in four regions: the eyes, the nose, the mouth, and the rest of the face. This semantic segmentation allows the user to choose whether to emphasize or not a facial part. In total, the method exposes ten knobs to the user: one scalar is used for the strength of the gradient EDFM and another one for the amount of deformation from the *k*NN regressor to be added. Those two weights are tunable for each of the five regions (four masks and full face). Segmenting the domain also allows to break the inherent linearity of PCA by learning different subspaces.

3.3 Curvature Exaggeration

To emphasize the small scale features of the input 3D face, the curvature exaggeration module performs EDFM on the mesh gradient. In the process, we use PCA as a mean to reduce high frequencies (**Figure 3**).

- Offline preprocessing. The edge-based gradient operator **E** (see **Supplementary Material**) is used to compute the gradients **g** of each face mesh **s** of our custom 3D face dataset (Section 3.1). Following the results of Mo et al. (2004) showing that low-variance features should be more taken into account, the gradients **G** are standardized: $\mathbf{G}^{std} = \frac{\mathbf{G}-\bar{\mathbf{g}}}{\sigma_{\mathbf{G}}}$. Then, a PCA is performed on the standardized gradients leading to the principal components **W** and each PCA scores **t** such that $\mathbf{t} = \mathbf{g}^{std} \cdot \mathbf{W}$.
- Runtime curvature exaggeration. The input face mesh **s** is standardized then projected into the PCA space learnt offline. EDFM technique is applied with a factor f_{grad} given by the user. To prevent noise, we weight the result by the normalized standard

deviation associated to each principal component $\boldsymbol{\sigma} = \sqrt{\frac{\lambda_{\mathrm{C}}}{\max(\lambda_{\mathrm{C}})}}$. The exaggerated PCA scores are obtained as $\hat{\mathbf{t}} = \mathbf{t} \cdot \max(f_{\mathrm{grad}} \cdot \boldsymbol{\sigma}, 1)$. The exaggerated gradient is then recovered as $\hat{\mathbf{g}} = \bar{\mathbf{g}} + \boldsymbol{\sigma}_{\mathbf{G}} \cdot (\hat{\mathbf{t}} \cdot \mathbf{W}^{\mathrm{T}})$. The gradients' exaggerated mesh $\hat{\mathbf{s}}$ is eventually reconstructed at the least squares sense by setting the border vertices fixed (the border of the eyes, the nostrils, the inner lips, and the contour of the head), as described in **Supplementary Material**.

3.4 Proportion Exaggeration

The proportion exaggeration module leverages the 3D caricatures (see Section 3.1) to sample a deformation that matches the input face difference from the mean using a *k*NN. Thus, it can be seen as an example-based version of EDFM. We argue that the sampled deformation contains mainly low frequencies and adding it to an input face will modify very little its surface curvatures. We observed that the 3D caricatures have more diverse global shapes than our 3D faces while being much smoother. In addition, the *k*NN regression also contributes to smooth out the deformation by averaging the *k* nearest neighbors. The process works as follows:

- Offline preprocessing. The 3D caricatures are first standardized using the standard deviation of our 3D faces to make the low-variance areas more important (Mo et al., 2004). Then, we fit a *k*NN regressor using a cosine distance metric, as we mainly seek to find directions of deformation rather than amplitudes of deformation. The amplitude tuning is reserved for the user.
- Runtime proportion exaggeration. The input face is standardized then projected into the 3D caricature space with the *k*NN regressor using barycentric weights. The obtained deformation $\boldsymbol{\delta}^{\mathrm{std}}$ is weighted by the 3D face standard deviation $\boldsymbol{\sigma}_{\mathbf{S}}$ and by a user-defined scalar f_{prop} for amplitude tuning. Eventually, we add this deformation to the curvature exaggerated face to get the vertex positions of the resulting caricature **c**:

$$\mathbf{c} = \hat{\mathbf{s}} + \boldsymbol{\delta} \qquad \text{with} \qquad \boldsymbol{\delta} = f_{\mathrm{prop}} \cdot \boldsymbol{\delta}^{\mathrm{std}} \cdot \boldsymbol{\sigma}_{\mathbf{S}} \tag{1}$$

3.5 Results

In this section, the results of both the curvature exaggeration module and the proportion exaggeration module are presented and compared to those of their most similar existing approaches. We compare the curvature exaggeration module to Sela et al. (2015) because they fix the positions of border vertices and therefore tend to preserve the proportions of the caricatured faces. Our proportion exaggeration module is compared to the baseline 3D position EDFM introduced in the seminal work of Blanz and Vetter (1999).

- Curvature exaggeration module. The benefit of the PCA-based denoising mechanism is visible in **Figure 3** between column b), and column c) and d). Without PCA, the EDFM technique magnifies the existing high frequencies of the face's difference from the mean. With PCA, the noise is removed but the exaggeration of facial lines remains. The use of a segmented model not only enables to provide more user-control, but also to emphasize the curvatures more locally. This effect can be noticed when comparing the results c) and d) in **Figure 3**. Sela et al. (2015)'s method successfully preserves the position of the eyes, the nostrils, the inner lips and the contour of the face. However other parts such as the nose, the lips and the chin seem greatly inflated and displaced which should not belong to facial lines enhancement. Conversely, our curvature exaggeration module modifies the vertex positions such that it only enhances the fine curvature details.
- Proportion exaggeration module. **Figure 4** shows the effect of modifying *k* on the results of our proportion exaggeration module. Visually, the parameter *k* of the *k*NN regressor has less impact than we expected. However, it appears that a small value of k (≤ 5) tends to introduce high-frequencies and vertex entanglement while larges values of k (≥ 1000) seem to produce less vivid results. We fixed $k = 40$ in our experiments.

The semantic segmentation has also an impact on our proportion exaggeration module. In **Figure 5**, the results with segmentation (column c) seem more caricatural but also more expressive than without segmentation (column b). Expressiveness is not intended by the proposed method since the focus is on neutral expression caricature generation. Nevertheless, we decided to conserve the segmentation scheme for the proportion exaggeration module. We also compare the proportion exaggeration algorithm to the baseline PCA-based EDFM on 3D coordinates proposed by Blanz and Vetter (1999) (column d). Our method clearly generates more diverse and inhomogeneous shapes than Blanz and Vetter (1999)'s approach. It is also noticeable that less high-frequency details are added than with the baseline method, which is what we aim at.

4 DEEP LEARNING BASED AUTOMATIC CARICATURIZATION

Rule-based methods allow the use of controllable and interpretable parameters, but are limited to capture information about caricature styles. Supervised learning based methods require a large paired mesh-to-caricature dataset, that are highly consuming in terms of both time and means to build. Instead, we consider the case of an unpaired learning-based approach, taking advantage of our 3D datasets of both neutral and caricatured faces (Cai et al., 2021) (cf. Section 3.1). Our network architecture is based on the shared content space assumption of Liu et al. (2019), that we adapt to the context of 3D data through the use of 3D convolutions of Gong et al. (2019), which define 3D convolution neighborhoods.

4.1 Framework Overview

Let us consider meshes of different styles (e.g. scans and caricatures), all sharing the same mesh topology. We represents our faces with raw 3D coordinates, and encode them using a recent 3D convolutional operator (Gong et al., 2019). Given a mesh $x \in X$ and an arbitrary style $y \in Y$, our goal is to train a single generator G that can generate diverse meshes of

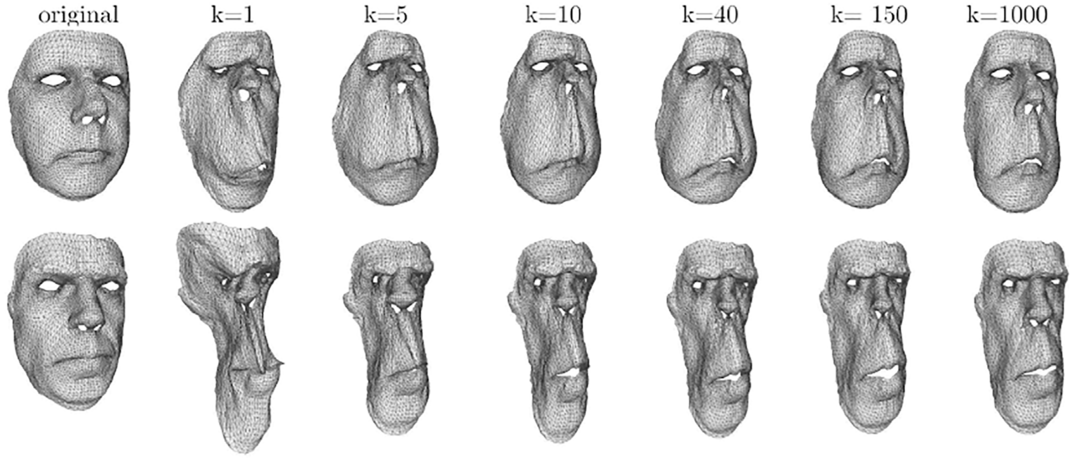

FIGURE 4 | A comparison of results with different values of *k* for the *k*NN algorithm of the proportion exaggeration module. The first column shows the original facial mesh. Here, the caricatures are generated with $f_{proportions} = 2$.

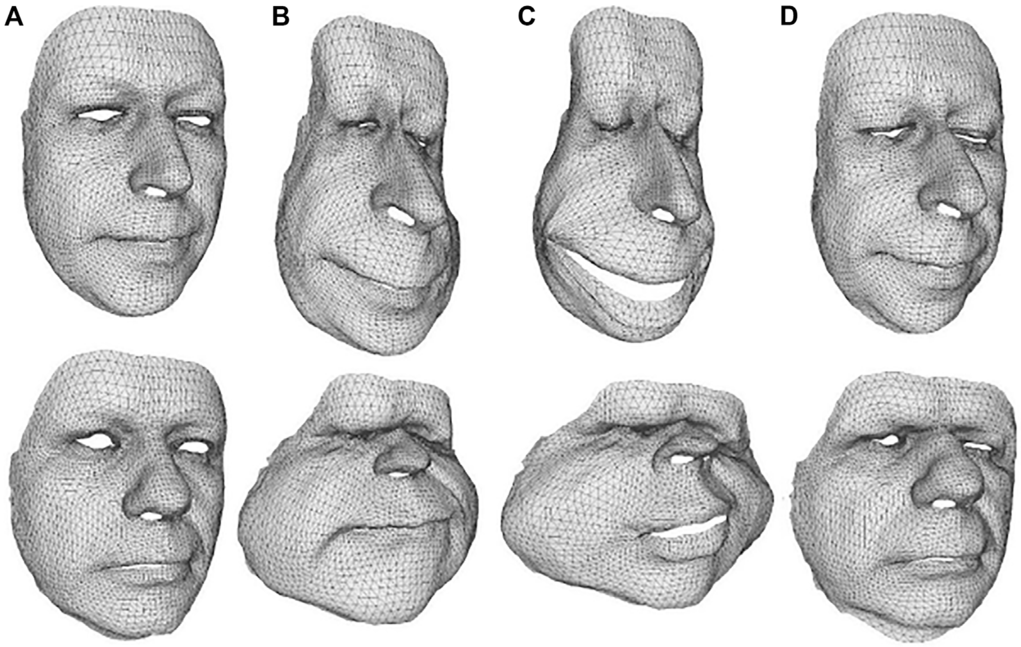

FIGURE 5 | A comparison between proportion exaggeration techniques on two facial meshes. **(A)** Original facial mesh. **(B)** Our proportion exaggeration algorithm without segmentation and **(C)** with segmentation. **(D)** Baseline PCA-based 3D positions EDFM (Blanz and Vetter, 1999).

each style *y* that corresponds to the mesh *x*. We generate style-specific vectors in the learned space of each style and train *G* to reflect these vectors. **Figure 6** illustrates an overview of our framework, which consists of three modules described below.

Generator. Our generator *G* translates an input mesh *x* into an output mesh *G* (*x*, *s*) reflecting a style-specific style code *s*, which is provided by the style encoder *E*. We use adaptive instance normalization (AdaIN) (Huang and Belongie, 2018a) to inject *s* into *G*. We observe that *s* can represent any style, which removes the necessity of providing *y* to *G* and allows *G* to synthesize meshes of all domains.

Style encoder. Given a mesh *x*, our encoder *E* extracts the style codes *s* = *E*(*x*). Similar to Liu et al. (2019), our style encoder benefits from the multi-task learning setup. E can produce diverse style codes using different reference meshes. This allows *G* to synthesize an output mesh reflecting the style code *s* of a reference mesh *x*.

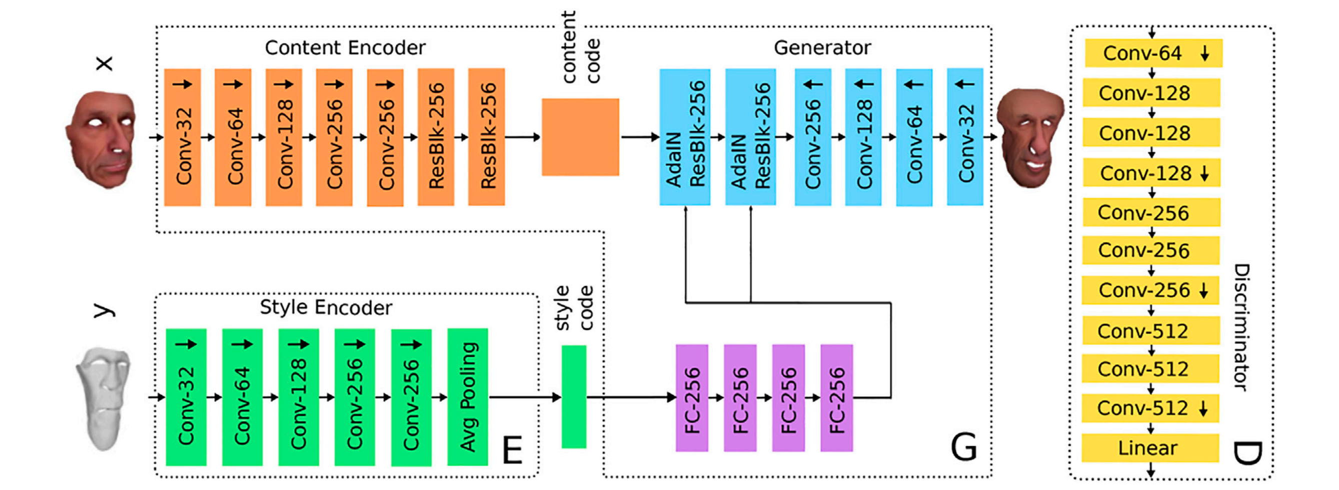

FIGURE 6 | Overview of UNGT. A facial scan's identity is encoded along with the style of a caricature mesh, in order to produce the caricatured face. Textures are not processed, and presented for illustration purpose only. E represent the Style Encoder, G the Generator, and D the Discriminator.

Discriminator. Our discriminator D is a multitask discriminator (Mescheder et al., 2018; Liu et al., 2019; Choi et al., 2020), which consists of multiple output branches. Each branch D_y learns a binary classification determining whether a mesh x is a mesh from the dataset of style y or a fake mesh $G(x, s)$ produced by G.

4.2 Training Objectives

Given a mesh $x \in X$ and its original style $y \in Y$, we train our framework using the following objectives:

- Adversarial objective. During training, we sample a mesh a and generate its style code $s = E(a)$. The generator G takes a mesh x and s as inputs and learns to generate an output mesh $G(x, s)$ that is indistinguishable from real meshes of the style y, via a classical adversarial loss (Arjovsky et al., 2017):

$$L_{adv} = E_{x,y}\left[logD_y(x)\right] + E_{x,\tilde{y}}\left[log\left(1 - D_{\tilde{y}}(G(x,s))\right)\right]$$

where $D_y(\cdot)$ denotes the output of D corresponding to the style y.

- Reconstruction and cycle losses. To guarantee that the generated mesh $G(x, s)$ properly preserves the style-invariant characteristics (e.g. identity) of its input mesh x, we employ the cycle consistency loss (Kim et al., 2017; Zhu et al., 2017; Choi et al., 2018)

$$L_{cyc} = E_{x,y,\tilde{y}}\left[\|x - G(G(x,\tilde{s}),\hat{s})\|^1\right]$$

where $\hat{s} = E_y(x)$ is the estimated style code of the input mesh x, $\tilde{y}$ and $\tilde{s}$ are the style and estimated style codes of another mesh than x. By encouraging the generator G to reconstruct the input mesh x with its estimated style code $\hat{s}$, G learns to preserve the original characteristics of x while changing its style faithfully. In a similar goal of preserving style invariant characteristics, we use a reconstruction loss

$$L_r = E_{x,y}\left[\|x - G(x,\hat{s})\|^1\right]$$

where $\hat{s} = E_y(x)$ is the estimated style code of the input mesh x.

- Full objective. Our objective function can be summarized as follows:

$$\min_{G,F,E} \max_{D}$$
$$L_{adv} + \lambda_{cyc} \cdot L_{cyc} + \lambda_r \cdot L_r$$

where λ_r and λ_{cyc} are hyper parameters for each term. We use the Adam Optimizer (Kingma and Ba, 2015).

4.3 Results

We trained the network for 50k iterations on a Titan X Pascal (4h, 8Go). Results of the approach are visible in **Figure 7**. The original faces (top row) are encoded using the network illustrated in **Figure 6** along with a random caricature of the dataset, producing the caricatured face (bottom row). Facial proportions are hence exaggerated according to the distribution of the neutral and caricatured faces learned during the training stage.

5 USER STUDY

In order to assess the subjective quality of the caricatures generated by the previously described methods, we have conducted a perceptual study. The goal of the perceptual study was to subjectively rank the generated caricatures based on the perceived quality of the caricatures. In addition to the two methods described in **Section 3** and **Section 4**, we also considered two baseline methods, the method from Sela et al. (2015) and a EDFM method (Blanz and Vetter (1999)).

5.1 Participants

Forty-nine participants took part in the experiment (9 females). They were between 18 and 63 years old (mean and STD age:

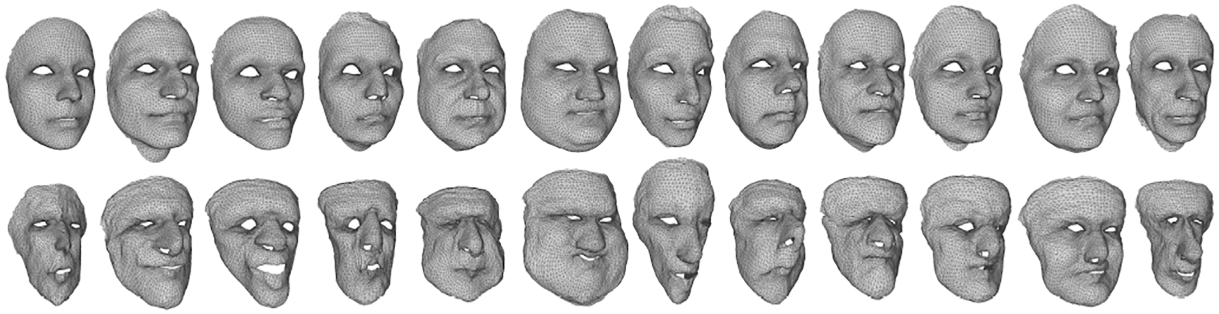

FIGURE 7 | Deep learning based caricatures for a number of facial scan examples.

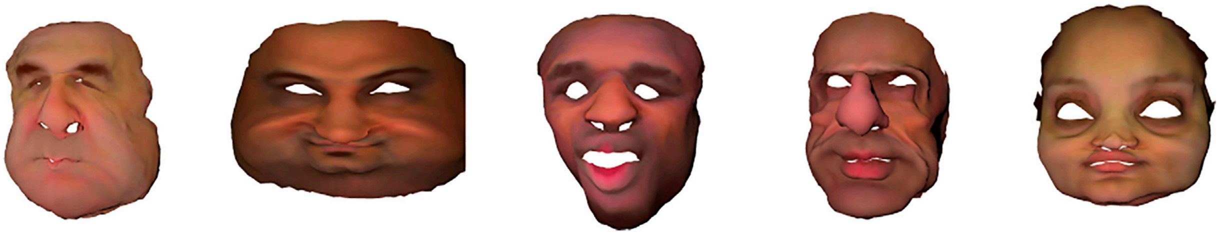

FIGURE 8 | The five best caricatures (with the best mean ranks; identities 7, 6, 9, 12, 2).

TABLE 1 | Parameters sets of the two variations of our rule-based method used in the user study (**Section 5**). The first variation targets the proportions more while the second strongly exaggerates the curvatures. These parameter sets aim at exploring the range of user control provided to the user. A number of other variations could have been proposed, but we meet complexity restrictions for the user study.

	Exaggeration type	Eyes	Nose	Mouth	Rest	Full face
Rule-based 1	curvatures	0.5	0	0	0	4
	proportions	1	0.5	0.75	0	0.75
Rule-based 2	curvatures	0	0	3	2	8
	proportions	0	1.75	0	0	0

31.0 ± 11.3), and were recruited from our laboratory among students and staff. They were all naive to the purpose of the experiment, had normal or correct-to-normal vision, and gave written and informed consent. The study conformed to the declaration of Helsinki. Participants were not compensated for their participation and none of the participants knew the human faces used in the study.

5.2 Stimuli

The top part of **Figure 7** presents the 12 human face scans (Identity factor) used in the study (4 females, eight males). They were caricatured using five different approaches (Method factor): the learning-based approach (Deep) presented in **Section 4**, two variations of the rule-based approach presented in **Section 3** (see **Table 1**), and two state-of-the-art caricaturization methods–EDFM (Blanz and Vetter, 1999) and Sela (Sela et al., 2015). For each face (original and caricatured), we used the cartoonization module presented in **Section 3**. The texture blurring is expected to reduce the mismatch of realism between the shape and the texture and therefore make the caricature more acceptable to human observers (Zell et al., 2015). The stimuli were rendered with a rotation of 30°around the vertical axis, with a fixed view. The angle was chosen as a common viewpoint between a frontal and profile view. We considered only the facial mask, hence other facial attributes such as eyes and hair were not displayed.

5.3 Protocol

The perceptual study consisted of two parts. The first part of the study assessed the results produced by each method for each face, according to participant's preferences. For each human facial scan, participants were presented with the original face and the caricatures generated with the five methods. They were asked to rank all five caricatures from the best to the worst caricature. The order of the scans and the presentation of the caricatures was randomized independently for each participant and each facial scan was only presented once, for a total of 12 trials. The second

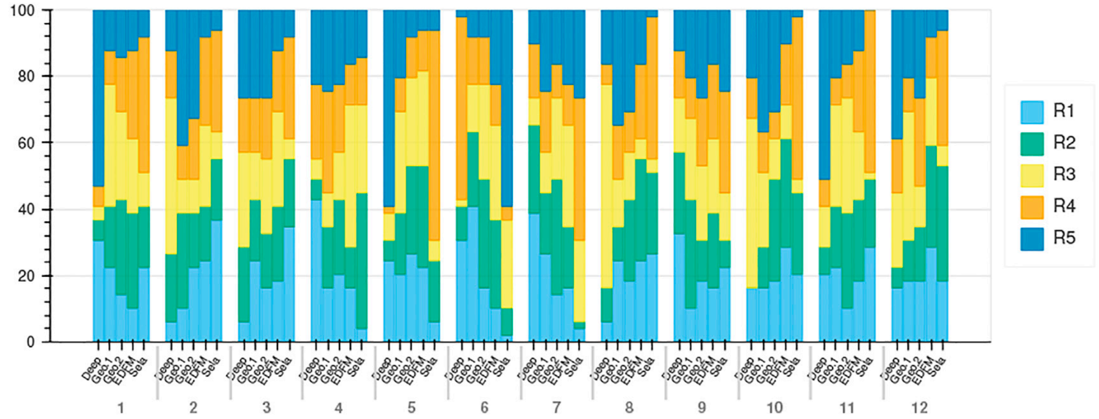

FIGURE 9 | Average rankings, per Method and Identity. R1 to R5 are the ranks 1 to rank 5. Note the high variance per face and method.

part of the study aimed at evaluating globally each of the five methods. For each method (in a random order), the caricaturization results (12 facial scans) were displayed at once. Participants were asked to indicate how much they agreed to three statements using 5-point Likert scales. The statements were “They preserve the identity of the person,” “They correspond to what would be expected of a caricature,” “I like the results”. There was no time limit for any of the two parts, and the evaluation was conducted online using the PsyToolkit software (Stoet, 2010, 2017). We include a sample view of the ranking task in **Supplementary Figure 22**. A render of all 12 caricatures for each method can be seen on **Supplementary Figures 17–19, 21**.

5.4 Results

We present in this section the statistical results of the user study.

5.4.1 Average Rankings

To analyze ranking distributions (**Figures 9–Figure 10**), we first performed a Friedman test with the within-subject factor Method (using the average rank between all 12 scans). We found an effect of the Method on average ranking (χ^2 = 12.21; $p < 0.05$). The effect is then explored further using a Wilcoxon post-hoc test for pair-wise comparisons. We found significant differences only between EDFM and Deep, Geo.1, Sela (all $p < 0.05$). We found that per method, average rankings vary between 2.81 (EDFM) and 3.12 (Deep) 10. In order to determine whether ranking distributions per method differed with identities, we used a Friedman test with within-subject factors Method and Identity. Out of 12 distinct identities, 6 (identities 2, 5, 6, 7, 11, 12) showed significantly different rankings between methods. This is in most cases (5 out of 6) due to worse than average performance from a set of methods, usually Deep or Sela.

5.4.2 Top Rankings

We measured Top-1, Top-2, and Top-3 rank differences per method, using Friedman tests, Top-X rankings being the number of times the techniques were ranked X or lower (lower is better, **Figure 11**). We found no significant differences for Top-1 (χ^2 = 4.14; $p = 0.38$) rankings, but an effect was found for both Top-2 (χ^2 = 9.74; $p < 0.05$) and Top-3 rankings (χ^2 = 34.60; $p < 0.001$). The effect for Top-2 and Top-3 rankings is then explored using a Wilcoxon post-hoc test. For Top-2 rankings, we found that EDFM was chosen significantly more often as first or second choice than Deep ($p < 0.05$) and Sela ($p < 0.01$). For Top-3 rankings, we found a similar preference for EDFM over Deep, Geo.1, and Sela ($p < 0.05$), as well as a significant lower preference for Sela over all others ($p < 0.05$).

5.4.3 Variations Between Participants

We looked into participant-wise preferences for caricature methods using a Friedman test on ranking choices of each participant, individually. Out of 49 participants, separate Friedman tests on their Top-1 rankings showed that only 12 had a significant preference towards a set of methods, and out of these only 4 towards a specific one. These numbers are too low to show anything conclusive in that regard.

5.4.4 Subjective Scores

Subjective ratings results were analyzed separately using a one-way ANOVA with within-subject factor Method on the data of each question. All subjective results differences between methods were found to be significant (p values of $5.7e-6$, $7.35e-6$, and $2.28e-5$). We conducted separate post-hoc analyses using Wilcoxon. For the statement “They preserve the identity of the person” (**Figure 12**), significantly different groups of method were Deep, Sela (mean = 3), and Geo.1, EDFM (mean = 2.3). The method Geo.2 (mean = 2.6) was not significantly different from others. For the statements “They correspond to what would be expected of a caricature” (**Figure 13**) and “I like the results” (**Figure 14**), the only significant differences were between the group of Geo.1, Geo.2, EDFM, and Sela, Deep being in between.

6 DISCUSSION

In this paper, we have proposed two novel caricaturization methods. One leveraging the capabilities of deep style transfer

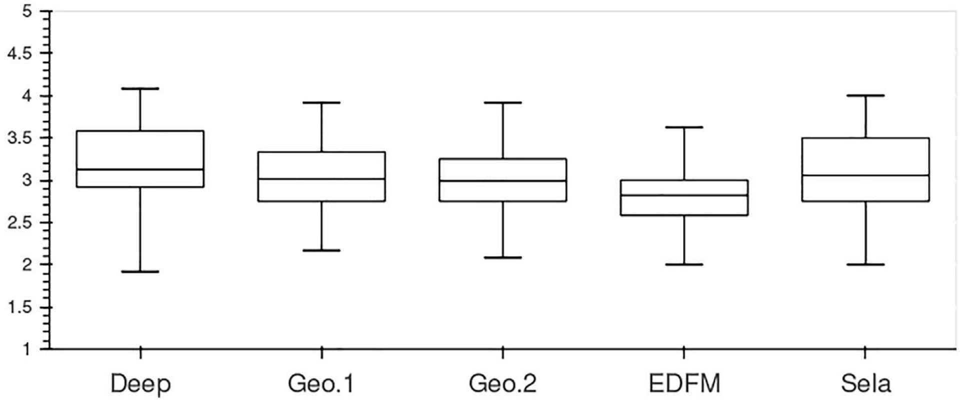

FIGURE 10 | Boxplot of the average rankings over participants, per method. Rankings range from 1 to 5. Overall, all methods achieve similar performances, averages being between 2.81 and 3.12 (lower is better).

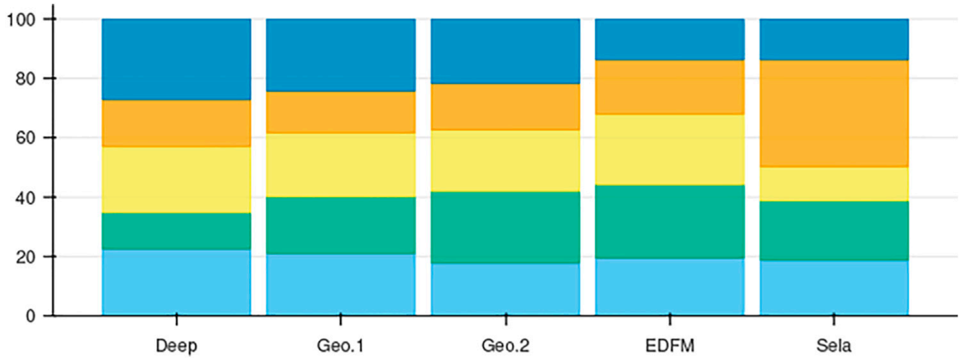

FIGURE 11 | Caricature ranking distribution across all participants, per method. Top-1 to Top-5 rankings respectively shown in light blue, green, yellow, orange, and blue.

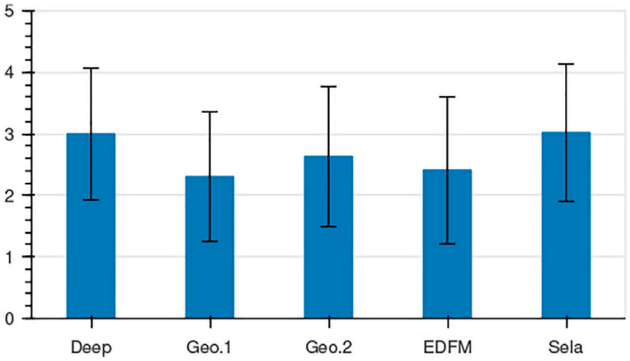

FIGURE 12 | Average Likert ratings for the statement “They preserve the identity of the person”. Deep and Sela are significantly different to Geo.1 and EDFM.

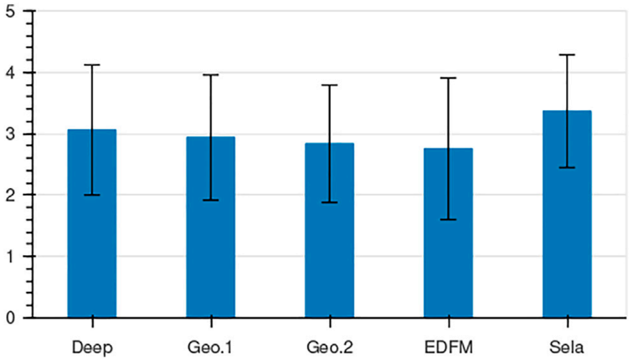

FIGURE 13 | Average Likert ratings for the statement “I like the results”. Geo.1, Geo.2, and EDFM are significantly different to Sela.

networks for caricaturization (Deep), and the two remaining are variations of a gradient-based EDFM, with and without the use of a data driven face shape stylization (Geo.1 and Geo.2).

The proposed methods, and two additional methods from the literature were evaluated through a user study considering 12 different facial scans and the corresponding caricature generated from these different methods. Overall, the results

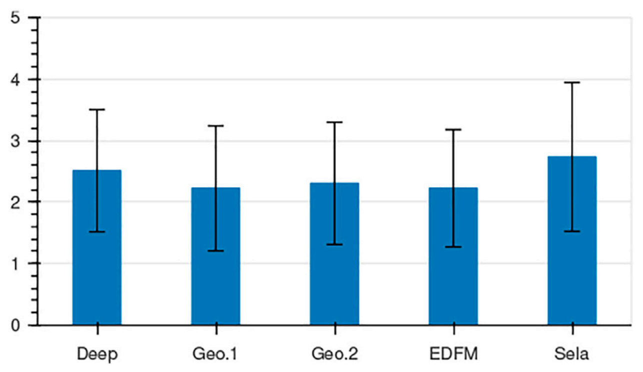

FIGURE 14 | Average Likert ratings for the statement "They correspond to what would be expected of a caricature". Geo.1, Geo.2, and EDFM are significantly different to Sela.

showed that all methods achieved similar performances, average ratings going from 2.82 to 3.12 (lower is better). An observation from the results is that in general, there was not a method which was significantly superior to the others. The results considering only the method (see **Figure 12**) show a fairly distributed results, although Deep and Sela approaches seem to generate a higher number of "badly ranked caricatures" (fourth and fifth ranks). This observation matches with the global appreciation from participants, as EDFM, Geo.1 and Geo.2 got slightly higher scores. While this result could suggest that some of the methods worked better from some facial scans than others, the results split by Identity do not totally support this hypothesis (see **Figure 11**). Looking at the top five worst ranked caricatures (**Figure 15**), we can identify several cases in which the method considered could have generated undesired results. The facial features of face six interpenetrate each other when using Sela, and the borders of face seven are spread too widely using the same method. On face 5, eye size difference is too greatly exaggerated with the method Deep. These generated faces rated significantly worse than others on average can be easily identified, opening possibilities of a manual or automatic filtering protocol. Nevertheless, these results seem to evidence that some methods had a particularly bad performance on some of the facial scans. Yet, this did not happen consistently. Each caricaturization method had a pre-defined set of meta-parameters. The chosen configuration could have suited better some faces than others, generating caricatures of different qualities. The top five best ranked caricatures can be seen on **Figure 8**.

Another potential explanation for the results is that the task was too hard and subjective, choices ending up being random. Using faces with no hair or eyes might have even increased the complexity of the task. Indeed, some participants explicitly stated that the task was difficult, especially as they were judging textured facial masks instead of full faces. Nevertheless, this potential user preference does not seem to be linked with any particular caricaturization method. Looking at participant preferences, only 12 participants out of 49 showed a significant rating variation between methods ranked first. Looking at results on subjective questions, the two worse rated (Deep and Sela) methods rank-wise (being also those with the worst rated specific caricatures) were rated higher both at "They correspond to what would be expected of a caricature" and "I like the result," where caricatures of each method were presented globally, suggesting that without their bad results on specific faces–which might be less visible when presented amongst all the others–they could actually have ranked higher than other methods. The conception of a perceptual metric reliably judging the quality of a caricature could help guide its creation, but the high variation of participant preferences in our study suggest that it would require a considerably larger study to be defined.

Considering these findings, we issue the following guidelines for choosing a method to generate caricatures automatically.

- If the main goal is to generate caricatures with a given set of parameters, no specific style, and as little variance as possible in quality, an EDFM-based method is the most suitable.
- If there is still no specific style required, but more tolerance to variance in quality (for instance if it is possible to tune the generated faces when they are unsatisfying), we recommend the approach of Sela, rated very similarly to EDFM on average in the rankings task, and significantly more on the subjective questionnaire.

- If a specific caricature style is required, the Deep approach will offer results comparable with Sela both in the ranking task and the questionnaire.

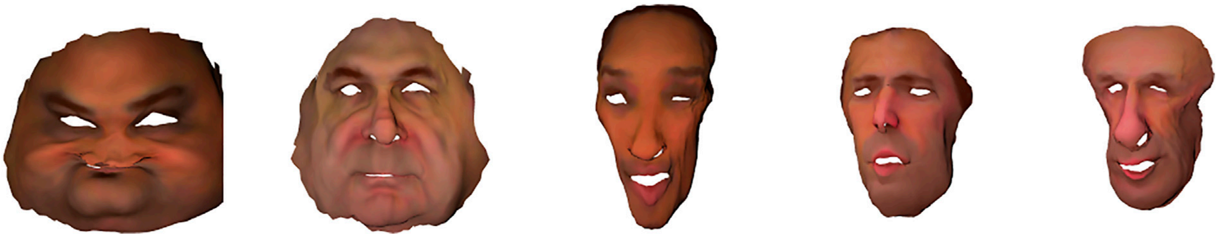

FIGURE 15 | The five worst caricatures (with the worst mean ranks; identities 6, 7, 5, 11, 12).

- Finally, if there is a need to target a specific user, the best solution is to use the panel of available methods, and leave the choice to them.

Caricatures provide a style whose notion can be understood as an "accentuation of facial features," allowing manually defined rules to achieve comparable performance to learning-based approaches. Other stylistic facial domains, such as aliens or anthropomorphic animals could have more to gain from learning. Such non-realistic 3D facial data is although currently very scarce.

7 CONCLUSION

In this paper we have introduced two novel approaches to automatically generate caricatures from 3D facial scans. T he first method mixes EDFM-based curvature deformation and data driven proportion deformation, while the second method is based on a domain-to-domain translation deep neural network. Then, we present and discuss a perceptual study aiming to assess the quality of the generated caricatures. Overall, the results showed that the different evaluated methods performed in a similar way, although their performance could vary with respect to the facial scan used. This result illustrates both the subjectivity of evaluating caricaturization performance, along with the complementarity of using different approaches, producing different styles of caricatures. Future work could involve looking into automatic detection of the worse cases of automatic caricaturization, to apply a correction or a filter, or exploring learned-based automatic caricaturization by learning on different caricature styles, and setting up a network able to generate faces of a given style. We believe this study of the extended state of the art have helped grow and precise the landscape of automatic caricaturization approaches, and 3D facial stylization in general, and that our work provides interesting insights and guidelines for the automatic generation of caricatures that will help practitioners and inspire future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

NO contributed to this work during his PhD, GK during his Master. They were supervised by FA, QA, FD, PG, LH, and FM.

ACKNOWLEDGMENTS

We wish to thank all the reviewers for their comments, and the participants in our experiment. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 952 147. This project has received funding from the Association Nationale de la Recherche et de la Technologie under CIFRE agreement No 2018/1656.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.785104/full#supplementary-material

REFERENCES

Akleman, E. (1997). Making Caricatures with Morphing. *ACM SIGGRAPH*. doi:10.1145/259081.259231

Akleman, E., Palmer, J., and Logan, R. (2000). Making Extreme Caricatures with a New Interactive 2d Deformation Technique with Simplicial Complexes. *Proc. Vis.*, 100–105.

Akleman, E., and Reisch, J. (2004). Modeling Expressive 3d Caricatures. *ACM SIGGRAPH*. doi:10.1145/1186223.1186299

Arjovsky, M., Chintala, S., and Bottou, L. (2017). "Wasserstein Generative Adversarial Networks," in International Conference on Machine Learning (ICML), 214–223. ArXiv: 1701.07875.

Blanz, V., and Vetter, T. (1999). A Morphable Model for the Synthesis of 3D Faces. *ACM SIGGRAPH*, 187–194. doi:10.1145/311535.311556

Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A., and Dunaway, D. (2016). "A 3D Morphable Model Learnt from 10,000 Faces," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.598

Brennan, S. E. (1985). Caricature Generator: The Dynamic Exaggeration of Faces by Computer. *Leonardo* 18, 170–178. doi:10.2307/1578048

Cai, H., Guo, Y., Peng, Z., and Zhang, J. (2021). Landmark Detection and 3d Face Reconstruction for Caricature Using a Nonlinear Parametric Model. *Graphical Models* 115, 101103. doi:10.1016/j.gmod.2021.101103

Cao, K., Liao, J., and Yuan, L. (2019). CariGANs. *ACM Trans. Graph.* 37, 1–14. doi:10.1145/3272127.3275046

Chen, H., Zheng, N.-N., Liang, L., Li, Y., Xu, Y.-Q., and Shum, H.-Y. (2002). PicToon. *ACM Multimedia*. doi:10.1145/641007.641040

Chen, Y.-L., Liao, W.-H., and Chiang, P.-Y. (2006). "Generation of 3d Caricature by Fusing Caricature Images," in IEEE International Conference on Systems, Man and Cybernetics. doi:10.1109/icsmc.2006.384498

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-To-Image Translation," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (Salt Lake City, Utah: CVPR), 8789–8797. doi:10.1109/CVPR.2018.00916

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. (2020). "StarGAN V2: Diverse Image Synthesis for Multiple Domains," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (Virtual Event (originally Seattle): CVPR), 8185–8194. doi:10.1109/CVPR42600.2020.00821

Cimen, G., Bulbul, A., Ozguc, B., and Capin, T. (2012). Perceptual Caricaturization of 3d Models. *Computer Inf. Sci.*, 201–207. doi:10.1007/978-1-4471-4594-3_21

Clarke, L., Min Chen, M., and Mora, B. (2011). Automatic Generation of 3d Caricatures Based on Artistic Deformation Styles. *IEEE Trans. Vis. Comput. Graphics* 17, 808–821. doi:10.1109/tvcg.2010.76

Danieau, F., Gubins, I., Olivier, N., Dumas, O., Denis, B., Lopez, T., et al. (2019). "Automatic Generation and Stylization of 3D Facial Rigs," in IEEE Conference on Virtual Reality and 3D User Interfaces (Osaka, Japan: VR), 784–792. doi:10.1109/VR.2019.8798208

Eigensatz, M., Sumner, R. W., and Pauly, M. (2008). Curvature-domain Shape Processing. *Computer Graphics Forum* 27, 241–250. doi:10.1111/j.1467-8659.2008.01121.x

Fujiwara, T., Koshimizu, H., Fujimura, K., Fujita, G., Noguchi, Y., and Ishikawa, N. (2002). A Method for 3d Face Modeling and Caricatured Figure Generation. *IEEE Int. Conf. Multimedia Expo* 2, 137–140. doi:10.1109/ICME.2002.1035531

Gong, S., Chen, L., Bronstein, M., and Zafeiriou, S. (2019). "Spiralnet++: A Fast and Highly Efficient Mesh Convolution Operator," in IEEE/CVF International Conference on Computer Vision Workshop (Seoul, South Korea: ICCVW), 4141–4148. doi:10.1109/ICCVW.2019.00509

Gooch, B., Reinhard, E., and Gooch, A. (2004). Human Facial Illustrations. *ACM Trans. Graph.* 23, 27–44. doi:10.1145/966131.966133

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative Adversarial Nets," in Conference on Neural Information Processing Systems (NIPS), 2672–2680. ArXiv: 1406.2661.

Gu, Z., Dong, C., Huo, J., Li, W., and Gao, Y. (2021). Carime: Unpaired Caricature Generation with Multiple Exaggerations. *IEEE Trans. Multimedia* 1, 1. doi:10.1109/TMM.2021.3086722

Guo, Y., Jiang, L., Cai, L., and Zhang, J. (2019). 3D Magic Mirror: Automatic Video to 3D Caricature Translation. *CoRR* abs/1906, 00544, 2019 . ArXiv: 1906.00544.

Hong Chen, H., Ying-Qing Xu, Y.-Q., Heung-Yeung Shum, H.-Y., Song-Chun Zhu, S.-C., and Nan-Ning Zheng, N.-N. (2001). "Example-based Facial Sketch Generation with Non-parametric Sampling," in IEEE/CVF International Conference on Computer Vision (ICCV). doi:10.1109/iccv.2001.937657

Huang, X., and Belongie, S. (2018a). "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization," in IEEE/CVF International Conference on Computer Vision (Munich, Germany: ICCV), 1510–1519. doi:10.1109/ICCV.2017.167

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018b). "Multimodal Unsupervised Image-To-Image Translation," in Proceedings of the European Conference on Computer Vision (Newcastle, UK: ECCV), 179179–196196. doi:10.1007/978-3-030-01219-9_11

Huo, J., Li, W., Shi, Y., Gao, Y., and Yin, H. (2018). "Webcaricature: a Benchmark for Caricature Recognition," in British Machine Vision Conference. ArXiv: 1703.03230.

Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J. (2017). "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks," in International Conference on Machine Learning (Sydney, NSW: ICML), 1857–1865. doi:10.5555/3305381.3305573

Kingma, D. P., and Ba, J. (2015). "Adam: A Method for Stochastic Optimization," in International Conference on Learning Representations, (ICLR). ArXiv: 1412.6980.

Lin Liang, L., Hong Chen, H., Ying-Qing Xu, Y.-Q., and Heung-Yeung Shum, H.-Y. (2002). "Example-based Caricature Generation with Exaggeration," in Pacific Conference on Computer Graphics and Applications. doi:10.1109/pccga.2002.1167882

Litany, O., Remez, T., Rodola, E., Bronstein, A., and Bronstein, M. (2017). "Deep Functional Maps: Structured Prediction for Dense Shape Correspondence," in IEEE/CVF International Conference on Computer Vision (ICCV). doi:10.1109/iccv.2017.603

Liu, J., Chen, Y., Miao, C., Xie, J., Ling, C. X., Gao, X., et al. (2009). Semi-supervised Learning in Reconstructed Manifold Space for 3d Caricature Generation. *Computer Graphics Forum* 28, 2104–2116. doi:10.1111/j.1467-8659.2009.01418.x

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). "Unsupervised Image-To-Image Translation Networks," in Conference on Neural Information Processing Systems (Long Beach, CA: NIPS), 700–708. doi:10.5555/3294771.3294838

Liu, M.-Y., Huang, X., Mallya, A., Karras, T., Aila, T., Lehtinen, J., et al. (2019). "Few-shot Unsupervised Image-To-Image Translation," in IEEE/CVF International Conference on Computer Vision (ICCV), 10550–10559. doi:10.1109/ICCV.2019.01065

Liu, S., Wang, J., Zhang, M., and Wang, Z. (2012). Three-dimensional Cartoon Facial Animation Based on Art Rules. *Vis. Comput.* 29, 1135–1149. doi:10.1007/s00371-012-0756-2

Maron, H., Galun, M., Aigerman, N., Trope, M., Dym, N., Yumer, E., et al. (2017). Convolutional Neural Networks on Surfaces via Seamless Toric Covers. *ACM Trans. Graph.* 36, 1–10. doi:10.1145/3072959.3073616

Mescheder, L. M., Nowozin, S., and Geiger, A. (2018). Which Training Methods for gans Do Actually Converge? *Int. Conf. Machine Learn. (Icml)* 80, 3478–3487. ArXiv: 1801.04406.

Mo, Z., Lewis, J. P., and Neumann, U. (2004). Improved Automatic Caricature by Feature Normalization and Exaggeration. *ACM SIGGRAPH*. doi:10.1145/1186223.1186294

Moschoglou, S., Ploumpis, S., Nicolaou, M. A., Papaioannou, A., and Zafeiriou, S. (2020). 3dfacegan: Adversarial Nets for 3d Face Representation, Generation, and Translation. *Int. J. Comput. Vis.* 128, 2534–2551. doi:10.1007/s11263-020-01329-8

Olivier, N., Hoyet, L., Danieau, F., Argelaguet, F., Avril, Q., Lecuyer, A., et al. (2020). "The Impact of Stylization on Face Recognition," in ACM Symposium on Applied Perception. doi:10.1145/3385955.3407930

Pengfei Li, P., Yiqiang Chen, Y., Junfa Liu, J., and Guanhua Fu, G. (2008). "3d Caricature Generation by Manifold Learning," in IEEE International Conference on Multimedia and Expo. doi:10.1109/icme.2008.4607591

Ranjan, R., Sankaranarayanan, S., Castillo, C. D., and Chellappa, R. (2017). "An All-In-One Convolutional Neural Network for Face Analysis," in IEEE International Conference on Automatic Face and Gesture Recognition (FG), 17–24. doi:10.1109/FG.2017.137

Redman, L. (1984). *How to Draw Caricatures*.

Sela, M., Aflalo, Y., and Kimmel, R. (2015). Computational Caricaturization of Surfaces. *Computer Vis. Image Understanding* 141, 1–17. doi:10.1016/j.cviu.2015.05.013

Shi, Y., Deb, D., and Jain, A. K. (2019). "WarpGAN: Automatic Caricature Generation," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.01102

Sorkine, O. (2005). "Laplacian Mesh Processing," in *Eurographics 2005 - State of the Art Reports*. doi:10.2312/egst.20051044

Stoet, G. (2017). PsyToolkit. *Teach. Psychol.* 44, 24–31. doi:10.1177/0098628316677643

Stoet, G. (2010). PsyToolkit: A Software Package for Programming Psychological Experiments Using Linux. *Behav. Res. Methods* 42, 1096–1104. doi:10.3758/brm.42.4.1096

Sumner, R. W., and Popović, J. (2004). Deformation Transfer for triangle Meshes. *ACM SIGGRAPH*. doi:10.1145/1186562.1015736

Taigman, Y., Polyak, A., and Wolf, L. (2017). "Unsupervised Cross-Domain Image Generation," in International Conference on Learning Representations (ICLR). ArXiv: 1611.02200.

Wang, C., Chai, M., He, M., Chen, D., and Liao, J. (2021). "Cross-Domain and Disentangled Face Manipulation with 3D Guidance," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). ArXiv: 2104.11228.

Wu, Q., Zhang, J., Lai, Y.-K., Zheng, J., and Cai, J. (2018). "Alive Caricature from 2d to 3d," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2018.00766

Xie, J., Chen, Y., Liu, J., Miao, C., and Gao, X. (2009). Interactive 3d Caricature Generation Based on Double Sampling. *ACM Multimedia*. doi:10.1145/1631272.1631403

Ye, Z., Yi, R., Yu, M., Zhang, J., Lai, Y., and Liu, Y. (2020). "3d-carigan: An End-To-End Solution to 3d Caricature Generation from Face Photos," in *Computing Research Repository (CoRR)*. ArXiv: 2003.06841.

Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). "DualGAN: Unsupervised Dual Learning for Image-To-Image Translation," in IEEE/CVF International Conference on Computer Vision (ICCV). doi:10.1109/iccv.2017.310

Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., McDonnell, R., et al. (2015). To Stylize or Not to Stylize? *ACM Trans. Graph.* 34, 1–12. doi:10.1145/2816795.2818126

Zhou, J., Tong, X., Liu, Z., and Guo, B. (2016). 3d Cartoon Face Generation by Local Deformation Mapping. *Vis. Comput.* 32, 717–727. doi:10.1007/s00371-016-1265-5

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," in IEEE/CVF

International Conference on Computer Vision (ICCV). doi:10.1109/ICCV.2017.244

Conflict of Interest: Authors NO, GK, QA, FD, and PG were employed by the company InterDigital.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.