Modeling Caricature Expressions by 3D Blendshape and Dynamic Texture

Keyu Chen
University of Science and Technology of China
cky95@mail.ustc.edu.cn

Jianmin Zheng
Nanyang Technological University
ASJMZheng@ntu.edu.sg

Jianfei Cai
Monash University
Jianfei.Cai@monash.edu

Juyong Zhang
University of Science and Technology of China
juyong@ustc.edu.cn

ABSTRACT

The problem of deforming an artist-drawn caricature according to a given normal face expression is of interest in applications such as social media, animation and entertainment. This paper presents a solution to the problem, with an emphasis on enhancing the ability to create desired expressions and meanwhile preserve the identity exaggeration style of the caricature, which imposes challenges due to the complicated nature of caricatures. The key of our solution is a novel method to model caricature expression, which extends traditional 3DMM representation to caricature domain. The method consists of shape modelling and texture generation for caricatures. Geometric optimization is developed to create identity-preserving blendshapes for reconstructing accurate and stable geometric shape, and a conditional generative adversarial network (cGAN) is designed for generating dynamic textures under target expressions. The combination of both shape and texture components makes the non-trivial expressions of a caricature be effectively defined by the extension of the popular 3DMM representation and a caricature can thus be flexibly deformed into arbitrary expressions with good results visually in both shape and color spaces. The experiments demonstrate the effectiveness of the proposed method.

CCS CONCEPTS

• **Applied computing** → **Media arts**; • **Computing methodologies** → **Image manipulation**.

KEYWORDS

Caricature; Expression; Deformation; Blendshape; Texture

ACM Reference Format:
Keyu Chen, Jianmin Zheng, Jianfei Cai, and Juyong Zhang. 2020. Modeling Caricature Expressions by 3D Blendshape and Dynamic Texture. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3394171.3413643

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3413643

1 INTRODUCTION

Caricatures are non-photorealistic images that exaggerate or simplify the features of subjects, serving as a descriptive art form in a wide range of applications such as social network, animation and entertainment industry. Like most other art forms, caricatures are commonly drawn by artists with subjectivity. Even for the same subject, different artists may focus on different features and thus create caricature portraits different in image styles and exaggeration forms. Such characteristics ensure the fascination of caricatures and meanwhile impose challenges in re-creation and editing such as manipulating the expressions of artist-created caricature paintings.

This paper considers the problem of deforming a 2D caricature image driven by an expression given in a normal face image. The deformed or re-created caricature is expected to exhibit the desired expression and also preserve the identity exaggeration style of the original caricature. In order to create caricatures by computer, many works have been developed to mimic artists' creative drawing process. They focus on automatic caricature generation in 2D [7, 17, 24, 34] and 3D [8, 29, 43] as well. Other related processes for caricatures such as art style classification [44], caricature landmark detection [36, 44] and caricature identity recognition [19] have also been developed. With the advance of machine learning techniques, generative adversarial networks (GANs) are designed to relate normal human face domain and caricature domain in caricature generation [7, 24, 34]. For example, the "face-to-caricature" generation builds an unpaired mapping function between two different sets. Different from these works, our work is a problem of deformation, a process on an existing caricature, rather than a creation from scratch. Moreover, our deformation is driven by a facial expression, which is a kind of semantic-guided manipulation natural to human's creation activity. To best of our knowledge, there is not much work done along this direction in the caricature domain.

Existing caricature techniques do not solve our problem well. Our problem has three technical challenges. First, caricatures generally do not have explicit expression definition or representation. Therefore when we extend the human facial expression representation [10] to the target caricature domain, we have to carefully and properly design the map. Second, the expression deformation often loses the preservation of the caricature's exaggeration styles. Third, large expression deformation often introduces artifacts in local regions such as the mouth and eyes' area.

Considering that the expression varies in both geometry and texture spaces, we propose a solution to overcome the challenges by constructing a series of 3D caricature expression blendshapes

for shape deformations and a conditional GAN for dynamic texture generation. Specifically, we map the source caricature landmarks to the human face domain using *CariGeoGAN* [7] and obtain a group of blendshapes via fitting a 3DMM model [2, 46]. Then we transfer the landmarks on the normal faces back to the caricature domain and reconstruct 3D dense meshes. In this way, the caricature expressions can be defined by the expression coefficients of 3DMM [2]. To maintain the identity consistency and the exaggeration style, we propose a geometric optimization procedure to reconstruct the caricature blendshapes from the mapped landmarks by constraining caricature expressions to keep the same exaggeration style. To eliminate artifacts generated in the deformation, we train a texture generation network conditioned on the expression coefficients. Inspired by [31], we adopt the attention module to learn which area should be modified in color space. In order to improve the robustness in dealing with different identities and color styles, we transfer many human face textures to caricature styles by adaptive instance normalization (AdaIN) [18]. Our network is trained in a supervised way on the adapted texture dataset and trained in an unsupervised way on a caricature texture dataset. The combined training manner helps to overcome the lack of a large caricature dataset. In the inference stage, the network takes as input the source texture with target expression parameters and outputs the desired texture for the target expression. As a result, we come up with a novel caricature expression generation method that consists of a shape component and a texture component.

The contribution of this paper lies in three aspects. First, we propose a framework consisting of shape modeling and texture generation to tackle the challenges occurred in expression driven caricature deformation. Second, we design an optimization-based method to construct and refine 3D caricature blendshapes, which extends the popular facial expression representation to the caricature domain. Third, we present a dynamic texture generation module for caricatures by training a GAN model conditioned on expression parameters.

2 RELATED WORK

This section briefly reviews some related works: caricature generation, blendshape representation and facial expression manipulation.

2.1 Caricature Generation

Caricature generation aims at creating exaggerated caricatures from given human portraits. Techniques in this area can be classified into two categories: 2D image-based and 3D geometry-based approaches. 2D image-based caricature generation was firstly introduced with interactive methods [4, 32]. With advances in image processing, some automatic approachs were developed. For example, Chiang [25] developed a caricature generation system that automatically analyzes the facial features of a subject and then determines how the face components should be altered and placed. Recently, deep learning is adopted into the generation process. Cao [7] proposed a shape and texture exaggeration process by using CycleGAN structure. Shi [34] developed an automatic network for generating caricatures, which learns how to warp a face photo into a caricature and transfer texture styles as well. Usually the shape and texture processes used in these methods to control the generation are independent.

Relatively, the work for 3D geometry-based caricature generation is much less. Despite some surface based methods [27–29, 33], only a few works associate the 3D shape generation process (i.e., caricature reconstruction) with 2D inputs. Wu [43] introduced an intrinsic deformation representation for constructing caricatures from 2D images. Han [17] proposed a method for generating personalized caricatures from 2D sketches. Compared to generating 2D caricatures, generating 3D caricatures requires more control of shapes and textures and thus is more difficult. On the other hand, 3D models provide more information that can benefit other processes such as animation or face editing.

2.2 Blendshape Representation

Blendshape is a prevalent parametric model in animation and movie industry, which represents facial variations as linear combinations of several given shapes [15]. The basic theory can be dated back to the facial action coding system (FACS) [10]. Recently developed 3D face reconstruction works take pre-defined blendshapes (e.g., FaceWarehouse model [6] or 3D morphable model-3DMM [2]) for expression and identity modelling [9, 12, 16, 22, 39]. It is also used in various applications such as facial retargeting [23, 42] and manipulation [47]. A blendshape model can be considered as the bases of the expression deformation space of a particular subject. There are several approaches to creating blendshapes. For a real actor, a template model can be registered to its scan and repeating the process for different expressions can generate a range of topology-consistent blendshapes [6]. For digital characters, a skilled modeling artist can deform a base mesh into different shapes to cover the expression range [23].

2.3 Facial Expression Manipulation

Many facial expression manipulation methods are based on 3D parametric facial models such as 3DMM. For example, Vlasic [41] proposed a multilinear model for tracking and retargeting expression information. Cao [5] extended this idea by performing co-tensor computation on FaceWarehouse dataset [6]. Thies [38] proposed video-to-video facial expression retargeting by fitting 3DMM with additional lighting parameters.

However, 3D parametric representation is not capable of representing fine-scale geometry and texture details such as teeth and wrinkles. Therefore, many recent approaches editing 2D facial images are based on generative adversarial networks (GANs) [14]. For instance, by considering the expression generation process as unpaired image-to-image translation [45], Pumarola [31] designed an unsupervised framework *GANimation* within the CycleGAN structure to train the face generation network conditioned on Face Action Units. Ververas [40] replaced the action unit vector with 3DMM coefficients for stable and accurate expression manipulations. Geng [13] combined the advantages of both geometry-based methods and image generation networks and presented a 3D guided texture and shape refinement approach.

3 METHODOLOGY

3.1 Overview

Our basic problem can be described as follows: given a caricature image $\mathbf{I}(\mathbf{e_x})$ with caricature expression $\mathbf{e_x}$ and labelled facial landmarks $\mathbf{L}(\mathbf{e_x})$ and any normal face image $\tilde{\mathbf{I}}(\mathbf{e_t})$ with expression $\mathbf{e_t}$,

our target is to change the caricature expression to $\mathbf{e_t}$, i.e. generating $\mathbf{I}(\mathbf{e_t})$. Following the common approach of 3D parametric modelling of normal faces, we aim to build up a parametric caricature expression model that shares the same identity with $\mathbf{I}(\mathbf{e_x})$ while covering different expression variations. With the parametric expression model, any caricature expression can be parametrically represented by the pre-defined bases. However, unlike normal faces, there is no explicit definition on the caricature expression and there is also no sufficient caricature expression data to build up the parametric model bases. Our idea is to associate the caricature expression with the popular normal face expression model *3D Morphable Model* (3DMM) [2] by adopting the cross-domain mapping *CariGeoGAN* from [7], which facilitates the mapping from a caricature face's landmarks to the corresponding normal face's landmarks that can be modelled by 3DMM or inversely. In this way, we can define all the expressions, e.g. $\mathbf{e_x}$ or $\mathbf{e_t}$, as the 3DMM expression coefficients, which is widely used in human face related applications.

In particular, we decompose the caricature new expression generation task into a shape modelling component and a texture generation component. Such disentanglement separates the expression deformation in geometry and color space and allows specific optimization for each component. To obtain the initial shape and texture from original caricature data $\{\mathbf{I}(\mathbf{e_x}), \mathbf{L}(\mathbf{e_x})\}$, we employ the 3D caricature reconstruction method in [43] to reconstruct the caricature shape $\mathbf{S}(\mathbf{e_x})$ from $\mathbf{L}(\mathbf{e_x})$. Then we apply classical rasterization and rendering pipeline in graphics to extract the texture map $\mathbf{T}(\mathbf{e_x})$ from $\mathbf{I}(\mathbf{e_x})$, based on a pre-defined ARAP parameterization map [26]. Note that the texture maps are essentially 2D meshes with fixed vertex locations but varying colors corresponding to different textures.

For the shape modelling component, our target is to construct 3D caricature blendshapes $\{\mathbf{S}(\mathbf{e_i})\}_{i=0}^{N}$ where $\{\mathbf{e_i}\}_{i=0}^{N}$ are expression parameters defined in 3DMM. To this end, we use the cycle consistent network *CariGeoGAN* [7] to translate caricature landmarks $\mathbf{L}(\mathbf{e_x})$ to normal face domain and fit a 3DMM model on the mapped landmarks. By changing the expression parameters to $\{\mathbf{e_i}\}_{i=0}^{N}$that are pre-defined in normal face domain, we get the blendshape landmarks, which are then mapped back to the caricature domain for reconstructing 3D caricature blendshapes. During this process, the caricature identity information is often disturbed by expression deformation. Thus, we propose an optimization approach to refine the blendshapes to preserve the identity of original shape $\mathbf{S}(\mathbf{e_x})$. The details are illustrated in Fig. 1 and described in Sec. 3.2.

Meanwhile, just modeling 3D geometric deformation is not enough for image-based caricature expression manipulation. Not only the shapes but also the textures should be deformed in expression changes [13]. For example, when a neutral face image is manipulated to be a smile, the teeth part should be added on the mouth area for better photorealistic. Based on this observation, we propose a dynamic texture module for generating different caricature textures under different expression conditions. Our texture generation model is a conditional GAN and importantly it still takes the same 3DMM expression representation as the shape modelling component. In the training stage, we overcome the issue of lacking large scale caricature dataset by adapting normal face textures to caricature styles using *Adaptive Instance Normalization* [18]

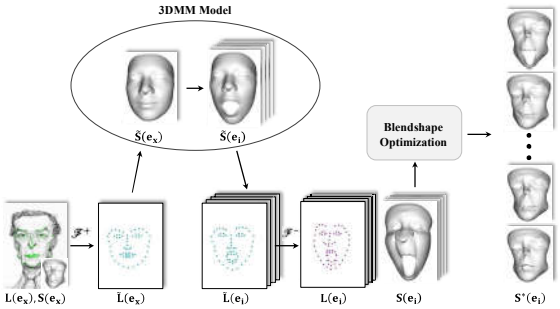

Figure 1: Process of constructing 3D caricature blendshapes. From left to right, we first fit 3DMM model on the mapped landmarks $\tilde{\mathbf{L}}(\mathbf{e_x})$ and then translate those manipulated landmarks $\tilde{\mathbf{L}}(\mathbf{e_i})$ back to the caricature domain, obtaining $\mathbf{L}(\mathbf{e_i})$. The initial caricature blendshapes $\mathbf{S}(\mathbf{e_i})$ are reconstructed based on [43] and further optimized by our optimization module for preserving the identity.

(AdaIN). In the inference stage, our texture model takes as input original texture $\mathbf{T}(\mathbf{e_x})$ and target expression $\mathbf{e_t}$, and outputs the desired texture map $\mathbf{T}(\mathbf{e_t})$. The texture generation can be formulated as $\{\mathbf{T}(\mathbf{e_x})|\mathbf{e_t}\} \rightarrow \mathbf{T}(\mathbf{e_t})$. The details are illustrated in Fig. 2 and described in Sec. 3.3.

3.2 Shape Modelling Component

In order to faithfully construct 3D blendshapes for the original caricature $\{\mathbf{I}(\mathbf{e_x}), \mathbf{L}(\mathbf{e_x})\}$, we combine two state-of-the-art methods: 3D caricature reconstruction [43] and photo-caricature landmark mapping [7] (See Sec. 3.2.1), to construct initial caricature blendshapes, which are then further refined by our carefully designed optimization module to generate identity-preserving 3D caricature blendshapes.

3.2.1 Preliminary. **3D Caricature Reconstruction** refers to the process of generating 3D caricature model from 2D information. Different from previous deformation-based techniques requiring existing 3D caricature template, Wu [43] proposed an optimization approach that can reconstruct 3D caricature from sparse 2D landmarks by using intrinsic deformation representation [11]. Assuming the original caricature landmarks $\mathbf{L}(\mathbf{e_x}) \in \mathbb{R}^{2\times K}$ labelled on image plane, the method [43] can output the caricature 3D mesh $\mathbf{S}(\mathbf{e_x}) \in \mathbb{R}^{3\times N}$ with associated camera parameters $\mathbf{\Pi}, \mathbf{R}, \mathbf{t}$. The orthographic relationship between 2D and 3D can be written as:

$$\mathbf{q} = \mathbf{\Pi R p} + \mathbf{t}, \tag{1}$$

where $\mathbf{\Pi} \in \mathbb{R}^{2\times 3}$ is the scaling matrix, $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is the rotation matrix, $\mathbf{t} \in \mathbb{R}^{2\times 1}$ is the translation vector in image plane, $\mathbf{p}$ and $\mathbf{q}$ are the locations of a 3D dense mesh vertex in the world coordinate system and the image plane, respectively. K and N are vertex numbers of 2D sparse landmarks and 3D dense mesh.

Photo-Caricature Landmark Mapping serves as cycle translation functions between normal face and caricature domain. In order to automatically generate caricature images from human portraits, Cao [7] proposed a cycleGAN based method to translate facial landmarks across two domains. In our problem, we find that such a cross mapping function bridging caricatures and normal faces can

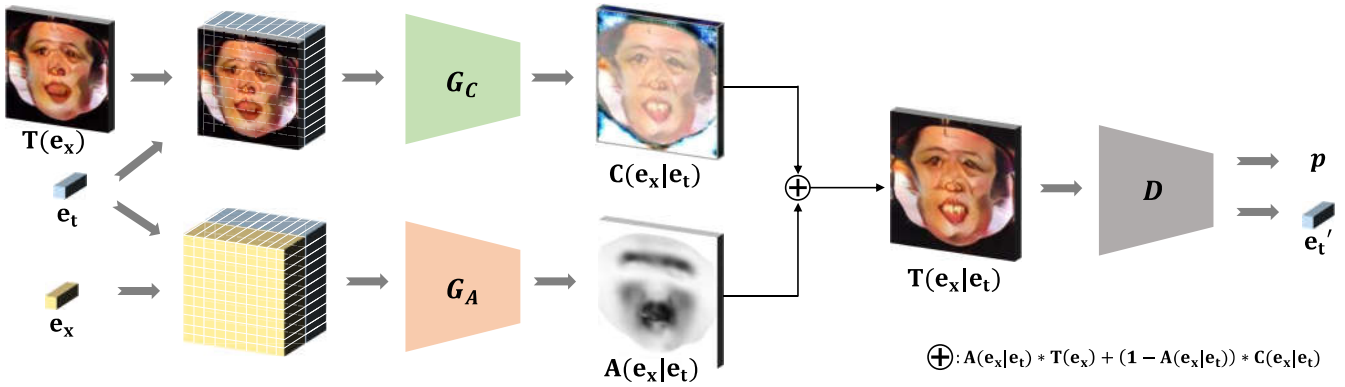

Figure 2: The conditional GAN based dynamic texture generation network. The inputs include the original caricature texture map $\mathrm{T}(\mathbf{e_x})$, the original expression $\mathbf{e_x}$ and the target expression $\mathbf{e_t}$, and the output is the generated new texture map $\mathrm{T}(\mathbf{e_x}|\mathbf{e_t})$. The two-branch generator estimates color and attention masks separately, where the former suggests what color values should be changed to and the latter indicates where to make changes. The discriminator is to judge whether the generated texture map looks real or not as well as predicting its expression $\mathbf{e}'_\mathbf{t}$, which is to be close to $\mathbf{e_t}$.

be used to extend the human facial expression representation like 3DMM. In particular, we firstly transfer the original caricature landmarks $\mathbf{L}(\mathbf{e_x})$ to normal face domain as $\tilde{\mathbf{L}}(\mathbf{e_x}) = \mathcal{F}^+(\mathbf{L}(\mathbf{e_x}))$. Then by fitting a 3DMM model on $\tilde{\mathbf{L}}(\mathbf{e_x})$ and manipulating expression parameters to the pre-defined expression as $\mathbf{e_i}$, we project the new 3D landmark points to image plane as $\tilde{\mathbf{L}}(\mathbf{e_i})$. Finally, we map the generated normal facial landmarks back to caricature domain as $\mathbf{L}(\mathbf{e_i}) = \mathcal{F}^-(\tilde{\mathbf{L}}(\mathbf{e_i}))$. $\mathcal{F}^+$ and $\mathcal{F}^-$ are the bi-directional translation networks adopted from [7]. Through this way, we can modify the underlying expression of the original caricature landmarks.

3.2.2 Blendshape Construction. Based on the mapped landmarks in normal face demain, i.e. $\tilde{\mathbf{L}}(\mathbf{e_x}) = \mathcal{F}^+(\mathbf{L}(\mathbf{e_x}))$, we firstly approximate the PCA-based expression and identity coefficients within 3DMM to fit the 3D face model as:

$$\tilde{\mathbf{S}}(\mathbf{e_x}) = \tilde{\mathbf{S}}_{mean} + \mathbf{U}_{id} \cdot \alpha_{id} + \mathbf{U}_{exp} \cdot \alpha_{exp}, \tag{2}$$

where $\tilde{\mathbf{S}}_{mean}$ is the mean face shape, and $\mathbf{U}_{id}$ and $\mathbf{U}_{exp}$ are the PCA bases for identity and expression, respectively. Accordingly, the expression of the source caricature can be directly defined as the expression parameters in Eq. 2, i.e., $\mathbf{e_x} = \alpha_{exp}$.

Next we select a group of existing blendshapes from FaceWarehouse [6] dataset and extract their expression parameters as $\{\mathbf{e_i}\}_{i=0}^{46}$. By replacing the expression parameters α_{exp} with $\mathbf{e_i}$, we deform $\tilde{\mathbf{S}}(\mathbf{e_x})$ to

$$\tilde{\mathbf{S}}(\mathbf{e_i}) = \tilde{\mathbf{S}}_{mean} + \mathbf{U}_{id} \cdot \alpha_{id} + \mathbf{U}_{exp} \cdot \mathbf{e_i}, \tag{3}$$

and project the landmarks back to image plane as $\{\tilde{\mathbf{L}}(\mathbf{e_i})\}_{i=0}^{46}$.

Finally, we use the inverse mapping function $\mathcal{F}^-$ to transfer the manipulated landmarks $\tilde{\mathbf{L}}(\mathbf{e_i})$ to caricature domain as $\mathcal{F}^-(\tilde{\mathbf{L}}(\mathbf{e_i})) = \mathbf{L}(\mathbf{e_i})$. Following the caricature reconstruction method [43], we obtain a group of 3D caricature models $\{\mathbf{S}(\mathbf{e_i})\}_{i=0}^{46}$ from landmarks $\{\mathbf{L}(\mathbf{e_i})\}_{i=0}^{46}$ that convey the faithful expression $\{\mathbf{e_i}\}_{i=0}^{46}$.

The full process of initializing the caricature blendshape model is depicted in Fig. 1. It can be observed that despite the expression semantics of the generated caricature model $\mathbf{S}(\mathbf{e_i})$, the identity consistency, however, might be violated significantly. The reasons could be: (a) In the photo-caricature landmark translation process, there is no guarantee that the landmarks of different expressions of the same person shall be mapped consistently along with the identity information; (b) Reconstructing 3D caricature model from sparse landmarks is an ill-posed problem, where the 3D identical information could be easily affected by expression deformation in a low-dimensional space. To address this issue, next we propose an optimization module to refine the initial caricature models.

3.2.3 Blendshape Optimization. In order to refine the initial caricature blendshapes, we design our optimization energies with two objectives: (i) reducing over-exaggeration; (ii) preserving the structural correlations among blendshape sets.

For the first objective, we adopt the *handle-based surface editing* [35] to formulate an energy term $\mathbb{E}_{def}$. The key idea is to penalize the over-exaggerated area by evaluating their corresponding displacements in normal face models. Denote the residual blendshapes of $\{\mathbf{S}(\mathbf{e_i})\}_{i=0}^{46}$ and $\{\tilde{\mathbf{S}}(\mathbf{e_i})\}_{i=0}^{46}$ by

$$\mathbf{D}(\mathbf{e_i}) = \mathbf{S}(\mathbf{e_i}) - \mathbf{S}(\mathbf{e_x}),\ i = 0, 1, ..., 46, \tag{4}$$

$$\tilde{\mathbf{D}}(\mathbf{e_i}) = \tilde{\mathbf{S}}(\mathbf{e_i}) - \tilde{\mathbf{S}}(\mathbf{e_x}),\ i = 0, 1, ..., 46. \tag{5}$$

Then we compute a displacement intensity mask $\mathbf{M}(\mathbf{e_i}) = \{\mathbf{m_i^n}\}_{n=0}^{N}$ by normalizing each vertex displacement vector $\tilde{\mathbf{d}}_i^n \in \tilde{\mathbf{D}}(\mathbf{e_i})$ into a scalar $\mathbf{m}_i^n \in [0, 1]$:

$$\mathbf{m}_i^n = \frac{\|\tilde{\mathbf{d}}_i^n\|}{\max_{1 \le n \le N} \|\tilde{\mathbf{d}}_i^n\|},\ n = 0, 1, ..., N, \tag{6}$$

where N is the number of dense mesh vertices for both $\tilde{\mathbf{S}}(\mathbf{e_i})$ and $\mathbf{S}(\mathbf{e_i})$, $i = 0, .., 46$.

Let $\mathbf{D}^\star = \{\mathbf{D}^\star(\mathbf{e_i}\}_{i=0}^{46}$ be the optimizing target. We define $\mathbb{E}_{def}$ as a weighted Laplacian deformation term:

$$\mathbb{E}_{def}(\mathbf{D}^\star) = \sum_{i=0}^{46} \|\Delta(\mathbf{D}^\star(\mathbf{e_i}) - \mathbf{M}(\mathbf{e_i}) \cdot \mathbf{D}(\mathbf{e_i}))\|^2. \tag{7}$$

In practice, we employ the standard cotangent weights [30] for the Laplacian operator $\Delta(*)$ defined on discrete mesh vertices. Eq.7

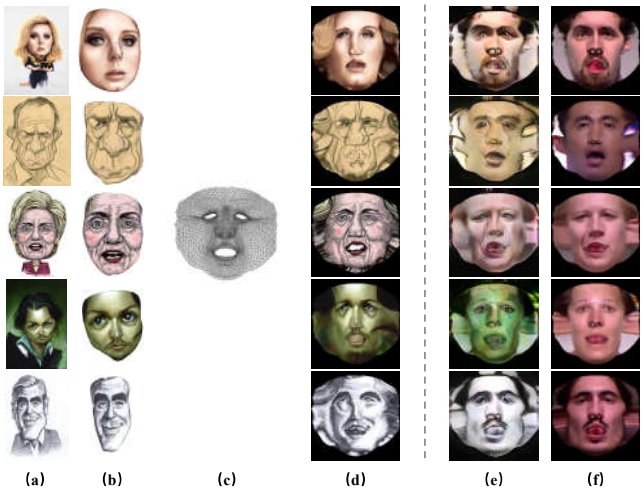

Figure 3: Selected texture training data. (a)-(d) Source caricature image, model, ARAP parameterization and extracted texture; (e)(f) Style-transferred and original human face textures. By transferring image styles, we augment the training dataset with normal face textures.

essentially is to use the deformation in normal face domain to guide the deformation in caricature domain so as to penalize undesired over-exaggeration.

For the second objective, our target is to constrain the two parallel blendshapes $\{\mathbf{S}(\mathbf{e_i})\}_{i=0}^{46}$ and $\{\tilde{\mathbf{S}}(\mathbf{e_i})\}_{i=0}^{46}$ to keep similar group structure. Inspired by [20], we use the cosine distance to evaluate the similarities among normal face blendshapes:

$$cos(\mathbf{e_i}, \mathbf{e_j}) = \frac{\tilde{\mathbf{D}}(\mathbf{e_i}) \cdot \tilde{\mathbf{D}}(\mathbf{e_j})}{\|\tilde{\mathbf{D}}(\mathbf{e_i})\| \cdot \|\tilde{\mathbf{D}}(\mathbf{e_j})\|}, \ i, j \in \{0, 1, ..., 46\}. \tag{8}$$

Then the correlation weights are applied on caricature blendshapes to enforce the same interior similarities, which leads to our second energy term:

$$\mathbb{E}_{str}(\mathbf{D}^\star) = \sum_{i=0}^{46}\sum_{j=0}^{46} cos(\mathbf{e_i}, \mathbf{e_j})\|\mathbf{D}^\star(\mathbf{e_i}) - \mathbf{D}^\star(\mathbf{e_j})\|^2. \tag{9}$$

Finally, we add energy term $\mathbb{E}_{smo}$ to improve the local smoothness for optimized caricature blendshapes:

$$\mathbb{E}_{smo}(\mathbf{D}^\star) = \sum_{i=0}^{46} \|\Delta(\mathbf{D}^\star(\mathbf{e_i}))\|^2. \tag{10}$$

Summing the three terms up gives our full objective function of caricature blendshape optimization:

$$\mathbb{E}_{total} = \lambda_{def}\mathbb{E}_{def} + \lambda_{str}\mathbb{E}_{str} + \lambda_{smo}\mathbb{E}_{smo}, \tag{11}$$

where λ_{def}, λ_{str} and λ_{smo} are trade-off weights. The entire optimization can be solved linearly. Once $\mathbf{D}^\star$ is optimized, the optimized caricature blendshapes $\mathbf{S}^\star$ can be obtained by adding the source model back according to Eq. 4.

3.3 Texture Generation Component

As pointed out in Sec.3.1, only manipulating the underlying geometric shapes is in general not sufficient to generate realistic caricature expressions. With expression changes, the face images will contain varying transformations in color space, especially in the areas of mouth and eyes. In order to model the specific color transformation, we need to modify the texture map according to the target expression. Let $\mathbf{T}(\mathbf{e_x}) \in \mathbb{R}^{H\times W\times 3}$ denote source caricature texture and $\mathbf{e_t} \in \mathbb{R}^D$ be the target expression. The texture model $\mathcal{M}$ seeks to estimate the target texture map as $\mathbf{T}(\mathbf{e_t}) = \mathcal{M}(\mathbf{T}(\mathbf{e_x}), \mathbf{e_t})$.

Toward this objective, we propose a conditional GAN model. However, training image-based generative networks requires a large-scale dataset, which we do not have for caricatures. To overcome this issue, we augment our caricature texture training data by adapting normal human face textures to various caricature styles. In this way, our network can be trained on both adapted texture dataset (in a supervised manner) and real caricature texture dataset (in an unsupervised manner).

3.3.1 Network Structure. Fig. 2 shows the overall network structure of our texture generation model $\mathcal{M}$, which is composed of a two-branch generator G and a discriminator D. To condition the texture generation results on different expressions, we randomly pair a source caricature and a target expression as $(\mathbf{T}(\mathbf{e_x}), \mathbf{e_t})$. Following the attention mechanism design in [31], the generator $G = (G_A, G_C)$ takes as input $(\mathbf{T}(\mathbf{e_x}), \mathbf{e_t})$ and estimates two masks: color mask $\mathbf{C} \in R^{H\times W\times 3}$ and attention mask $\mathbf{A} \in [0, 1]^{H\times W\times 1}$, where the former suggests what color values should be changed to and the latter indicates where to make changes. The final regressed target texture $\mathbf{T}(\mathbf{e_x}|\mathbf{e_t})$ is computed by:

$$\mathbf{T}(\mathbf{e_x}|\mathbf{e_t}) = \mathbf{A} \cdot \mathbf{T}(\mathbf{e_x}) + (1 - \mathbf{A}) \cdot C. \tag{12}$$

Different from the *GANimation* model [31], which operates in image domain, our texture generation operates on texture map $\mathbf{T}$, which is universally aligned and focuses on texture attributes only. Note that our attention estimation branch is based on source and target expression coefficients and does not need to consider the color information of $\mathbf{T}(\mathbf{e_x})$.

The discriminator D is responsible for evaluating the expression of the generated texture map and discriminating real and fake textures. In one branch of D, a Patch-GAN [21] instance will map an input image to overlapping patches and output real probability for each patch. The other branch of D will regress the output texture to estimate its expression.

3.3.2 Training. **Data Acquisition**. Our training data comes from two sources. The first one is from real caricature images. We collect thousands of caricature paintings from internet and reconstruct their 3D shapes with manually labeled landmarks. Then by applying rasterization pipeline on a pre-defined ARAP parameterization map, we obtain their individual textures. Some selected examples are shown in Fig. 3. Due to the big data requirement of deep learning methods, we also explore the possibility of augmenting our training data from normal face textures. We use the *Adaptive Instance Normalization* [18] approach to transfer normal face textures in FaceWarehouse [6] dataset into random caricature style. The augmented data samples are also given in Fig. 3. It can be observed that in this way, the transferred textures look very similar to the original samples.

Supervised Training. In FaceWarehouse [6] dataset, there are multiple expressions of the same person. We make use of this feature to train the texture model in a supervised manner with source

texture map $\mathbf{T}(\mathbf{e_x})$, target texture map $\mathbf{T}(\mathbf{e_t})$ and 'ground-truth' attention map $\mathbf{A}(\mathbf{e_x}, \mathbf{e_t})$, which is obtained by mapping the vertex displacement map computed from the corresponding 3D shapes to the texture map space. In other words, a large 3D vertex displacement at a 3D vertex suggests likely texture color change at the corresponding texture map location.

We train the generator G with two loss functions. The first one is attention mask regression loss. The generated attention $\mathbf{A}(\mathbf{e_x}|\mathbf{e_t})$ should be regressed close to the pre-computed map $\mathbf{A}(\mathbf{e_x}, \mathbf{e_t})$:

$$\mathbb{L}_{att} = \|\mathbf{A}(\mathbf{e_x}|\mathbf{e_t}) - \mathbf{A}(\mathbf{e_x}, \mathbf{e_t})\|_1. \tag{13}$$

The second loss is color transformation loss. The generated texture map $\mathbf{T}(\mathbf{e_x}|\mathbf{e_t})$ combined by Eq. 12 should be close to the groundtruth $\mathbf{T}(\mathbf{e_t})$:

$$\mathbb{L}_{color} = \|\mathbf{T}(\mathbf{e_x}|\mathbf{e_t}) - \mathbf{T}(\mathbf{e_t})\|_1. \tag{14}$$

Unsupervised Training. For real caricature data, there is no paired expressions of the same identity. Thus, we do not have groundtruth for either attention mask or output texture. So we replace the supervised training loss with cycle reconstruction loss:

$$\mathbb{L}_{cycle} = \|G(G(\mathbf{T}(\mathbf{e_x})|\mathbf{e_t})|\mathbf{e_x}) - \mathbf{T}(\mathbf{e_x})\|_1, \tag{15}$$

which means that the generator G is expected to transform source texture map $\mathbf{T}(\mathbf{e_x})$ under target condition $\mathbf{e_t}$ and then transform it back under source condition $\mathbf{e_x}$.

Discriminator. The generators in both scenarios need to cooperate with the discriminator for adversarial training. D is trained with two discriminative loss functions.

The first one is image adversarial loss $\mathbb{L}_{adv}$,

$$\mathbb{L}_{adv} = \mathbb{E}_{\mathbf{T}_f \sim \mathbf{P}_f}[D(\mathbf{T}_f)] - \mathbb{E}_{\mathbf{T}_r \sim \mathbf{P}_r}[D(\mathbf{T}_r)] + \lambda_{gp}\mathbb{E}_{GP}, \tag{16}$$

which aims to maximize the probability of classifying the real sample $\mathbf{T}_r \sim \mathbf{P}_r$ and the fake generation $\mathbf{T}_f \sim \mathbf{P}_f$ based on the continuous Earth Mover Distance [1], and λ_{gp} is the weight for gradient penalty term $\mathbb{E}_{GP}$ in [1]. The generator G tries to fool D simultaneously.

The second one is expression condition loss $\mathbb{L}_{exp}$, which forces the network to minimize the error between target expression coefficients and regressed ones:

$$\mathbb{L}_{exp} = \|D(G(\mathbf{T}(\mathbf{e_x})|\mathbf{e_t})) - \mathbf{e_t}\|^2 + \|D(\mathbf{T}(\mathbf{e_x})) - \mathbf{e_x}\|^2. \tag{17}$$

Finally, the total loss function becomes:

$$\mathbb{L} = \lambda_{att}\mathbb{L}_{att} + \lambda_{color}\mathbb{L}_{color} + \lambda_{cycle}\mathbb{L}_{cycle} + \lambda_{adv}\mathbb{L}_{adv} + \lambda_{exp}\mathbb{L}_{exp}. \tag{18}$$

All λ are hyper-parameters to balance training weights and switch supervised/unsupervised training mode.

4 EXPERIMENTS

4.1 Implementation Details

The computation of our shape modelling component is conducted on a PC with Intel Xeon W-2133 CPU, 16GB RAM. For caricature blendshape optimization, we set λ_{def} as 1.0, λ_{str} as 0.1 and λ_{smo} as 0.05. The full computation of the shape modelling component costs about 1 second. The dynamic texture model is trained on an NVIDIA TITAN V graphics card for about 10 hours. We first train the model in a supervised manner with $\lambda_{att} = 10.0, \lambda_{color} = 10.0, \lambda_{cycle} = 0.0, \lambda_{adv} = 1.0$ and $\lambda_{exp} = 100.0$. After the attention mask loss converges, we switch to unsupervised training with $\lambda_{att} = 0.0, \lambda_{color} = 0.0$ and $\lambda_{cycle} = 10.0$. In total, we trained the network for 600 epochs with batch size of 40 and learning rate of 1e-5.

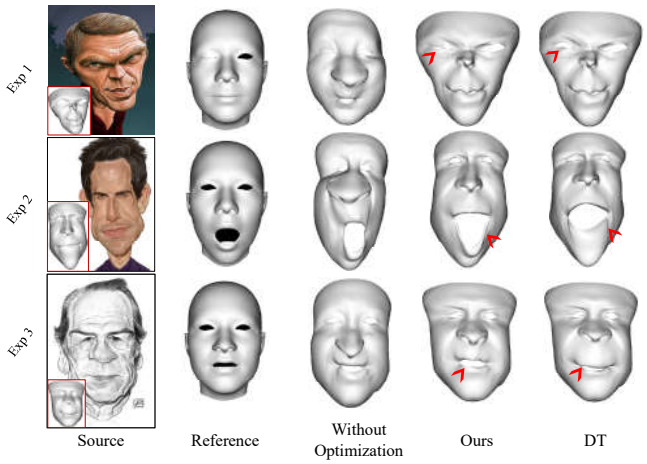

Figure 4: Ablation study on our blendshape optimization method and comparison with *Deformation transfer* [37] (DT). For each source caricature shape in the left column, we show the generated blendshapes of different methods in columns 3-5 with reference to human expressions in the second column. The results clearly indicate that without optimization, the caricature identity will be lost. The red arrows amplify the key different areas between *deformation transfer* and our results. Our method achieves more exaggerated and accurate expression deformations than DT.

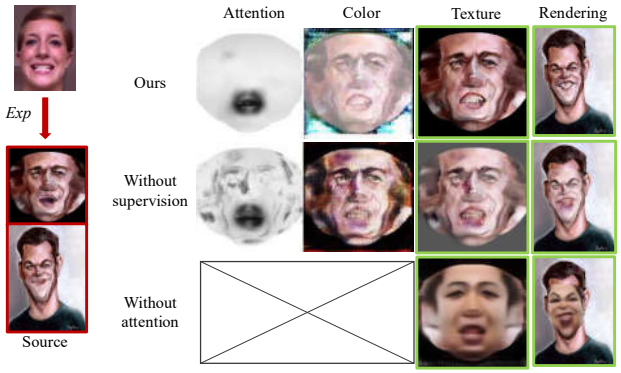

Figure 5: Ablation study on the texture generation component. We compare our method with models trained without fully supervised data or without the attention module. The target expression of mouth-open requires the texture model to generate the teeth part in the mouth area.

4.2 Ablation Study

4.2.1 Caricature Blendshapes. To evaluate the caricature shape model, we compare the blendshapes generated with and without

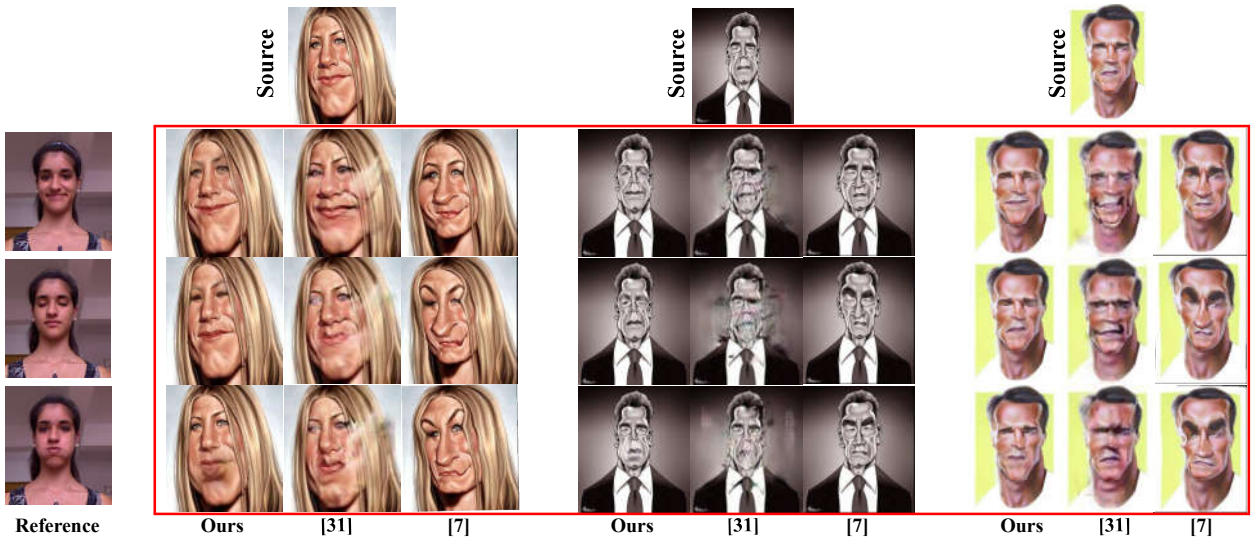

Figure 6: Comparing our method with GANimation [31] and CariGeoGAN [7]. The reference photos guide the caricature generation by giving 3DMM expression coefficients to [7] and our model or giving Action Unit code to GANimation [31]. The comparisons show that our results preserve better image quality than [31] and maintain more stable identity than [7]. Please zoom in for details.

our optimization approach. Besides, we also compare with *deformation transfer* [37], which uses template model of source and target objects to compute the shape correspondence. Then by directly transferring the deformation of source object, it can generate deformed results for target. In experiments, we make the fitted 3DMM model $\tilde{S}(\mathbf{e_x})$ as source template and caricature model $S(\mathbf{e_x})$ as target template. The other 3DMM expressions $\{\tilde{S}(\mathbf{e_i})\}$ give the deformation reference to drive target caricature shapes.

Fig. 4 shows some comparison results generated with/without optimization and *deformation transfer*. It can be observed that without optimization, although the generated caricature shapes can maintain reference expressions, they lose source identity information significantly. In contrast, our method preserves the identitiy information well. Compared with *deformation transfer*, our method achieves more exaggerated and accurate expression deformations. This is because *deformation transfer* only copies deformations of normal face shapes and they are not strong enough to animate caricature shapes. In contrast, our method adopts the cross-domain translation so that we can handle the exaggeration by learning with various expressions of different caricatures as domain knowledge.

4.2.2 Dynamic Texture Generation. We evaluate our trained conditional texture generation network with two ablation studies. First, we remove the adapted texture dataset from normal face textures, which leaves the entire training to be purely unsupervised. Specifically, we keep the unsupervised losses $\mathbb{L}_{cycle}, \mathbb{L}_{adv}$ and $\mathbb{L}_{exp}$ and train the network only on original caricature texture dataset. The visual comparison in Fig. 5 proves the significance of the adapted dataset and the combining training scheme. Without the supervised training, the conditional GAN model is hard to predict satisfactory color transformation and attention mask.

The second study is on the attention mechanism. We compare our model with and without the attention branch. Note that removing the attention branch makes our model degenerate to a vanilla cGAN network. Fig. 5 shows that the results of our model without the attention is poor.

The main reason is that the generator is hard to handle various image styles and color transformations simultaneously and the attention mechanism can well ease that difficulty by estimating color and attention mask in a separate way.

4.3 Comparisons

In this part, we evaluate our full model through comparison with other competitive methods. Since there is no face editing works proposed specifically for caricatures, we choose two feasible methods, GANimation [31] and CariGeoGAN [7], as our baselines.

4.3.1 Comparison with GANimation [31]. As one of the *state-of-the-art* facial expression editing works, GANimation [31] is capable of generating plausible human portraits conditioned on facial Action Unit [10], which is similar to our problem settings. So we compare with GANimation [31] to see if it can be generalized to caricature images. We directly use the pre-trained model of [31] and make the GANimation and our model target for the same expression reference. Fig. 6 shows some samples generated by both methods. It can be clearly seen that our results are of much better image quality. The reason is that we process caricature images more meticulously with disentangled 3D geometry and texture components, while GANimation is built entirely on 2D domain.

4.3.2 Comparison with CariGeoGAN [7]. CariGeoGAN [7] is cycle consistent mapping which translates landmarks between normal face and caricature. As aforementioned, the caricature landmarks can be translated to normal face landmarks by [7], which can be used to fit 3DMM model with expression parameters. The 3DMM-based blendshapes are then translated back to caricature domain to form a group of caricature landmarks, which are capable of

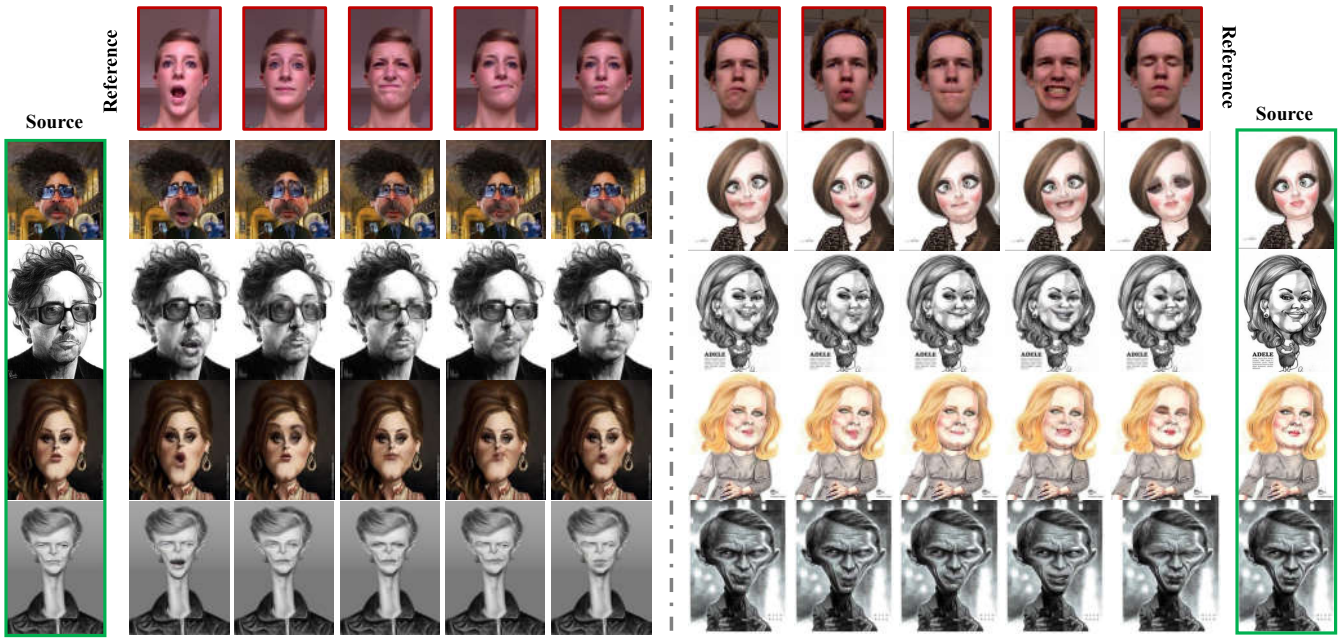

Figure 7: Rendering expression editing examples. Images in first row (red block) are references that specify target expressions. For each caricature image, we only need to generate its optimized blendshapes once and they can be used to produce many caricature images with new expressions by simple blending specified in Eq. 19. Our model can handle various expressions such as open-mouth, pouting, contempt and closing-eyes. Please zoom in for details.

composing arbitrary new expressions given 3DMM parameters. Consequently, we use 2D thin-plate spline warping [3] to deform source caricature images guided by target landmarks. The examples illustrated in Fig. 6 show that this baseline could not preserve the identity information compared with our method. The underlying explanation is that our shape model is optimized on dense mesh vertices but CariGeoGAN [7] only warps the image with sparse landmarks, which likely results in losing identity information in such low-dimensional space.

4.4 More Results

Our caricature expression model can be used to edit the expressions of any given caricature painting. Since the shape and texture components are conditioned upon the same facial expression representation (3DMM model), our method can be easily integrated with previous 3DMM-based human face reconstruction works by extracting their expression coefficients and approximating blendshape weights.

For a caricature, once its blendshapes $\{\mathbf{S}(\mathbf{e_i})\}_{i=0}^{46}$ are optimized, it can be used to easily compute another shape under random input expression parameters $\mathbf{e_r}$ by regressing blendshape weights $\mathbf{w_r} \in \mathbb{R}^{46}$ on corresponding normal face blendshapes $\{\tilde{\mathbf{S}}(\mathbf{e_i})\}_{i=0}^{46}$ and mapping $\mathbf{w_r}$ to caricature as:

$$\begin{aligned} &\arg\min_{\mathbf{w_r}} \|\tilde{\mathbf{S}}(\mathbf{e_r}) - \tilde{\mathbf{S}}(\mathbf{e_0}) - \sum_{i=1}^{46} \mathbf{w_r^i}(\tilde{\mathbf{S}}(\mathbf{e_i}) - \tilde{\mathbf{S}}(\mathbf{e_0}))\|^2 \\ &\mathbf{S}(\mathbf{e_r}) = \mathbf{S}(\mathbf{e_0}) + \sum_{i=1}^{46} \mathbf{w_r^i}(\mathbf{S}(\mathbf{e_i}) - \mathbf{S}(\mathbf{e_0})) \end{aligned} \tag{19}$$

The texture map can be quickly inferred by applying the trained texture generation network as $\mathbf{T}(\mathbf{e_r}) = G(\mathbf{T}(\mathbf{e_x})|\mathbf{e_r})$. Finally, we reuse the camera projection matrix approximated in Sec. 3.2.1 to render new caricature images. Fig. 7 shows a gallery of expression editing results generated from real caricature paintings. More results can be found in the supplementary material.

5 CONCLUSION

We have presented a method to model caricature expression, which extends traditional 3DMM representation to caricature domain. The method has shape and texture components, for which geometric optimization and deep learning methods are developed, respectively. By the method, a group of 3D blendshapes and a dynamic texture generator are cooperated so that one can transform a caricature image into arbitrary expressions. Hence one can easily manipulate the expressions of artist-drawn caricature painting while preserving the exaggerated identity. In future, more applications such as video-driven caricature animation will be developed with sequential consistency constraints.

ACKNOWLEDGMENTS

This research is partially supported by the National Natural Science Foundation of China (No. 61672481), Youth Innovation Promotion Association CAS (No. 2018495), Zhejiang Lab (NO. 2019NB0AB03), NTU DSAIR grant (No. 04INS000518C130), the National Research Foundation, Singapore under its International Research Centres in Singapore Funding Initiative, and Monash FIT Start-up Grant.

REFERENCES

[1] Martín Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein Generative Adversarial Networks. *ICML, 2017.*

[2] Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. *SIGGRAPH 1999.*

[3] Fred L. Bookstein. 1989. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989), 567–585.

[4] Susan E. Brennan. 1985. Caricature Generator: The Dynamic Exaggeration of Faces by Computer. *Leonardo* 18 (1985), 170 – 178.

[5] Chen Cao, Qiming Hou, and Kun Zhou. 2014. Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Transactions on graphics (TOG)* 33 (2014), 43:1–43:10.

[6] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2014. Facewarehouse: A 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2014), 413–425.

[7] Kaidi Cao, Jing Liao, and Lu Yuan. 2018. CariGANs: unpaired photo-to-caricature translation. *ACM Transactions on graphics (TOG)* 37 (2018), 244:1–244:14.

[8] Lyndsey Clarke, Min Chen, and Benjamin Mora. 2011. Automatic Generation of 3D Caricatures Based on Artistic Deformation Styles. *IEEE Transactions on Visualization and Computer Graphics* 17 (2011), 808–821.

[9] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. 2019. Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set. *IEEE Computer Vision and Pattern Recognition Workshops, 2019.*

[10] Paul Ekman and Wallace V. Friesen. 1978. Facial action coding system: a technique for the measurement of facial movement. *Consulting Psychologists Press* (1978).

[11] Lin Gao, Yu-Kun Lai, Jie Yang, Ling-Xiao Zhang, Leif Kobbelt, and Shihong Xia. 2019. Sparse Data Driven Mesh Deformation. *IEEE transactions on visualization and computer graphics* (2019).

[12] Baris Gecer, Stylianos Ploumpis, Irene Kotsia, and Stefanos Zafeiriou. 2019. GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019* (2019), 1155–1164.

[13] Zhenglin Geng, Chen Cao, and Sergey Tulyakov. 2019. 3D Guided Fine-Grained Face Manipulation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).

[14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. *NIPS 2014.*

[15] Brian Guenter, Cindy Grimm, Daniel Wood, Henrique Malvar, and Fredric Pighin. 1998. Making Faces. *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques.*

[16] Yudong Guo, Juyong Zhang, Jianfei Cai, Boyi Jiang, and Jianmin Zheng. 2019. CNN-Based Real-Time Dense Face Reconstruction with Inverse-Rendered Photo-Realistic Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2019), 1294–1307.

[17] Xiaoguang Han, Kangcheng Hou, Dong Du, Yuda Qiu, Yizhou Yu, Kun Zhou, and Shuguang Cui. 2018. CaricatureShop: Personalized and Photorealistic Caricature Sketching. *IEEE Transactions on Visualization and Computer Graphics* (2018).

[18] Xun Huang and Serge J. Belongie. 2017. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. *IEEE International Conference on Computer Vision (ICCV)* (2017), 1510–1519.

[19] Jing Huo, Wenbin Li, Yinghuan Shi, Yang Gao, and Hujun Yin. 2018. WebCaricature: a benchmark for caricature recognition. *BMVC 2018.*

[20] Roger Blanco i Ribera, Eduard Zell, John P. Lewis, Jun yong Noh, and Mario Botsch. 2017. Facial retargeting with automatic range of motion alignment. *ACM Transactions on graphics (TOG)* 36 (2017), 154:1–154:12.

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 5967–5976.

[22] Luo Jiang, Juyong Zhang, Bailin Deng, Hao Li, and Ligang Liu. 2018. 3D Face Reconstruction With Geometry Details From a Single Image. *IEEE Transactions on Image Processing* 27 (2018), 4756–4770.

[23] John P. Lewis, Ken ichi Anjyo, Taehyun Rhee, Mengjie Zhang, Frédéric H. Pighin, and Zhigang Deng. 2014. Practice and Theory of Blendshape Facial Models. *Eurographics, 2014.*

[24] Wenbin Li, Wei Xiong, Haofu Liao, Jing Huo, Yang Gao, and Jiebo Luo. 2018. CariGAN: Caricature Generation through Weakly Paired Adversarial Learning. *ArXiv* abs/1811.00445 (2018).

[25] Pei-Ying Chiang Wen-Hung Liao and Tsai-Yen Li. 2004. Automatic caricature generation by analyzing facial features. *Asian Conference on Computer Vision (ACCV), 2004* 2.

[26] Ligang Liu, Lei Zhang, Yin Xu, Craig Gotsman, and Steven J. Gortler. 2008. A Local/Global Approach to Mesh Parameterization. *Comput. Graph. Forum* 27 (2008), 1495–1504.

[27] Ghulam Mustafa, Hao Li, Juyong Zhang, and Jiansong Deng. 2015. l1-Regression based subdivision schemes for noisy data. *Computer-Aided Design* 58 (2015), 189–199.

[28] Alice J O'Toole, Theodore Price, Thomas Vetter, James C Bartlett, and Volker Blanz. 1999. 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing* 18, 1 (1999), 9–19.

[29] Alice J. O'Toole, Thomas Vetter, Harald Volz, and Elizabeth M. Salter. 1997. Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age. *Perception* 26 6 (1997), 719–32.

[30] Ulrich Pinkall and Konrad Polthier. 1993. Computing Discrete Minimal Surfaces and Their Conjugates. *Experimental Mathematics* 2 (1993), 15–36.

[31] Albert Pumarola, Antonio Agudo, Aleix M. Martínez, Alberto Sanfeliu, and Francesc Moreno-Noguer. 2018. GANimation: Anatomically-aware Facial Animation from a Single Image. *European Conference on Computer Vision (ECCV)* 11214 (2018), 835–851.

[32] Suriati Bte Sadimon, Mohd Shahrizal Sunar, Dzulkifli Bin Mohamad, and Habibollah Haron. 2010. Computer Generated Caricature: A Survey. *International Conference on Cyberworlds, 2010* (2010), 383–390.

[33] Matan Sela, Yonathan Aflalo, and Ron Kimmel. 2015. Computational caricaturization of surfaces. *Computer Vision and Image Understanding* 141 (2015), 1–17.

[34] Yichun Shi, Debayan Deb, and Anil K. Jain. 2018. WarpGAN: Automatic Caricature Generation. *CVPR 2018.*

[35] Olga Sorkine-Hornung, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and Hans-Peter Seidel. 2004. Laplacian surface editing. *SGP '04.*

[36] Marco Stricker, Olivier Augereau, Koichi Kise, and Motoi Iwata. 2018. Facial Landmark Detection for Manga Images. *ArXiv* abs/1811.03214 (2018).

[37] Robert W. Sumner and Jovan Popovic. 2004. Deformation transfer for triangle meshes. *ACM Trans. Graph.* 23 (2004), 399–405.

[38] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2018. Face2Face: real-time face capture and reenactment of RGB videos. *Commun. ACM* 62 (2018), 96–104.

[39] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard G. Medioni. 2017. Regressing Robust and Discriminative 3D Morphable Models with a Very Deep Neural Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 1493–1502.

[40] Evangelos Ververas and Stefanos Zafeiriou. 2019. SliderGAN: Synthesizing Expressive Face Images by Sliding 3D Blendshape Parameters. *ArXiv* abs/1908.09638 (2019).

[41] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popovic. 2005. Face transfer with multilinear models. *ACM Trans. Graph.* 24 (2005), 426–433.

[42] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2011. Realtime performance-based facial animation. *ACM transactions on graphics (TOG)* 30, 4, 77.

[43] Qianyi Wu, Juyong Zhang, Yu-Kun Lai, Jianmin Zheng, and Jianfei Cai. 2018. Alive Caricature from 2D to 3D. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), 7336–7345.

[44] Jordan Yaniv, Yael Newman, and Ariel Shamir. 2019. The face of art: landmark detection and geometric style in portraits. *ACM Transactions on graphics (TOG)* 38 (2019), 60:1–60:15.

[45] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision (ICCV)* (2017), 2242–2251.

[46] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z. Li. 2016. Face Alignment Across Large Poses: A 3D Solution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 146–155.

[47] C. Çetinaslan, J. Lewis, and V. Orvalho. 2017. Transposition Based Blendshape Direct Manipulation. *International Conf. on Computer Graphics Theory and Applications 2017.*