Computers & Graphics (2024)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

AnaConDaR: Anatomically-Constrained Data-Adaptive Facial Retargeting

Nicolas Wagner[a], Ulrich Schwanecke[b], Mario Botsch[a]

[a] *TU Dortmund University, Dortmund, Germany*
[b] *University of Applied Sciences RheinMain, Wiesbaden, Germany*

ARTICLE INFO

Article history:

Keywords: Facial Animation, Offline Performance Retargeting, Physics-based Simulation

ABSTRACT

Offline facial retargeting, i.e., transferring facial expressions from a source to a target character, is a common production task that still regularly leads to considerable algorithmic challenges. This task can be roughly dissected into the transfer of sequential facial animations and non-sequential blendshape personalization. Both problems are typically solved by data-driven methods that require an extensive corpus of costly target examples. Other than that, geometrically motivated approaches do not require intensive data collection but cannot account for character-specific deformations and are known to cause manifold visual artifacts.

We present AnaConDaR, a novel method for offline facial retargeting, as a hybrid of data-driven and geometry-driven methods that incorporates anatomical constraints through a physics-based simulation. As a result, our approach combines the advantages of both paradigms while balancing out the respective disadvantages. In contrast to other recent concepts, AnaConDaR achieves substantially individualized results even when only a handful of target examples are available. At the same time, we do not make the common assumption that for each target example a matching source expression must be known. Instead, AnaConDaR establishes correspondences between the source and the target character by a data-driven embedding of the target examples in the source domain. We evaluate our offline facial retargeting algorithm visually, quantitatively, and in two user studies.

1. Introduction

Creating high-fidelity facial expressions for human or humanoid characters is one of the most challenging problems in computer graphics applications. To that end, it is common practice to record a source actor with high-resolution motion capture technology and subsequently transfer the scanned expressions to the targeted character either frame-by-frame or via blendshapes [1]. A comprehensive corpus of research focuses on the latter step, the so-called offline facial performance retargeting. While deep learning predominates in various facial animation tasks, here, *more traditional* approaches retain distinct advantages and are commonly used in production [2]. Particularly, due to the still limited availability of high-resolution facial expression meshes for training, the risk of generalization gaps is ubiquitous [3]. The reliance on implicit representations within current neural telepresence applications [4, 5] underscores the lack of suitable training data.

Two main streams of work can be identified within which most of the current non-learning methodologies can be categorized. On the one hand, there are data-driven methods that have access to numerous exemplary facial expressions of the target character and form new expressions by combining these [2, 6]. On the other hand, there are geometry-driven methods that try to

e-mail: nicolas.wagner@tu-dortmund.de (Nicolas Wagner), nicolas.wagner@tu-dortmund.de (Ulrich Schwanecke), mario.botsch@tu-dortmund.de (Mario Botsch)

transfer the geometric deformations of the source actor's face to the target character [7, 8, 9]. Both methodologies offer complementary advantages and disadvantages. For instance, data-driven methods can consider anatomy-specific differences between the source and the target, whereas geometry-driven methods force deformations regardless of the structure of the respective heads. In return, geometry-driven methods do not rely on elaborately recorded or artistically sculpted examples of the target character and are, therefore, usually more efficient than data-driven methods.

Generally, there is a trade-off between the cost and complexity of data acquisition and retargeting quality. When time and effort are not a constraint, establishing extensive corresponding linear blendshape (LBS) systems [1] between the source and target character can be the most reasonable approach to facial retargeting. As such situations rarely occur in reality, the current state-of-the-art Anatomical Local Model (ALM) [2] has been developed. ALM requires a significantly reduced amount of blendshapes due to replacing plain LBS with more expressive patchwise LBS (PLBS). However, the authors point out that insufficiently comprehensive PLBS nonetheless result in severe retargeting artifacts and recognize the limitation that non-corresponding source and target blendshapes are not supported. Similar shortcomings in LBS can partially be overcome by employing example-based facial rigging (EBFR) [6], which supplements the data-driven retargeting with a geometry-driven deformation transfer [7]. Unfortunately, there has not been an adaption to ALM so far.

In this work, we improve on ALM and fill this very gap by introducing AnaConDaR, an anatomically-constrained and data-adaptive facial retargeting. Here, corresponding PLBS systems are derived from the available target examples and used for an initial retargeting in a data-driven manner. The parts that are not explainable by PLBS are retargeted by a novel anatomical deformation transfer (ADT). In a final step, both the PLBS and ADT results are added together and a physics-based simulation ensures anatomical plausibility, also with combined retargeting. Moreover, this simulation enables artistic interventions on material properties, can incorporate external forces, and preserves expression-specific characteristics.

We evaluate AnaConDaR in two user studies and a quantitative comparison. In one user study, we asked the participants to benchmark the state-of-the-art peer group against AnaConDaR, while the other focused on the necessity of individual algorithmic components. Quantitative comparisons of facial retargeting algorithms are generally challenging, as the subjective nature of perceiving facial expressions makes it difficult to establish a definitive ground truth. Therefore, we quantitatively showcase the advantages of AnaConDaR over ALM in a particularly construed retargeting scenario.

The *key novelties and contributions* we present in this paper can be summarized as follows:

- A novel hybrid approach for offline facial performance retargeting that can leverage a small number of target examples.
- A new, fully volumetric deformation transfer for faces, which respects anatomical and physical constraints. During the deformation transfer, expression-specific characteristics are retained.
- Two user studies, a quantitative analysis, and various visual examples that evaluate and showcase AnaConDaR.

2. Related Work

2.1. Facial Retargeting in General

Besides offline performance targeting, there are several other variants of facial retargeting, which are all related but can also be clearly distinguished.

First, the 2D variant in which so-called deep fakes [10, 11, 12, 13, 14, 15, 16, 17, 18] swap faces directly in images almost entirely independent of the underlying geometry [16, 19]. While these works can generate outstanding results, they are hardly artist-controllable, cannot integrate physics-based effects, and lose mesh-based advantages like shading adjustments. Our approach offers all of the features mentioned above.

Second, online performance retargeting algorithms that animate characters in real time. Usually, such methods are either of low quality [1, 20] or need time consuming training on extensive datasets [21, 22, 23, 24, 25, 26, 27]. Our approach can handle high resolutions, is applicable without training, and only requires a handful of expression examples.

Third, more general (neural) face models [28, 29, 30, 31, 32, 3] that capture both human identities and facial expressions in latent spaces. Unfortunately, their generalization capabilities usually do not meet the quality requirements of sophisticated CGI productions [30, 32]. Moreover, many models can only perform the facial retargeting task for low-resolution geometries [29, 33, 28]. Starting from a reversed perspective, the neural physics-based facial animation of Yang et al. [26] has recently been extended into a more comprehensive face model [3]. Nonetheless, this model is severely limited to only a handful of identities and adding a novel identity requires five days of retraining [26]. Further, they expect access to 30 seconds of performance capture per identity while the captured expressions must be semantically aligned. The likewise neural approach Animatomy [25] faces similar problems. Neither of the latter two algorithms [3, 25] was evaluated concerning facial retargeting.

Finally, image-based face avatars primarily work on low-resolution geometries [4, 5] and, hence, do not meet production requirements, as well. Overall, we follow the recent assessment of Chandran et al. [2] that deep learning for facial retargeting still cannot fully compete with *more traditional* techniques.

2.2. Offline Facial Performance Retargeting

As the introduction notes, offline facial performance retargeting without learning can be divided mainly into data-driven and geometry-driven methods. For data-driven methods, linear blendshapes [1] are still the gold standard due to their simplicity and computational speed. Since the nonlinear aspects of facial expressions have a significant influence, a variety of extensions [34, 35, 36] have been developed over the years. Nonetheless, only minor improvements have been achieved, and it remains common practice to model or scan a large number of linear blendshapes to account for nonlinearity. In an effort

to reduce costs, methods have been developed that generate extensive blendshape rigs from just a few exemplary expressions [6, 33]. Often, however, these only exhibit weak personalization. Recently, Chandran et al. [2] demonstrated how to gain more expressiveness from expression samples using piecewise linear blendshapes. To the best of our knowledge, none of the aforementioned data-driven techniques deals with missing information due to insufficient training data. The method we present in this work addresses this problem by combining piecewise linear blendshapes with a geometry-driven approach.

The most widely used geometry-driven facial retargeting approach is deformation transfer [7, 8, 9]. This approach extracts deformation gradients from a source expression and applies them to the neutral target face. Closely related is delta transfer, which transfers deformations in the form of (scaled) per-vertex displacements. However, neither deformation nor delta transfer can prevent the retargeting of character-specific details. Further, many known artifacts arise, such as loss of volume, self-collisions, and incorrectly transmitted deformation amplitudes. A body of related work is therefore concerned with explicitly distinguishing expression-specific from character-specific details [9, 37, 38]. For instance, Onizuka et al. [9] propose a locally scaled deformation transfer to keep facial contours consistent, Xu et al. [37] use an adapted deformation transfer for edges to focus on lip and eye contours, and Bhat et al. [38] show how to transfer lip contours to humanoid aliens. In contrast to previous work, we design facial features that aim to retain not only contours but also other facial proportions. Furthermore, we use a fully volumetric approach to avoid artifacts like volume loss and self-collisions.

Notation	Description
M	Surface mesh *and* stacked vertex positions
$\mathcal{S},\mathcal{T}$	Source and retargeted animation
$\mathcal{S}_\mathcal{E},\mathcal{T}_\mathcal{E}$	Source and target examples
S,T	Neutral head surfaces
S_i,T_i	Source expression and AnaConDaR retargeting
S_i^L,T_i^L	Reconstruction and retargeting of S_i with LBS
S_i^P,T_i^P	Reconstruction and retargeting of S_i with PLBS
$\mathbf{w}_i^L,\mathbf{w}_i^P$	Optimal LBS and PLBS reconstruction weights
$\hat{S}_i^M,\hat{T}_i^M$	Missing delta blendshapes
S_i^M,T_i^M	Missing blendshapes
$\mathbb{S},\mathbb{M}$	Template soft and muscle tissue tetrahedra meshes
$\mathrm{H}_S,\mathrm{H}_T$	Source and target heads
F_i	Facial characteristics

Table 1: An overview of the notation of AnaConDaR.

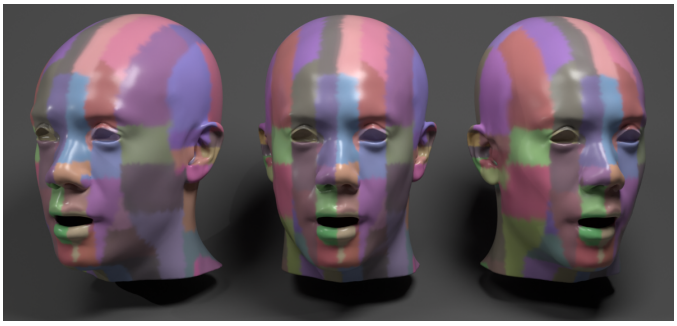

Fig. 1: The patch layout (80 patches) we use has been automatically determined with METIS [39].

3. Method

3.1. Problem Statement & Method Overview

The input to *offline facial performance retargeting* is a facial animation of a source character captured as a set $\mathcal{S}=\{S_i\}_{i=0}^N$ of $N+1$ surface meshes with identical tessellation. The overall goal is to curate a corresponding set of surface meshes $\mathcal{T}=\{T_i\}_{i=0}^N$ for a different target character, such that each expression T_i exhibits the same characteristics as S_i. These characteristics are primarily rooted in human perception and, therefore, difficult to capture through formal means.

To achieve this goal, we present AnaConDaR (Section 3.2), a mainly data-driven approach to facial retargeting, which is supplemented by a geometry-driven component (Section 3.3) whenever the available data is not sufficiently expressive. Moreover, anatomical plausibility and expression characteristics are ensured through a quasi-static physics-based simulation (Section 3.4).

In the ensuing formal derivation of AnaConDaR, we follow a top-down scheme in which we first explain the fundamental functionality of our approach (Section 3.2). Afterward, individual constituents are explained in more detail (Sections 3.3, 3.4, and 3.5). To ease the reading flow, Table 1 gives a summary of the notation. We slightly abuse the notation by denoting a surface mesh and the corresponding vector of stacked vertex positions with the same symbol.

3.2. Anatomically-Constrained Data-Adaptive Facial Retargeting

3.2.1. Data-Driven Component

For the derivation of the data-driven component of AnaConDaR, we initially assume to have access to a set of target examples $\mathcal{T}_\mathcal{E}$ with corresponding expressions $\mathcal{S}_\mathcal{E}\subset\mathcal{S}$. This assumption will be lifted in Section 3.5. Further, we expect the neutral head surfaces S and T of both characters to be known. In such situations, a variety of blendshape concepts can be applied for data-driven facial retargeting. For example, plain linear blendshapes (LBS) [1] first *approximate* each source expression $S_i\in\mathcal{S}$ by a linear combination

$$S_i^L = S + \sum_{S_j\in\mathcal{S}_\mathcal{E}} w_{ij}^L (S_j - S) \tag{1}$$

of the source examples $\mathcal{S}_\mathcal{E}$. The optimal blending weights $\mathbf{w}_i^L=\left(\dots,w_{ij}^L,\dots\right)$ are the solution of the linear least squares problem

$$\mathbf{w}_i^L = \arg\min_{\mathbf{w}_i} \left\| S + \sum_{S_j\in\mathcal{S}_\mathcal{E}} w_{ij}(S_j - S) - S_i \right\|^2 + \lambda_{reg}\left\|\mathbf{w}_i\right\|^2, \tag{2}$$

where the first term draws the blended surface S_i^L to the targeted expression S_i. Since this reconstruction is underconstrained,

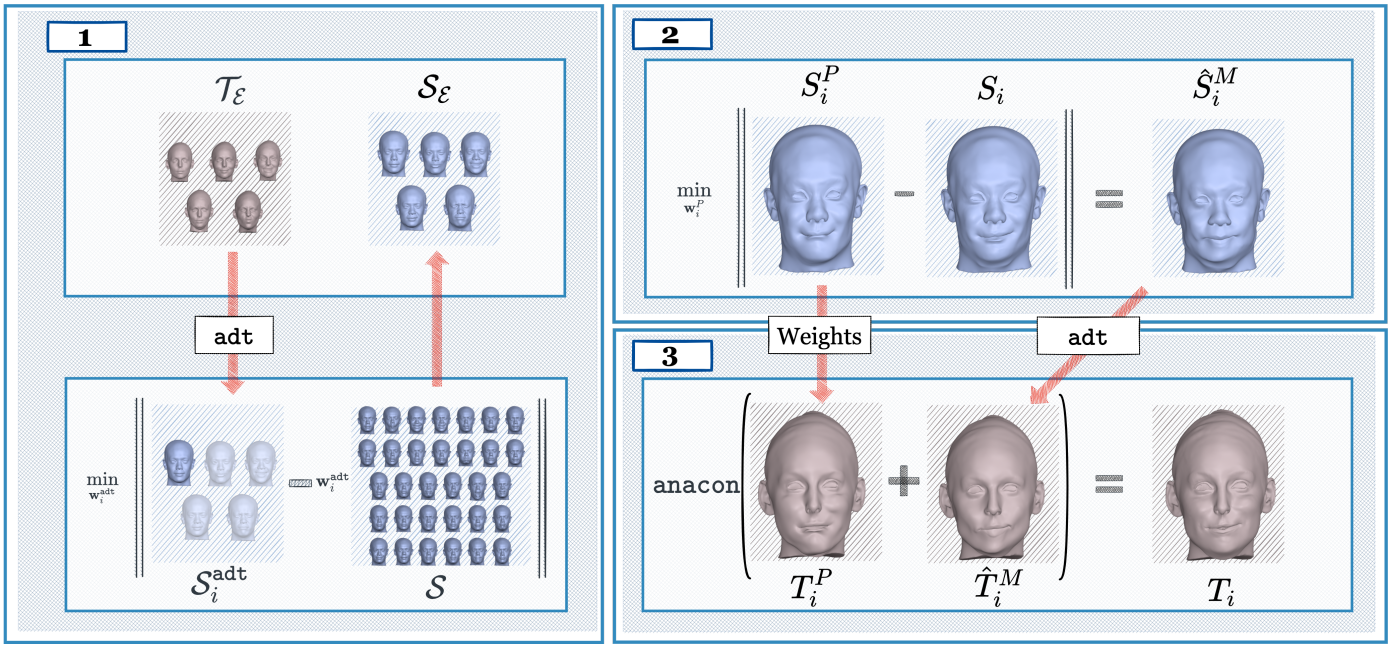

Fig. 2: An overview of AnaConDaR (Section 3.2). In the first step, the target examples $\mathcal{T}_\mathcal{E}$ are mapped into the source domain (Section 3.5). In the second step, the embedded expressions $\mathcal{S}_\mathcal{E}$ are used to form a PLBS approximation S_i^P of the targeted expression S_i by optimizing the patchwise blending weights $\mathbf{w}_i^P$ (Section 3.2.1). In step three, the evaluation of the same patchwise weights in the target domain T_i^P is supplemented with the `adt` result (Section 3.3) of the *missing* blendshape T_i^M (Section 3.2.2). Lastly, `anacon` ensures anatomical plausibility of the final AnaConDaR retargeting T_i (Section 3.4).

the second term adds the squared norm $\|\mathbf{w}_i\|^2$ of the blending weights to regularize them to be close to zero. The factor $\lambda_{reg} \in \mathbb{R}$ controls the strength of the regularization. Subsequently, the LBS *retargeting*

$$T_i^L = T + \sum_{T_j \in \mathcal{T}_\mathcal{E}} w_{ij}^L (T_j - T) \tag{3}$$

is obtained by simply applying the optimized weights $\mathbf{w}_i^L$ to the target examples $\mathcal{T}_\mathcal{E}$.

Patchwise linear blendshapes (PLBS) outperform the classical LBS in efficiency and expressiveness [2, 40]. Our implementation partitions all vertices consistently into a set of small (non-overlapping) patches (Figure 1) and performs the LBS retargeting defined in Equations (1–3) independently for each patch. We refer to the resulting source approximation of PLBS as S_i^P and to the retargeted expression as T_i^P.

The PLBS retargeting T_i^P is the *data-driven component* of AnaConDaR.

3.2.2. Geometry-Driven Component

Although variants of PLBS are the foundation of the current state-of-the-art in facial retargeting [2], errors in the source approximation

$$\hat{S}_i^M = S_i - S_i^P \tag{4}$$

are inevitably retargeted, as well. Seen from a different perspective, $\hat{S}_i^M$ is a missing delta blendshape for which no corresponding blendshape $T_i^M = \hat{T}_i^M + T$ is known. We approximate

$$T_i^M = \texttt{adt}\left(S_i^M, S, T\right) \tag{5}$$

with a novel (geometry-driven) deformation transfer `adt` (Section 3.3), which transfers the deformations of the missing blendshape $S_i^M = \hat{S}_i^M + S$ from the source to the target character. As opposed to the original deformation transfer [7], `adt` is physics-based, volumetric, and anatomically-constrained. Moreover, `adt` preserves expression-specific characteristics from S_i^M in T_i^M.

The retargeted missing delta blendshape $\hat{T}_i^M = T_i^M - T$ is the *geometry-driven component* of AnaConDaR.

3.2.3. Assembling the Components

AnaConDaR processes the sum of both the actual patchwise blendshapes T_i^P (data-driven component) and the missing delta blendshape $\hat{T}_i^M$ (geometry-driven component) with the physics-based simulation `anacon` (Section 3.4) to form the final retargeting

$$T_i = \texttt{anacon}\left(T_i^P + \hat{T}_i^M, S_i, S, T\right). \tag{6}$$

Conceptually, `anacon` is similar to `adt` and also enhances the retargeting plausibility through anatomical constraints as well as expression-specific characteristics. Additionally, visible patch boundaries are eliminated, which can occur in the PLBS result T_i^P.

Summarized in words, AnaConDaR retargets as extensively as possible through exemplary data but does not lose valuable information due to source approximation errors, since these are corrected with the geometry-driven component. The overview of AnaConDaR described so far is also visualized in steps 2 and 3 of Figure 2.

Next, we will depict `adt` and `anacon` in more detail. As both only differ slightly, we will explain them using the example of `adt` (Section 3.3) and then discuss the differences to `anacon` (Section 3.4). Finally, we will resolve the initial assumption of corresponding source and target examples $\mathcal{S}_\mathcal{E}$ and $\mathcal{T}_\mathcal{E}$ (Section 3.5).

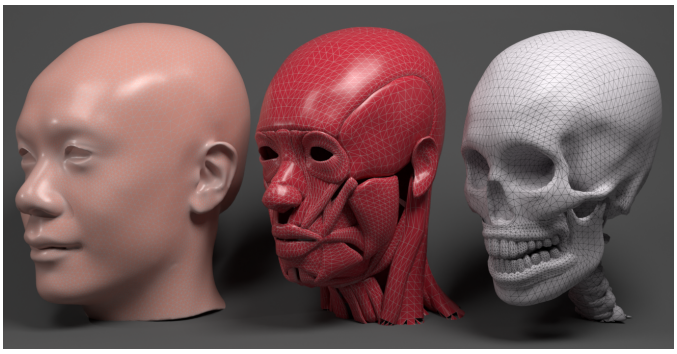

Fig. 3: The surfaces of all anatomical structures that are part of the volumetric head template we use for AnaConDaR. From left to right, the surface of the soft tissue, the surface of the muscle tissue, and the skull surface. The soft tissue includes the neutral head surface, the muscle tissue is connected to the skull as well as the soft tissue, and the skull is separated into jaw and cranium.

Algorithm 1 Anatomical Deformation Transfer

Input

S_i^M *The missing blendshape*
S, T *The neutral head surfaces*

Function $\texttt{adt}(S_i^M, S, T)$

// Section 3.3.2 Template Fitting.
$\mathrm{H}_S = \texttt{fitHead}(S, \mathrm{H})$, $\mathrm{H}_T = \texttt{fitHead}(T, \mathrm{H})$

// Section 3.3.2 Inverse Simulation.
$(\nabla\mathbb{S}, \nabla\mathbb{M}, \nabla B) = \texttt{invSim}(S_i^M, \mathrm{H}_S)$

// Section 3.3.2 Facial Characteristics.
$\mathrm{F}_i = \texttt{fc}\big(S_i^M, S, T\big)$

// Section 3.3.2 Forward Simulation.
$T_i^M = \texttt{fwdSim}((\nabla\mathbb{S}, \nabla\mathbb{M}, \nabla B), \mathrm{F}_i, \mathrm{H}_T)$

// Return the retargeted missing blendshape.
return T_i^M

3.3. Anatomical Deformation Transfer

3.3.1. Overview

Given the neutral head surfaces S and T of the source and target character, `adt` executes four fundamental functions for retargeting the missing blendshape S_i^M to T_i^M as outlined in Algorithm 1. To facilitate the introduction of `adt`, we again follow a top-down scheme and first give a brief overview of every function in this section. The subsequent Section 3.3.2 provides the corresponding detailed descriptions, each of which can be found in an identically named paragraph.

Template Fitting. As a first step, the function `fitHead` creates volumetric head representations for the source and target character by fitting a template head $\mathrm{H} = (\mathbb{S}, \mathbb{M}, B)$ to the neutral surfaces S and T. The template comprises a soft tissue tetrahedra mesh $\mathbb{S}$, a muscle tissue tetrahedra mesh $\mathbb{M}$, and a skull surface mesh B. Please refer to Figure 3 for a visualization of the corresponding surfaces and more details. The resulting heads

$$\begin{aligned} \mathrm{H}_S &= (\mathbb{S}_S, \mathbb{M}_S, B_S) = \texttt{fitHead}(S, \mathrm{H}) \\ \mathrm{H}_T &= (\mathbb{S}_T, \mathbb{M}_T, B_T) = \texttt{fitHead}(T, \mathrm{H}) \end{aligned} \tag{7}$$

consist of the fitted components.

Inverse Simulation. After fitting the template, the inverse physics-based simulation

$$(\nabla\mathbb{S}, \nabla\mathbb{M}, \nabla B) = \texttt{invSim}(S_i^M, \mathrm{H}_S) \tag{8}$$

identifies volumetric changes of the source head H_S to form the targeted missing blendshape S_i^M while respecting bio-mechanical and physical properties. Here, $\nabla\mathbb{S}$ and $\nabla\mathbb{M}$ are stacked per tetrahedron 3×3 deformation gradients that capture changes in soft and muscle tissue $\mathbb{S}_S$ and $\mathbb{M}_S$, respectively. For the jaw and cranium parts of B_S, rigid movements are individually captured by ∇B.

Facial Characteristics. Alongside the volumetric changes, the function `fc` identifies expression-specific facial characteristics

$$\mathrm{F}_i = \texttt{fc}\big(S_i^M, S, T\big) \tag{9}$$

in the missing blendshape S_i^M and adapts them to the target character. These characteristics are our answer to the following thought experiment:

> "If you are given a picture of an expression to mimic and a mirror to look at yourself, what do you use as guidance?"

We assume that human perception is guided by relative changes of face openings and facial contours which can be influenced through muscle activation. More specifically, we assume that the eyes in the missing and the retargeted blendshape should open and close by almost the same relative proportions while the skin around the eye sockets is assumed to move in a consistent manner. Furthermore, we expect the lips to form similar contours in both, since these can be manipulated by humans with a great degree of control.

Forward Simulation. Finally, the forward physics-based simulation `fwdSim` generates the retargeted missing blendshape

$$T_i^M = \texttt{fwdSim}((\nabla\mathbb{S}, \nabla\mathbb{M}, \nabla B), \mathrm{F}_i, \mathrm{H}_T) \tag{10}$$

by applying the previously calculated volumetric changes $(\nabla\mathbb{S}, \nabla\mathbb{M}, \nabla B)$ and facial characteristics F_i to the target head H_T.

3.3.2. Constituents

In the remainder of this section the four `adt` functions `fitHead`, `invSim`, `fc`, and `fwdSim` are precisely described.

Template Fitting. The template fitting `fitHead`, which fits the volumetric head template $\mathrm{H} = (\mathbb{S}, \mathbb{M}, B)$ to the neutral source and target head surfaces S and T, performs two steps.

1. The skull B is placed by a dense linear model trained on the computed tomography dataset of Achenbach et al. [41]. This model maps from the vertex positions of the head surface to the vertex positions of the skull surface.
2. Soft and muscle tissue $\mathbb{S}, \mathbb{M}$ are positioned by a radial basis function (RBF) space warp [42] calculated from the template to the targeted head and skull surfaces. By the construction of RBFs, the vertices of $\mathbb{S}$ and $\mathbb{M}$ are warped to a similar semantic position as in the template.

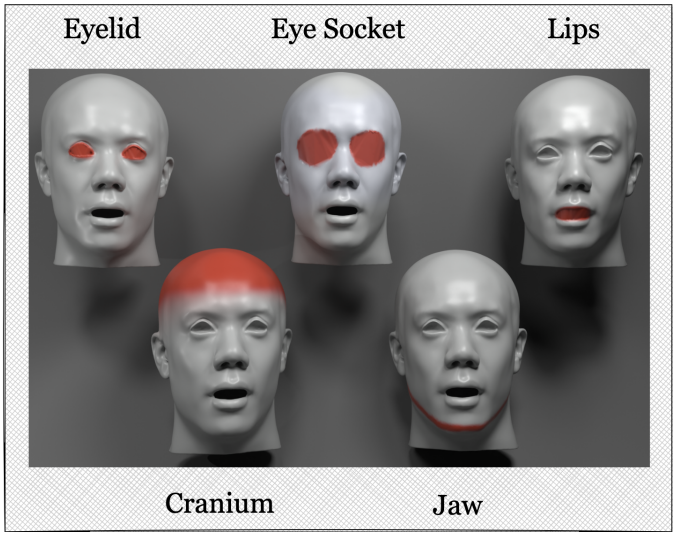

Fig. 4: The supplementary meshes that are used to determine the position of the jaw and cranium bones, as well as the expression-specific characteristics.

Inverse Simulation. The inverse simulation `invSim`, which aligns the source head $\mathrm{H}_S = (\mathbb{S}_S, \mathbb{M}_S, B_S)$ with the targeted blendshape S_i^M, is composed of two steps.

1. Both skull parts of B_S, cranium and jaw, are each directly positioned by independent rigid transformations ∇B that are calculated between respective subsets of S and S_i^M. The subsets are visualized in Figure 4 Cranium and Jaw.
2. Soft tissue and muscle tissue are deformed by minimizing the following energies that reflect anatomical properties. The energy for soft tissue is defined as

$$E_{\mathbb{S}}(\mathbb{S}_S) = \sum_{t \in \mathbb{S}_S} \Big(\min_{\mathbf{R} \in SO(3)} \|\nabla(t, \mathbb{S}_S) - \mathbf{R}\|_F^2 + (\det(\nabla(t, \mathbb{S}_S)) - 1)^2 \Big), \tag{11}$$

which for each soft tissue tetrahedron t penalizes changes in volume and strain. Here, $\mathbf{R} \in SO(3)$ denotes the optimal rotation, $\nabla(t, \mathbb{S}_S) \in \mathbb{R}^{3\times 3}$ the deformation gradient of t, and $\|\cdot\|_F$ the Frobenius norm.
For the muscle tetrahedra, only a volume-preservation term

$$E_{\mathbb{M}}(\mathbb{M}_S) = \sum_{t \in \mathbb{M}_S} (\det(\nabla(t, \mathbb{M}_S)) - 1)^2 \tag{12}$$

is applied to allow for muscle contractions.
Finally, the source head surface $S \subset \mathbb{S}_S$ is drawn to the targeted missing blendshape S_i^M via

$$E_{\text{tar}}(\mathbb{S}_S, S_i^M) = \left\| S - S_i^M \right\|^2. \tag{13}$$

In total, we minimize the weighted energy

$$E_{\text{inv}}(\mathbb{S}_S, \mathbb{M}_S, S_i^M) = w_{\mathbb{S}} E_{\mathbb{S}}(\mathbb{S}_S) + w_{\mathbb{M}} E_{\mathbb{M}}(\mathbb{M}_S) + w_{\text{tar}} E_{\text{tar}}(\mathbb{S}_S, S_i^M) \tag{14}$$

with respect to the vertex positions of the soft and muscle tissue meshes $\mathbb{S}_S, \mathbb{M}_S$ in the projective dynamics framework [43]. The values of all weights can be found in Table 3.

Paired with the rigid transformations of the skull ∇B, the deformations caused by the simulation are passed on to the forward simulation `fwdSim` in the form of stacked per tetrahedron deformation gradients $\nabla\mathbb{S}$ (soft tissue), $\nabla\mathbb{M}$ (muscle tissue).

Facial Characteristics. In correspondence to the facial characteristics described above, `fc` is composed of three methods

$$\texttt{fc} = (\texttt{fc}_{\text{eo}}, \texttt{fc}_{\text{es}}, \texttt{fc}_{\text{lc}}) \tag{15}$$

which specify objectives for the eye opening ($\texttt{fc}_{\text{eo}}$), the eye sockets ($\texttt{fc}_{\text{es}}$), and the lip contour ($\texttt{fc}_{\text{lc}}$) in the forward simulation `fwdSim`.

To capture the eye characteristics with $\texttt{fc}_{\text{eo}}$ and $\texttt{fc}_{\text{es}}$, we add supplementary triangles between the upper and lower eyelids (Eyelid) and between the upper and lower boundaries of the eye sockets (Eye Socket) as visualized in Figure 4. Hereafter, we refer to these triangles as EO and ES, respectively. Since the eye characteristics are intended to transfer relative movements, we define them such that the scaling of the surface area of the previously added triangles is identical in both the targeted and the retargeted blendshape. More formally, the characteristic $\mathrm{F}_{\text{eo}} = \texttt{fc}_{\text{eo}}(S_i^M, S, T)$ is a vector which contains the surface area of each EO triangle in T, scaled by the ratio of the corresponding triangle areas in the targeted S_i^M and the neutral S. $\mathrm{F}_{\text{es}} = \texttt{fc}_{\text{es}}(S_i^M, S, T)$ is defined accordingly.

We define the characteristic of the lip contour F_{lc} on a set of vertices LC as visualized in Figure 4 Lips. Here, we intend to transfer the vertex positions of the contour from the targeted S_i^M to the retargeted blendshape T_i^M as similar as possible. To that end, we first apply the original deformation transfer `dt` [7] to determine the coarse shape and position of the targeted lip contour in the retargeting result. Afterward, we correct `dt` by finding an optimal similarity mapping. Formally, we define

$$\mathrm{F}_{\text{lc}} = \texttt{fc}_{\text{lc}}(S_i^M, S, T) = s\mathbf{R}\left(S_i^M\right)^{LC} + \mathbf{t}, \tag{16}$$

where $s \in \mathbb{R}$ (scaling), $\mathbf{R} \in SO(3)$ (rotation), $\mathbf{t} \in \mathbb{R}^3$ (translation) represent the optimal similarity mapping regarding

$$\min_{s,\mathbf{R},\mathbf{t}} \left\| \texttt{dt}(S_i^M, S, T)^{\text{LC}} - s\mathbf{R}\left(S_i^M\right)^{LC} - \mathbf{t} \right\|^2 \tag{17}$$

and $(\cdot)^{\text{LC}}$ selects the vertices of the lip contour.

Forward Simulation. The forward simulation `fwdSim`, which applies the previously identified deformations ($\nabla\mathbb{S}$, $\nabla\mathbb{M}$, ∇B) and expression-specific facial characteristics F_i to the target head $\mathrm{H}_T = (\mathbb{S}_T, \mathbb{M}_T, B_T)$, consists of three steps.

1. As for `invSim`, the skull B_T is directly positioned by applying the rigid transformations ∇B. However, to align the range of motions, we scale the translational components by $\frac{\text{BB}(T)}{\text{BB}(S)}$, where BB calculates diameters of the respective bounding boxes.
2. The weighted energy

$$E_{\text{fwd}}(\mathbb{S}_T, \mathbb{M}_T, \nabla\mathbb{S}, \nabla\mathbb{M}, \mathrm{F}_i) = w_{\nabla\mathbb{S}} E_{\nabla\mathbb{S}}(\mathbb{S}_T, \nabla\mathbb{S}) + w_{\nabla\mathbb{M}} E_{\nabla\mathbb{M}}(\mathbb{M}_T, \nabla\mathbb{M}) + w_{\mathrm{F}} E_{\mathrm{F}}(\mathbb{S}, \mathrm{F}_i) \tag{18}$$

is minimized, which applies the deformation gradients $\nabla\mathbb{S}, \nabla\mathbb{M}$ to the respective tissue while adhering to the facial characteristics F_i. All energies in Equation (18) act similar to Equation (13) and are formally defined in the Appendix. Again, we rely on projective dynamics for solving the minimization problem.

3. Finally, we resolve self-collisions between lips similar to Komaritzan et al. [44]. Here, each collided lower lip point and the closest upper lip point in vertical direction on the head surface are resolved to the average position of both. The average position is enforced in an additional run of the second step.

After both optimizations, the retargeted missing blendshape $T_i^M \subset \mathbb{S}_T$ can be extracted.

3.4. Anatomical Plausibility

Based on the functions of `adt`, we can now implement `anacon`, the final physics-based simulation of AnaConDaR (Equation (6)). By setting

$$\begin{aligned} T_i &= \texttt{anacon}\big(T_i^P + \hat{T}_i^M, S_i, S, T\big) \\ &= \texttt{fwdSim}\Big(\texttt{invSim}\Big(T_i^P + \hat{T}_i^M, \mathrm{H}_T\Big), \mathrm{F}_i, \mathrm{H}_T\Big), \end{aligned} \tag{19}$$

the anatomical constraints involved in `invSim` (Section 3.3.2 Inverse Simulation) improve the anatomical plausibility of the combined retargeting $T_i^P + \hat{T}_i^M$ while preventing visible patch boundaries. Moreover, expression-specific facial characteristics $\mathrm{F}_i = \texttt{fc}(S_i, S, T)$ (Section 3.3.2 Facial Characteristics) derived from the targeted expression S_i are also reflected in the final AnaConDaR result T_i.

3.5. Target Example Embedding

Although all components are now specified, AnaConDaR is still unable to handle situations where the target examples $\mathcal{T}_\mathcal{E}$ lack corresponding expressions in the source animation $\mathcal{S}$, a common limitation of other data-driven facial retargeting approaches [1, 2]. We remove this initial assumption (Section 3.2.1) by embedding the target examples in the source domain.

To that end, we first retarget $\mathcal{T}_\mathcal{E}$ with `adt` to create an initial embedding

$$\mathcal{S}_\mathcal{E}^{\texttt{adt}} = \big\{\texttt{adt}(T_j, T, S)\big\}_{T_j \in \mathcal{T}_\mathcal{E}}. \tag{20}$$

As `adt` is geometry-driven, $\mathcal{S}_\mathcal{E}^{\texttt{adt}}$ might still exhibit character-specific details of the target character. In a second step, we therefore exploit the observation that, in most cases, the source animation $\mathcal{S}$ is extensive and expressive in linear combinations.

More precisely, we reconstruct each $S_i^{\texttt{adt}} \in \mathcal{S}_\mathcal{E}^{\texttt{adt}}$ by solving linear least squares problems as in Equations (1–3). This time, however, all source expressions $S_j \in \mathcal{S}$ act as blendshapes. The resulting optimal blending weights $\mathbf{w}_i^{\texttt{adt}} = \left(..., w_{ij}^{\texttt{adt}}, ...\right)$ are then used to form the data-driven embedding

$$\mathcal{S}_\mathcal{E} = \left\{ S + \sum_{S_j \in \mathcal{S}} w_{ij}^{\texttt{adt}} (S_j - S) \right\}_{S_i^{\texttt{adt}} \in \mathcal{S}_\mathcal{E}^{\texttt{adt}}}. \tag{21}$$

By construction, $\mathcal{S}_\mathcal{E}$ is fully embedded in the source domain and no longer includes details of the target character. The embedding process is also illustrated in step 1 of Figure 2, which completes the visual overview of AnaConDaR.

4. Experiments

Before visually demonstrating AnaConDaR's capabilities for offline facial performance retargeting in Section 4.2, we discuss implementation details and runtimes in Section 4.1. Thereafter, in Section 4.3, a user study investigates the human perception of AnaConDaR in comparison to the most relevant peers. In Section 4.4, a quantitative analysis demonstrates the advantages of AnaConDaR over the state-of-the-art ALM [2] algorithm. However, we also elaborate on why quantitative evaluations only have limited meaningfulness for facial retargeting. Section 4.5 focuses on an extensive ablation study, while Section 4.6 showcases selected AnaConDaR features in more detail.

4.1. Implementation & Runtimes

We implement all projective dynamics simulations with the CPU-based ShapeOp framework [45] and exploit parallelism wherever applicable. Table 2 gives the dimensions of all template components, and Table 3 states the weights of all experiments. All runtimes were determined on an AMD Ryzen Threadripper PRO 3995WX processor.

Overall, once the simulations are initialized ($\approx 9\,\mathrm{s}$), AnaConDaR can be run at either approximately 10 fps (without collision resolving) or 0.3 fps (with collision resolving). There are many GPU-based solvers available (e.g., `http://suitesparse.com`) that can optimize the runtime in general. Collision resolution could also be accelerated, as most of the time spent on collision resolution is due to the refactorisation of the projective dynamics solver. In Wang et al. [46], for instance, an efficient alternative is proposed. However, as our focus has been on methodological improvements, not on inference speed, we leave computationally more efficient implementations as future work.

4.2. Qualitative Evaluation

Figure 5 and Figure 6 display representative retargeting results of AnaConDaR. All shown 3D models are part of the commercial `3Dscanstore.com` database and have been acquired with a high-resolution optical multi-view scanner. We manually established a common topology using `faceform.com`. The retargeted expressions are either facial movements like *cheek puffer* and *mouth stretch* or emotions like *sad*, *happy*, and *surprise*.

Figure 5 displays results obtained from a $\mathcal{T}_\mathcal{E}$ composed of only 5 target examples, whereas for the results from Figure 6, an extensive set of 30 examples has been available. For each retargeting result, a different $\mathcal{T}_\mathcal{E}$ has been randomly drawn. Please refer to the attached video for a demonstration of the temporal consistency of AnaConDaR.

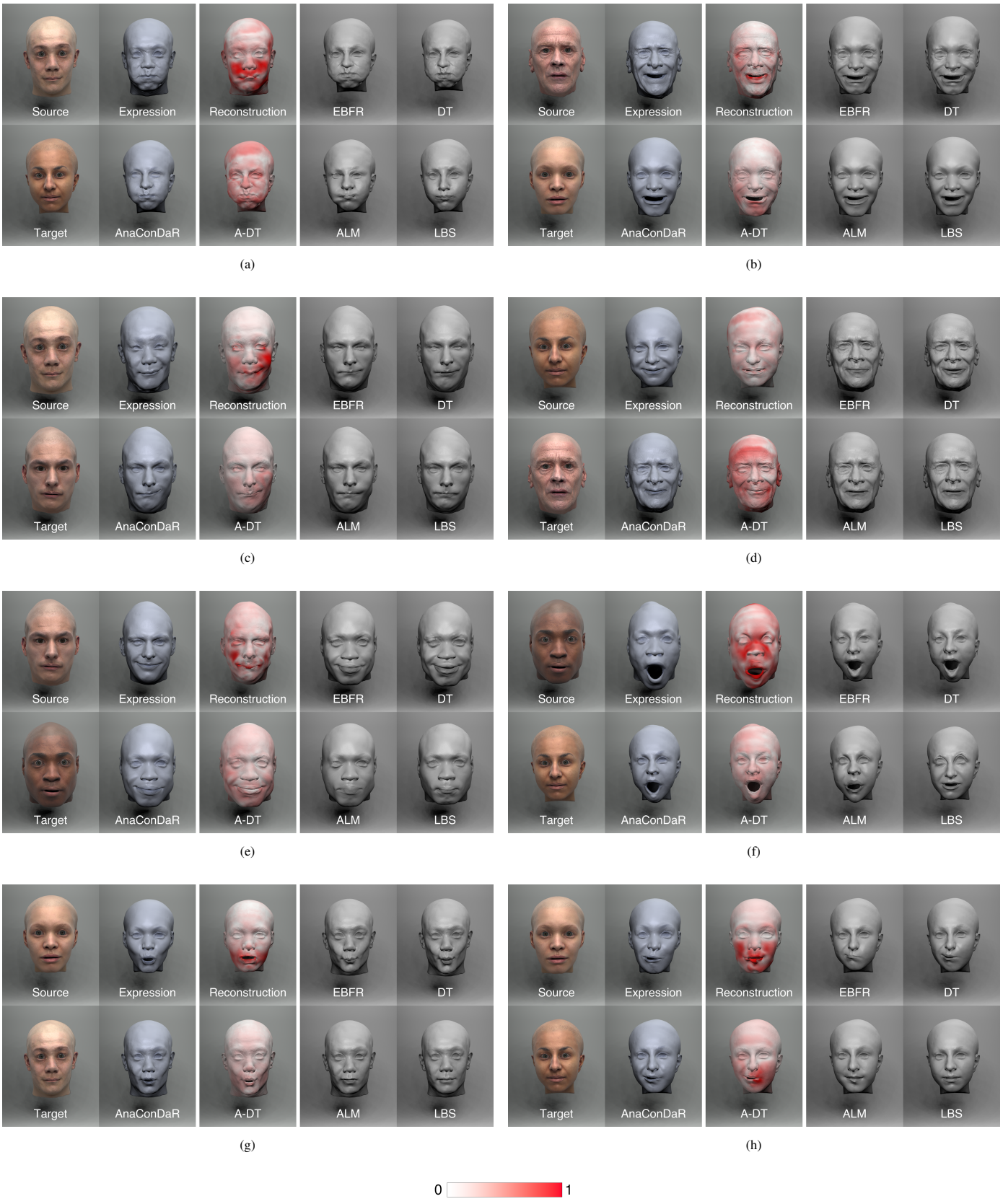

Fig. 5: AnaConDaR in comparison to the state-of-the-art peer group EBFR [6], ALM [2], DT [9], LBS [1], and to our ADT. Furthermore, the source reconstruction after applying anatomical constraints (i.e., `anacon` w/o facial characteristics) is shown for a reasonable comparison. Plotted on the reconstruction is the PLBS reconstruction error (in centimeters). The difference between ADT and AnaConDaR is plotted on the ADT expression. All results have been achieved with **five randomly drawn examples** of the target character. Especially in this setting, with only a few target examples, AnaConDaR leads to considerable improvements.

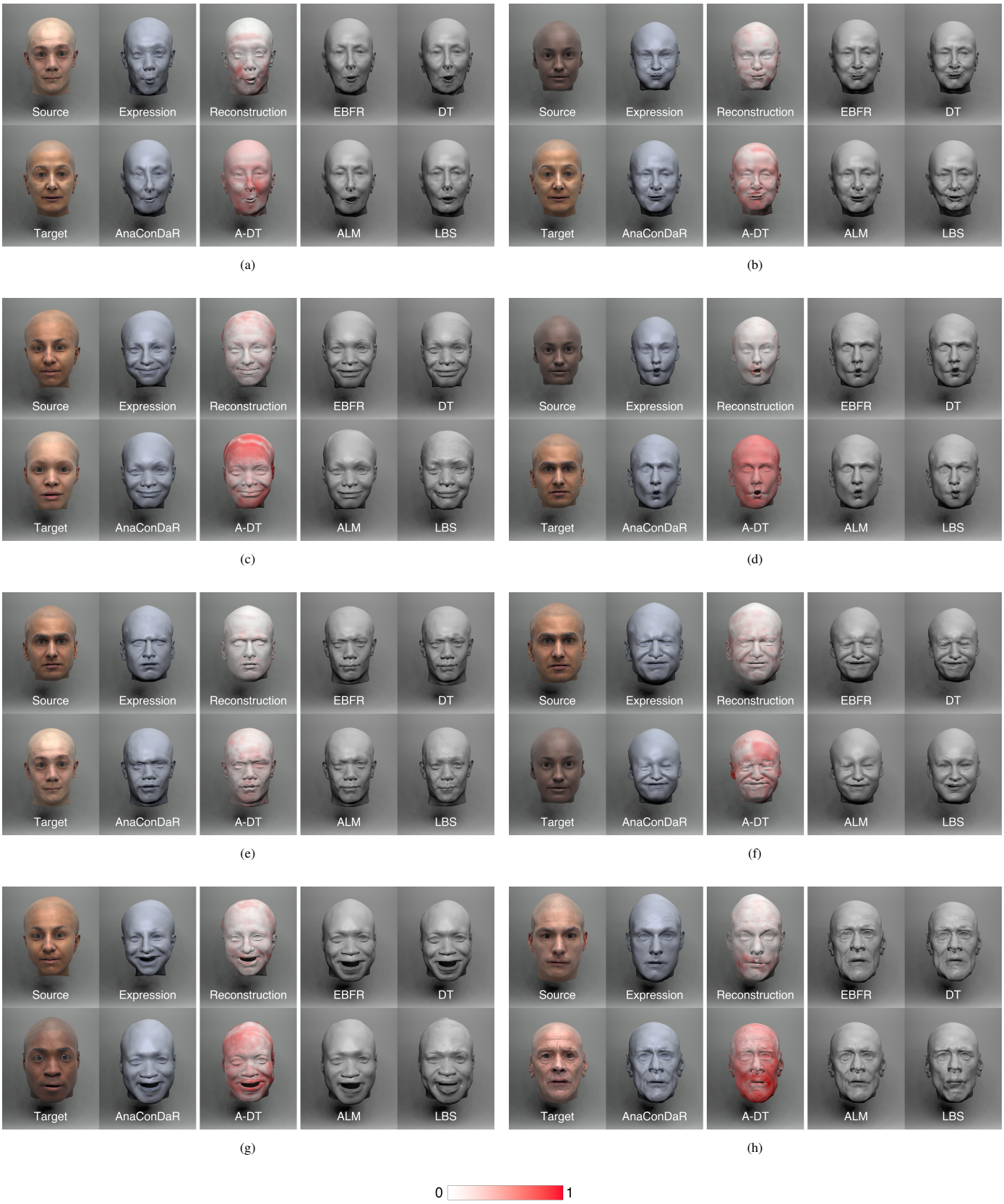

Fig. 6: The same experiment as in Figure 5, however, **30 examples** of the target character have been available. In principle, reconstruction errors decrease with more target examples, reducing the influence of the *missing* blendshape. Nevertheless, considerable benefits of AnaConDaR can also be recognized in this setting, as even with more examples, a complete reconstruction is not guaranteed. Moreover, other advantages, such as the fully volumetric simulation or the preservation of facial features, weigh in.

4.2.1. Peer Group

We compare AnaConDaR to Example-Based Facial Rigging [6] (EBFR), Anatomical Local Models [2] (ALM), Deformation Transfer [9] (DT), Linear Blendshapes [1] (LBS), and our own Anatomical Deformation Transfer (ADT).[1] Generally, ALM and LBS require expressions in the source animation $\mathcal{S}$ that correspond to the target examples $\mathcal{T}_\mathcal{E}$. Therefore, we follow the suggestion by the authors of ALM to use EBFR as preprocessing if this requirement is not fulfilled.

4.2.2. Discussion

The subsequent discussion of the presented outcomes follows along the structural varieties of all compared algorithms. For easier traceability of our analysis, Figure 7 provides a visual overview.

- The DT implementation we investigate [9] is the most recent adaption to faces. Here, locally adapted delta transfers for predefined landmarks are additionally incorporated. Previous findings [2] already indicated minimal distinctions between delta transfer and DT. Our results consistently demonstrate that also this DT variant transfers character-specific details and not only deformations related to the targeted expressions.
- EBFR seeks a target animation $\mathcal{T}$ such that a linear combination of $\mathcal{T} \setminus \mathcal{T}_\mathcal{E}$ can approximate the target examples $\mathcal{T}_\mathcal{E}$. Since this is a strongly underdetermined optimization problem, the DT results are used for regularization. As a consequence, depending on the solver, either only a few expressions of $\mathcal{T}$ contribute to the explanation of each example in $\mathcal{T}_\mathcal{E}$ or all contribute only a small fraction to the explanation. In any case, this leads to only minor personalization beyond DT, especially when only 5 examples are available.
- The state-of-the-art ALM approach [2] is closely related to LBS [1]. After establishing a corresponding set of source examples $\mathcal{S}_\mathcal{E}$ to the target examples $\mathcal{T}_\mathcal{E}$ with EBFR, both form the target animation $\mathcal{T}$ by blending $\mathcal{T}_\mathcal{E}$. The blending weights are found by rebuilding the source animation $\mathcal{S}$ with $\mathcal{S}_\mathcal{E}$. ALM mainly differs from LBS in that the blending is conducted on small patches and not on complete meshes (please refer to Section 3.2.1 for more details).
  In our experiments, LBS suffers from a strong bias which prevents an adequate reconstruction of source expressions, leading to the retargeting of *different* expressions. Put differently, the poor results primarily stem from the *missing* blendshape as described in Section 3.2.2. The PLBS of ALM significantly mitigate this issue, especially when having access to 30 target examples $\mathcal{T}_\mathcal{E}$. Nonetheless, notable reconstruction errors remain.
- Our AnaConDaR approach exhibits a high degree of personalization even with only a few examples from the target character and can still achieve more appealing results than ALM, LBS, EBFR, DT, or ADT if the number of examples is high.
  In contrast to DT, the data-driven AnaConDaR component abstains from transferring character-specific details wherever feasible. Unlike ALM and LBS, additionally retargeting the *missing* blendshape ensures that AnaConDaR does not lose information due to informational gaps of the exemplary target data. Lastly, different from EBFR, in our approach the target examples $\mathcal{T}_\mathcal{E}$ explain each expression to be transferred in $\mathcal{S}$, and not all expressions to be transferred explain the target examples. This strategy effectively avoids the *explanation problem* associated with EBFR, as discussed before.
  In summary, AnaConDaR achieves convincing visual results by compensating the conceptual weaknesses of other algorithms while adopting their respective advantages.

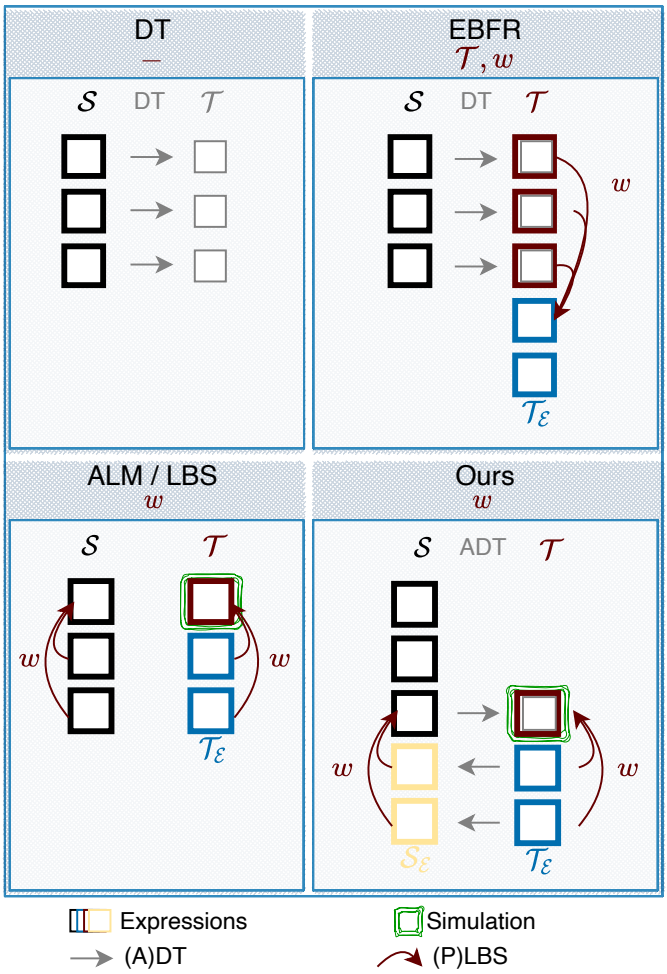

Fig. 7: A structural overview of AnaConDaR and other state-of-the-art facial retargeting approaches. Below each algorithm name, the variables under direct optimization are stated. Source and target expressions correspond only if depicted in the same row.

4.3. User Study

In a user study, we presented the following task.

> "Please rank the images according to how natural the transfer of the expression seems to you from best to worst."

[1] We compare to our own implementations of the peer group.

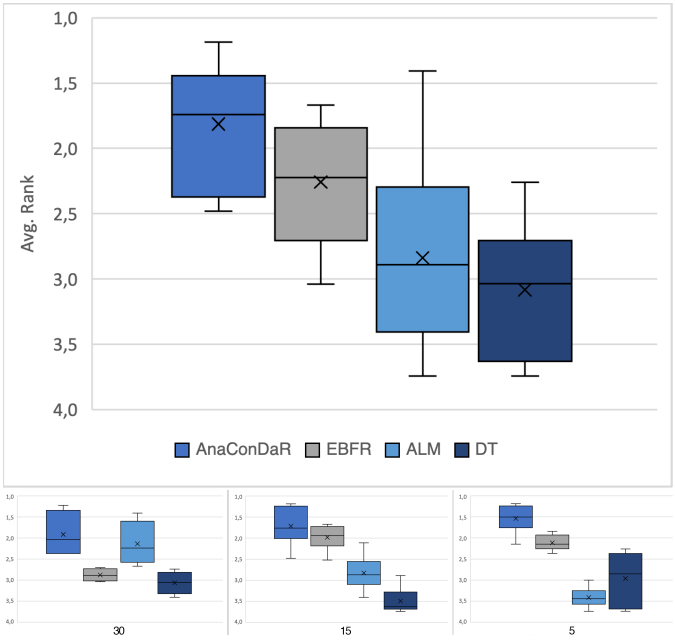

Fig. 8: A user study among university members and computer graphics students from two universities supports that AnaConDaR is perceived as a more natural facial retargeting. The combined results (top) as well as the results per number of target examples (bottom) are shown.

The study involves 15 randomly selected retargeting instances, with five each produced using 5, 15, and 30 examples of the target character. Participants ranked the results of AnaConDaR, DT, EBFR, and ALM, respectively. The design of the study is aligned with Chandran et al. [2], an illustration can be found in Appendix B. To ensure independent documentation, we used `survio.com` for the technical implementation.

The outcome shown in Figure 8 summarizes 33 responses by university members and computer graphics students from two universities who were not familiar with facial retargeting algorithms. We performed Wilcoxon Signed-Rank tests to inspect if the tendency of the AnaConDaR mean rank in comparison to the other peers is significant. We can confirm this hypothesis for all peers on a significance level of 0.05.

The user study emphasizes that AnaConDaR is perceived as a more natural facial retargeting. Nonetheless, perception variations are evident from the ranking variances shown in Figure 8. Probably by construction, the *data-infused* EBFR outperforms the solely geometry-based DT. Interestingly, EBFR also outperforms ALM, while ALM exhibits the highest variance. This was to be expected to some extent, as ALM is the only method lacking a geometry-based component and its retargeting quality, therefore, heavily depends on the number of target examples. The latter observation is further supported by the separated representation of the user study in Figure 8.

4.4. Quantitative Evaluation

In previous work, quantitative evaluations have only been conducted in cherry-picked individual cases but not in empirically comprehensive investigations [2, 6]. This is mostly due to ambiguities in facial expressions (see Figure 9 and Wu et al. [40] for examples) as well as varying human perception. Our

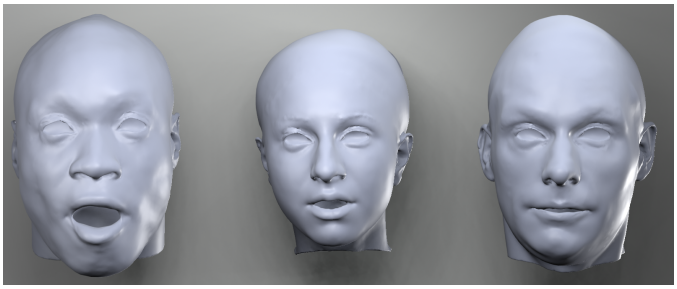

Fig. 9: An example of the diverse ways in which individuals interpret the same expression (here, *Surprise*). For more examples please refer to Wu et al. [40].

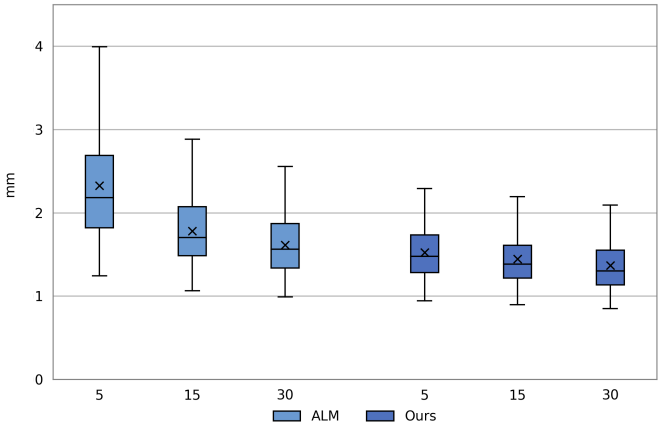

Fig. 10: A quantitative comparison of AnaConDaR against ALM [2] based on synthesized ground truth (Section 4.4). The results are grouped by the number of available target examples and reported in terms of the L2 error (mm). AnaConDar outperforms ALM in each scenario.

user study (Section 4.3) underscores the latter issue. Although the perceived qualities of individual retargeting methods differ significantly, the variances are not negligible. Sometimes cyclic errors, i.e., mapping from the source to the target and back, are considered as a suitable evaluation protocol. Nonetheless, they only validate how well geometric transformations are preserved in the cycle. By construction, deformation transfer [7] would be unsurpassed in this evaluation, while the flaws of geometry-driven approaches are well known.

To nonetheless quantitatively compare AnaConDaR to ALM, we first synthesize an appropriate dataset. More precisely, we use EBFR [6] to create personalized ARKit [2] blendshape rigs for each identity of the `3Dscanstore.com` database. Subsequently, we create the same 250 random facial expressions for all identities through linear blending of the ARKit rigs with blending weights recorded in dyadic conversations [47]. In the resulting dataset, corresponding facial expressions exhibit reduced ambiguities and, hence, can rather be regarded as ground truth.

Therefore, we conduct the following experiment on this dataset. To begin with, five source expressions $\mathcal{S}$ as well as either 5, 15, or 30 target examples $\mathcal{T}_\mathcal{E}$ are randomly drawn for all source-target identity combinations. Afterward, we run Ana-

[2] `https://developer.apple.com/augmented-reality/arkit/`

Mesh	S	B	$\mathbb{S}$	$\mathbb{M}$
# Vertices	29826	14572	61875	
# Faces / Tetrahedra	59648	28727	126612	107437

Table 2: The dimensions of all template components in our experiments.

$w_{\nabla\mathbb{S}}$	$w_{\nabla\mathbb{M}}$	w_F	$w_{\mathbb{S}}$	$w_{\mathbb{M}}$	w_{tar}	λ_{reg}
1.0	1.0	10.0	1.0	1.0	100.0	0.01

Table 3: The weights of the physics-based simulations and the PLBS reconstruction.

ConDaR and ALM for each source expression and measure the average of the vertex-wise L2 differences to the ground truth in mm. The findings of this experiment, reported in Figure 10, indicate that AnaConDaR outperforms ALM, especially when only a few target examples are available. A moderate improvement can still be recognized when many target examples are available. This quantitative evaluation ignores human perception but is nonetheless consistent with the previously discussed user study (Section 4.3).

4.5. Ablation Study

We examine the main components of AnaConDaR in another user study, which is summarised in Figure 11 and visualised in Figure 12. Particularly, we compare the regular AnaConDaR to AnaConDaR without expression-specific facial characteristics, without the missing blendshape, and with the standard deformation transfer `dt` [7] instead of our `adt`. The design of the user study is mostly as described in Section 4.3. However, no similar example has been provided to the 29 participants. Please refer to Appendix B for an exemplary question from this study.

Deformation Transfer. The most noticeable visual differences arise in the setting in which `dt` is used rather than `adt`. Here, the artifacts caused by PLBS patch boundaries are transferred by `dt`, and the strain constraint in `anacon` does not provide a sufficient countermeasure. An increased strain weight $w_{\mathbb{S}}$ could potentially compensate for this but would also remove high-frequency details. Since `adt`, unlike `dt`, also applies anatomical constraints, similar artifacts do not occur in the regular AnaConDaR results. The user study supports this visual observation as the `dt` variant is ranked last.

Instead of an amplified strain, another option would be to eliminate the patch boundaries directly in the source estimation S_i^P before calculating the missing blendshape S_i^M. For this, there are at least two obvious solutions. The first solution is to set up anatomical models as described in ALM [40, 2]. However, this adds considerable unnecessary complexity, mainly due to additional optimization steps. Since these models only use data-driven anatomical surface constraints, they also cannot be used as an alternative to `anacon`. Particularly, they are not applicable to unseen expressions, cannot enforce facial characteristics, and cannot resolve collisions. The second solution is to apply `anacon` to the source estimation S_i^P. Essentially, this means applying the same physics-based simulations as in `adt` to a different input. Nevertheless, we decided to favor `adt` for theoretical reasons. Particularly, as `adt` applies anatomical constraints and expression-specific facial characteristics during the retargeting and not before. Neither of the two variants was visually superior in our experiments.

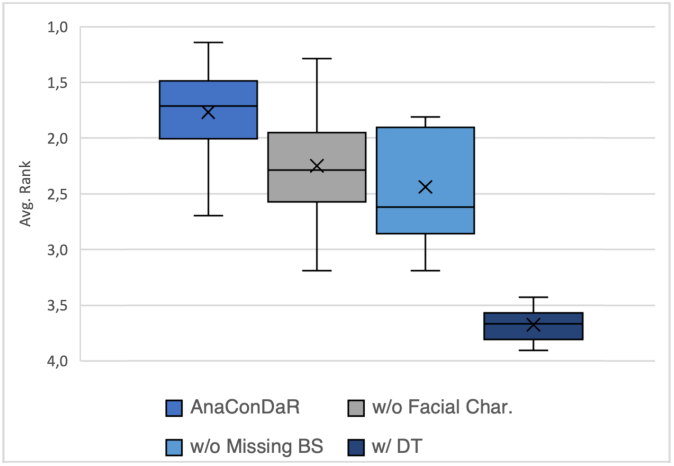

Fig. 11: A user study among university members and computer graphics students from two universities proves the benefits of each AnaConDaR component.

Facial Characteristics & Missing Blendshape. The AnaConDaR modifications without facial characteristics and missing blendshape demonstrate that both components are essential, although their importance varies depending on the retargeting scenario. For instance, in the first row of Figure 12, the expression-specific facial characteristics are especially important, whereas in the second row, the missing blendshape has a strong impact. Again, the user study confirms this visual observation, in which AnaConDaR is ranked ahead of both modifications.

The influence of the facial characteristics and the missing blendshape can also be observed in Figure 13, in which each retargeting is performed once with 5 and once with 30 target examples. Although the relevance of both components is most evident when only a few target examples are available, the effects of both are still not negligible, even when there are many available target examples.

4.6. Collisions and Artistic Control

In this paragraph, we will briefly highlight two features that become feasible through the physics-based simulations involved in AnaConDaR.

First, Figure 14 displays our approach to resolving lip collisions. Not only do the upper and lower lip get disentangled, but the final volumetric simulation `anacon` of AnaConDaR (Section 3.4) propagates the displacements through the soft tissue.

Second, Figure 15 shows an example of artistic intervention into `anacon`. To that end, we manually modify vertices of the lip contour and add corresponding soft Dirichlet constraints to the forward component `fwdSim`. For streamlining the process, we move only a few control points and govern the remaining lip contour points through an RBF space warp [42].

This example only serves as one illustration of applicable artistic interventions. For instance, material properties, the weight of a character, or external forces, like varying gravity directions

Fig. 12: A visual ablation study that illustrates the individual components of AnaConDaR. In particular, the effects of enforcing expression-specific facial characteristics, adding the missing blendshape, and using ADT over DT become apparent. Additionally, the PLBS result and the missing blendshape are depicted. Please note, that we show the PLBS results after applying anatomical constraints (i.e., `anacon` w/o facial characteristics) for a reasonable comparison.

Fig. 13: AnaConDaR retargetings with 5 and 30 available target examples. For each instance, the PLBS component (after imposing anatomical constraints) and the missing blendshape are shown.

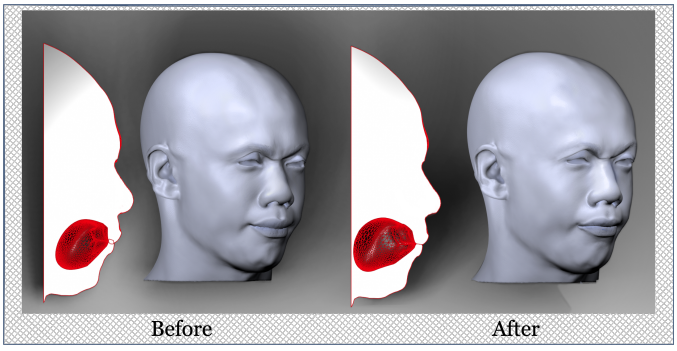

Fig. 14: An example of our method for resolving collisions. The lips get disentangled and the arising forces propagate through the soft tissue.

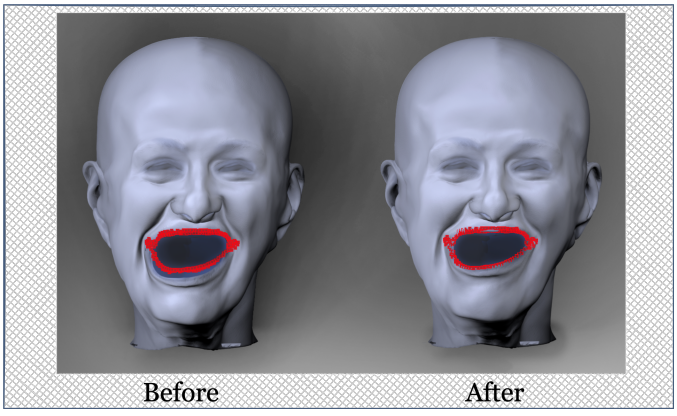

Fig. 15: An example of an artistic intervention into `anacon`. The targeted contour (red) is realized and the surrounding tissue is moved appropriately.

[47], can also be manipulated. Furthermore, artistic interventions into the patchwise blending weights of PLBS are inherited from ALM. For a more detailed description, please refer to [2].

5. Limitations

We assume that no global rigid motion occurs in facial expressions. Effective methods to achieve this prerequisite are available [48]. Moreover, AnaConDaR requires a shared mesh/patch topology of all source and target expressions. If this is not the case, a mapping can be found with unsupervised approaches [49, 50] or manually, for instance, using `faceform.com`. Concerning the physics-based simulations, we focus on the projective dynamics simulator [43] and do not add dynamic effects to obtain temporal independence. We chose projective dynamics because of its simplicity and sufficient efficiency, but other simulators can be used as drop-in replacements. Finally, we only handle self-collisions of the lips, while lip-teeth and eyelid collisions might also occur.

6. Conclusion

In this work, we introduced AnaConDaR, a method that integrates data-driven and geometry-driven facial retargeting algorithms. More precisely, the geometry-driven approach bridges informational gaps resulting from insufficiently expressive target examples within the data-driven approach. As a result, we enhance the current state-of-the-art ALM [2] to attain superior retargeting outcomes, particularly in situations where only a minimal number of target examples is available.

Due to the usage of patchwise linear blendshapes and the volumetric head representation, the user can readily guide and tailor AnaConDaR. The presented visually convincing qualitative examples of our approach are supported by two user studies and a quantitative analysis.

Promising future directions for improving AnaConDaR are to employ even more anatomically precise physics-based simulations and fully volumetric blendshapes [51]. Also, a more in-depth user study that queries rationales may facilitate targeted improvements. Finally, an accelerated GPU implementation of AnaConDaR could potentially achieve real-time operating speeds.

References

[1] Lewis, JP, Anjyo, K, Rhee, T, Zhang, M, Pighin, FH, Deng, Z. Practice and theory of blendshape facial models. Eurographics (State of the Art Reports) 2014;1(8):2.

[2] Chandran, P, Ciccone, L, Gross, M, Bradley, D. Local anatomically-constrained facial performance retargeting. ACM Transactions on Graphics (ToG) 2022;41(4):1–14.

[3] Yang, L, Zoss, G, Chandran, P, Gotardo, P, Gross, M, Solenthaler, B, et al. An Implicit Physical Face Model Driven by Expression and Style-Supplemental 2023;.

[4] Zielonka, W, Bolkart, T, Thies, J. Instant volumetric head avatars. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023, p. 4574–4584.

[5] Qian, S, Kirschstein, T, Schoneveld, L, Davoli, D, Giebenhain, S, Nießner, M. GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians. arXiv preprint arXiv:231202069 2023;.

[6] Li, H, Weise, T, Pauly, M. Example-based facial rigging. Acm Transactions on Graphics (ToG) 2010;29(4):1–6.

[7] Sumner, RW, Popović, J. Deformation transfer for triangle meshes. ACM Transactions on Graphics (ToG) 2004;23(3):399–405.

[8] Botsch, M, Sumner, R, Pauly, M, Gross, M. Deformation transfer for detail-preserving surface editing. In: Vision, Modeling & Visualization. Citeseer; 2006, p. 357–364.

[9] Onizuka, H, Thomas, D, Uchiyama, H, Taniguchi, Ri. Landmark-guided deformation transfer of template facial expressions for automatic generation of avatar blendshapes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019, p. 0–0.

[10] Chen, R, Chen, X, Ni, B, Ge, Y. Simswap: An efficient framework for high fidelity face swapping. In: Proceedings of the 28th ACM International Conference on Multimedia. 2020, p. 2003–2011.

[11] Garrido, P, Valgaerts, L, Rehmsen, O, Thormahlen, T, Perez, P, Theobalt, C. Automatic face reenactment. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 4217–4224.

[12] Kim, H, Elgharib, M, Zollhöfer, M, Seidel, HP, Beeler, T, Richardt, C, et al. Neural style-preserving visual dubbing. ACM Transactions on Graphics (ToG) 2019;38(6):1–13.

[13] Nirkin, Y, Keller, Y, Hassner, T. Fsgan: Subject agnostic face swapping and reenactment. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 7184–7193.

[14] Perov, I, Gao, D, Chervoniy, N, Liu, K, Marangonda, S, Umé, C, et al. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. arXiv preprint arXiv:200505535 2020;.

[15] Ren, Y, Li, G, Chen, Y, Li, TH, Liu, S. Pirenderer: Controllable portrait image generation via semantic neural rendering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, p. 13759–13768.

[16] Wang, Y, Chen, X, Zhu, J, Chu, W, Tai, Y, Wang, C, et al. Hififace: 3d shape and semantic prior guided high fidelity face swapping. arXiv preprint arXiv:210609965 2021;.

[17] Zhang, J, Zeng, X, Wang, M, Pan, Y, Liu, L, Liu, Y, et al. Freenet: Multi-identity face reenactment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, p. 5326–5335.
[18] Moser, L, Chien, C, Williams, M, Serra, J, Hendler, D, Roble, D. Semi-supervised video-driven facial animation transfer for production. ACM Transactions on Graphics (ToG) 2021;40(6):1–18.
[19] Hong, Y, Peng, B, Xiao, H, Liu, L, Zhang, J. Headnerf: A real-time nerf-based parametric head model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, p. 20374–20384.
[20] Bouaziz, S, Wang, Y, Pauly, M. Online modeling for realtime facial animation. ACM Transactions on Graphics (ToG) 2013;32(4):1–10.
[21] Chen, L, Cao, C, De la Torre, F, Saragih, J, Xu, C, Sheikh, Y. High-fidelity face tracking for ar/vr via deep lighting adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, p. 13059–13069.
[22] Lombardi, S, Saragih, J, Simon, T, Sheikh, Y. Deep appearance models for face rendering. ACM Transactions on Graphics (ToG) 2018;37(4):1–13.
[23] Cao, C, Simon, T, Kim, JK, Schwartz, G, Zollhoefer, M, Saito, SS, et al. Authentic volumetric avatars from a phone scan. ACM Transactions on Graphics (ToG) 2022;41(4):1–19.
[24] Garbin, SJ, Kowalski, M, Estellers, V, Szymanowicz, S, Rezaeifar, S, Shen, J, et al. VolTeMorph: Realtime, Controllable and Generalisable Animation of Volumetric Representations. arXiv preprint arXiv:220800949 2022;.
[25] Choi, B, Eom, H, Mouscadet, B, Cullingford, S, Ma, K, Gassel, S, et al. Animatomy: an Animator-centric, Anatomically Inspired System for 3D Facial Modeling, Animation and Transfer. In: SIGGRAPH Asia 2022 Conference Papers. 2022, p. 1–9.
[26] Yang, L, Kim, B, Zoss, G, Gözcü, B, Gross, M, Solenthaler, B. Implicit neural representation for physics-driven actuated soft bodies. ACM Transactions on Graphics (ToG) 2022;41(4):1–10.
[27] Kim, S, Jung, S, Seo, K, i Ribera, RB, Noh, J. Deep Learning-Based Unsupervised Human Facial Retargeting. In: Computer Graphics Forum; vol. 40. Wiley Online Library; 2021, p. 45–55.
[28] Feng, Y, Feng, H, Black, MJ, Bolkart, T. Learning an animatable detailed 3D face model from in-the-wild images. ACM Transactions on Graphics (ToG) 2021;40(4):1–13.
[29] Li, T, Bolkart, T, Black, MJ, Li, H, Romero, J. Learning a model of facial shape and expression from 4D scans. ACM Trans Graph 2017;36(6):194–1.
[30] Chandran, P, Bradley, D, Gross, M, Beeler, T. Semantic deep face models. In: 2020 international conference on 3D vision (3DV). IEEE; 2020, p. 345–354.
[31] Zhang, J, Chen, K, Zheng, J. Facial expression retargeting from human to avatar made easy. IEEE Transactions on Visualization and Computer Graphics 2020;28(2):1274–1287.
[32] Yang, H, Zhu, H, Wang, Y, Huang, M, Shen, Q, Yang, R, et al. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, p. 601–610.
[33] Li, J, Kuang, Z, Zhao, Y, He, M, Bladin, K, Li, H. Dynamic facial asset and rig generation from a single scan. ACM Trans Graph 2020;39(6):215–1.
[34] Kim, PH, Seol, Y, Song, J, Noh, J. Facial Retargeting by Adding Supplemental Blendshapes. In: PG (Short Papers). 2011,.
[35] Song, J, Choi, B, Seol, Y, Noh, J. Characteristic facial retargeting. Computer Animation and Virtual Worlds 2011;22(2-3):187–194.
[36] Ribera, RBI, Zell, E, Lewis, JP, Noh, J, Botsch, M. Facial retargeting with automatic range of motion alignment. ACM Transactions on Graphics (ToG) 2017;36(4):1–12.
[37] Xu, F, Chai, J, Liu, Y, Tong, X. Controllable high-fidelity facial performance transfer. ACM Transactions on Graphics (ToG) 2014;33(4):1–11.
[38] Bhat, KS, Goldenthal, R, Ye, Y, Mallet, R, Koperwas, M. High fidelity facial animation capture and retargeting with contours. In: Proceedings of the 12th ACM SIGGRAPH/eurographics symposium on computer animation. 2013, p. 7–14.
[39] Karypis, G, Kumar, V. METIS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices 1997;.
[40] Wu, C, Bradley, D, Gross, M, Beeler, T. An anatomically-constrained local deformation model for monocular face capture. ACM Transactions on Graphics (ToG) 2016;35(4):1–12.
[41] Achenbach, J, Brylka, R, Gietzen, T, zum Hebel, K, Schömer, E, Schulze, R, et al. A multilinear model for bidirectional craniofacial reconstruction. In: Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine. 2018, p. 67–76.
[42] Botsch, M, Kobbelt, L. Real-time shape editing using radial basis functions. In: Computer graphics forum; vol. 24. Blackwell Publishing, Inc Oxford, UK and Boston, USA; 2005, p. 611–621.
[43] Bouaziz, S, Martin, S, Liu, T, Kavan, L, Pauly, M. Projective dynamics: Fusing constraint projections for fast simulation. ACM Transactions on Graphics (ToG) 2014;33(4):1–11.
[44] Komaritzan, M, Botsch, M. Projective skinning. Proceedings of the ACM on Computer Graphics and Interactive Techniques 2018;1(1):1–19.
[45] Deuss, M, Deleuran, AH, Bouaziz, S, Deng, B, Piker, D, Pauly, M. ShapeOp—a robust and extensible geometric modelling paradigm. In: Modelling Behaviour. Springer; 2015, p. 505–515.
[46] Wang, Q, Tao, Y, Brandt, E, Cutting, C, Sifakis, E. Optimized processing of localized collisions in projective dynamics. In: Computer Graphics Forum; vol. 40. Wiley Online Library; 2021, p. 382–393.
[47] Wagner, N, Botsch, M, Schwanecke, U. Softdeca: Computationally efficient physics-based facial animations. In: Proceedings of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games. 2023, p. 1–11.
[48] Beeler, T, Bradley, D. Rigid stabilization of facial expressions. ACM Transactions on Graphics (ToG) 2014;33(4):1–9.
[49] Achenbach, J, Zell, E, Botsch, M. Accurate Face Reconstruction through Anisotropic Fitting and Eye Correction. In: VMV. 2015, p. 1–8.
[50] Schmidt, P, Pieper, D, Kobbelt, L. Surface Maps via Adaptive Triangulations. In: Computer Graphics Forum; vol. 42. Wiley Online Library; 2023, p. 103–117.
[51] Ichim, AE, Kavan, L, Nimier-David, M, Pauly, M. Building and animating user-specific volumetric face rigs. In: Symposium on Computer Animation. 2016, p. 107–117.

Appendix A. Energies

In the following, we formally state the individual energies of the forward simulation `simFwd` (Equation (18)).

Appendix A.1. Facial Characteristics

The energy for the facial characteristics

$$E_{\mathrm{F}}(\mathbb{S}_T, \mathrm{F}_i) = E_{\mathrm{eo}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{eo}}) + E_{\mathrm{es}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{es}}) + E_{\mathrm{lc}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{lc}}) \quad \text{(A.1)}$$

is composed of terms for the eye openings, the eye sockets, and the lip contour.

The energy for eye openings

$$E_{\mathrm{eo}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{eo}}) = \sum_{f \in \mathrm{EO}} \left(A(\mathbb{S}_T, f) - a_{\mathrm{eo}}(f)\right)^2 \quad \text{(A.2)}$$

penalizes for each triangular face $f \in \mathrm{EO}$ deviations in the surface area $A(\mathbb{S}_T, f)$ from the corresponding targeted surface area $a_{\mathrm{eo}}(f) \in \mathrm{F}_{\mathrm{eo}}$.

The energy for the eye sockets

$$E_{\mathrm{es}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{es}}) = \sum_{f \in \mathrm{ES}} \left(A(f) - a_{\mathrm{es}}(f)\right)^2 \quad \text{(A.3)}$$

penalizes for each triangular face $f \in \mathrm{ES}$ deviations in the surface area $A(\mathbb{S}_T, f)$ from the corresponding targeted surface area $a_{\mathrm{es}}(f) \in \mathrm{F}_{\mathrm{es}}$.

The energy for the lip contours

$$E_{\mathrm{lc}}(\mathbb{S}_T, \mathrm{F}_{\mathrm{lc}}) = \left\| (\mathbb{S}_T)^{LC} - \mathrm{F}_{\mathrm{lc}} \right\|^2 \quad \text{(A.4)}$$

draws the vertices $(\mathbb{S}_T)^{LC} \in \mathbb{S}_T$ to the corresponding vertices F_{lc}.

Appendix A.2. Tissue Deformations

The energy for the soft tissue

$$E_{\nabla\mathbb{S}}(\mathbb{S}_T, \nabla\mathbb{S}) = \sum_{t\in\mathbb{S}_T} \|\nabla(t, \mathbb{S}_T) - \mathbf{DG}_{\nabla\mathbb{S}}(t)\|_F^2 \tag{A.5}$$

penalizes for each tetrahedron $t \in \mathbb{S}_T$ deviations from the deformation gradient $\nabla(t, \mathbb{S})$ to the corresponding targeted deformation gradient $\mathbf{DG}_{\nabla\mathbb{S}}(t) \in \nabla\mathbb{S}$.

The energy for the muscle tissue

$$E_{\nabla\mathbb{M}}(\mathbb{M}_T, \nabla\mathbb{M}) = \sum_{t\in\mathbb{M}_T} \|\nabla(t, \mathbb{M}_T) - \mathbf{DG}_{\nabla\mathbb{M}}(t)\|_F^2 \tag{A.6}$$

penalizes for each tetrahedron $t \in \mathbb{M}_T$ deviations from the deformation gradient $\nabla(t, \mathbb{M})$ to the corresponding targeted deformation gradient $\mathbf{DG}_{\nabla\mathbb{M}}(t) \in \nabla\mathbb{M}_T$.

Appendix B. User Studies

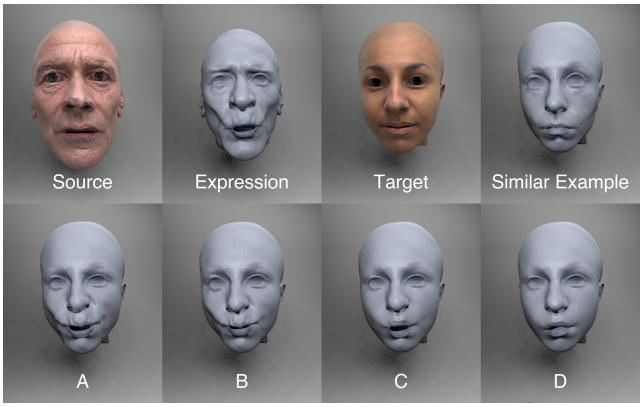

Fig. B.16: An instance of the user study, wherein 33 participants ranked the peer group of AnaConDaR. Consistent with the user study conducted by Chandran et al. [2], a real target example supported the participants in ranking. In this illustration, A-D are the results of EBFR, DT, AnaConDaR, and ALM. Generally, the results were placed in a random order.

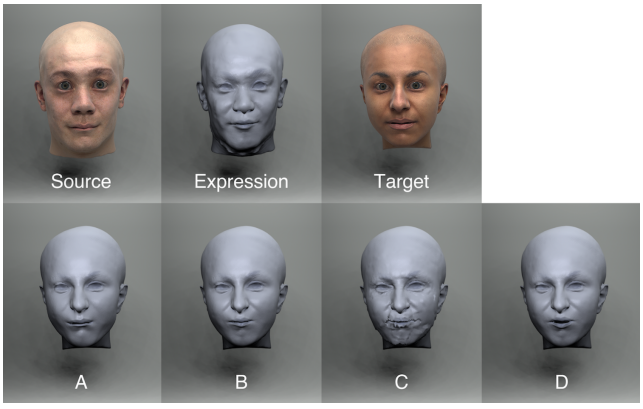

Fig. B.17: An instance of the user study, wherein 29 participants ranked individual components of AnaConDaR. In this illustration, A-D are the results of AnaConDaR without the missing blendshape, AnaConDaR, AnaConDaR with DT, and AnaConDaR without facial features. Generally, the components were placed in a random order.