Anatomically Constrained Implicit Face Models

Prashanth Chandran
DisneyResearch|Studios
prashanth.chandran@disneyresearch.com

Gaspard Zoss
DisneyResearch|Studios
gaspard.zoss@disneyresearch.com

Abstract

Coordinate based implicit neural representations have gained rapid popularity in recent years as they have been successfully used in image, geometry and scene modeling tasks. In this work, we present a novel use case for such implicit representations in the context of learning anatomically constrained face models. Actor specific anatomically constrained face models are the state of the art in both facial performance capture and performance retargeting. Despite their practical success, these anatomical models are slow to evaluate and often require extensive data capture to be built. We propose the anatomical implicit face model; an ensemble of implicit neural networks that jointly learn to model the facial anatomy and the skin surface with high-fidelity, and can readily be used as a drop in replacement to conventional blendshape models. Given an arbitrary set of skin surface meshes of an actor and only a neutral shape with estimated skull and jaw bones, our method can recover a dense anatomical substructure which constrains every point on the facial surface. We demonstrate the usefulness of our approach in several tasks ranging from shape fitting, shape editing, and performance retargeting.

1. Introduction

Deformable face models are an important tool in the arsenal of visual effects artists dealing with facial animation. As they are ubiquitously used both in high-end production workflows and lightweight consumer applications, building expressive face models for various applications continues to remain an active area of research [17]. Face models today can range from simple linear global shape models [4, 27, 29] to highly complex local models that incorporate the underlying facial anatomy through physical simulation [15, 44, 48] or through anatomical constraints [47].

In this work, we concern ourselves primarily with the high-quality facial animation workflow where actor specific linear blendshape models [27] continue to remain the most commonly used tool for creating facial animations [10, 33, 47]. We propose a new class of actor specific shape models named the *Anatomical Implicit face Model* (AIM) which provides several unique advantages over the existing actor specific face models, and can be used as a drop-in replacement for traditional blendshape models.

An actor specific blendshape model is a collection of 3D shapes of the given actor performing a number of facial expressions, usually created by face scanning [2] or by an artist. While the user-friendliness of such actor specific blendshape models contributes to their wide adoption, it is a well known limitation that such models often require hundreds of shapes to accurately model complex facial deformation [27]. To address these shortcomings, local blendshape models [10, 42, 47] were proposed. By splitting the face into regions, and allowing the individual regions to deform independently, local shape models are able to capture complex deformations with a limited number of shapes.

While local models address the lack of expressivity in global shape models, state-of-the-art methods in facial performance capture [47] and retargeting [10] often incorporate anatomical constraints on the facial surface to plausibly restrict the range of the skin deformations. The anatomical constraints employed by these models [10, 47] provide a few hidden advantages that end up contributing towards their practical success. For example, in the context of facial performance capture, Wu *et al*. [47] demonstrated that including anatomical constraints derived from the relationship between the facial skin and underlying bones (skull and mandible) helps to separate the rigid and non-rigid components of facial deformation, leading to better face performance capture. In the context of facial performance retargeting, Chandran *et al*. [10] made use of such an anatomically constrained local face model to restrict a retargeted shape to lie within the space of anatomically plausible shapes of the target actor.

Despite their practical success, anatomical constraints are often formulated in practice as regularization terms that have to be satisfied as part of complex optimization problems involving several objectives. As a result, fitting these anatomical face models to a target scan or an image for instance, is a computationally intensive procedure taking several minutes per frame on a CPU, or requires hand crafted

GPU solvers [20]. Furthermore anatomy constraints are enforced only in sparse regions of the face, whereas in reality the facial skin surface is more densely constrained by the underlying anatomy, and simulating this dense interaction between the anatomy and facial skin through physical simulation can be even more computationally intensive [39, 48].

In this paper, we propose the *Anatomical Implicit face Model*; a framework that allows for a holistic representation of both the facial anatomy and the skin surface using simple implicit neural networks and facilitates the learning of a continuous anatomical structure that densely constrains the skin surface. Our model formulation, inspired by the anatomical local model (ALM) of Wu *et al*. [47], can further disentangle deformation arising from rigid bone motion (jaw motion) and non-rigid deformations created by muscle activations. Our model also addresses the computational bottleneck of the ALM model by explicitly deriving the skin surface from the anatomy, instead of formulating it as a constrained optimization problem. By ensuring that a point on the skin surface is always reconstructed through the underlying anatomy, our method provides several unique features in comparison to existing implicit face models, such as anatomy based face manipulation (see Section 5). Before describing the details of our anatomical formulation in Section 3, we discuss related work in Section 2.

2. Related Work

3D Morphable Models Facial models used in animation make up for an extremely well studied body of work with the earliest works dating back to the late 1970s [18]. We therefore refer to the survey of Egger *et al*. [17] for an in-depth review of the state-of-the-art methods, and provide only a concise summary in this section. Facial blendshapes [18, 27] have been conventionally used as a standard tool by artists to navigate the geometric space of human faces. The seminal 3D linear morphable model proposed by Blanz and Vetter [4] used principal component analysis to describe the variation in facial geometry and texture, which was later extended to multilinear models, jointly modeling identity and expression by Vlasic *et al*. [43] and later by Cao *et al*. [7]. Today a very commonly used morphable face model is the FLAME model [29] which incorporates identity, expression and corrective blendshapes in addition to modeling bone motion with linear blend skinning. Due to its flexible nature, the FLAME model is widely used by face reconstruction algorithms today [19]. Finally Chai *et al*. [8] recently created the HIFI3D++ morphable model which is built from a union of scans from several previously proposed models.

In the past few years, numerous face models leveraging the power of deep neural networks to model the non-linear deformation of the human face have also been proposed. While the initial work in this area by Ranjan *et al*. [38] focused on the use of specialized graph convolutional networks to operate on shapes, several later approaches proposed further modifications to the network architecture to improve the accuracy in shape representation [5, 14, 22, 55]. To make these deep morphable models intuitive to use, Chandran *et al*. [9] subsequently proposed the Semantic Deep Face Model which treats a collection of neural networks like a multilinear model to achieve identity-expression disentanglement. Extensions of such a semantically controllable model to deal with topology changes [12] and temporal sequences of geometry [11] have also been proposed. Deep neural models that jointly model the facial geometry and appearance with semantic controls have also been proposed [28].

Implicit Face Models Owing to the success of coordinate based neural networks in representing images [30, 40], 3D shapes [35] and arbitrary scenes [31], today's research on parametric face models primarily focuses on implicit representations. Yenamandra *et al*. [49] proposed *i3DMM* as an initial exploration of using coordinate based networks for modeling full head geometries. This was followed by IMFace [51] which disentangled facial geometry into separate identity and expression embeddings with the help of individual deformation fields. More recently, Neural Parametric Head Models (NPHM) [21] proposed a method which improves the fidelity of neural implicit representations by jointly training an ensemble of local neural fields centered around anchor points. Implicit neural representations have also successfully been employed in learning an animatable avatar of a human face from only monocular video as demonstrated by IMAvatar [52] and Point Avatar [53]. Wang *et al*. [45] also proposed MoRF, which is a Neural Radiance Field [31] conditioned on an identity code allowing for photorealistic free viewpoint rendering of the full head in a fixed expression. Recently Buhler *et al*. [6] also explored how such multi-identity radiance fields can be fit to sparse images to recover a volumetric head model. Finally coordinate based neural networks have also been successfully employed in creating animatable human body models [3, 16, 23, 34].

Anatomically Constrained Face Models The anatomical local model proposed in the context of monocular facial performance capture by Wu *et al*. [47], first introduced the coupling of the anatomical bone structure to the skin surface and modeled the effect of skin patches sliding over the bone through soft anatomical constraints. This formulation was later adapted by Chandran *et al*. [10] for facial performance retargeting. Qiu *et al*. proposed *SCULPTOR* [37], a multi-identity joint morphable model of facial anatomy and skin learned from a database of computed tomography (CT) scans. Recently Choi *et al*. proposed *Animatomy* [15], a muscle fiber based anatomical basis for animator friendly face modeling applications. Lastly we recognize several

physically based face models [39, 41, 44, 48] which inherently have the ability to model anatomy constraints through simulation.

We draw inspiration from the three classes of facial morphable models discussed above and propose the *Anatomical Implicit face Model*: a blendshape based, implicit, anatomically constrained face model targeted towards high-quality actor specific face modeling. Our method can be seen as general extension of local blendshape models [10] to a continuously evaluable implicit function, and represents a set of actor blendshapes through a novel anatomical formulation. Unlike traditional patch-based models, our framework allows us to approximate complex shapes without requiring the user to specify patch layouts and other hyperparameters. Our solution is based on simple coordinate based MLPs enabling efficient training and inference, and provides computational benefits over previous anatomically formulated face models [47]. Finally to the best of our knowledge, our method is the first to explore anatomical constraints inside an implicit facial blendshape model.

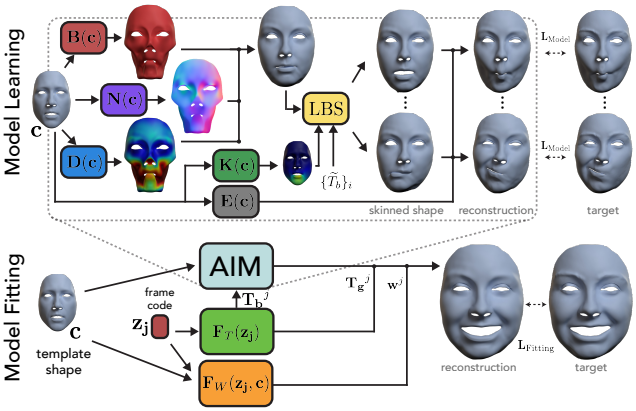

Figure 1. Our approach consists of a model learning stage (Section 4.1) and a model fitting stage (Section 4.2). In the model learning stage, a set of an actor's blendshapes are memorized by an ensemble of MLPs by our *Anatomical Implicit face Model* (AIM). In the second model fitting stage, the memorized model can be used as power shape prior to fit the actor model to target shapes.

3. Anatomical Model Formulation

The core idea of our approach is to formulate a learning scheme for an implicit neural representation that can reproduce an actor blendshape model while automatically learning the underlying facial anatomy and constraining the skin surface to this learned anatomy. Crucial to our learning scheme is our anatomically constrained face model that geometrically couples the underlying facial anatomy to the enclosing skin surface which we describe next.

We assume that we are given a set of N 3D scans $(S_0, S_1, S_2, .., S_{N-1})$ of an actor represented as meshes. Without loss of generality, let S_0 be the shape with a neutral expression (or the rest pose). Each shape S_i consists of V vertices, and all shapes share the same vertex connectivity. For simplicity we exclude the index of the vertex in a shape in our notation and present our formulation as operating on surface points $\mathbf{s} \in \mathbf{R}^3$. Let $\mathbf{s_0} \in \mathbf{R}^3$ and $\mathbf{s_i} \in \mathbf{R}^3$ be corresponding points on the skin surface for the neutral expression and expression i respectively. In most previous methods for learning neural face models, a skin surface point $\mathbf{s}$ is learned as a displacement from a base face surface [9, 12, 21] or simply as points lying in an arbitrary 3D space [45, 51, 52]. Contrary to such approaches, we propose to learn the skin surface $\mathbf{s}$ using implicit neural representations that arrives at the facial skin surface through a formulation that combines anatomical constraints, linear blend skinning (LBS), and expression blendshapes into a single framework.

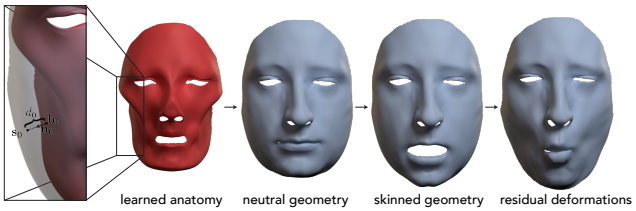

Figure 2. We show the break down of how we anatomically build up the facial skin surface. Starting from a learned anatomy surface (left), and learned anatomical properties like the soft tissue thickness, and anatomical surface normals, we reconstruct the neutral skin geometry. The neutral anatomy is skinned, and non-rigidly deformed with residual displacements to result in the final shape.

For our model formulation, we take inspiration from the anatomical constraints first proposed for non-neural face models [1, 47], particularly that of Wu *et al.* [47]. They establish a link between the skin surface and the anatomical bones by modeling the thickness $d_i \in \mathbf{R}$ of the soft tissue between a bone point $\mathbf{b_i} \in \mathbf{R}$ and the skin surface $\mathbf{s_i}$. These constraints are defined in sparse regions of the face where a skin point can be trusted to have bone underneath. We draw inspiration from their simple formulation and make some important deviations that enable us to jointly learn both the surface of the underlying skin anatomy and the enclosing skin surface for *every* point on the skin through end-to-end learning. Specifically, we arrive at a point on the skin surface as follows

$$\mathbf{s_0} = \mathbf{b_0} + d_0\mathbf{n_0} \tag{1}$$

where $\mathbf{s_0}$ is the position of a surface point corresponding to $\mathbf{s}_i$ but on the neutral shape S_0, $\mathbf{b_0}$, d_0, and $\mathbf{n_0}$ are the bone point, soft tissue thickness and the bone normal at $\mathbf{s_0}$. While Eq. 1 allows us to reconstruct points on the neutral face geometry, to adequately represent skin surfaces under arbitrary facial expressions, we need to account for surface deformation arising from the rigid motion of underlying facial bones (skull and mandible), and the non-rigid skin motion arising from muscle activations, skin sliding, and

self collisions. To accommodate these additional degrees of freedom in skin deformation, we incorporate standard linear blend skinning, and expression blendshapes similar to the FLAME model [29]. Therefore given an anatomically reconstructed point on the neutral skin surface $\mathbf{s_0}$, we can now compute the position of the same point in an arbitrary expression $\mathbf{s_i}$ as

$$\mathbf{s_i} = \text{LBS}(\mathbf{s_0}, T_b, k) + \mathbf{e_i} \tag{2}$$

where LBS refers to the standard linear blend skinning operator that rigidly transforms the anatomically reconstructed neutral surface point $\mathbf{s_0}$ with a transformation T_b and a skinning weight k, $\mathbf{e_i} \in \mathbf{R}^3$ is the corrective displacement that is added on top of the skinned result to account for deformations that cannot be explained by skinning alone. A visual overview of our approach to anatomically build up the facial skin surface is shown in Fig. 2.

At this point we have established how to arrive at points on the skin surface $\mathbf{s_i}$ for a shape in an arbitrary facial expression S_i by starting from the underlying anatomy $\mathbf{b_i}$. It is important to note that the anatomical constraints as defined by Wu *et al*. [47] can only be computed on regions with an underlying bone, and thus, regions like the cheeks are not anatomically constrained in their approach. An essential feature of our approach that distinguishes it from all previous works is that we enforce anatomical constraints for every point on the skin surface; even in regions where there is no underlying biological bone structure. For this purpose we redefine the anatomy in our work as a rigidly deforming region underneath the skin surface that is not restricted to only the manifold of the skull and mandible bones. Since this structure does not exist in reality and is, therefore, not available for supervised learning, we formulate a learning framework where such rigidly deforming surface can be learned only from the sparse set of anatomical constraints computed between the skin and the underlying bones. As we will see in Section 5, learning this anatomical surface from data leads to several interesting applications in shape manipulation and performance retargeting that were previously challenging to obtain without expensive physical simulation [48] or extensive volumetric data capture [37].

4. Anatomical Implicit Face Model

At a high level, our method is comprised of two stages: first, a model learning stage (Section 4.1) and second, a model fitting stage (4.2). In the model learning stage, we bake a collection of expression blendshapes from an actor into an implicit neural network that uses the anatomical model formulation described in Section 3. Our model fitting stage uses this learned *Anatomical Implicit face Model* (AIM) and optimizes for coefficients that deform the model to match test time constraints like 3D shapes, 2D landmarks and so on. The overview of our approach is shown in Fig. 1.

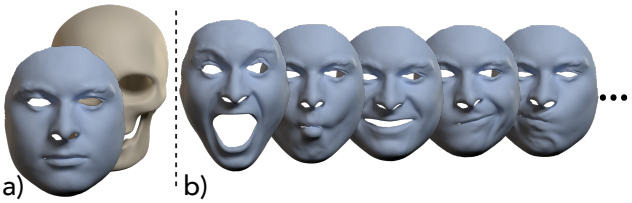

Figure 3. a) We assume we are given the neutral geometry of an actor along with an rough estimate of the skull and jaw bone [56]. b) We additionally use a collection of N 3D shapes of the actor performing expressions. Unlike Wu *et al*. [47], we do not require the tracked anatomy (skull [1], jaw [57]) for the expression shapes.

4.1. Model Learning

To learn our anatomical implicit face model, we assume we are given a template shape C, a registered set of N shapes $(S_0, S_1, S_2, .., S_{N-1})$ of a single actor in the same topology of the canonical shape. Additionally we fit a template skull and jaw *only* to the neutral shape using the method of Zoss *et al*. [56]. The template shape C can either be the neutral shape of the actor or a generic face shape, and the number of shapes provided can be arbitrary. We use a collection of 20 shapes in our work. A visual summary of our training data is shown in Fig. 3. Our objective in the learning stage is to use a coordinate based neural network to memorize the given shapes through the anatomical formulation in Section 3. Given the high representation power of periodic implicit neural networks [40], we use the SIREN coordinate network; an MLP with sinusoidal activation functions, as our base architecture. An ablation study on alternate network choices is provided in our supplemental.

Given a point $\mathbf{c} \in \mathbf{R}^3$ on the template shape C, we use three independent MLPs denoted by $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{N}$ to predict the anatomy point $\widetilde{\mathbf{b}}_0 \in \mathbf{R}^3$, the soft tissue thickness $\widetilde{d}_0 \in \mathbf{R}$, and the anatomy normal $\widetilde{\mathbf{n}}_0 \in \mathbf{R}^3$. These predicted anatomical properties are then used to reconstruct the position of a point on the neutral skin surface $\widetilde{\mathbf{s}}_0$ as

$$\widetilde{\mathbf{b}}_0 = \mathbf{B}(\mathbf{c}) \tag{3}$$

$$\widetilde{d}_0 = \mathbf{D}(\mathbf{c}) \tag{4}$$

$$\widetilde{\mathbf{n}}_0 = \mathbf{N}(\mathbf{c}) \tag{5}$$

$$\widetilde{\mathbf{s}}_0 = \widetilde{\mathbf{b}}_0 + \widetilde{d}_0 \widetilde{\mathbf{n}}_0. \tag{6}$$

As discussed in Section 3, to further account for the rigid and non-rigid deformations of the skin surface, the anatomically constructed neutral skin point $\widetilde{\mathbf{s}}_0$ has to be skinned and further displaced with residual expression deformations. We therefore employ two additional MLPs $\mathbf{K}$ and $\mathbf{E}$ that predict the skinning weight $\widetilde{k} \in \mathbf{R}$ and the corrective displacements basis $\mathcal{B}_\mathbf{e} \in \mathbf{R}^{(N-1)\times 3}$ respectively. Note here that, as an implementation detail, we predict the expression displacements for all $N-1$ blendshapes (excluding the neutral) at once from $\mathbf{E}$. The corrective expression

displacement $\widetilde{\mathbf{e_i}} \in \mathbf{R}^3$ for shape i can be extracted from this output by indexing $\mathcal{B}_\mathbf{e}$ appropriately.

$$\widetilde{k} = \mathbf{K}(\mathbf{c}) \tag{7}$$

$$\mathcal{B}_\mathbf{e} = \mathbf{E}(\mathbf{c}) \tag{8}$$

$$\widetilde{\mathbf{e_i}} = \mathcal{B}_\mathbf{e}[i] \tag{9}$$

$$\widetilde{\mathbf{s}}_i = \text{LBS}\left(\widetilde{\mathbf{s}}_0, \widetilde{T}_b, \widetilde{k}\right) + \widetilde{\mathbf{e_i}} \tag{10}$$

Here $\widetilde{T}_b \in \mathbf{R}^9$ is a 6-DOF jaw bone transformation optimized along with the training of the MLPs to account for rigid motion of the mandible. Here we parameterize the jaw bone rotation $\widetilde{T}_b$ following the continuous 6D representation [54].

4.1.1 Training Objectives

We next describe the training objectives to learn actor expression blendshapes along with the underlying anatomy structure for each skin surface point.

Skin Position Loss The skin position loss penalizes the difference between the estimated skin point $\widetilde{\mathbf{s}}_i$ and the ground truth skin point $\mathbf{s}_i$.

$$\mathbf{L}_\text{S} = \lambda_\text{S} ||\widetilde{\mathbf{s}}_\mathbf{i} - \mathbf{s}_\mathbf{i}||_\mathbf{2}^\mathbf{2} \tag{11}$$

We set $\lambda_\text{S} = 1.0$ for all our experiments.

Anatomy Regularizer Since we can roughly estimate the skull and jaw geometry on the neutral shape using the method of Zoss *et al*. [56], we compute sparse anatomical constraints [47] and loosely regularize the learned anatomical properties to stay close to these estimates only in regions where the constraints can be accurately computed (i.e. skin regions with an underlying bone).

$$\mathbf{L}_\text{A} = \lambda_b ||\widetilde{\mathbf{b}}_\mathbf{0} - \mathbf{b}_\mathbf{0}||_\mathbf{2}^\mathbf{2} + \lambda_\mathbf{d} ||\widetilde{\mathbf{d}}_\mathbf{0} - \mathbf{d}_\mathbf{0}||_\mathbf{2}^\mathbf{2} + \lambda_\mathbf{n} ||\widetilde{\mathbf{n}}_\mathbf{0} - \mathbf{n}_\mathbf{i}||_\mathbf{2}^\mathbf{2} \tag{12}$$

We set $\lambda_\text{b} = \lambda_\text{d} = \lambda_\text{n} = 1.0$ for all our experiments, and observe that this constraint only regularizes 5-10% of all the vertices generated by the model on average (see Supplemental).

Thickness Regularizer We regularize the soft tissue thickness $\widetilde{d}$ predicted by the model in unconstrained regions to remain as small unless dictated otherwise by the skin position loss.

$$\mathbf{L}_\text{D} = \lambda_D^{Reg} ||\widetilde{d}_0||_2^2 \tag{13}$$

We set $\lambda_D^{Reg} = 7.5\text{e}{-4}$ for all our experiments.

Symmetry Regularizer To exploit the symmetry of the face, we regularize the predictions of the anatomy MLP $\mathbf{B}$ to be symmetric. We achieve this by requiring that reflecting the input points $\mathbf{c}$ along the plane of symmetry provides the same result as reflecting the predicted anatomy points $\widetilde{\mathbf{a}}$.

$$\mathbf{L}_\text{Sym} = \lambda_{sym} ||\mathbf{B}(\mathbf{R}(\mathbf{c})) - \mathbf{R}(\mathbf{B}(\mathbf{c}))||_\mathbf{2}^\mathbf{2} \tag{14}$$

where R is an operator that reflects a point along the plane of symmetry. We set $\lambda_{sym} = 1\text{e}{-4}$ for all our experiments. Note that we do not regularize symmetry on the predicted thickness or anatomy normals thereby allowing the model to still be able to represent asymmetric faces.

Optional **Skinning Weight Regularizer** Finally inspired by [52], we use an *optional* loss that encourages the estimated skinning weights $\widetilde{k}$ in regions like the forehead that are guaranteed to not be affected by the rigid deformation of the jaw bone to be zero.

$$\mathbf{L}_\text{K} = \lambda_k ||\mathbf{K}(\mathbf{c}^*)||_2^2 \tag{15}$$

here $\mathbf{c}^*$ refers to a small region on the canonical shape C which includes the forehead. We set $\lambda_K = 1\text{e}2$ for all our experiments.

Our final model energy $\mathbf{L}_\text{Model}$ is a summation of the above losses and is minimized using gradient decent [26] to train our ensemble of coordinate MLPs end-to-end.

$$\mathbf{L}_\text{Model} = \mathbf{L}_\text{S} + \mathbf{L}_\text{A} + \mathbf{L}_\text{D} + \mathbf{L}_\text{Sym} + \mathbf{L}_\text{K} \tag{16}$$

4.2. Model Fitting

While the aforementioned model can recover interesting anatomical properties of the face with only sparse supervision, it is not very useful unless it can be deformed to match user constraints and serve as a shape prior for an actor facial geometry.

After training our anatomical implicit face model on a collection of N shapes, the coefficients that are required to deform it include a jaw bone transformation $\mathbf{T_b}^* \in \mathbf{R}^9$, coefficients $\mathbf{w}^* \in \mathbf{R}^{N-1}$ that can be used to blend the corrective expression displacements $\mathcal{B}_e \in \mathbf{R}^{(N-1)\times 3}$, and an optional global head transformation $\mathbf{T_g}^* \in \mathbf{R}^9$. Following Equation (10), we can therefore evaluate our anatomical implicit face model as

$$\mathbf{s}^* = \mathbf{T_g}^* \left(\text{LBS}\left(\widetilde{\mathbf{s}}_0, \mathbf{T_b}^*, \widetilde{k}\right) + \sum_{N-1} \mathbf{w}^* \mathbf{B}_e \right) \tag{17}$$

where $\mathbf{T_g}^*$, $\mathbf{T_b}^*$ and $\mathbf{w}^*$ are the only unknowns, and the rest can be queried from a pre-trained AIM. We consider two scenarios for model fitting which include i) fitting our model to a sequence of 3D scans *e.g*. from a facial performance, and ii) fitting our model to 2D landmarks detected on a video [13, 46].

For both scenarios, inspired by the state-of-the-art findings of Kim *et al*. [50], we employ neural reparameterized optimization [25] and solve for the weights of a simple MLP that predicts the unknown parameters instead of directly optimizing for them. Specifically when given a sequence of J frames with 3D/2D constraints, we optimize for J frame codes $\mathbf{z_j} \in \mathbf{R}^f$ which, when fed as input to a simple 4-layer

MLP $\mathbf{F}_T$ with GeLU [24] activations, predicts the head ${\mathbf{T_g}}^j$ and jaw ${\mathbf{T_b}}^j$ poses for each frame. Additionally as the coefficients $\mathbf{w}^j$ are local and spatially varying depending on the template query point $\mathbf{c}$, we use a separate 4-layer MLP $\mathbf{F}_W$ which predicts the coefficients $\mathbf{w}^j$ by taking both the frame code $\mathbf{z_j}$ and the query point $\mathbf{c}$ as input.

$$[{\mathbf{T_g}}^j, {\mathbf{T_b}}^j] = \mathbf{F}_T(\mathbf{z_j}) \tag{18}$$

$$\mathbf{w}^j = \mathbf{F}_W(\mathbf{z_j}, \mathbf{c}) \tag{19}$$

Unlike the method of Kim *et al*. [50] where the reparameterized optimization was used mainly for improved performance, this neural optimization is even necessary in our case to restrict the number of optimized variables as the number of spatially varying coefficients $\mathbf{w}^*$ used to evaluate our anatomical implicit face model can vary drastically depending on the number of constraint points (see Section 5).

4.2.1 Fitting Objectives

3D Position Constraint For fitting our trained model to 3D constraints coming from a facial performance of an actor, we minimize the euclidean distance between the estimated skin point $\mathbf{s}^*$ and the ground truth skin point $\mathbf{s}^{GT}$. However by constraining only the final skin surface, expression displacements could overcompensate for the skinned geometry. To prevent this from happening, we additionally require the skinned shape without corrective displacements $\mathbf{s}^*_{lbs}$ to be as close as possible to the ground truth skin point.

$$\mathbf{L}^{3D}_{\text{Pos}} = \lambda_{3D} \left(||\mathbf{s}^* - \mathbf{s}^{GT}||^2_2 + ||\mathbf{s}^*_{lbs} - \mathbf{s}^{GT}||^2_2 \right) \tag{20}$$

2D Position Constraint For fitting our model to 2D constraints such as facial landmarks estimated by a pre-trained landmark detector [13, 46], we project the estimated skin point $\mathbf{s}^*$ to screen space using known camera intrinsics ψ and calculate the euclidean distance in 2D between the project point $\psi(\mathbf{s}^*)$ and the corresponding landmark.

$$\mathbf{L}^{2D}_{\text{Pos}} = \lambda_{2D} \left(||\psi(\mathbf{s}^*) - \mathbf{p}||^2_2 + ||\psi(\mathbf{s}^*_{lbs}) - \mathbf{p}||^2_2 \right) \tag{21}$$

$\mathbf{p} \in \mathbf{R}^2$ is a detected landmark corresponding to point $\mathbf{s}^*$.

Coefficient Regularizer As the complexity of our implicit anatomical face model can be arbitrarily large, we regularize the estimated blending coefficients $\mathbf{w}^*$ to be small with a weak L2 regularizer.

$$\mathbf{L}_W = \lambda^w_{Reg} ||\mathbf{w}^*||^2_2 \tag{22}$$

We set $\lambda^w_{Reg} = 0.75$ for all our experiments.

Temporal Regularizer Finally when optimizing for coefficients on sequential data, we regularize the optimized frame codes $\mathbf{z_j}$ to remain similar between adjacent frames.

$$\mathbf{L}_T = \lambda^t_{Reg} ||\mathbf{z_j} - \mathbf{z_{j-1}}||^2_2 \tag{23}$$

We set $\lambda^t_{Reg} = 0.05$ for all our experiments. Our final fitting energy $\mathbf{L}_{\text{Fitting}}$ is therefore

$$\mathbf{L}_{\text{Fitting}} = \mathbf{L}^{3D}_{\text{Pos}} + \mathbf{L}^{2D}_{\text{Pos}} + \mathbf{L}_{\text{W}} + \mathbf{L}_{\text{T}} \tag{24}$$

For 3D/2D fitting, we set λ_{2D} and λ_{3D} to 0 respectively.

4.3. Implementation Details

In the model learning stage, we optimize our implicit coordinate networks for $1e4$ iterations with a learning rate of $2e{-}3$. This takes approximately 10 minutes to converge on a single Nvidia RTX 3090 for an actor model with 40,000 vertices and 20 blendshapes. In the model fitting stage, we use a learning rate of $1e{-}3$ and optimize the fitting MLPs $\mathbf{F}_T$ and $\mathbf{F}_W$ for $1e4$ iterations. This process takes 1 second per frame on a single Nvidia RTX 3090. We implement all our MLPs in PyTorch [36]. In our supplementary material we discuss the performance implications of replacing our current python backend with the well engineered fused MLP implementation [32].

5. Results

We now present several results, applications and evaluations of our *Anatomical Implicit face Model* (AIM).

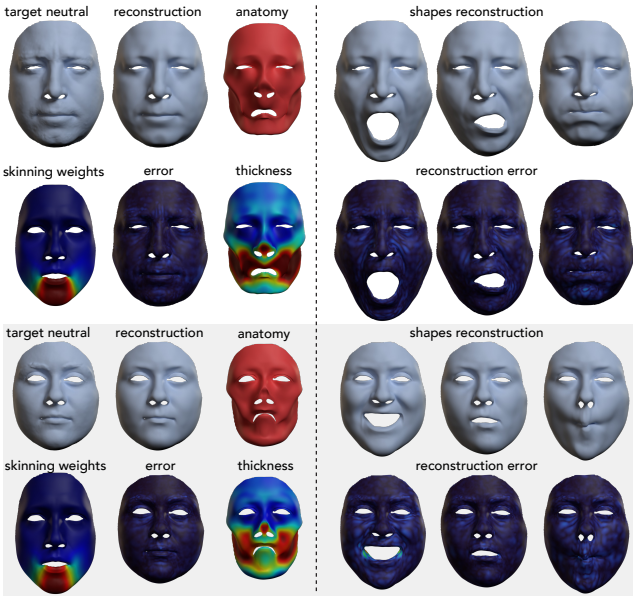

Figure 4. We demonstrate the ability of our *Anatomical Implicit face Model* to recover plausible anatomic features of the face, while also modeling the skin surface with very high fidelity. A subset of 3 expressions from 2 different actor specific models are shown here. The errors are displayed with a scale of 0mm 5mm.

5.1. Learning Actor Specific Anatomy

We begin by showing the reconstruction accuracy of our AIM on facial blendshapes of multiple actors. As seen in

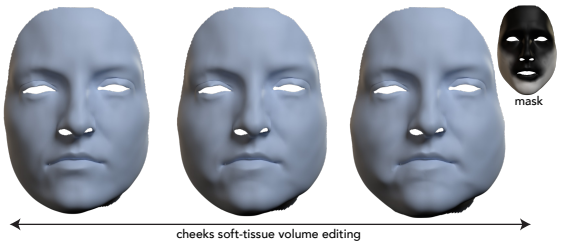

Figure 5. Once the AIM is learned for an actor, it can be used to intuitively deform a face using the learned anatomic properties, as demonstrated here by scaling the soft tissue thickness in a hand painted cheek region, and by propagating the change to the skin surface thanks to our formulation.

Fig. 4 on 2 different actors, our method can faithfully represent facial shapes with high fidelity while capturing both the low and high frequency features of facial shape and expression. We also show the anatomic features recovered by our new formulation which includes the dense underlying facial anatomy (shown in red), the soft tissue thickness at every point on the anatomy (visualized as heatmap), and the optimized subject specific skinning weights. These results highlight the new abilities introduced by our method in recovering plausible anatomy features while jointly learning to model surface deformations.

5.2. Anatomy Manipulation

Our ability to estimate the underlying anatomy that densely constrains the skin surface opens up new, yet computationally inexpensive ways to edit facial geometry using our learned anatomic properties. For example, as illustrated in Fig. 5, by simply scaling the learned soft tissue thickness d in desired regions of the face (denoted by the hand drawn mask), an artist can interactively sculpt/deform an actor's face shape to match their requirements.

5.3. Expression Reconstruction

We next evaluate the expressiveness of our model by fitting it to unseen 3D performances of multiple actors. Given a sequence of J dynamic 3D shapes from a studio scanner [2], we first deform our template mesh C to match the scanned shapes using standard mesh registration techniques such that the dynamic 3D scans are in full vertex correspondence with our AIM. We then follow the fitting procedure described in Section 4.2 and obtain per-frame transformations $[\mathbf{T_g}^j, \mathbf{T_b}^j]$ and shape coefficients $\mathbf{w}^j$ that explain the captured ground truth shape. For this experiment, we use the 3D position constraint from Eq. (24) and set $\mathbf{L_{2D}}$ to 0. We densely constrain the fitting procedure at every vertex of the ground truth shape. In Fig. 6 we provide both a qualitative and quantitative comparison of fitting to novel performance from an actor against global blendshapes (GBS) [27], a patch blendshape model (PBS) [13], and the anatomical local model (ALM) [47]. In this experiment, we use 20 ground truth actor blendshapes to build the GBS, PBS, and ALM models, and the anatomically reconstructed blendshapes for our method. Even under this slight disadvantage, our method outperforms both GBS, and PBS and provides visually comparable results to the ALM model. Table 1 shows the average fitting error of each method across 819 frames from 5 sequences of 5 different actors. Ours (G) refers to a variant of our fitting algorithm where the expression coefficients are applied globally to obtain a face shape. Our method converges in a few seconds for each frame, while the ALM algorithm consistently requires several minutes per frame. While the continuous nature of AIM enables us to evaluate it with coefficients of arbitrary locality, it could result in situations where our fitting is underconstrained in the absence of dense constraints leading to broken shapes. To illustrate that this does not happen in our reparameterized optimization, we show the result of fitting the AIM to sparse constraints in Fig. 7. While increasing the density of constraints improves the fitting accuracy, fitting our model to sparse landmarks also provides plausible results. Note that we do not compare fitting accuracy against generic morphable models like FLAME [29] or NPHM [21] as ours is actor specific and therefore a quantitative comparison might be unfair to the other methods. We kindly refer to our supplemental material for more results.

Table 1. Average fitting error (in mm) across 819 frames from 5 sequences of 5 different actors.

GBS [27]	**PBS** [13]	**ALM** [47]	**Ours (G)**	**Ours**
0.83	0.51	0.09	0.86	0.31

5.4. 3D Performance Retargeting

Another important application of our method is in the area of 3D performance retargeting, where the goal is to transfer a facial animation from a source to a target character while respecting the identity and anatomical characteristics of the target character. To accomplish this using our model, we learn two separate instances of our model for the source and target character respectively from a sparse set of 20 blendshapes in correspondence. We then fit our source model to the facial animation of the source target character to obtain per-frame transformations $[\mathbf{T_g}^j, \mathbf{T_b}^j]$ and shape coefficients $\mathbf{w}^j$. These coefficients can simply be played back on the target model to achieve facial performance retargeting. In Fig. 8, we provide a qualitative comparison to the state-of-the-art 3D retargeting algorithm of Chandran *et al*. [10] by retargeting the performance from a source to a target character. Our method provides competitive results to state of the art, while also allowing users to disentangle the rigid jaw motion and the soft tissue deformations of the skin surface. Our method additionally provides a substantial runtime benefit here and retargets each frame in a few (2-3) seconds, while the method of Chandran *et al*. requires

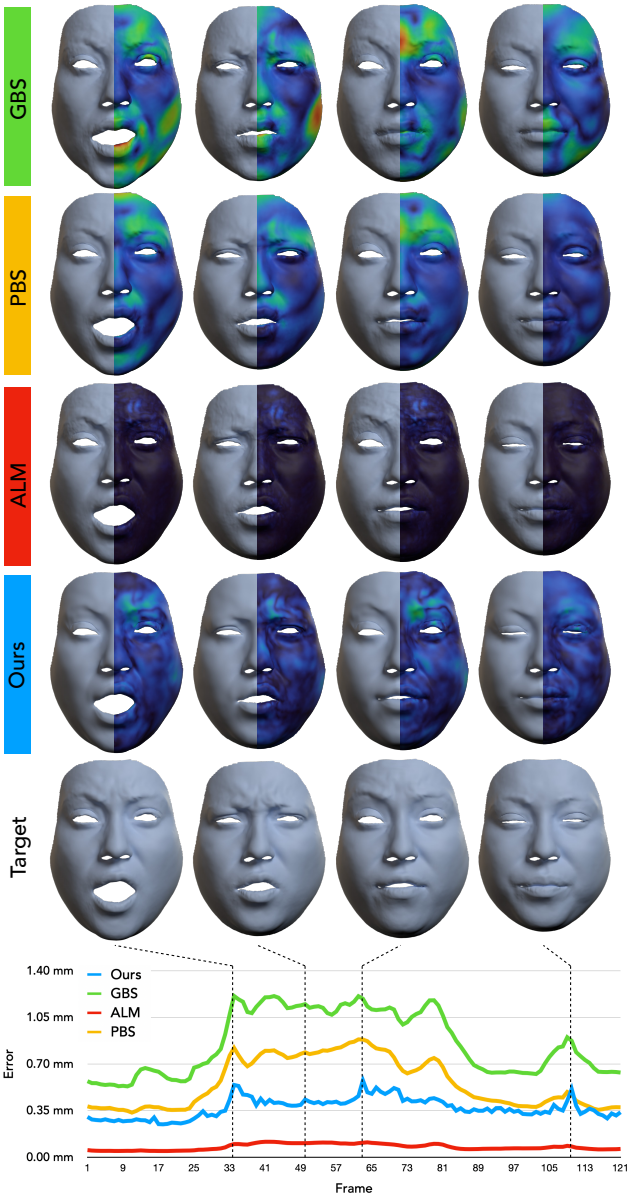

Figure 6. We show qualitative and quantitative comparisons of fitting 3D performances with various actor specific models. All the errors are displayed with a scale of 0mm 5mm.

several minutes per frame due to a costly anatomical solve. Finally unlike the approach of Chandran *et al*., our method provides all of above benefits without having to manually choose design parameters such as the patch layout, number of overlaps etc.

5.5. Limitations

Due to the sparse supervision on the facial anatomy, sometimes artifacts can appear on the learned anatomical surface especially in areas surrounding the lip region. Another limitation of our work is we current do not skin the facial

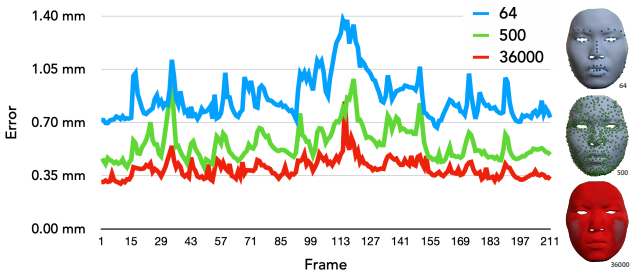

Figure 7. Our continuous anatomical face model can be fit to 3D scans with varying density of constraints and still provide valid results due to our fitting algorithm.: all the errors are displayed with a scale of 0mm 5mm.

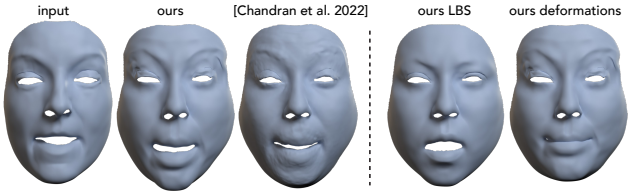

Figure 8. We show the result of facial performance transfer in 3D from an input actor (left) to a different actor as produced by our method (2nd column) and the local retargeting model of Chandran *et al*. [10]. While providing qualitatively similar results, our model implicitly disentangles the performance into rigid jaw motion (3rd column), and nonrigid soft tissue deformations (4th column).

anatomy to rigidly deform it along with facial expressions. Addressing these limitations through additional anatomical regularization or by predicting expression specific normals and thickness maps could be interesting future work. Some temporal jitter could also occur in our fitting step for challenging performances if the optimization is terminated too early. Finally extending our model to support facial appearance could be valuable future work.

6. Conclusion

In this paper we propose a new anatomically constrained implicit face model which provides a holistic representation of both facial anatomy and the enclosing skin surface using an ensemble of coordinate neural networks. Given an arbtrary set of skin surface meshes and only a neutral shape with estimated skull and jaw bones, our method recovers a dense anatomical substructure to constrain each point on the skin surface, and can model complex skin deformations with high fidelity. While we have explored the use of such a model in the context of actor specific blendshape models, future work could analyze it's implications as a generic morphable model, by extending our formulation to handle multiple identities at once. Our new *Anatomical Implicit face Model* (AIM) has applications in shape representation and manipulation, retargeting and more, and we hope that our method encourages exciting future research.

References

[1] Thabo Beeler and Derek Bradley. Rigid stabilization of facial expressions. *ACM TOG*, 33(4), 2014. 3, 4

[2] Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graphics Proc SIGGRAPH*, 30, 2011. 1, 7

[3] Sourav Biswas, Kangxue Yin, Maria Shugrina, Sanja Fidler, and Sameh Khamis. Hierarchical neural implicit pose network for animation and motion retargeting. *CoRR*, abs/2112.00958, 2021. 2

[4] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Siggraph*, 1999. 1, 2

[5] G. Bouritsas, S. Bokhnyak, S. Ploumpis, S. Zafeiriou, and M. Bronstein. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *ICCV*, 2019. 2

[6] Marcel C Bühler, Kripasindhu Sarkar, Tanmay Shah, Gengyan Li, Daoye Wang, Leonhard Helminger, Sergio Orts-Escolano, Dmitry Lagun, Otmar Hilliges, Thabo Beeler, et al. Preface: A data-driven volumetric prior for few-shot ultra high-resolution face synthesis. In *ICCV*, 2023. 2

[7] Chen Cao, Yanlin Weng, Shun Zhou, Y. Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20, 2014. 2

[8] Zenghao Chai, Haoxian Zhang, Jing Ren, Di Kang, Zhengzhuo Xu, Xuefei Zhe, Chun Yuan, and Linchao Bao. Realy: Rethinking the evaluation of 3d face reconstruction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 2

[9] Prashanth Chandran, Derek Bradley, Markus Gross, and Thabo Beeler. Semantic deep face models. In *TDV*, 2020. 2, 3

[10] Prashanth Chandran, Loïc Ciccone, Markus Gross, and Derek Bradley. Local anatomically-constrained facial performance retargeting. *ACM Trans. Graph.*, 41(4), 2022. 1, 2, 3, 7, 8

[11] Prashanth Chandran, Gaspard Zoss, Markus Gross, Paulo Gotardo, and Derek Bradley. Facial Animation with Disentangled Identity and Motion using Transformers. *Computer Graphics Forum*, 2022. 2

[12] Prashanth Chandran, Gaspard Zoss, Markus Gross, Paulo Gotardo, and Derek Bradley. Shape transformers: Topology-independent 3d shape models using transformers. 41(2), 2022. 2, 3

[13] P. Chandran, G. Zoss, P. Gotardo, and D. Bradley. Continuous landmark detection with 3d queries. In *CVPR*. IEEE Computer Society, 2023. 5, 6, 7, 1

[14] Zhixiang Chen and Tae-Kyun Kim. Learning feature aggregation for deep 3d morphable models. In *CVPR*, 2021. 2

[15] Byungkuk Choi, Haekwang Eom, Benjamin Mouscadet, Stephen Cullingford, Kurt Ma, Stefanie Gassel, Suzi Kim, Andrew Moffat, Millicent Maier, Marco Revelant, Joe Letteri, and Karan Singh. Animatomy: An animator-centric, anatomically inspired system for 3d facial modeling, animation and transfer. In *SIGGRAPH Asia 2022 Conference Papers*, 2022. 1, 2

[16] Boyang Deng, J. P. Lewis, Timothy Jeruzalski, Gerard Pons-Moll, Geoffrey Hinton, Mohammad Norouzi, and Andrea Tagliasacchi. Nasa neural articulated shape approximation. In *ECCV*, 2020. 2

[17] Bernhard Egger, William A. P. Smith, Ayush Tewari, Stefanie Wuhrer, Michael Zollhoefer, Thabo Beeler, Florian Bernard, Timo Bolkart, Adam Kortylewski, Sami Romdhani, Christian Theobalt, Volker Blanz, and Thomas Vetter. 3d morphable face models - past, present and future. *ACM TOG*, 39(5), 2020. 1, 2

[18] Paul Ekman and Wallace V. Friesen. Facial action coding system: a technique for the measurement of facial movement. 1978. 2

[19] Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. Learning an animatable detailed 3D face model from in-the-wild images. *TOG*, 40(4), 2021. 2

[20] Marco Fratarcangeli, Derek Bradley, Aurel Gruber, Gaspard Zoss, and Thabo Beeler. Fast Nonlinear Least Squares Optimization of Large-Scale Semi-Sparse Problems. *Computer Graphics Forum*, 2020. 2

[21] Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Learning neural parametric head models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3, 7

[22] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou. Spiralnet++: A fast and highly efficient mesh convolution operator. In *ICCV*, 2019. 2

[23] Chen Guo, Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges. Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[24] Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415, 2016. 6

[25] Stephan Hoyer, Jascha Sohl-Dickstein, and Sam Greydanus. Neural reparameterization improves structural optimization. *CoRR*, abs/1909.04240, 2019. 5

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 5

[27] J. P. Lewis, K. Anjyo, Taehyun Rhee, M. Zhang, Frédéric H. Pighin, and Z. Deng. Practice and theory of blendshape facial models. In *Computer Graphics Forum (Proc. Eurographics*, 2014. 1, 2, 7

[28] Ruilong Li, Karl Bladin, Yajie Zhao, Chinmay Chinara, Owen Ingraham, Pengda Xiang, Xinglei Ren, Pratusha Prasad, Bipin Kishore, Jun Xing, and Hao Li. Learning formation of physically-based face attributes. *CoRR*, abs/2004.03458, 2020. 2

[29] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM*, 2017. 1, 2, 4, 7, 3, 5

[30] Julien N. P. Martel, David B. Lindell, Connor Z. Lin, Eric R. Chan, Marco Monteiro, and Gordon Wetzstein. ACORN: adaptive coordinate networks for neural scene representation. *CoRR*, abs/2105.02788, 2021. 2

[31] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2

[32] Thomas Müller. tiny-cuda-nn, 2021. 6, 2

[33] Verónica Orvalho, Pedro Bastos, Frederic Parke, Bruno Oliveira, and Xenxo Alvarez. A Facial Rigging Survey. In *Eurographics 2012 - State of the Art Reports*. The Eurographics Association, 2012. 1

[34] Pablo R. Palafox, Aljaz Bozic, Justus Thies, Matthias Nießner, and Angela Dai. Npms: Neural parametric models for 3d deformable shapes. abs/2104.00702, 2021. 2

[35] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 2

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*. 2019. 6

[37] Zesong Qiu, Yuwei Li, Dongming He, Qixuan Zhang, Longwen Zhang, Yinghao Zhang, Jingya Wang, Lan Xu, Xudong Wang, Yuyao Zhang, and Jingyi Yu. Sculptor: Skeleton-consistent face creation using a learned parametric generator. *ACM Trans. Graph.*, 41(6), 2022. 2, 4

[38] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3d faces using convolutional mesh autoencoders. In *ECCV*, 2018. 2

[39] Eftychios Sifakis, Andrew Selle, Avram Robinson-Mosher, and Ronald Fedkiw. Simulating speech with a physics-based facial muscle model. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 2006. 2, 3

[40] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Proc. NeurIPS*, 2020. 2, 4

[41] Sangeetha Grama Srinivasan, Qisi Wang, Junior Rojas, Gergely Klár, Ladislav Kavan, and Eftychios Sifakis. Learning active quasistatic physics-based models from data. *ACM Trans. Graph.*, 40(4), 2021. 3

[42] J. Rafael Tena, Fernando De la Torre, and Iain Matthews. Interactive region-based linear 3d face models. *ACM Trans. Graphics Proc SIGGRAPH*, 30(4), 2011. 1

[43] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. Face transfer with multilinear models. *ACM TOG*, 24(3), 2005. 2

[44] Nicolas Wagner, Mario Botsch, and Ulrich Schwanecke. Softdeca: Computationally efficient physics-based facial animations. In *Proceedings of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games*, 2023. 1, 3

[45] Daoye Wang, Prashanth Chandran, Gaspard Zoss, Derek Bradley, and Paulo Gotardo. Morf: Morphable radiance fields for multiview neural head modeling. In *ACM SIGGRAPH 2022 Conference Proceedings*, 2022. 2, 3

[46] Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Matthew Johnson, Jingjing Shen, Nikola Milosavljevic, Daniel Wilde, Stephan Garbin, Chirag Raman, Jamie Shotton, Toby Sharp, Ivan Stojiljkovic, Tom Cashman, and Julien Valentin. 3d face reconstruction with dense landmarks, 2022. 5, 6

[47] Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. An anatomically-constrained local deformation model for monocular face capture. *ACM TOG*, 35(4), 2016. 1, 2, 3, 4, 5, 7

[48] Lingchen Yang, Byungsoo Kim, Gaspard Zoss, Baran Gözcü, Markus Gross, and Barbara Solenthaler. Implicit neural representation for physics-driven actuated soft bodies. *ACM Trans. Graph.*, 41(4), 2022. 1, 2, 3, 4

[49] Tarun Yenamandra, Ayush Tewari, Florian Bernard, Hans-Peter Seidel, Mohamed Elgharib, Daniel Cremers, and Christian Theobalt. i3dmm: Deep implicit 3d morphable model of human heads. In *CVPR*, 2021. 2

[50] Kim Youwang, Lee Hyun, Kim Sung-Bin, Suekyeong Nam, Janghoon Ju, and Tae-Hyun Oh. A large-scale 3d face mesh video dataset via neural re-parameterized optimization. *arXiv*, 2023. 5, 6

[51] Mingwu Zheng, Hongyu Yang, Di Huang, and Liming Chen. Imface: A nonlinear 3d morphable face model with implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 3

[52] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C. Bühler, Xu Chen, Michael J. Black, and Otmar Hilliges. Im avatar: Implicit morphable head avatars from videos. In *Computer Vision and Pattern Recognition (CVPR)*, 2022. 2, 3, 5

[53] Yufeng Zheng, Wang Yifan, Gordon Wetzstein, Michael J. Black, and Otmar Hilliges. Pointavatar: Deformable point-based head avatars from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[54] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[55] Yi Zhou, Chenglei Wu, Zimo Li, Chen Cao, Yuting Ye, Jason Saragih, Hao Li, and Yaser Sheikh. Fully convolutional mesh autoencoder using efficient spatially varying kernels. In *NeurIPS*, 2020. 2

[56] Gaspard Zoss, Derek Bradley, Pascal Bérard, and Thabo Beeler. An empirical rig for jaw animation. *ACM TOG*, 37 (4), 2018. 4, 5, 1

[57] Gaspard Zoss, Thabo Beeler, Markus Gross, and Derek Bradley. Accurate markerless jaw tracking for facial performance capture. *ACM TOG*, 38(4), 2019. 4