

LORAMI: Low-Rank Adaptation for Multi-Identity Physics-Based Face Rigs

D. Corigliano¹ , D. Peter^{1,2} , B. Solenthaler¹ , and B. Thomaszewski¹ 

¹ETH Zurich, Switzerland
²Align Technology, Switzerland

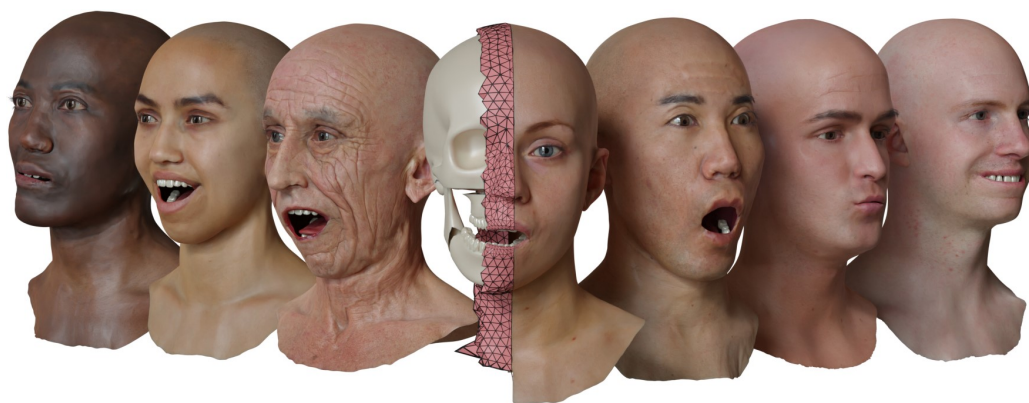


Figure 1: LORAMI produces physically-based facial deformations across a continuous space of identities. By combining rig controls and identity encodings, our method generalizes across identity variations without the need for retraining or fine-tuning.

Abstract

Physics-based simulation can augment facial rigs with high-quality deformations, but at steep computational costs. While neural surrogates can substantially reduce computation time, existing methods do not generalize across identity variations and instead require expensive per-case retraining. In this work, we present LORAMI—a method for learning physics-based face rigs over a continuous space of identities using low-rank adaptation. LORAMI addresses the challenge of identity variation through a novel architecture that combines a shared neural surrogate with low-rank weight adaptations. Instead of training a dense model across identity space, we modulate a shared base network using low-rank factors. These factors are scaled by diagonal matrices predicted from identity parameters. This design enables efficient modeling of identity-dependent variations while preserving the generic deformation behavior of the underlying physics-based rig. Our experiments show that LORAMI achieves deformation accuracy on par with single-identity models and outperforms fully dense identity-conditioned networks. As a result, our method enables real-time physics-based facial animation with continuous identity control.

CCS Concepts

• **Computing methodologies** → Physical simulation; Neural networks;

1. Introduction

Digital humans—virtual doubles of real-world individuals with realistic appearance and behavior—are widely used in film production, game technology, and medical applications. Nevertheless, achieving realistic soft-tissue deformation during speech, facial expressions, and other dynamic motion remains a central challenge. Physics-based simulation can greatly enhance realism by augment-

ing static facial geometry with deformations that capture interactions between skin, muscle, and skeletal structures. When integrated into facial rigs, such physics-based models enable expressive, anatomically grounded motion that is difficult to achieve with purely kinematic or blendshape-based approaches.

Despite their visual fidelity, physics-based facial simulations come at a steep computational cost. Accurately modeling elastic

deformation requires iterative numerical solvers, which typically preclude real-time applications. This tension between physical accuracy and practical usability has motivated a growing body of work on learning-based surrogates for facial deformation.

Recent advances in learning-based deformation models have shown that neural surrogates can approximate physics-based facial simulations at a fraction of the computational cost of a conventional simulator [CPH*25]. By learning a direct mapping from rig controls to surface deformation, these methods enable real-time inference while retaining much of the visual fidelity of the underlying simulation. However, existing physics-based neural approaches are largely restricted to a single identity and do not generalize across variations in facial geometry or anatomy. Extending such methods to new identities typically requires costly retraining, precluding their employment in applications that require instant generalization to unseen characters or continuous identity variations.

Generalization across identities presents a unique challenge for physics-based facial rigs. Identity variation affects not only the shape of the outside surface (i.e., the skin) but also the volumetric tissue layer and bone structure underneath. As a result, the mechanical response of the human face to applied loads and constraints can vary substantially between individuals. Naively training a single neural model across multiple identities often leads to either underfitting, where character-specific details are lost, or to prohibitively large dense models with poor time and data efficiency during training. Furthermore, unconstrained interpolation or extrapolation across identity space can quickly produce implausible deformations, undermining the physical credibility that motivates simulation-based approaches in the first place. Addressing identity variation thus requires models that can share structure across identities while retaining the ability to adapt to character-specific deformation behavior.

In this work, we present LORAMI, a method for learning physics-based face rigs over a continuous space of identities using low-rank adaptations. LORAMI extends learning-based surrogate models beyond single characters by building identity-dependent adaptation directly into the network architecture. Instead of training a generic dense model across identity space, we adopt an identity-conditioned design that combines a shared neural surrogate with low-rank weight adaptations. The shared base network captures deformation behavior common to all identities, while compact, identity-specific adaptations modulate this behavior to account for individual anatomical differences. We represent identity-dependent updates using shared low-rank factors whose influence is scaled by a set of learned, subject-specific coefficients predicted from identity parameters. By conditioning the network in this manner, LORAMI efficiently captures identity-dependent variation while preserving the inductive bias and stability of the underlying physics-based rig. Compared to other modulation strategies, this design has the advantage of significantly reducing the number of network weights that must be learned to model identity variation. Furthermore, the separation between shared deformation behavior and identity-specific modulation promotes generalization across identities, enabling smooth interpolation within a continuous identity space.

We train and evaluate LORAMI on a dataset of facial mod-

els constructed from topologically consistent skin, tissue, and bone meshes. Our learned multi-identity rig closely matches ground-truth physics-based simulations and achieves accuracy comparable to single-identity surrogate models. In particular, our experiments demonstrate that LORAMI achieves substantially lower error than fully dense alternatives and other conditioning strategies. In summary, our work makes the following contributions:

- A multi-identity learning framework for physics-based facial rigs, extending real-time neural surrogates beyond single individuals.
- An identity-conditioned low-rank adaptation architecture that efficiently models identity-dependent deformation while sharing structure across individuals.
- A practical method for continuous identity control that preserves the physical plausibility and deformation behavior of simulation-based rigs.
- Experimental evidence showing improved training efficiency and high accuracy relative to both dense multi-identity models and single-identity baselines.

Together, these contributions enable scalable, real-time physics-based facial animation across a continuous space of identities, bridging the gap between high-quality simulation and practical multi-character deployment.

2. Related Work

2.1. Facial Animation

Blendshape-based and Data-driven Animation Facial animation has been a central topic in computer graphics for decades [EST*20]. A widely adopted paradigm is based on blendshape rigs [LA10,LCF00], where facial expressions are constructed by combining a set of artist-designed deformation shapes. This representation enables intuitive control and fast evaluation, which has made it the standard choice in many practical systems. However, since the deformation space is defined a priori by a fixed set of shapes, the resulting motion is inherently restricted in its ability to reproduce complex, nonlinear behavior. To address these limitations, more recent approaches augment blendshape models with learned corrective deformations and anatomically inspired components [CEM*22], improving expressiveness while retaining their efficiency. In the context of speech animation, data-driven methods have been proposed to learn deformation behavior from images, text, or video [CBL*19,RZW*21,MSS*21,PLX*25,WSGW25].

Physics-Based Animation Physically-based approaches model facial deformation by simulating the underlying biomechanical behavior of soft tissue, typically using numerical solvers such as Newton methods or projective dynamics [BML*14,IKKP17]. By incorporating elastic material models [ZBLJ20] and anatomical structure [MSE*11], these methods can reproduce complex deformation effects beyond the capabilities of purely kinematic techniques. However, these methods require minimizing the mechanical energy for the physical system at runtime, which leads to high computational cost and limits their applicability in practical settings.

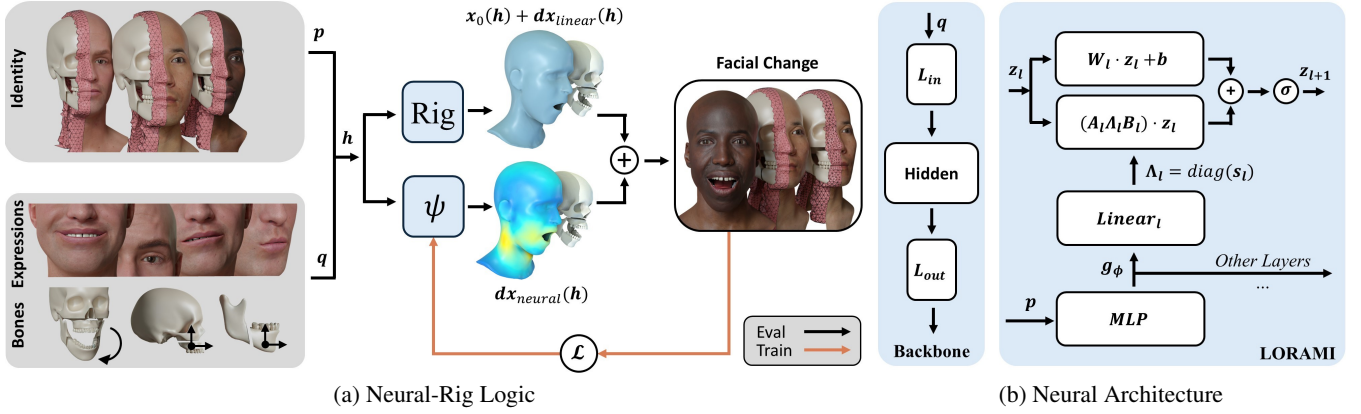


Figure 2: (a) LORAMI integrates identity encodings $\mathbf{p}$ and rig controls $\mathbf{q}$ into a unified architecture, where identity codes define anatomical head models and rig controls define facial deformations. For any pair $\mathbf{h} = (\mathbf{q}, \mathbf{p})$, a linear rig computes bone and teeth positions alongside an initial soft-tissue deformation $\mathbf{x}_0(\mathbf{h}) + \mathbf{dx}_{linear}(\mathbf{h})$. A neural network, parameterized by ψ , then predicts a physically-based correction $\mathbf{dx}_{neural}(\mathbf{h})$ to reach mechanical equilibrium. Our self-supervised training strategy minimizes the system's total mechanical energy. Details on the rig logic are provided in Section 4. (b) LORAMI extends the original NeuRiPhy network (backbone, left) by incorporating identity modulation via low-rank adaptations (right). For each hidden layer, the linear parameters $(\mathbf{W}_l, \mathbf{b})$ and the adaptation matrices $(\mathbf{A}_l, \mathbf{B}_l, \Lambda_l)$ are trained jointly. Here, $(\mathbf{z}_l, \mathbf{z}_{l+1})$ denote the input and output of each hidden layer. See Section 6.2 for more architectural details.

Learning-Based and Neural Simulation Methods Recent work has explored the use of neural networks to approximate the behavior of physically-based systems, providing efficient alternatives to traditional numerical solvers. These approaches aim to learn mappings to equilibrium configurations or deformation states, either from simulation data or by incorporating physical principles into the training process [FMD*19, HDDN19, RCPO21, SYS*21, SRJ*23, WDCT24]. In the context of facial animation, data-driven methods have been proposed to learn deformation behavior from physically-based simulations, captured performances, or artist-crafted animations, enabling fast evaluation at runtime [BODO18, BODO20, RdGM20, MSS*21, SWR*21, QSA*23].

Among existing approaches, NeuRiPhy [CPH*25] is particularly relevant, as it learns a neural surrogate for physics-based facial rigs by mapping control parameters directly to soft-tissue deformations. By incorporating physical energy terms into the training objective, the method produces deformations that are consistent with an underlying biomechanical model while achieving real-time performance and an intuitive control through a set of rig parameters. However, such approaches are trained for a single identity, and extending them to new subjects requires retraining, limiting their applicability in settings with varying anatomy.

Recent work has started to address this limitation by incorporating identity variation into neural deformation models. [YZC*24] learns a data-driven model of a generalized physical face that can be adapted to new identities through a learned latent representation. Once fitted to a target character, the method produces animation-ready, physics-based deformations and supports retargeting across identities. While this enables generalization across characters, it relies on supervised training with precomputed data and necessitates both per-identity fitting and FEM simulation at evaluation time.

2.2. Network Finetuning and Low-Rank Adaptation

Parameter-efficient adaptation methods aim to modify a pretrained neural model using a small number of additional parameters while keeping the base network fixed [DQY*23]. A prominent approach is Low-Rank Adaptation (LoRA), which represents weight updates as the product of two low-rank matrices [HSW*22]. While originally proposed in the context of large language models, this formulation has since been applied across a wide range of domains, including image and video diffusion models [Ryu23, YMD*25, LTJ24, ZCYZ25] and geometric foundation models [LYX*24].

The standard principle of LoRA is to take a pretrained model and specialize it to a target domain through low-rank weight updates. However, the low-rank formulation itself does not rely on the presence of large pretrained models and can also be applied to instance-specific adaptation. For example, [TMKLS25] use low-rank updates to efficiently modify trained neural fields, enabling structured perturbations of geometric representations. Similarly, [WBZ22] apply low-rank updates to generative models in order to match edited outputs. These approaches demonstrate that low-rank parameterizations provide a flexible mechanism for encoding structured variations in neural networks.

In facial animation, models must capture identity-dependent variation while preserving shared control semantics across subjects. This motivates representations that can encode structured variations efficiently. Low-rank parameterizations provide a natural mechanism for this, as they enable compact updates to a shared model while preserving its overall structure.

3. Face Model and Identity Representation

3.1. Anatomical Model and Rig Controls

Our simulation-ready head models follow the anatomical rig formulation of [CPH*25]. Each identity is represented using a multi-

layer anatomical model compatible with physics-based simulation, including soft tissue, skull, mandible, and teeth.

Facial motion is controlled by a set of rig parameters $\mathbf{q} \in \mathbb{R}^m$, with $m = 14$, which induce transformations of the anatomical components while respecting biomechanical limits. The controls are structured into distinct groups corresponding to different modes of actuation. In particular, we consider mandible controls $\mathbf{q}_m \in \mathbb{R}^4$, representing jaw articulation via three rotations and one translation, and orthognathic controls $\mathbf{q}_o \in \mathbb{R}^3$, which model clinically inspired bone displacements within anatomically feasible bounds. These skeletal transformations define boundary conditions for the soft tissue, which deforms elastically in response.

Realistic facial expressions are incorporated through a dual-representation approach that combines surface-level blendshapes $\{\mathcal{B}_i\}$ with soft-tissue expression operators $\{\mathbf{F}_{E_i}\}$. The expression parameters $\mathbf{q}_e \in \mathbb{R}^7$ control these operators, including six expression controls and an additional parameter representing the rest configuration, which acts as the neutral reference state of the model. Together, the full control vector $\mathbf{q} = (\mathbf{q}_m, \mathbf{q}_o, \mathbf{q}_e)$ defines the actuation of the anatomical model and provides a consistent, semantically meaningful control space across identities.

Compared to [CPH*25], we increase the number of expression controls from 3 to 7 and provide greater flexibility by enabling mandible and surgical controls within the same rig. We refer to the supplementary material for details on the expression operators, the anatomical model, and its components, such as teeth, soft-tissue, and rendering layers.

3.2. Identity Representation

Each identity is represented by a collection of anatomically consistent components, including soft tissue, skull, mandible, and teeth. Specifically, we define an identity vector

$$\phi = [\mathcal{ST}_{Rest}, \mathcal{ST}_{Neutral}, \mathcal{S}, \mathcal{M}, \mathcal{T}_{upper}, \mathcal{T}_{lower}], \quad (1)$$

which concatenates the linearized meshes of the corresponding anatomical structures. Here, $\mathcal{S}$ and $\mathcal{M}$ represent the skull and mandible, $\mathcal{T}_{upper}$ and $\mathcal{T}_{lower}$ correspond to the upper and lower teeth geometries, and $\mathcal{ST}_{Rest}$ and $\mathcal{ST}_{Neutral}$ share the same topology and denote two configurations of the soft-tissue meshes. $\mathcal{ST}_{Rest}$ defines the rest configuration used in simulation, where slightly intersecting lips ensure realistic rest stresses. Conversely, $\mathcal{ST}_{Neutral}$ represents the unactuated pose with touching lips, used for skinning and as the simulation's starting point.

We construct a dataset of anatomical head models using scans and fitting techniques. A shared soft tissue template is then registered to these models using rigid and non-rigid alignments. To model identity variation, we construct a low-dimensional statistical shape model by performing principal component analysis (PCA) on a set of identity vectors $\{\phi_i\}$. Each identity is then encoded by a coefficient vector $\mathbf{p} \in \mathbb{R}^P$, where P denotes the number of retained principal components. The coefficients are obtained by projecting ϕ onto the learned subspace. Given a coefficient vector $\mathbf{p}$, the full anatomical configuration can be reconstructed via the inverse PCA transform.

This representation captures the dominant modes of anatomical

variation across subjects while preserving structural consistency required for simulation. The resulting identity space is continuous and enables interpolation between subjects, allowing the generation of novel identities within the span of the training data. During training, identity parameters are sampled within this space to expose the model to a diverse range of anatomies.

We refer to the supplementary material for additional details on the anatomical model, dataset construction, and statistical model.

3.3. Physics-based Face Model

Our method builds on the physics-based facial rig introduced by NeuRiPhy [CPH*25], which computes facial deformations as equilibrium states of a biomechanical system. A set of rig controls $\mathbf{q}$ induce boundary conditions on the anatomy, and the resulting soft-tissue deformation is obtained by minimizing the total potential energy. We extend this formulation by introducing identity encodings $\mathbf{p}$, granting control over the geometry of the anatomical structures.

Our simulation-ready head models closely follow the definition of NeuRiPhy. The soft tissue is represented as a tetrahedral finite element mesh, and its deformation is governed by elastic forces derived from a nonlinear material model. The total energy combines several physically motivated components. The elastic energy models the stress response of the soft tissue to strain, while attachment constraints enforce coupling between soft tissue and rigid structures such as the skull and mandible. Our attachment formulation differs from the model in NeuRiPhy, as we opt for hard constraints instead of soft penalties. Contact interactions account for collisions between anatomical components, including lip–lip and lip–teeth interactions, and prevent interpenetration during articulation.

Facial expressions are modeled through expression operators that modify the rest configuration of the soft tissue via deformation gradients, following [CPH*25]. These operators are driven by the expression controls $\mathbf{q}_e$ and act as kinematic constraints within the physical system, enabling the synthesis of complex facial deformations while maintaining physical consistency. In our framework, these operators are adapted to each identity to ensure consistent semantic behavior across varying anatomies, following [YZC*24], as detailed in the supplementary material.

The equilibrium configuration $\mathbf{x}$ of the soft tissue is obtained as the solution to the minimization problem

$$\mathbf{x}(\mathbf{h}) = \arg \min_{\mathbf{x}} E(\mathbf{x}, \mathbf{h}). \quad (2)$$

The above expression defines a mapping $\mathbf{h} \mapsto \mathbf{x}(\mathbf{h})$, where $\mathbf{h} = (\mathbf{q}, \mathbf{p})$, to equilibrium facial configurations, and $E = E_{\text{elastic}} + E_{\text{contact}}$ is the total mechanical energy. The identity parameters $\mathbf{p}$ determine the geometry of the model and therefore influence the resulting deformation under fixed material parameters. We refer to [CPH*25] and the supplementary material for details on the individual energy terms.

4. Two-Layer Rig Representation

We approximate the mapping $\mathbf{h} \mapsto \mathbf{x}(\mathbf{h})$ defined by the physics-based model using a structured rig composed of a linear and a neutral component.

Linear Rig Our multi-identity rig consists of two components. The linear rig provides an initial estimate of facial deformation by sequentially applying bone transformations via linear blend skinning, followed by expression blendshapes,

$$\mathbf{x}_{\text{linear}}(\mathbf{h}) = \mathbf{x}_0(\mathbf{h}) + \mathbf{dx}_{\text{linear}}(\mathbf{h}), \quad (3)$$

$$\mathbf{dx}_{\text{linear}}(\mathbf{h}) = \mathcal{B}(\mathbf{h}) + \mathcal{LBS}(\mathbf{h}), \quad (4)$$

where $\mathbf{x}_0(\mathbf{h})$ is the neutral configuration of the soft tissue for the selected identity. The blendshapes $\mathcal{B}(\mathbf{h})$ vary across identities. To adapt the scan-derived base expression blendshapes to unseen identities, we compute template-to-target deformation gradients and use them to deform these shapes. The skinning function $\mathcal{LBS}(\mathbf{h})$ is similarly personalized using the target identity's neutral configuration. We further detail $\mathcal{B}(\mathbf{h})$ in the supplemental material.

Neural Rig The neural layer adds a per-vertex correction that ensures physical equilibrium. The result is a neural architecture that combines linear rig and neural surrogate

$$\mathcal{N}_{\psi}(\mathbf{h}) = \mathbf{x}_{\text{linear}}(\mathbf{h}) + \mathbf{dx}_{\text{neural}}(\mathbf{h}). \quad (5)$$

5. Training Strategy

Our method aims to achieve real-time performance for both identity sampling and deformation evaluation. To this end, we learn a neural approximation of the mapping from identity and control parameters to equilibrium configurations.

5.1. Unsupervised Learning

We learn the mapping $\mathbf{h} \mapsto \mathbf{x}(\mathbf{h})$ in a self-supervised fashion. Following the approach of [CPH*25], we incorporate the physical energy directly into the loss function.

To ensure coverage of the identity and control space, we sample new values of identity and control parameters at each training epoch. Given M sampled inputs $\{\mathbf{h}_i\}_{i=1}^M$, the network parameters ψ are optimized as

$$\hat{\psi} = \arg \min_{\psi} \frac{1}{M} \sum_{i=1}^M E(\mathcal{N}_{\psi}(\mathbf{h}_i), \mathbf{h}_i), \quad (6)$$

where $\mathcal{N}_{\psi}(\mathbf{h}_i)$ denotes the predicted soft tissue configuration.

5.2. Controls and Identity Sampling

We extend the training strategy of the baseline NeuRiPhy framework [CPH*25]. While it traditionally processes only rig controls $\mathbf{q}$, we augment it to incorporate identity encodings $\mathbf{p}$.

Rig Controls During training, the neural simulator is given an arbitrary configuration of the boundary conditions, and if the teeth penetrate the soft tissue, the training process might not be able to recover. To address this problem, the baseline implementation [CPH*25] gradually expands the control values from an initial seed $\mathbf{q}_0$ during the training procedure:

$$\mathbf{q}(e, D) = \mathbf{q}_0 + g(e, D) \cdot (\mathbf{q} - \mathbf{q}_0), \quad (7)$$

where $g \in \mathbb{N}^2 \rightarrow [0, 1]$ maps the current epoch e to a progressive control scale, $\mathbf{q}_0$ is the seed, and D is the time to reach 1. The initial seed state $\mathbf{q}_0$ is assumed to contain no intersections, allowing the network to learn increasingly difficult contact interactions. Rig controls are sampled uniformly in the selected ranges. To boost accuracy on individual rig controls, we sample individual directions progressively with probability p , setting all other controls to zero.

Identity Identity sampling poses an additional challenge, as the PCA latent space is unbounded: naively drawing from a Gaussian $\mathbf{p} \sim \mathcal{N}(0, \{\lambda_i\})$, using the PCA standard deviations λ_i , makes the network encounter anatomically implausible extrapolations far from any real training subject. Those identities can greatly destabilize training, causing large spikes in the gradient and loss profiles. To avoid this issue, we restrict sampling to convex combinations of identities within our dataset. We first define the convex hull encompassing the training identity codes. For each training identity, we solve a linear program to determine if the identity can be expressed as a convex combination of the remaining identities. If the linear program is infeasible, the identity is a vertex of the convex hull. During training, we sample identity codes as convex combinations of the convex hull vertices. Mixture weights $\gamma \sim \text{Dir}(\alpha)$ are drawn from a symmetric Dirichlet distribution, and the resulting identity code is computed as

$$\mathbf{p} = \gamma^{\top} \mathbf{P}, \quad (8)$$

where $\mathbf{P} \in \mathbb{R}^{N \times P}$ represents identity codes of the training set. By construction, every sampled identity is a convex combination of the input anatomies, with the concentration parameter α governing the sparsity of the mixture. During training, we also extend the control expansion to our identity codes, learning deformations for increasingly diverse identities.

6. Identity-Conditioned Neural Architecture

Identity information is incorporated into the neural rig. Rather than committing to a single formulation, we consider a design space of identity-conditioned architectures and derive our final model from this perspective.

6.1. Design Space for Identity Adaptation

A central question is how identity should influence the neural architecture $\mathcal{N}_{\psi}$. We consider three complementary strategies, corresponding to different points in the design space.

Input-Level Conditioning (Dense) The simplest multi-identity baseline is a plain MLP that receives both the rig controls and the identity code as a single concatenated input vector. The network then follows the same architecture as the single-identity NeuRiPhy [CPH*25] with GELU activations. No explicit modulation mechanism is introduced; the network must learn implicitly, from data alone, how the identity code interacts with each rig control to produce the correct deformation. This formulation incorporates identity at the input level, requiring the network to jointly model control and identity effects in an unstructured manner.

Feature Modulation (Conditioning) Rather than relying on implicit interactions, identity can be incorporated as a continuous conditioning signal (in the style of [YZC*24]). An identity PCA code $\mathbf{p}$ is first projected by a shared conditioning MLP to a feature vector $g_\phi(\mathbf{p}) \in \mathbb{R}^{d_c}$. At each hidden layer ℓ , $g_\phi(\mathbf{p})$ is fed into an activated linear layer to produce the per-layer scaling vector $\mathbf{s}_\ell(g_\phi(\mathbf{p})) \in \mathbb{R}^{d_h}$ that modulates the layer's weight matrix after activation,

$$\mathbf{y}_\ell = \mathbf{s}_\ell(g_\phi(\mathbf{p})) \odot \sigma(\mathbf{W}_\ell \mathbf{y}_{\ell-1} + \mathbf{b}_\ell), \quad (9)$$

where $\mathbf{W}_\ell$ and $\mathbf{b}_\ell$ are the weight matrix and bias vector for the layer ℓ , $\odot$ is the element-wise product of two vectors, and σ is the GELU activation function. A single network is thus shared across all training subjects, operating on novel identity codes without retraining. Compared to dense conditioning, this formulation introduces structure by modulating intermediate features, but identity still affects the network indirectly through activation scaling.

Weight-Space Adaptation (Low-Rank) A more direct approach is to incorporate identity into the weights of the network. Motivated by the observation that deformation behavior is largely shared across identities, while identity variation is low-dimensional, we consider low-rank adaptations of the network weights. We learn, for each hidden layer ℓ , two low-rank matrices $\mathbf{A}_\ell \in \mathbb{R}^{r \times d_h}$ and $\mathbf{B}_\ell \in \mathbb{R}^{d_h \times r}$, shared across all identities. The identity code $\mathbf{p}$ is encoded by a shared MLP to produce a latent vector $g_\phi(\mathbf{p})$.

A general formulation introduces identity-dependent scaling of both low-rank factors (VeRA-style adaptation [KBA24]). For each hidden layer ℓ , we define two diagonal scaling matrices $\mathbf{s}_\ell^A = \mathbf{s}_\ell^A(g_\phi(\mathbf{p})) \in \mathbb{R}^r$ and $\mathbf{s}_\ell^B = \mathbf{s}_\ell^B(g_\phi(\mathbf{p})) \in \mathbb{R}^{d_h}$, where $(\mathbf{s}_\ell^A, \mathbf{s}_\ell^B)$ results from two separate single-layer MLPs. The weight matrix for the hidden layer becomes

$$\hat{\mathbf{W}}_\ell = \mathbf{W}_\ell + \text{diag}(\mathbf{s}_\ell^B) \cdot \mathbf{B}_\ell \cdot \text{diag}(\mathbf{s}_\ell^A) \cdot \mathbf{A}_\ell. \quad (10)$$

This formulation allows identity to directly modulate the network in weight space while maintaining a shared low-rank structure across identities.

6.2. LORAMI

Within the above design space, we seek a formulation that preserves the benefits of weight-space adaptation while minimizing the number of identity-dependent parameters. We observe that scaling both low-rank factors is unnecessary. Instead, identity variation can be captured by scaling a shared low-rank subspace with a single diagonal matrix. We therefore define the effective weight at inference as

$$\hat{\mathbf{W}}_\ell = \mathbf{W}_\ell + \mathbf{B}_\ell \cdot \text{diag}(\mathbf{s}_\ell) \cdot \mathbf{A}_\ell. \quad (11)$$

Here, $\mathbf{A}_\ell$ and $\mathbf{B}_\ell$ are shared across all identities while $\mathbf{s}_\ell = \mathbf{s}_\ell(g_\phi(\mathbf{p})) \in \mathbb{R}^r$ is an identity-dependent scaling vector. In this formulation, identity influences the network by scaling each of the r low-rank directions, rather than constructing the full weight update. For both VeRA-style adaptation and LORAMI, we jointly optimize the base network weights, adaptation matrices, and diagonal scaling parameters, rather than training adaptation components on top of frozen pretrained weights or fixed low-rank adaptations. This design provides a compact and structured representation of identity-dependent deformation behavior, forming the core of our method.

6.3. Single-Identity and Transfer Baselines

For completeness, we also consider single-identity models and transfer-based approaches. These approaches provide strong per-identity accuracy, but require per-subject optimization and do not support continuous identity control.

Single-Identity Network (NeuRiPhy) The baseline architecture is NeuRiPhy, a fully-connected MLP with $L = 4$ hidden layers of width $d_h = 1024$ and GELU activations. It maps a rig-control vector $\mathbf{z}_c \in \mathbb{R}^{n_c}$ directly to a soft-tissue displacement field $d\mathbf{x} \in \mathbb{R}^{n_v}$, where $n_v = 3|\mathcal{V}|$ and $\mathcal{V}$ is the set of tetrahedral mesh vertices. Following [CPH*25], a frozen *zero-trick* bias ensures that the network's predicted deformation vanishes given zero rig controls and the mean identity, thereby yielding an intersection-free configuration at the start of training. This is implemented either by setting the backbone network biases to zero or by applying a subtractive bias $\tau = f(\mathbf{0})$. Because no identity information enters the network, a separate instance must be trained from scratch for every new subject.

LoRA Identity Transfer. To adapt a pre-trained source network to a new target identity without full retraining, we wrap each hidden linear layer with a LoRA [HSW*22] module. Each hidden layer's effective weight matrix becomes $\hat{\mathbf{W}}_\ell = \mathbf{W}_\ell + \mathbf{B}_\ell \mathbf{A}_\ell$, where $\mathbf{A}_\ell \in \mathbb{R}^{r \times d_h}$ and $\mathbf{B}_\ell \in \mathbb{R}^{d_h \times r}$ are the low-rank factors with rank r . The backbone weights $\mathbf{W}$ are fixed and only $\mathbf{A}_\ell$ and $\mathbf{B}_\ell$ receive gradients. This yields a parameter-efficient transfer from source to target with only $2Lrd_h$ trainable parameters (excluding biases), roughly two orders of magnitude fewer than the full backbone.

Full Finetuning As an alternative, all hidden weights $\mathbf{W}$ and biases $\mathbf{b}$ are fine-tuned with standard gradient descent, giving full expressive capacity (Ld_h^2 trainable parameters, excluding biases).

7. Implementation

Network and Training We provide a detailed description of our architectures, in terms of layer size and number of parameters, in Table 1. Additionally, we provide a visual representation of the architectures in the supplementary material. To ensure fair comparisons in our experiments, all the architectures use the same backbone with 6 hidden layers of size 1024 and GeLU activations. The neural network receives a rig-control vector of size $n_q = 10$ as input. Each network is trained using the Adam optimizer using learning rates in the range $[1, 5] \times 10^{-5}$. For the identity and control range expansion during training, we set the growth duration to 5000 epochs and use an increasing cosine scheduler with values in $[0, 1]$. Our PCA model for identity is trained on a dataset of 93 anatomical head models. We furthermore use an isotropic elastic material with Poisson ratio $\nu = 0.4$ and Young modulus $E = 5000$ along with a penalty weight of $w_c = 10^5$ for contact. Eyebrow controls are implemented only as blendshapes, as they do not affect areas where elastic deformations and contacts happen. The rest configuration is also excluded from the neural control vector, as it represents the default state in the absence of other expressions.

Interactive Rig Application Our rig can be queried using an interactive application. To enhance usability, related rig controls are

Table 1: We report parameter counts for each architecture using the following dimensions: neural rig-control input size $n_q = 10$, hidden width $d_h = 1024$, output size (nodal displacements) $n_v = 24,108$, and identity-code dimension $n_p = 40$. The network consists of $L = 6$ hidden layers. Where modulation is employed, the relevant layers have size $d_c = 128$ and low-rank adaptations have order $r = 12$. For clarity, bias terms are omitted from the expressions. The shared output layer, which accounts for approximately 24.7M parameters ($d_h n_v$), is excluded from the count of trainable parameters. The ID-logic metric specifically denotes parameters used to encode or process identity information. For a fair comparison, all architectures maintain identical depth and hidden layer dimensions. The first section of the table details single-identity models, whereas the second section focuses on multi-identity models.

Architecture	Trainable (with output layer)	Frozen	Trainable	ID-logic
NeuRiPhy	$n_q d_h + L d_h^2 + d_h n_v$	—	$\sim 6.3\text{M}$	—
LoRA Finetune	$2r L d_h$	$n_q d_h + L d_h^2 + d_h n_v$	$\sim 147\text{K}$	$\sim 147\text{K}$
Full Finetune	$L d_h^2$	$n_q d_h + d_h n_v$	$\sim 6.3\text{M}$	$\sim 6.3\text{M}$
Dense	$(n_q + n_p) d_h + L d_h^2 + d_h n_v$	—	$\sim 6.3\text{M}$	$\sim 40\text{K}$
Condition	$n_q d_h + d_c (n_p + d_c) + L d_h (d_h + d_c) + d_h n_v$	—	$\sim 7.1\text{M}$	$\sim 808\text{K}$
VeRA	$n_q d_h + d_c (n_p + d_c) + L (d_h^2 + 2r d_h + r d_c + d_h d_c) + d_h n_v$	—	$\sim 7.3\text{M}$	$\sim 965\text{K}$
LORAMI	$n_q d_h + d_c (n_p + d_c) + L (d_h^2 + 2r d_h + r d_c) + d_h n_v$	—	$\sim 6.5\text{M}$	$\sim 178\text{K}$

grouped into unified sliders. For instance, a single UI element may map to the lateral transitions of the lips or the vertical range between a frown and raised eyebrows. While these controls are coupled for the user, the underlying network treats each expression control as a distinct input.

Simulation To evaluate LORAMI, we compare the neural predictions with ground-truth simulation. Our baseline simulator is implemented in Python, using the L-BFGS optimizer with line-search from PyTorch. We guarantee consistency between LORAMI and the simulator by using the same mechanical energy. To obtain reliable ground-truth data, our baseline simulator makes use of double-precision arithmetic. Simulations are considered converged when the norm of the residual force falls below 10^{-5} .

Performance We train all networks for 12000 epochs, each consisting of 1000 random samples. For the single-identity NeuRiPhy baseline, training lasts approximately 7 hours; however, for the multi-identity models, training requires about 2 days. This increased duration is due to the expression personalization process, as detailed in the supplementary material. Since the identities are randomly generated during training, personalization must be repeated for every new combination. We observe that this overhead can be eliminated by fixing the dataset size and pre-computing the expression blendshapes and actuations for each training sample, at the cost of exposing the network to a smaller subset of identities. We note that the expression operators are only required during training; for evaluation, the underlying deformations are already embedded in the neural architecture.

Detailed evaluation runtimes for each architecture are provided in Table 2. The timings are produced using an NVIDIA RTX 4080 Super.

8. Results

We evaluate our method through experiments, baseline comparisons, and ablation studies, which we present in the following sections.

8.1. Experiments

Continuous Interpolation in Identity Space We evaluate the behavior of the model across identity space by interpolating between PCA identity codes while keeping the rig controls fixed. In Figure 3, we select three random identities and linearly interpolate between them in latent space, maintaining the same expression across all samples. Our method produces smooth transitions in both geometry and deformation. The effect of the controls is preserved along the interpolation path, while adapting to identity-specific anatomical features. In particular, nonlinear deformation patterns and contact interactions evolve coherently across identities. The model supports real-time variation of identity parameters, enabling interactive exploration of the identity space under fixed or varying controls.

Applications of Continuous Identity Control In Figure 4, we simulate orthognathic surgery by varying the shape of the skull and mandible, resulting in different bite configurations (underbite, overbite, normal bite). Our anatomical model supports bone cuts, enabling the simulation of pre- and post-treatment states within a unified framework. Using a single neural model, LORAMI captures these variations across identities while maintaining stable deformations and consistent contact interactions between tissue, lips, and teeth. The model remains robust under large anatomical changes and supports jaw roto-translations and expressions, allowing visualization of both static outcomes and dynamic motion after surgical interventions. While this model serves as a promising prototype, translating it to medical applications requires further validation.

In Figure 5, we present the simulation of an edentulous case, comparing configurations with and without teeth. The absence of teeth induces a forward translation and upward rotation of the facial structure, affecting the contact region between the lips. The model reproduces these changes while maintaining physically plausible expressions, including lip-lip interactions under smile activation.

In addition to structural variations, we show facial expressions across identities in Figure 1. Expressions are driven by different control configurations for each identity, showcasing our architecture's ability to produce diverse facial deformations. Our method

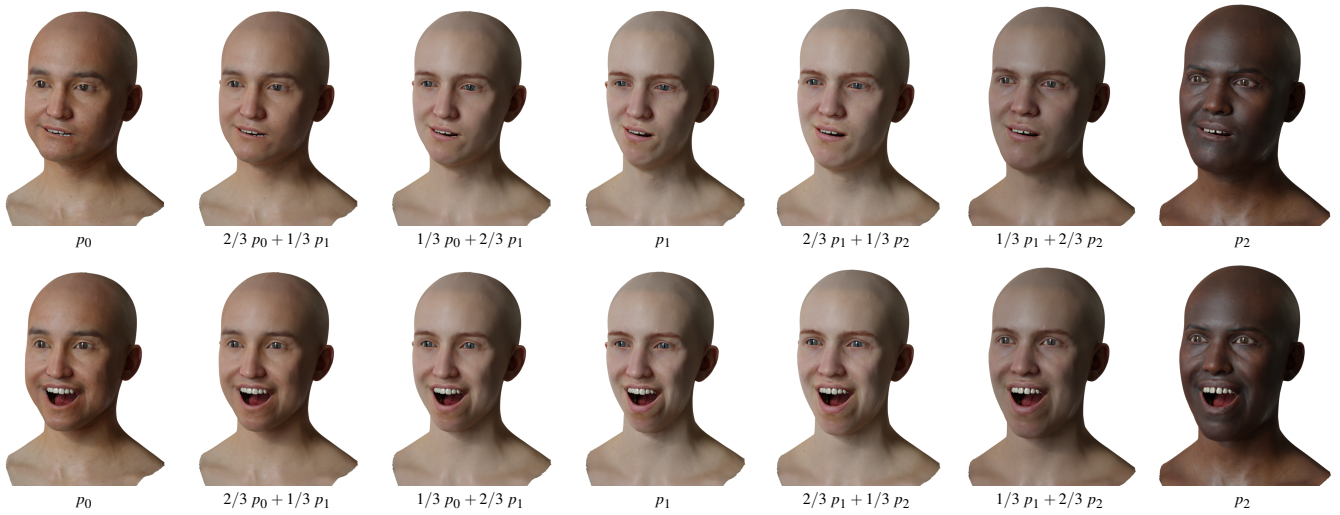


Figure 3: Latent-space interpolation between randomly chosen identities for fixed controls. LORAMI achieves smooth transitions.

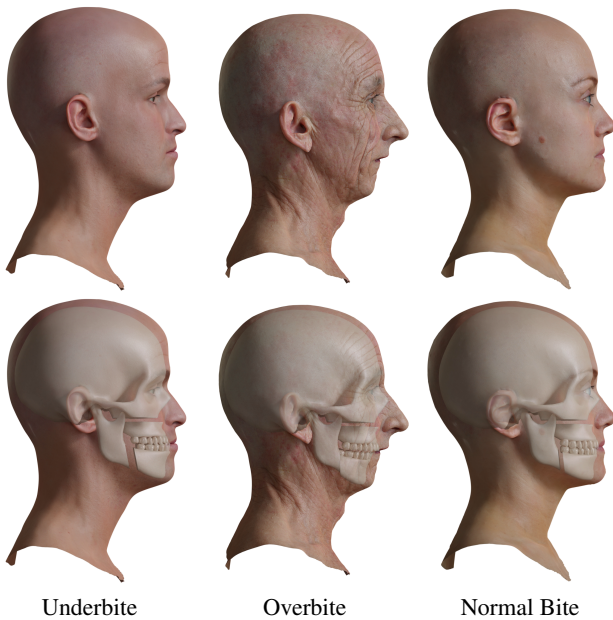


Figure 4: Orthognathic surgery simulation. We show facial deformations predicted by multi-identity LORAMI on different identities.

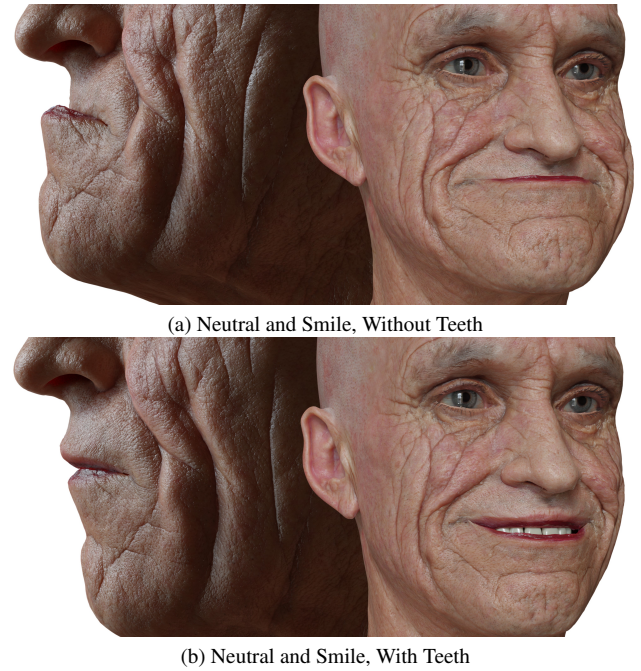


Figure 5: Edentulous case. We show the effect of teeth presence on facial appearance in the neutral state and during smiling.

enables the application of the same set of rig controls across varying facial geometries within a single, unified model.

8.2. Comparisons with Baselines

We compare LORAMI with both multi-identity and single-identity baselines, evaluating the effect of different architectural choices on the physical plausibility of the predicted deformations. We measure the deviation from the corresponding physics-based (FEM) simu-

lation (GT) in terms of vertex error, mechanical energy, and force norm.

Multi-Identity Model Comparison We first compare LORAMI with alternative multi-identity architectures. In Figure 6 and Table 2, we compare LORAMI with Dense and Conditioning baselines. In Figure 6, we consider an expression combining jaw opening with lip compression. The Dense and Conditioning models fail to re-

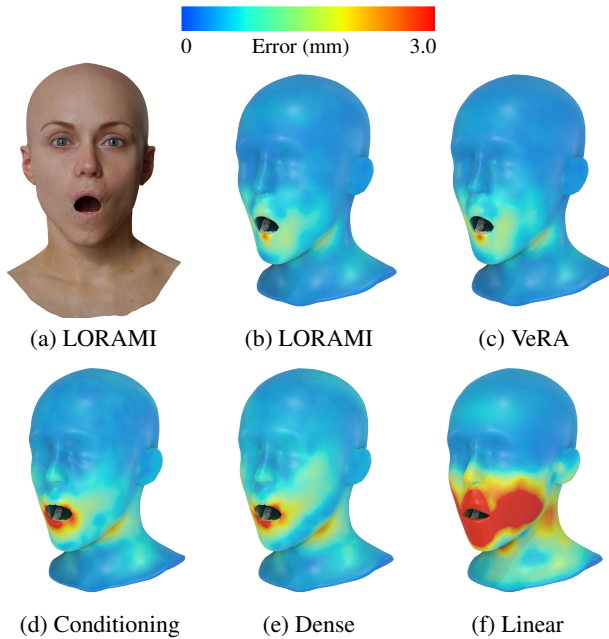


Figure 6: Comparison of multi-identity models in terms of distance from the ground-truth simulation.

produce lip compression, generating unnaturally stretched lips. In contrast, LORAMI and VeRA produce results closer to the ground-truth simulation, with LORAMI achieving a better balance in terms of number of parameters and evaluation time. The Blendshape-LBS linear rig fails to represent the complexity of physically based deformations. These observations are supported by the numerical metrics in Table 2, where we evaluate 100 randomly sampled identity and control configurations. LORAMI achieves the lowest error, energy, and residual force norm across all metrics. Our method introduces a small runtime overhead compared to simpler baselines (approximately 0.008–0.011 s per evaluation). Nevertheless, it still allows for real-time evaluation and is thus several orders of magnitude faster than the baseline FEM simulation.

Comparison with Single-Identity Models We further compare LORAMI with NeuRiPhy trained on a single identity and with finetuning approaches applied to a new identity. Given a training identity and an unseen evaluation identity from capture, we evaluate the accuracy of LORAMI when using the same control configuration. The target identity is reconstructed using our 40-dimensional PCA model. Both training and evaluation identities are subjected to a mandible opening combined with a slightly skewed expression. In Figure 7, we show that LORAMI achieves results on par with NeuRiPhy for the training identity, with both models reaching a mean error of 0.5 mm and a maximum error of ~ 6 mm. For the target identity, we compare LORAMI with finetuning of the hidden layers. LORAMI reaches similar accuracy, with mean and maximum error of 0.4 mm and 6 mm, respectively. After approximately one hour of finetuning, both Full Finetune and LoRA Finetune achieve lower accuracy than LORAMI, with larger errors concentrated around the lips, corresponding to approximately 0.8

mm mean error and 8 mm maximum error. We observe that Full Finetuning converges faster than LoRA Finetuning, although individual training epochs are slower. Additional numerical results are provided in Table 3. We detail the accuracy of our PCA model in the reconstruction of unseen identities in the supplementary material.

We further expand the comparison across multiple identities sampled from the PCA space. In Figure 8, we consider three randomly sampled identities and compare LORAMI with NeuRiPhy trained separately for each identity. LORAMI produces results that are on par with NeuRiPhy across all configurations. We observe that LORAMI achieves lower error on the upper lips, whereas NeuRiPhy shows lower error on the lower lips. These observations are further supported by the numerical results in Table 4, where LORAMI exhibits similar values in energy and error, and achieves lower maximum error in some scenarios. In general, LORAMI performs better on complex expressions involving large deformations, while NeuRiPhy shows slightly better accuracy in contact regions around the interfaces. We attribute this difference to the training procedure, as LORAMI leverages multiple identities during training, favoring large-scale deformation behavior over localized contact interactions.

Contact Interactions We further evaluate contact interactions at anatomical interfaces. In Figure 9, we show the ability of LORAMI to represent subtle teeth–lip and lip–lip interactions. The top row illustrates contacts during a smile combined with mandible rotation, while the bottom row shows lip–lip interactions for the edentulous case. LORAMI produces results that are visually comparable to the single-identity baseline, while significantly improving over the linear rig. Numerically, as shown in Table 4, LORAMI exhibits a small increase in maximum error for the smile compared to NeuRiPhy while maintaining comparable mean error. Results remain on par for the edentulous case.

8.3. Ablations

We investigate the impact of the identity sampling strategy on training stability. As shown in Figure 10, the choice of distribution for sampling identity codes is critical. When using unbounded normal sampling, $\mathbf{p} \sim \mathcal{N}(0, \lambda_i)$, the sampled identities may be implausible and outside the data manifold, destabilizing the gradient flow and affecting the training procedure. This manifests as large loss spikes and can lead to complete training failure. In contrast, using a Dirichlet distribution, $\mathbf{p} \sim \text{Dir}(\alpha)$, restricts sampling to the convex hull of the data. This prevents the occurrence of invalid geometries and ensures stable convergence. As illustrated in Figure 10, Dirichlet sampling avoids the instability observed with normal sampling, while preserving similar loss trends. Except for the spikes, the two training curves largely agree, indicating that the sampling strategy preserves good variability, while restricting sampling to feasible identities. We report only the training curve, as the validation curve is nearly identical due to random sampling.

9. Discussion

In this work, we introduced LORAMI, a novel technique designed for real-time, identity-aware, physics-based facial anima-

Table 2: Comparison of multi-identity models. The metrics are evaluated on a validation dataset of 100 random samples, with the mean, maximum, and standard deviation reported for each measure. Notably, 'Linear' represents the baseline blendshape-LBS linear rig. For all the neural models, the timings are obtained by running the networks in isolation, excluding the timings for the linear rig. We omit the force at the ground-truth, as the simulation is assumed to be converged when the force norm is $\sim 10^{-5}$.

Model	Error (mm) ↓		Energy ↓		Force Norm ↓		Time (s) ↓	
	mean	max	mean	max	mean	max	mean	max
Simulation (GT)	—	—	0.03 ± 0.04	0.17	—	—	60 ± 19	97
Linear	0.71 ± 0.43	2.0	0.16 ± 0.22	1.13	16.1 ± 22.5	114	0.050 ± 0.001	0.060
Dense	0.31 ± 0.16	0.70	0.04 ± 0.04	0.19	1.14 ± 1.04	9.37	0.005 ± 0.0001	0.005
Condition	0.31 ± 0.16	0.70	0.04 ± 0.04	0.18	1.90 ± 8.36	84.1	0.006 ± 0.0002	0.006
VeRA	0.21 ± 0.13	0.66	0.03 ± 0.04	0.17	0.80 ± 0.78	5.24	0.009 ± 0.0004	0.010
LORAMI	0.21 ± 0.13	0.64	0.03 ± 0.04	0.17	0.77 ± 0.72	4.64	0.008 ± 0.0003	0.009

Table 3: Comparison of LORAMI with single-identity and finetune models on a dataset of 100 random controls. We employ the same metrics as Table 2. The single-identity comparison utilizes an identity from the training set, whereas the finetune comparison employs an unseen identity.

Experiment	Model	Error (mm) ↓		Energy ↓		Force Norm ↓		Time (s) ↓	
		mean	max	mean	max	mean	max	mean	max
Single-Identity	Simulation (GT)	—	—	0.03 ± 0.03	0.10	—	—	58 ± 15	103
	NeuRiPhy	0.26 ± 0.14	0.59	0.03 ± 0.03	0.11	0.68 ± 0.40	1.73	0.005 ± 0.0003	0.006
	LORAMI	0.23 ± 0.11	0.54	0.03 ± 0.03	0.11	0.93 ± 0.96	8.5	0.008 ± 0.0003	0.009
Finetune	Simulation (GT)	—	—	0.03 ± 0.04	0.13	—	—	64 ± 21	103
	Full Finetune	0.50 ± 0.18	1.04	0.05 ± 0.05	0.20	2.30 ± 1.42	8.16	0.005 ± 0.002	0.006
	LoRA Finetune	0.54 ± 0.15	1.03	0.08 ± 0.04	0.19	6.66 ± 9.14	60.19	0.008 ± 0.0003	0.009
	LORAMI	0.26 ± 0.15	0.84	0.04 ± 0.04	0.15	0.94 ± 0.38	2.17	0.008 ± 0.0003	0.009

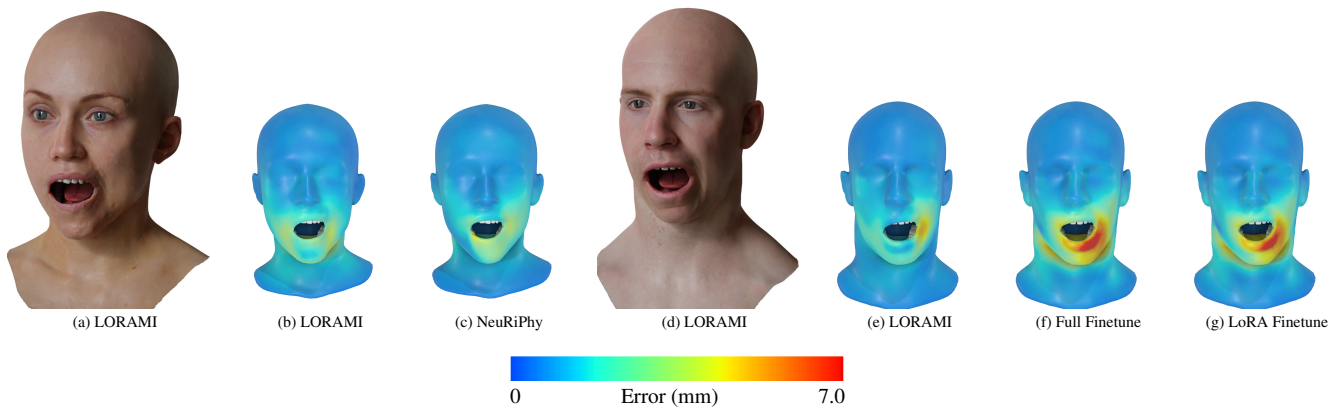


Figure 7: Comparison of LORAMI with single-identity NeuRiPhy and fine-tuning techniques, in terms of distance from the ground-truth simulation. On the left, we test an identity from the training dataset. On the right, we show generalization on an unseen identity.

Table 4: Numerical metrics for the selected comparisons for LORAMI and our NeuRiPhy (N.Phy) baseline. The Error measures the mean distance in mm from the ground truth simulation (GT), and E and $|f|$ are the energy and the force norm of the prediction. The ground-truth simulation is converged when the force norm is $\sim 1e-5$ for each example.

Scenario	Mean Error (mm)			Max Error (mm)			Energy				Force Norm		
	Lin.	LORAMI	N.Phy	Lin.	LORAMI	N.Phy	Lin.	LORAMI	N.Phy	GT	Lin.	LORAMI	N.Phy
Fig. 7, left	1.7	0.5	0.5	14.0	6.3	6.0	0.466	0.112	0.109	0.083	63.66	2.234	2.352
Fig. 9, upper	0.8	0.2	0.2	5.7	5.3	3.8	0.126	0.029	0.029	0.027	6.173	0.494	0.482
Fig. 9, lower	0.7	0.2	0.3	8.0	3.0	3.0	0.472	0.017	0.017	0.013	261.8	1.636	1.475
Fig. 8, left	1.2	0.2	0.3	10.7	2.9	3.8	0.056	0.050	0.052	0.047	73.82	1.011	0.844
Fig. 8, center	1.3	0.3	0.5	13.0	9.0	9.2	0.402	0.131	0.135	0.122	27.51	1.038	1.249
Fig. 8, right	0.5	0.3	0.3	5.7	5.4	5.4	0.186	0.004	0.004	0.002	107.2	0.454	0.460

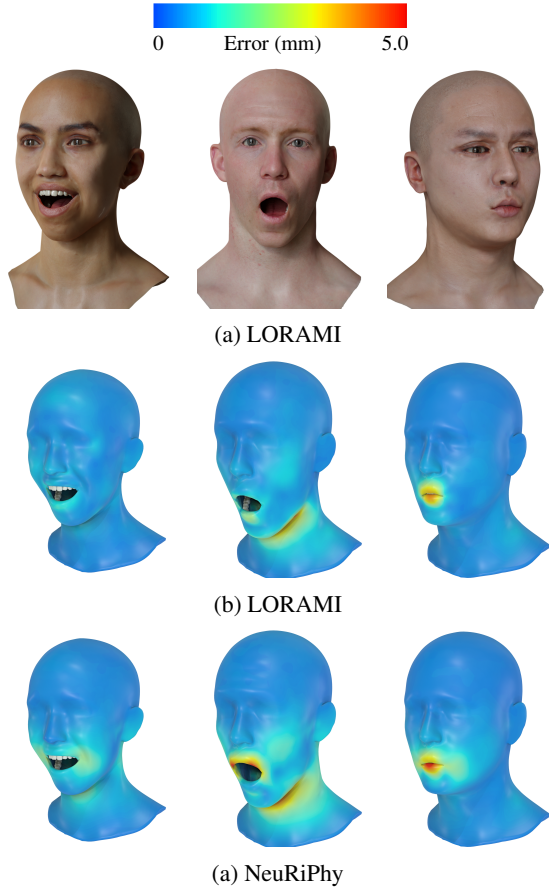


Figure 8: Comparison of multi-identity models with NeuRiPhy and Ground-Truth Simulation. The color plots illustrate the distance from the ground-truth simulation.

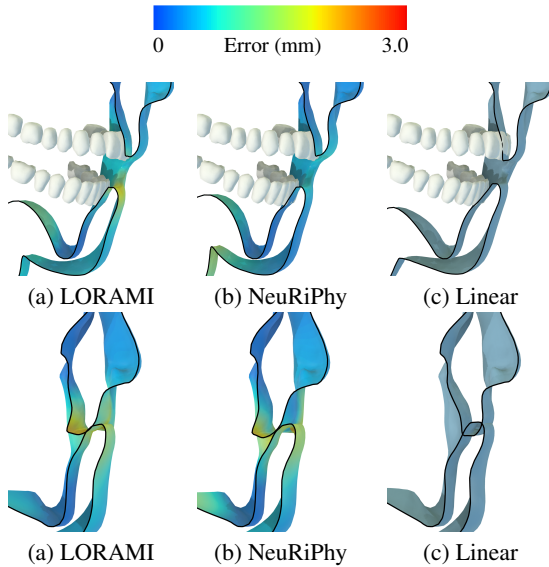


Figure 9: Comparison of teeth-lips contact interactions in terms of distance from the ground-truth simulation.

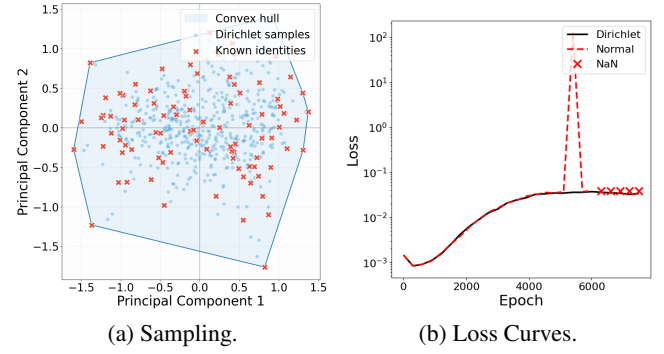


Figure 10: Identity Sampling. (a) Visualization of convex combinations produced by a Dirichlet distribution. (b) Comparison of training losses with Normal $\mathbf{p} \sim \mathcal{N}(0, \{\lambda_i\})$ and Dirichlet sampling $\mathbf{p} \sim \text{Dir}(\alpha)$. Dirichlet sampling ensures stability, whereas Normal sampling causes the appearance of NaN values.

tion. By leveraging low-rank factors shared across identities that are modulated by learned diagonal matrices, LORAMI bypasses the need for per-case retraining while maintaining the accuracy of single-identity models. This design reflects an inductive bias in which deformation behavior is largely shared across identities, while identity-specific effects can be captured through low-dimensional variations. Our approach effectively bridges the gap between physically-based facial deformations and identity modeling, enabling high-fidelity simulations across a continuous identity manifold.

In its current form, the reliance on shared rig controls implies that consistent control semantics must be maintained across identities, which is a common but non-trivial assumption in practical animation pipelines. At the same time, the ability of our model to represent identities is bounded by the underlying PCA model, and our use of convex combinations restricts the identity manifold to interpolations of the training set. As a result, the method is well-suited for interpolating between known identities, but does not aim to extrapolate to previously unseen anatomical configurations. Furthermore, Dirichlet sampling does not entirely eliminate the risk of implausible identity configurations, and the resulting head models may exhibit self-intersections. This highlights a trade-off between flexibility of the identity representation and the enforcement of geometric validity. Future research could explore non-linear embedding spaces for identity, where constraints on anatomical fidelity are explicitly enforced while still allowing for variability across subjects.

Overall, LORAMI achieves its goal of extending the accuracy of single-identity, physically-principled character rigs to continuous identity spaces. By eliminating the need for per-case retraining, our method facilitates the creation of sophisticated digital humans and streamlines the transfer of complex deformations across diverse identity spaces.

Acknowledgements

This research was funded by Align Technology.

References

- [BML*14] BOUAZIZ S., MARTIN S., LIU T., KAVAN L., PAULY M.: Projective dynamics: fusing constraint projections for fast simulation. *ACM Trans. Graph.* 33, 4 (July 2014). URL: <https://doi.org/10.1145/2601097.2601116>, doi:10.1145/2601097.2601116. 2
- [BODO18] BAILEY S. W., OTTE D., DILORENZO P., O'BRIEN J. F.: Fast and deep deformation approximations. *ACM Transactions on Graphics* 37 (2018). doi:10.1145/3197517.3201300. 3
- [BODO20] BAILEY S. W., OMENS D., DILORENZO P., O'BRIEN J. F.: Fast and deep facial deformations. *ACM Transactions on Graphics* 39 (7 2020). URL: <https://dl.acm.org/doi/10.1145/3386569.3392397>, doi:10.1145/3386569.3392397/SUPPL_FILE/A94-BAILEY.MP4. 3
- [CBL*19] CUDEIRO D., BOLKART T., LAIDLAW C., RANJAN A., BLACK M. J.: Capture, Learning, and Synthesis of 3D Speaking Styles. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA, June 2019), IEEE Computer Society, pp. 10093–10103. URL: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.01034>, doi:10.1109/CVPR.2019.01034. 2
- [CEM*22] CHOI B., EOM H., MOUSCADET B., CULLINGFORD S., MA K., GASSEL S., KIM S., MOFFAT A., MAIER M., REVELANT M., LETTERI J., SINGH K.: Anatomomy: an animator-centric, anatomically inspired system for 3d facial modeling, animation and transfer. *Proceedings - SIGGRAPH Asia 2022 Conference Papers* (11 2022). URL: <https://dl.acm.org/doi/10.1145/3550469.3555398>, doi:10.1145/3550469.3555398. 2
- [CPH*25] CORIGLIANO D., PETER D., HUBER N. B., THOMASZEWSKI B., SOLENTHALER B.: Neuriphy: Neural baking of physics-based deformations for facial rigs. *Proc. ACM Comput. Graph. Interact. Tech.* 8, 4 (Aug. 2025). URL: <https://doi.org/10.1145/3747866>, doi:10.1145/3747866. 2, 3, 4, 5, 6
- [DQY*23] DING N., QIN Y., YANG G., WEI F., YANG Z., SU Y., HU S., CHEN Y., CHAN C.-M., CHEN W., ET AL.: Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature machine intelligence* 5, 3 (2023), 220–235. 3
- [EST*20] EGGER B., SMITH W. A., TEWARI A., WUHRER S., ZOLLHOEFER M., BEELER T., BERNARD F., BOLKART T., KORTYLEWSKI A., ROMDHANI S., ET AL.: 3d morphable face models—past, present, and future. *ACM Transactions on Graphics (ToG)* 39, 5 (2020), 1–38. 2
- [FMD*19] FULTON L., MODI V., DUVENAUD D., LEVIN D. I., JACOBSON A.: Latent-space dynamics for reduced deformable simulation. *Computer Graphics Forum* 38 (5 2019), 379–391. doi:10.1111/CGF.13645. 3
- [HDDN19] HOLDEN D., DUONG B. C., DATTA S., NOWROUZEZAHRAI D.: Subspace neural physics: fast data-driven interactive simulation. In *Proceedings of the 18th Annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (New York, NY, USA, 2019), SCA '19, Association for Computing Machinery. URL: <https://doi.org/10.1145/3309486.3340245>, doi:10.1145/3309486.3340245. 3
- [HSW*22] HU E. J., SHEN Y., WALLIS P., ALLEN-ZHU Z., LI Y., WANG S., WANG L., CHEN W., ET AL.: Lora: Low-rank adaptation of large language models. *Iclr* 1, 2 (2022), 3. 3, 6
- [IKKP17] ICHIM A. E., KADLEČEK P., KAVAN L., PAULY M.: Phace: Physics-based face modeling and animation. *ACM Transactions on Graphics* 36 (2017). doi:10.1145/3072959.3073664. 2
- [KBA24] KOPICZKO D. J., BLANKEVOORT T., ASANO Y. M.: Vera: Vector-based random matrix adaptation, 2024. URL: <https://arxiv.org/abs/2310.11454>, arXiv:2310.11454. 6
- [LA10] LEWIS J. P., ANJYO K. I.: Direct-manipulation blendshapes. *IEEE Computer Graphics and Applications* 30 (2010), 42–50. doi:10.1109/MCG.2010.41. 2
- [LCF00] LEWIS J. P., CORDNER M., FONG N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, New Orleans, LA, USA, July 23–28, 2000* (2000), Brown J. R., Akeley K., (Eds.), ACM, pp. 165–172. URL: <https://doi.org/10.1145/344779.344862>, doi:10.1145/344779.344862. 2
- [LTJ24] LIU C., TAKIKAWA T., JACOBSON A.: A lora is worth a thousand pictures. *arXiv preprint arXiv:2412.12048* (2024). 3
- [LYX*24] LU Z., YANG H., XU D., LI B., IVANOVIC B., PAVONE M., WANG Y.: Lora3d: Low-rank self-calibration of 3d geometric foundation models. *arXiv preprint arXiv:2412.07746* (2024). 3
- [MSE*11] MCADAMS A., SELLE A., EMPEY M., TAMSTORF R., TERAN J., SIFAKIS E., ZHU Y., MCADAMS A., TERAN J., SIFAKIS E.: Efficient elasticity for character skinning with contact and collisions. *ACM Transactions on Graphics* 30 (7 2011), 1–12. doi:10.1145/2010324.1964932. 2
- [MSS*21] MA S., SIMON T., SARAGIH J., WANG D., LI Y., TORRE F. D. L., SHEIKH Y.: Pixel codec avatars. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2021), 64–73. doi:10.1109/CVPR46437.2021.00013. 2, 3
- [PLX*25] PAN Y., LIU C., XU S., TAN S., YANG J.: Vasa-rig: Audio-driven 3d facial animation with 'live' mood dynamics in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 2416–2425. doi:10.1109/TVCG.2025.3549168. 2
- [QSA*23] QIN D., SAITO J., AIGERMAN N., GROUEIX T., KOMURA T.: Neural face rigging for animating and retargeting facial meshes in the wild. *Proceedings - SIGGRAPH 2023 Conference Papers* (7 2023). doi:10.1145/3588432.3591556. 3
- [RCPO21] ROMERO C., CASAS D., PÉREZ J., OTADUY M.: Learning contact corrections for handle-based subspace dynamics. *ACM Trans. Graph.* 40, 4 (jul 2021). URL: <https://doi.org/10.1145/3450626.3459875>, doi:10.1145/3450626.3459875. 3
- [RdGM20] RADZIHOVSKY S., DE GOES F., MEYER M.: Facebaker: Baking character facial rigs with machine learning. In *ACM SIGGRAPH 2020 Talks* (New York, NY, USA, 2020), SIGGRAPH '20, Association for Computing Machinery. URL: <https://doi.org/10.1145/3388767.3407340>, doi:10.1145/3388767.3407340. 3
- [Ryu23] RYU S.: Low-rank adaptation for fast text-to-image diffusion fine-tuning. <https://github.com/cloneofsimon/lora>, 2023. Accessed 2026. 3
- [RZW*21] RICHARD A., ZOLLHÖFER M., WEN Y., DE LA TORRE F., SHEIKH Y.: Meshtalk: 3d face animation from speech using cross-modality disentanglement. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 1153–1162. doi:10.1109/ICCV48922.2021.00121. 2
- [SRJ*23] SHARP N., ROMERO C., JACOBSON A., VOUGA E., KRY P., LEVIN D. I., SOLOMON J.: Data-free learning of reduced-order kinematics. *Proceedings - SIGGRAPH 2023 Conference Papers* (7 2023). doi:10.1145/3588432.3591521. 3
- [SWR*21] SRINIVASAN S. G., WANG Q., ROJAS J., KLÁR G., KAVAN L., SIFAKIS E.: Learning active quasistatic physics-based models from data. *ACM Transactions on Graphics* 40 (7 2021). doi:10.1145/3450626.3459883. 3
- [SYS*21] SHEN S., YANG Y., SHAO T., WANG H., JIANG C., LAN L., ZHOU K.: High-order differentiable autoencoder for nonlinear model reduction. *ACM Trans. Graph.* 40, 4 (jul 2021). URL: <https://doi.org/10.1145/3450626.3459754>, doi:10.1145/3450626.3459754. 3
- [TMKLS25] TRUONG A., MAHMOUD A. H., KONAKOVIĆ LUKOVIĆ M., SOLOMON J.: Low-rank adaptation of neural fields. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers* (New York, NY, USA, 2025), SA Conference Papers '25, Association for Computing Machinery. URL: <https://doi.org/10.1145/3757377.3763882>, doi:10.1145/3757377.3763882. 3

- [WBZ22] WANG S.-Y., BAU D., ZHU J.-Y.: Rewriting geometric rules of a gan. *ACM Trans. Graph.* 41, 4 (July 2022). URL: <https://doi.org/10.1145/3528223.3530065>, doi:10.1145/3528223.3530065. 3
- [WDCT24] WANG J., DU Y., COROS S., THOMASZEWSKI B.: Neural modes: Self-supervised learning of nonlinear modal subspaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 23158–23167. 3
- [WSGW25] WITZIG P., SOLENTHALER B., GROSS M., WAMPFLER R.: Phonemet: A transformer pipeline for text-driven facial animation. In *The 18th ACM SIGGRAPH Conference on Motion, Interaction, and Games* (New York, NY, USA, 2025), MIG '25, Association for Computing Machinery. URL: <https://doi.org/10.1145/3677388.3696336>, doi:10.1145/3677388.3696336. 2
- [YMD*25] YANG J., MA Y., DI D., CUI J., LI H., CHEN W., XIE Y., YANG X., ZUO W.: Qr-lora: Efficient and disentangled fine-tuning via qr decomposition for customized generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2025), pp. 17587–17597. 3
- [YZC*24] YANG L., ZOISS G., CHANDRAN P., GROSS M., SOLENTHALER B., SIFAKIS E., BRADLEY D.: Learning a generalized physical face model from data. *ACM Trans. Graph.* 43, 4 (July 2024). URL: <https://doi.org/10.1145/3658189>, doi:10.1145/3658189. 3, 4, 6
- [ZBLJ20] ZHANG J. E., BANG S., LEVIN D. I. W., JACOBSON A.: Complementary dynamics. *ACM Trans. Graph.* 39, 6 (Nov. 2020). URL: <https://doi.org/10.1145/3414685.3417819>, doi:10.1145/3414685.3417819. 2
- [ZCYZ25] ZHANG Y., CAO C., YU C., ZHU J.: Lion-lora: Rethinking lora fusion to unify controllable spatial and temporal generation for video diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2025), pp. 14569–14579. 3