Virtual Reality & Intelligent Hardware

•Article•

October 2024 Vol. 6 No. 5: 383–395
10.1016/j.vrih.2024.08.006

Mesh representation matters: investigating the influence of different mesh features on perceptual and spatial fidelity of deep 3D morphable models

Robert KOSK[1,2*], Richard SOUTHERN[1], Lihua YOU[1], Shaojun BIAN[2,3], Willem KOKKE[2], Greg MAGUIRE[4]

1. *Centre for Digital Entertainment, National Centre for Computer Animation, Bournemouth University, Poole BH*12 5*BB, UK;*
2. *Humain Ltd., Belfast BT*1 2*LA, UK;*
3. *School of Creative and Digital Industries, Buckinghamshire New University, High Wycombe HP*11 2*JZ, UK;*
4. *Belfast School of Art, Ulster University, Belfast BT*15 1*ED, UK*

Received 29 March 2024; **Revised** 20 May 2024; **Accepted** 30 August 2024

Abstract: Background Deep 3D morphable models (deep 3DMMs) play an essential role in computer vision. They are used in facial synthesis, compression, reconstruction and animation, avatar creation, virtual try-on, facial recognition systems and medical imaging. These applications require high spatial and perceptual quality of synthesised meshes. Despite their significance, these models have not been compared with different mesh representations and evaluated jointly with point-wise distance and perceptual metrics. **Methods** We compare the influence of different mesh representation features to various deep 3DMMs on spatial and perceptual fidelity of the reconstructed meshes. This paper proves the hypothesis that building deep 3DMMs from meshes represented with global representations leads to lower spatial reconstruction error measured with $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ norm metrics and underperforms on perceptual metrics. In contrast, using differential mesh representations which describe differential surface properties yields lower perceptual FMPD and DAME and higher spatial fidelity error. The influence of mesh feature normalisation and standardisation is also compared and analysed from perceptual and spatial fidelity perspectives. **Results** The results presented in this paper provide guidance in selecting mesh representations to build deep 3DMMs accordingly to spatial and perceptual quality objectives and propose combinations of mesh representations and deep 3DMMs which improve either perceptual or spatial fidelity of existing methods.

Keywords: Shape modelling; Deep 3D morphable models; Representation learning; Feature engineering; Perceptual metrics

Supported by the Centre for Digital Entertainment at Bournemouth University by the UK Engineering and Physical Sciences Research Council (EPSRC) EP/L016540/1 and Humain Ltd.

Citation: Robert KOSK, Richard SOUTHERN, Lihua YOU, Shaojun BIAN, Willem KOKKE, Greg MAGUIRE. Mesh representation matters: investigating the influence of different mesh features on perceptual and spatial fidelity of deep 3D morphable models. Virtual Reality & Intelligent Hardware, 2024, 6(5): 383–395.

*Corresponding author, rkosk@bournemouth.ac.uk

Peer review under the responsibility of Beijing Zhongke Journal Publishing Co. Ltd.

1 Introduction

Parametric face models are extensively used in computer vision tasks. Most previous publications on deep 3D morphable models (deep 3DMMs) evaluate the proposed methods against only one mesh representation, most commonly standardised Euclidean coordinates. Consequently, the performance of these models with other mesh representations is unknown, which provides an opportunity to evaluate these models from mesh feature representation perspective. Furthermore, it allows to distinguish the combinations of models and feature representations which outperform the existing methods.

In literature on deep 3D morphable models, evaluation metrics which compare ground truth meshes with reconstructed meshes are usually limited to Euclidean distance (L_2 norm) or Manhattan distance (L_1 norm). However, *L*-norm metrics poorly correlate with perceptual quality of reconstructed meshes. Given that the perceptual quality of meshes is an essential factor in assessing the method's usability in visual applications, the lack of perceptual evaluation of deep 3D morphable models is a noticeable gap in the literature. This paper investigates the effects of using different mesh feature representations in deep 3DMMs, as well as the standardisation and normalisation of these features, with an objective of improving the perceptual quality and spatial fidelity of facial meshes output from these models.

An observation by Sorkine at al. is particularly relevant in this work[1]. Its authors focus on suppressing the visual effects of quantisation. The authors demonstrate that quantisation of meshes in differential representation introduces perceptually insignificant, low-frequency error, whilst quantisation of 3D meshes represented in 3D Euclidean coordinates space introduces noticeable, high-frequency discrepancies. Differential representation explicitly encodes the surface properties of the mesh, and the position of vertices is encoded implicitly in this representation. In contrast, 3D Euclidean coordinates describe the position of vertices in explicit form, and the surface properties can be only implicitly derived from this representation. It can be observed that after the representation is subjected to quantisation, the properties explicitly encoded in this representation are best preserved. Conversely, the properties which are only implicitly encoded in the representation can be more severely affected by quantisation. Since encoding 3D meshes in a compact parametric space is lossy, deep 3D morphable models have similar compression side-effects to quantisation. Based on these observations, it can be hypothesised that using different mesh feature representations can improve the perceptual or spatial fidelity of deep 3D morphable models. More specifically, it can be hypothesised that using differential mesh representations which explicitly encode surface properties improves the perceptual quality of the reconstructed meshes, whilst using global mesh representations which explicitly encode vertex positions in 3D space improves the spatial fidelity.

To test this hypothesis, combinations of global and differential representations with different deep 3DMMs are evaluated using metrics which measure spatial fidelity and perceptual quality of meshes reconstructed by these models. Euclidean coordinates and standardised Euclidean coordinates are the most commonly used global mesh representations in deep 3DMMs and therefore are used in the comparisons. Analogically, normalised deformation representation (DR Norm.) is selected, as it is the most commonly used differential representation in deep 3DMMs. Additionally, not normalised version of this representation (DR) is included in comparisons to verify the effect of normalisation. FeaStNet[2], Neural 3DMM[3], SpiralNet++[4], Mesh Autoencoder[5] and LSA-3DMM[6] deep 3D morphable models are compared in this paper. L_1 and L_2 norms are used to evaluate spatial fidelity of the reconstructed meshes when compared against the ground truth meshes. Dihedral angle mesh error (DAME)[7] and fast mesh perceptual distance (FMPD)[8] metrics are employed to measure perceptual quality of output meshes.

This paper proved the hypothesis that, across different deep 3DMMs, using Euclidean coordinates or standardised Euclidean coordinates representations yields lower L_1 and L_2 reconstruction error and higher

DAME and FMPD perceptual error in comparison with using DR and DR Norm. representations, which result in higher L_1 and L_2 reconstruction error and lower DAME and FMPD perceptual error.

Our main contributions are:

• Demonstrating that differential mesh representations in deep 3DMMs yield higher perceptual quality of the reconstructed meshes, whilst global mesh representations result in higher spatial fidelity.

• Identification of strengths and weaknesses of standardising Euclidean coordinates features and normalising DR features in deep 3D morphable models from spatial fidelity and perceptual quality perspectives.

• The improved quality of meshes generated with current deep 3DMM methods by matching them with best performing mesh representations when evaluated from spatial fidelity and perceptual quality perspectives.

2 Related work

2.1 3D shape representations

3D shapes can be represented in various ways, and the choice of a suitable representation depends on the application[9]. Representations can explicitly expose different geometric properties of 3D shapes. Global representations encode positional information of vertices in 3D space, whilst differential representations explicitly encode first-order surface properties, such as curvature or deformation gradient.

Global shape representations are widely used in geometric deep learning. Standardised Euclidean coordinates are the most common representation of 3D shapes in deep 3DMMs, utilised in [3, 4, 6, 10, 11]. The non-standardised version, Euclidean coordinates, is used in [5, 12, 13]. Therefore, standardised and not standardised Euclidean coordinates mesh representations are included in our comparisons.

Laplacian coordinates[14] were employed in mesh editing and shape approximation. In the realm of deep 3DMMs, [15] extends previous differential representations and introduces a rotation-invariant mesh difference (RIMD). Unlike Laplacian coordinates, RIMD is invariant to rigid transformations. Nevertheless, it is incompatible with most deep 3DMMsas as its signal is defined on vertices and edges. [16] improves computational limitations of RIMD and introduces as-consistent-as-possible (ACAP) representation, which allows to encode large rotations over 2π. [17] proposes a simpler normalised deformation representation (DR Norm.) based on the deformation gradient. It is also used in deep 3DMMs of [18, 19].

2.2 Parametric models with graph neutral networks

Deep learning approaches have been extended to irregular graphs [9, 20] and allowed for learning parametric models of 3D meshes. [11] utilises an autoencoder with spectral graph convolutional operations to develop the first deep 3DMM of 3D faces. More recent approaches move away from isotropic convolutional operators in spectral domain in favour of the anisotropic ones defined in the spatial domain. [3, 4, 5, 6, 10] improve deep 3DMMs with custom convolutional and aggregation operators. While these models use global mesh representations such as Euclidean coordinates, [21] convolves pre-processed ACAP deformation features, while [18, 19] use normalised deformation representation (DR Norm.).

All these deep 3D morphable models use either global or differential mesh representations and the influence of different mesh representations on these models has not been evaluated before. Moreover, the reconstruction results from all these models have been evaluated solely from spatial fidelity perspective. This paper addresses these gaps in literature and investigates the influence of global and differential mesh representations in different deep 3DMMs on spatial and perceptual fidelity of the reconstructed meshes.

3 Comparative framework

3.1 Overview

In the proposed framework, a full factorial experiment is conducted. The study uses L_1 and L_2 norms to assess spatial fidelity and DAME with FMPD to evaluate perceptual quality of all combinations of global representations (Euclidean coordinates and standardised Euclidean coordinates) and differential representations (deformation representation and normalised deformation representation) with five deep 3DMMs: FeaStNet[2], Neural 3DMM[3], SpiralNet++[4], Mesh Autoencoder[5] and LSA-3DMM[6]. This experimental setup results in forty combinations. Calculation of mesh representations is covered in Section 3.2. Section 3.3 describes implementation and training of deep 3DMMs compared in this work. Finally, details about evaluation with L-norms and perceptual metrics are provided in Section 3.4.

3.2 Mesh features

In this study, the comparisons are conducted on datasets of triangle meshes which share the same connectivity. All meshes in each dataset are translated so that their centroid coincides with the origin. Subsequently, the meshes are rigidly registered to their mean $\bar{P}$.

Global representation

3D shapes are most commonly represented with points in Euclidean space. Let P be Euclidean XYZ coordinates of these points, and $F_{\text{Eucl.}} \equiv P$ be the feature in Euclidean coordinates representation input to a deep 3DMM. The standardised Euclidean coordinates is calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=0}^{n}\left(F_{\text{Eucl.}} - \bar{P}\right)^2}{n}}$$
$$F_{\text{Eucl.Std.}} = \frac{F_{\text{Eucl.}} - \bar{P}}{\sigma} \tag{1}$$

The inverse to standardisation is calculated as follows:

$$F_{\text{Eucl.}} = \sigma F_{\text{Eucl.Std.}} + \bar{P} \tag{2}$$

Differential representation

Deformation representation (DR) encodes a per-vertex deformation gradient T_i between the position of a vertex p_i on a mean $\bar{P}$ and the position of a deformed vertex p'_i on P. Following [22], the deformation gradient T_i is calculated by minimising energy $E\left(T_i\right)$ in least-squares sense:

$$E\left(T_i\right) = \sum_{j \in \mathcal{N}_i} c_{ij} \left\| \left(p'_i - p'_j\right) - T_i\left(p_i - p_j\right) \right\|^2 \tag{3}$$

where c_{ij} are cotangent weights calculated on a mean of the training meshes.

Following [16], matrices T_i are decomposed using polar decomposition into a rotational part R_i and a scale/shear part S_i. Then, the rotation matrix $\mathbf{R_i}$ is transformed to $\log R_i$ and an identity matrix I is subtracted from S_i. The final DR feature f_i at i-th vertex is constructed of 6 non-trivial elements of S_i and 3 non-trivial elements of R_i, so that $\left| f_i \right| = 9$. The deformation representation feature is denoted as F_{DR}.

Normalised deformation representation scaled and shifted to range $[-1, 1]$ is calculated as:

$$F_{\text{DR Norm.}} = 2\,\frac{\boldsymbol{F}_{\text{DR}} - \boldsymbol{F}_{\min}}{\boldsymbol{F}_{\max} - \boldsymbol{F}_{\min}} + \boldsymbol{F}_{\min} \tag{4}$$

where $\boldsymbol{F}_{\min}$ and $\boldsymbol{F}_{\max}$ are DR features with minimum and maximum values across dataset, respectively. The denormalised feature is:

$$\boldsymbol{F}_{\mathrm{DR.}} = \frac{(\boldsymbol{F}_{\mathrm{DR\,Norm.}} + 1)(\boldsymbol{F}_{\max} - \boldsymbol{F}_{\min})}{2 + \boldsymbol{F}_{\min}} \tag{5}$$

3.3 Implementation and training of deep 3DMMs

Deep 3D morphable models compared in this study are either autoencoders (FeaStNet[2], Neural 3DMM[3], SpiralNet++[4] and LSA-3DMM[6]) or a variational autoencoder (Mesh Autoencoder[5]). At each training iteration, meshes from a training set transformed into features $F_{\mathrm{Eucl.}}$, $F_{\mathrm{Eucl.\,Std.}}$, F_{DR} or $F_{\mathrm{Norm.}}$ are encoded to latent parameters Z and subsequently decoded to a reconstructed feature $F'_{\mathrm{Eucl.}}$, $F'_{\mathrm{Eucl.\,Std.}}$, F'_{DR} or $F'_{\mathrm{Norm.}}$. In autoencoders, Z is directly output by the encoder. In variational autoencoders, Z is stochastically sampled from normal distribution output by the encoder. Learnable parameters of the network are updated at each iteration in terms of L_1 norm loss calculated in feature space. Importantly, features $F'_{\mathrm{Eucl.\,Std.}}$ are destandardised and features $F'_{\mathrm{Norm.}}$ are denormalised before loss calculation. In the variational autoencoder, an additional weighted KL divergence term with weight of 10^{-6} is added to the loss function.

All models are trained with Adam optimiser[23] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, latent space size of 64, learning rate of 10^{-3}, learning rate decay of 0.99 and batch size of 16. ELU activation functions are used. The models are trained for 450 epochs.

In experiments, the Mesh Autoencoder[5] has 5 upscaling convolutional blocks and 5 downscaling convolutional blocks. The residual rates are at 0.5. Due to the nature of subsampling method used in this approach, additional graph convolutional layer was added to bring the latent space size to 64 (8 latent vertices × 8-dimensional features). Therefore, the channel dimensions are [$|f|$, 32, 64, 128, 128, 8], where $|f|$ is the size of a per-vertex input feature. In Euclidean coordinates-based representations $|f| = 3$, whilst in differential coordinates-based representations $|f| = 9$. The convolutional operators have stride of 2, kernel radius of 2 and 35 weight bases.

In the case of FeaStNet[2], Neural 3DMM[3], SpiralNet++[4] and LSA-3DMM[6], their encoders are built of convolutional layers, each followed by a pooling layer with a pooling factor = 4. The last layer of the encoder is fully connected, with an output size equal to a latent space size of 64. The encoder channel dimensions = [$|f|$, 16, 32, 64, 128]. The decoder mirrors the encoder. Additionally, in Neural 3DMM[2], the encoder's first two layers and the decoder's last two layers are dilated convolutions with step size of 2 and dilation ratio of 2.

3.4 Evaluation strategy

Unlike loss functions, which are calculated in feature space, the reconstructed meshes P' are evaluated against their ground truth counterparts P in Euclidean space.

Spatial fidelity metrics

Two spatial fidelity metrics, point-wise L_1 and L_2 norms are used in the comparisons. The commonly used L_1 norm $||P - P'||_1$ and L_2 norm $||P - P'||_2$ between reconstructed Euclidean coordinates P' and the ground truth P are selected to evaluate spatial fidelity. Despite their popularity, L_1 and L_2 have low correlation with the human visual system[24].

Perceptual quality metrics

Previous work on deep 3DMMs does not address perceptual mesh quality in evaluation of reconstructed meshes. In this work, DAME[7] and FMPD[8] are used to measure perceptual discrepancy between ground truth meshes and meshes reconstructed by deep 3DMMs.

We select DAME because it has one of the highest correlation scores with human visual system on the compression task[24]. DAME is restricted to datasets with shared topology as it is based on the difference between oriented dihedral angles in meshes. The metric takes into account the masking effect and the visibility weighting. The last one depends on the camera view and resolution, thus we replace this term with triangle areas, as suggested in [7] method.

FMPD perceptual metric has achieved the highest overall correlation with human visual system in [24] and therefore has been selected as the second perceptual metric in the proposed comparative framework. FMPD measures discrepancy between local and global roughness of meshes. Similarly to DAME, it also accounts for the masking effect. However, unlike DAME, the metric is capable of measuring perceptual discrepancy between meshes with different connectivity.

4 Experimental results and discussion

The analysis of comparisons of mesh representations with different datasets provides valuable insights into the performance of Euclidean and differential coordinates-based representations used in deep 3D morphable models. Our comperative framework is applied to two datasets: FaceWarehouse[25] (150 meshes, 11510 vertices each) and Facsimile™[26] (202 meshes, 14921 vertices each). The ratio between training, validation and test subsets is 85:5:10. Tables 1 and 2 show quantitative evaluation of the reconstruction results, while Figures 1 and 2 give a qualitative insight.

4.1 Influence of global and differential mesh representations

On the Facsimile training dataset, Euclidean coordinates-based representations demonstrate superior spatial fidelity, as evidenced by lower L_1 and L_2 norm errors compared to differential coordinates-based representations. However, the latter outperform in perceptual quality metrics such as DAME and FMPD. This trend persists in the Facsimile test dataset, where Euclidean coordinates-based representations consistently excel in spatial fidelity, whilst differential coordinates-based representations show superiority in perceptual quality metrics.

Similar trends are observed in the FaceWarehouse dataset, where Euclidean coordinates-based representations exhibit lower L_1 and L_2 norm errors and higher DAME and FMPD errors on both training and test subsets. Conversely, differential coordinates-based representations indicate better perceptual quality, as they consistently yield lower DAME and FMPD with higher L_1 and L_2 norm errors.

4.2 Influence of normalisation and standardisation

Overall, standardisation applied to Euclidean coordinates representation improves the spatial and perceptual fidelity of the output meshes. On training sets, standardisation decreased the L_1 norm error by the factors of $3.65^{+1.54}_{-2.55}$ and $5.02^{+2.07}_{-2.75}$ on Facsimile and FaceWarehouse datasets, respectively. Standardisation results in $1.45^{+0.58}_{-0.42}$ times lower L_2 norm error on the Facsimile dataset with all the models except SpiralNet++[4], with which standardisation practically does not affect the L_2 norm error. Perceptual metrics, such as DAME and FMPD, were also consistently improved by standardisation, indicating its positive impact on overall quality. DAME decreased by $1.91^{+0.42}_{-0.62}$ and $4.99^{+2.54}_{-2.81}$ on Facsimile and FaceWarehouse, respectively. Moreover, standardisation improved FMPD by $3.30^{+2.65}_{-2.26}$, $4.87^{+2.43}_{-3.24}$, $3.89^{+3.53}_{-2.71}$, respectively. On test sets, standardisation of Euclidean coordinates representation generally improved mesh quality. Notably, it decreased L_1 norm error by $1.14^{+0.29}_{-0.20}$ and $1.15^{+0.25}_{-0.39}$ on the Facsimile and FaceWarehouse datasets, respectively. Similarly, L_2 norm error reductions were observed, alongside improvements in perceptual metrics, albeit with occasional exceptions.

In contrast, the impact of normalising features in the deformation representation varied across models and

Table 1 Quantitative comparison of the reconstruction results on the Facsimile ™[26] dataset. In each column, for each metric, best results are in bold

Facsimile™ Dataset-Training							
			FeaStNet	Neural 3DMM	Spiral Net++	Mesh Autoenc.	LSA-3DMM
Euclidean	L_1 *norm*	$\times10^{-3}$	9.27	6.595	4.777	2.602	6.881
	L_2 *norm*	$\times10^{-5}$	71.102	28.908	**13.498**	47.742	29.272
	FMPD	$\times10^{-2}$	100	93.873	100	51.962	**6.433**
	DAME	$\times10^{-2}$	31.542	15.306	22.990	6.036	2.85
Euclidean Std	L_1 *norm*	$\times10^{-3}$	**2.022**	**1.714**	**1.353**	**2.364**	**1.325**
	L_2 *norm*	$\times10^{-5}$	38.36	**23.012**	13.614	42.129	14.394
	FMPD	$\times10^{-2}$	35.369	25.553	44.532	43.544	8.971
	DAME	$\times10^{-2}$	5.296	3.839	5.349	4.9	2.719
DR	L_1 *norm*	$\times10^{-3}$	6.889	3.127	5.218	4.012	3.55
	L_2 *norm*	$\times10^{-5}$	**26.124**	57.278	14.933	**9.54**	71.399
	FMPD	$\times10^{-2}$	**20.671**	**18.254**	**10.659**	**6.795**	18.873
	DAME	$\times10^{-2}$	**2.27**	**2.092**	**2.296**	**2.767**	**2.104**
DR Norm	L_1 *norm*	$\times10^{-3}$	16.538	13.291	11.767	4.769	3.843
	L_2 *norm*	$\times10^{-5}$	135.619	83.129	68.504	13.078	87.011
	FMPD	$\times10^{-2}$	49.14	48.031	45.708	12.479	6.463
	DAME	$\times10^{-2}$	7.735	7.109	6.66	3.045	2.57

Facsimile™ Dataset-Test							
			FeaStNet	Neural 3DMM	Spiral Net++	Mesh Autoenc.	LSA-3DMM
Euclidean	L_1 *norm*	$\times10^{-3}$	9.45	6.983	**6.004**	8.320	7.194
	L_2 *norm*	$\times10^{-5}$	71.84	30.832	**19.428**	39.529	30.315
	FMPD	$\times10^{-2}$	100	94.536	100	48.925	6.862
	DAME	$\times10^{-2}$	31.813	15.414	22.875	5.814	2.806
Euclidean Std	L_1 *norm*	$\times10^{-3}$	**9.07**	**5.954**	6.38	**5.782**	**6.367**
	L_2 *norm*	$\times10^{-5}$	**49.656**	**20.974**	24.434	**19.54**	**24.894**
	FMPD	$\times10^{-2}$	34.374	23.965	39.42	30.846	8.396
	DAME	$\times10^{-2}$	5.354	3.807	4.869	3.885	2.939
DR	L_1 *norm*	$\times10^{-3}$	9.525	13.275	12.465	12.049	15.634
	L_2 *norm*	$\times10^{-5}$	50.34	107.757	84.487	84.892	129.809
	FMPD	$\times10^{-2}$	**19.875**	**17.118**	**12.445**	**2.653**	17.603
	DAME	$\times10^{-2}$	**2.228**	**2.456**	**2.278**	3.065	**2.393**
DR Norm	L_1 *norm*	$\times10^{-3}$	17.507	17.406	12.791	9.290	13.825
	L_2 *norm*	$\times10^{-5}$	158.564	158.41	85.656	49.449	102.034
	FMPD	$\times10^{-2}$	49.784	48.68	45.26	3.609	**4.603**
	DAME	$\times10^{-2}$	7.734	7.309	6.627	**3**	3.167

Table 2 Quantitative comparison of the reconstruction results on the FaceWarehouse [25] dataset. In each column, for each metric, best results are in bold

Facsimile™ Dataset-Training							
			FeaStNet	Neural 3DMM	Spiral Net++	Mesh Autoenc.	LSA-3DMM
Euclidean	L_1 *norm*	$\times10^{-3}$	11.552	7.202	6.267	2.541	7.275
	L_2 *norm*	$\times10^{-5}$	153.49	30.02	23.225	4.403	29.636
	FMPD	$\times10^{-2}$	100	100	100	47.637	28.462
	DAME	$\times10^{-2}$	49.7	28.935	44.265	5.266	2.943
Euclidean Std	L_1 *norm*	$\times10^{-3}$	**1.92**	**1.578**	**1.212**	**1.116**	**1.026**
	L_2 *norm*	$\times10^{-5}$	**3.193**	**1.902**	**1.041**	**1.091**	**0.896**
	FMPD	$\times10^{-2}$	45.656	38.517	50.578	32.636	16.847
	DAME	$\times10^{-2}$	6.81	4.48	6.062	3.23	1.767
DR	L_1 *norm*	$\times10^{-3}$	6.867	7.762	4.088	2.57	4.472
	L_2 *norm*	$\times10^{-5}$	25.609	31.06	9.575	3.326	10.881
	FMPD	$\times10^{-2}$	**1.580**	**1.549**	**0.971**	3.833	2.073
	DAME	$\times10^{-2}$	**0.904**	**0.827**	**0.866**	1.171	**0.810**
DR Norm	L_1 *norm*	$\times10^{-3}$	9.131	7.616	6.541	2.208	3.489
	L_2 *norm*	$\times10^{-5}$	41.392	27.944	20.778	2.467	6.562
	FMPD	$\times10^{-2}$	10.292	10.75	10.813	**3.668**	**1.101**
	DAME	$\times10^{-2}$	2.277	2.233	2.161	**1.137**	1.097

Facsimile™ Dataset-Test							
			FeaStNet	Neural 3DMM	Spiral Net++	Mesh Autoenc.	LSA-3DMM
Euclidean	L_1 *norm*	$\times10^{-3}$	12.177	8.043	6.996	**5.332**	8.295
	L_2 *norm*	$\times10^{-5}$	161.514	36.895	27.695	15.503	37.719
	FMPD	$\times10^{-2}$	100	100	100	48.566	26.868
	DAME	$\times10^{-2}$	49.314	29.068	44.252	5.499	2.687
Euclidean Std	L_1 *norm*	$\times10^{-3}$	8.169	**4.017**	**4.694**	5.337	**4.472**
	L_2 *norm*	$\times10^{-5}$	37.138	**9.072**	**12.218**	**15.337**	**11.919**
	FMPD	$\times10^{-2}$	43.419	37.621	47.91	24.573	17.532
	DAME	$\times10^{-2}$	6.299	4.458	5.633	2.528	1.89
DR	L_1 *norm*	$\times10^{-3}$	**7.806**	10.569	11.090	10.023	6.764
	L_2 *norm*	$\times10^{-5}$	**32.325**	58.098	64.124	50.386	23.561
	FMPD	$\times10^{-2}$	**1.577**	**1.629**	**1.123**	**3.096**	**2.189**
	DAME	$\times10^{-2}$	**0.867**	**0.81**	**0.857**	**1.19**	**0.8**
DR Norm	L_1 *norm*	$\times10^{-3}$	10.171	8.474	7.890	7.424	6.787
	L_2 *norm*	$\times10^{-5}$	52.726	36.706	31.728	28.66	24.309
	FMPD	$\times10^{-2}$	10.272	10.536	10.173	3.208	11.962
	DAME	$\times10^{-2}$	2.24	2.209	2.084	1.216	1.132

datasets. While normalisation demonstrated some benefits in Mesh Autoencoder[5] and LSA-3DMM[6], it also led to increased spatial and perceptual errors in the remaining cases. In FeaStNet[2], Neural 3DMM[3],

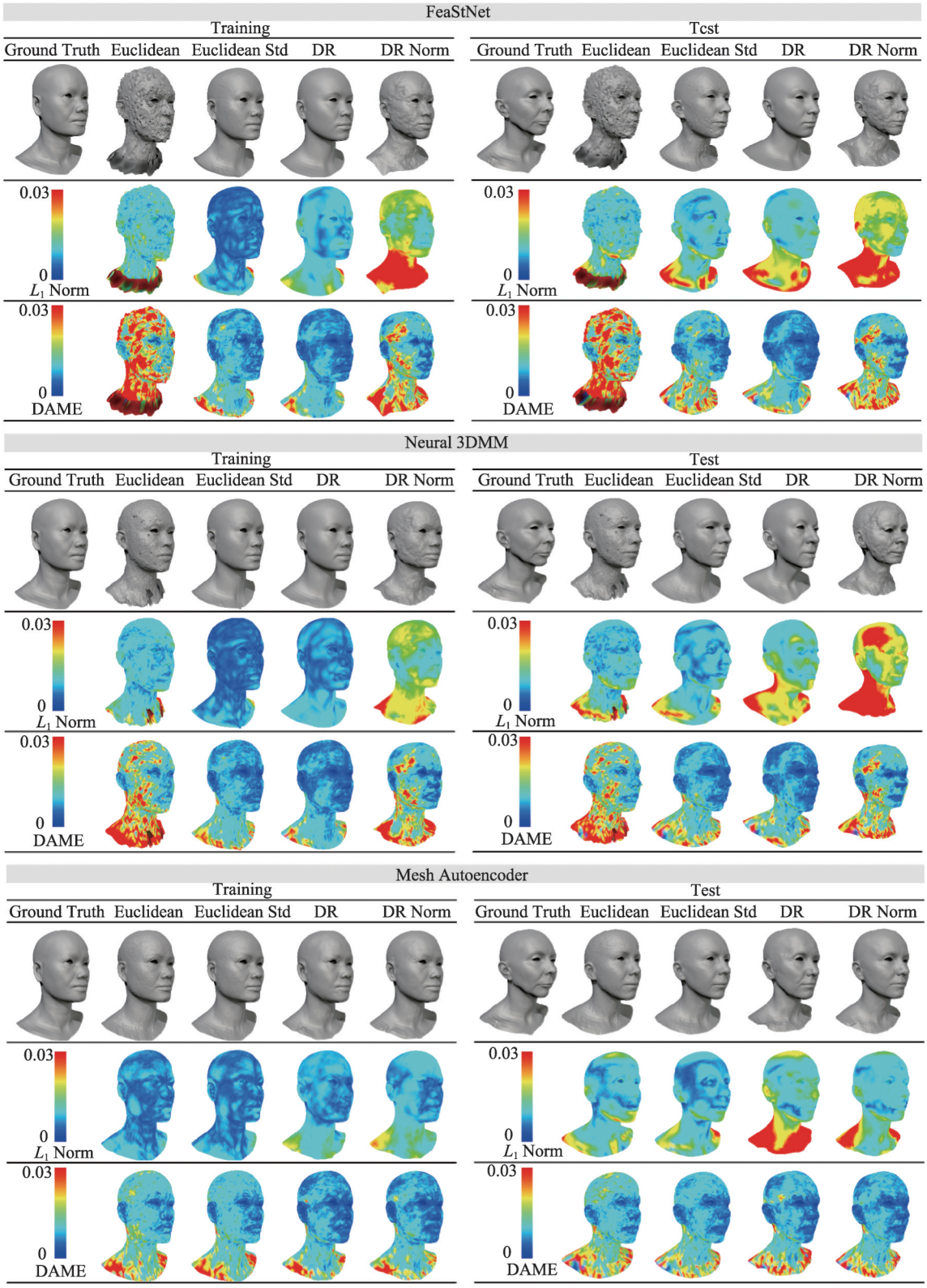

Figure 1 Qualitative evaluation on Facsimile™[26] training and test meshes output from the FeaStNet[2] (top), Neural 3DMM[3] (middle) and MeshAutoencoder[5] (bottom) using 4 feature representations (in columns). Per-vertex L_1 norm error and per-vertex DAME are rendered as colour.

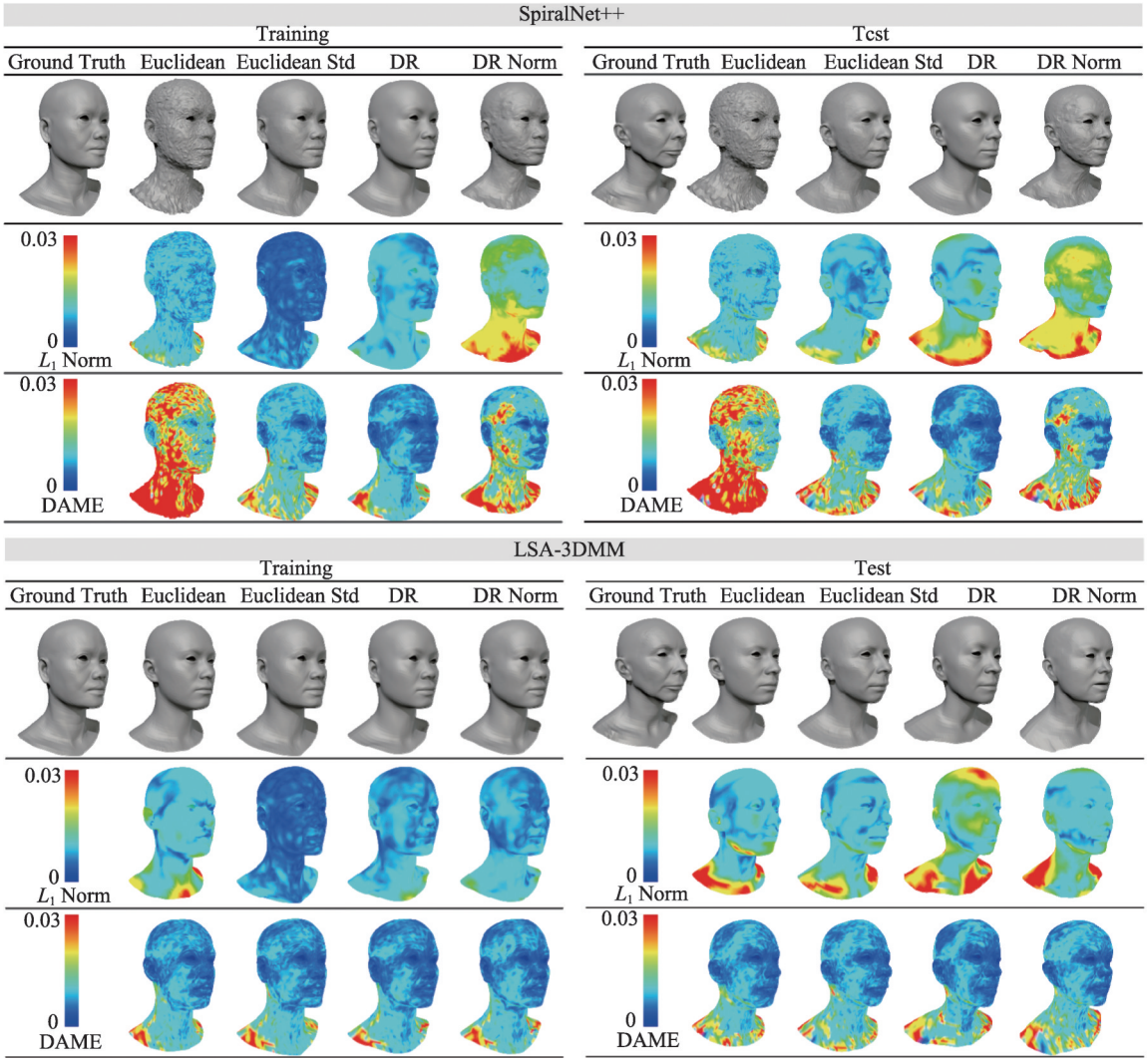

Figure 2 Qualitative evaluation on Facsimile™[26] training and test meshes output from the SpiralNet++[4] (top) and LSA-3DMM[6] (bottom) using 4 feature representations (in columns). Per-vertex L1 norm error and per-vertex DAME are rendered as colour.

SpiralNet++[4] models, normalisation tends to increase L_1 and L_2 norm errors across datasets. Similarly, perceptual metrics like DAME and FMPD are adversely affected by normalisation, with few exceptions. These observations suggest that current practice of normalisation of DR features should be reconsidered, as its positive or negative influence depends on deep 3DMM architecture and the dataset.

4.3 Optimal combinations

The combination of a model and feature representation, which achieves the lowest L-norm error, tends to have higher perceptual error than many other combinations. Analogically, the combination that achieves the lowest perceptual error tends to have significantly higher L-norm error than other combinations. This phenomenon partly stems from the properties of feature representations, as well as each model's varying ability to learn different representations of data, either global or differential coordinates. As demonstrated in Figure 3, the choice of models and mesh feature representations depends on different user requirements of maximising either perceptual or spatial fidelity, or balancing these two objectives.

On Facsimile training set, SpiralNet++ with $F_{\text{Eucl. Std.}}$ feature has the lowest L_1 norm error of 1.353, despite its high DAME. Neural3DMM with DR feature achieve the lowest DAME of 2.092, albeit high spatial error.

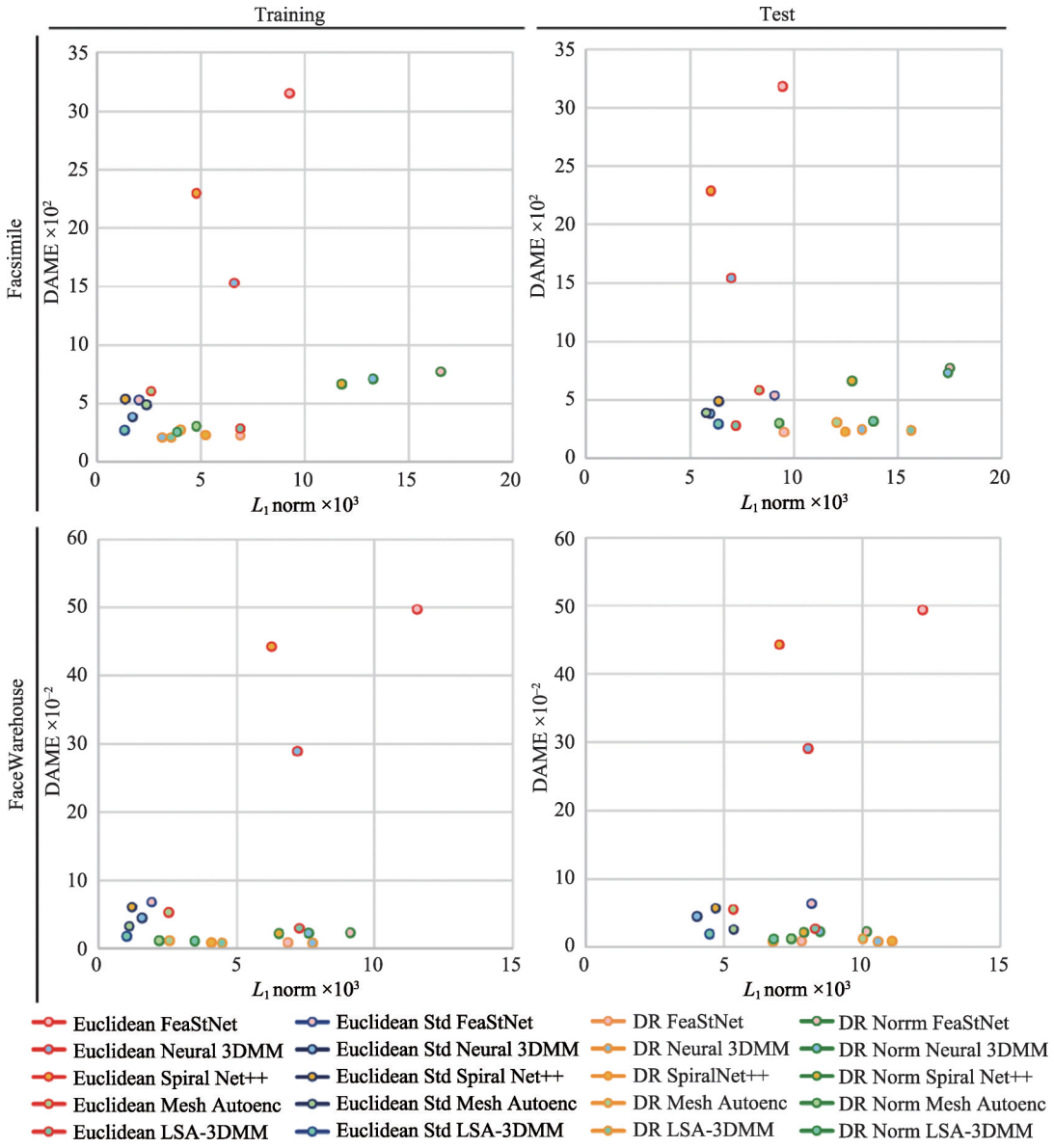

Figure 3 Comparative results plot against the L_1 and DAME metrics. This visualisation allows to simultaneously assess the models' performance in terms of spatial and perceptual fidelity.

LSA-3DMM with $F_{\text{Eucl. Std.}}$ feature (slightly in favour of spatial quality) and Neural3DMM with DR feature (slightly in favour of perceptual quality) minimise both of the objectives simultaneously.

On the Facsimile test set, Mesh Autoencoder with $F_{\text{Eucl. Std.}}$ representation results in the lowest L_1 norm error of 6.38. FeaStNet with the DR feature minimises DAME, while FMPD is minimised by the Mesh Autoencoder with DR feature. Both of these combinations perform poorly on spatial metrics. Considering both perceptual and geometric quality objectives, Mesh Autoencoder with $F_{\text{Eucl. Std.}}$ and LSA-3DMM with $F_{\text{Eucl. Std.}}$ balance both objectives.

Based on the FaceWarehouse training set results, LSA-3DMM with $F_{\text{Eucl. Std.}}$ feature minimises the L_1 and L_2 norm error. However, this combination results in high DAME. On the other hand, LSA-3DMM with DR feature minimises DAME for the price of spatial fidelity. When both objectives are considered, Mesh Autoencoder with DR Norm, LSA-3DMM with $F_{\text{Eucl. Std.}}$ and LSA-3DMM with DR feature provide the best overall balance.

On the FaceWarehouse test set, Neural3DMM with $F_{\text{Eucl. Std.}}$ has the lowest L_1 and L_2 norm error. Importantly, this combination also has one of the highest DAME and FMPD errors. LSA-3DMM with DR feature has the lowest DAME of all combinations. Although 5 other combinations outperform it on L_1 norm, this combination has a good overall performance.

5 Conclusion

Forty models from comparative experiments were evaluated from a spatial fidelity perspective using L_1 and L_2 norm metrics and from a perceptual quality perspective using DAME and FMPD metrics. It was demonstrated that using global, Euclidean coordinates-based feature representations outperforms differential coordinates-based feature representations in spatial fidelity, whilst differential coordinates-based feature representations achieve better results on perceptual DAME and FMPD metrics.

Standardisation of a feature in Euclidean coordinates representation improves the spatial and perceptual fidelity of meshes output by deep 3D morphable models. There are a few exceptions to this observation. Furthermore, the findings of this work prove that the common practice of normalisation of the deformation representation is not suitable in FeaStNet[2], Neural 3DMM[3] and SpiralNet++[4]. At the same time, it can be beneficial on some datasets in Mesh Autoencoder[5] and LSA-3DMM[6].

The proposed use of standardised Euclidean coordinates representation improved spatial and perceptual fidelity of Mesh Autoencoder[5] method, which originally used the Euclidean coordinates feature. Additionally, the proposed use of the DR feature improved the perceptual quality of all the compared methods. Among the proposed combinations, LSA-3DMM[6] and Mesh Autoencoder[5] achieved the best perceptual quality and spatial fidelity when these two objectives were considered simultaneously.

Declaration of competing interest

Robert Kosk, Shaojun Bian, Willem Kokke, and Greg Maguire, are employees of Humain Ltd. Richard Southern is an employee of The Foundry Visionmongers Ltd. The authors declare no conflicts of interest.

CRediT authorship contributions statement

Robert Kosk,Conceptualisation, Methodology, Software, Validation, Visualisation, Writing–original draft preparation, Writing–review and editing; **Richard Southern:**Writing–review and editing, Supervision; **Lihua You:**Writing–review and editing, Supervision; **Shaojun Bian:** Writing – review and editing; **Willem Kokke:** Writing---review and editing, Supervision, Resources and data curation; **Greg Maguire:** Writing–review and editing, Supervision.

Reference

1 Sorkine O, Cohen-Or D, Toledo S. High-pass quantization for mesh encoding. DBLP, 2003

2 Verma N, Boyer E, Verbeek J. FeaStNet: feature-steered graph convolutions for 3D shape analysis. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, IEEE, 2018, 2598–2606
DOI: 10.1109/cvpr.2018.00275

3 Bouritsas G, Bokhnyak S, Ploumpis S, Zafeiriou S, Bronstein M. Neural 3D morphable models: spiral convolutional networks for 3D shape representation learning and generation. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South), IEEE, 2019, 7212–7221
DOI: 10.1109/iccv.2019.00731

4 Gong S W, Chen L, Bronstein M, Zafeiriou S. SpiralNet: a fast and highly efficient mesh convolution operator. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South), IEEE, 2019, 4141–4148
DOI: 10.1109/iccvw.2019.00509

5 Zhou Y, Wu C, Li Z, Cao C,Sheikhet Y. Fully convolutional mesh autoencoder using efficient spatially varying kernels. 2020
DOI:10.48550/arXiv.2006.04325

6 Gao Z P, Yan J C, Zhai G T, Zhang J Y, Yang Y Y, Yang X K. Learning local neighboring structure for robust 3D shape representation. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 1397–1405
DOI: 10.1609/aaai.v35i2.16229

7 Va´sa L, Rus J. Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. Eurographics Symposium on Geometry Processing, 2012, 31(5):1715–1724
8 Wang K, Torkhani F, Montanvert A. A fast roughness-based approach to the assessment of 3D mesh visual quality. Computers & Graphics, 2012, 36(7): 808–818
DOI: 10.1016/j.cag.2012.06.004
9 Xiao Y P, Lai Y K, Zhang F L, Li C P, Gao L. A survey on deep geometry learning: from a representation perspective. Computational Visual Media, 2020, 6(2): 113–133
DOI: 10.1007/s41095-020-0174-8
10 Chen Z X, Kim T K. Learning feature aggregation for deep 3D morphable models. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA, IEEE, 2021, 13159–13168
DOI: 10.1109/cvpr46437.2021.01296
11 Ranjan A, Bolkart T, Sanyal S, Black M J. Generating 3D faces using convolutional mesh autoencoders. In: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, 725–741
DOI: 10.1007/978-3-030-01219-9_43
12 Cheng S Y, Bronstein M, Zhou Y X, Kotsia I, Pantic M, Zafeiriou S. MeshGAN: non-linear 3D morphable models of faces. 2019: 1903.10384. https://arxiv.org/abs/1903.10384v1
13 Hanocka R, Fleishman S, Hertz A, Fish N, Giryes R, Cohen D. MeshCNN: a network with an edge. ACM Transactions on Graphics (TOG), 2019, 38(4):1–12
14 Gao L, Lai Y K, Liang D, Chen S Y, Xia S H. Efficient and flexible deformation representation for data-driven surface modeling. ACM Transactions on Graphics, 2016, 35(5): 1–17
DOI: 10.1145/2908736
15 Gao L, Lai Y K, Yang J, Zhang L X, Xia S H, Kobbelt L. Sparse data driven mesh deformation. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(3): 2085–2100
DOI: 10.1109/tvcg.2019.2941200
16 Wu Q Y, Zhang J Y, Lai Y K, Zheng J M, Cai J F. Alive caricature from 2D to 3D. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2018
17 Jiang Z H, Wu Q Y, Chen K Y, Zhang J Y. Disentangled representation learning for 3D face shape. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2019
18 Zheng X Q, Jiang B Y, Zhang J Y. Deformation representation based convolutional mesh autoencoder for 3D hand generation. Neurocomputing, 2020, 444(4): 356–365
DOI: 10.1016/j.neucom.2020.01.122
19 Wu Z H, Pan S R, Chen F W, Long G D, Zhang C Q, Yu P S. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4–24
DOI: 10.1109/tnnls.2020.2978386
20 Yuan Y J, Lai Y K, Yang J, Duan Q, Fu H B, Gao L. Mesh variational autoencoders with edge contraction pooling. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA, IEEE, 2020, 1105–1112
DOI: 10.1109/cvprw50498.2020.00145
21 Sorkine O, Alexa M. As-Rigid-As-Possible surface modeling. In: Proceedings of the Fifth Eurographics Symposium on Geometry Processing. 2007
DOI:10.1145/1281991.1282006
22 Kingma D P, Ba J L. Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings, 2015
23 Corsini M, Larabi M C, Lavoue G, Petřík O, Váša L, Wang K. Perceptual metrics for static and dynamic triangle meshes. Computer Graphics Forum, 2013, 32(1): 101–125
24 Humain Limited. Humain Limited-Research & Development, 2022
25 Cao C, Weng Y L, Zhou S, Tong Y Y, Zhou K. FaceWarehouse: a 3D facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(3): 413–425
DOI: 10.1109/tvcg.2013.249