Geometric Deep Learning for 3D Facial Shape Analysis

--focus on target and bias

Xianjing Liu[1], Bo Li[1], Esther Bron[1], Wiro Niessen[1,2], Eppo Wolvius[1] and Gennady Roshchupkin[1]

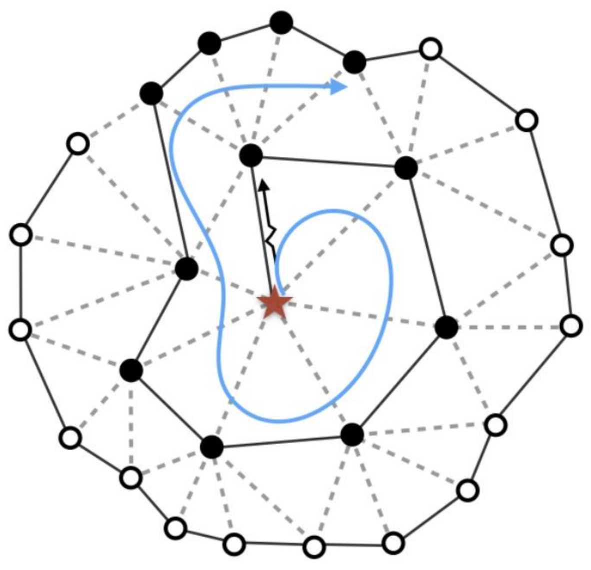
[1] Erasmus MC, Rotterdam, The Netherlands
[2] Delft University of Technology, Delft, The Netherlands

Scan for full paper

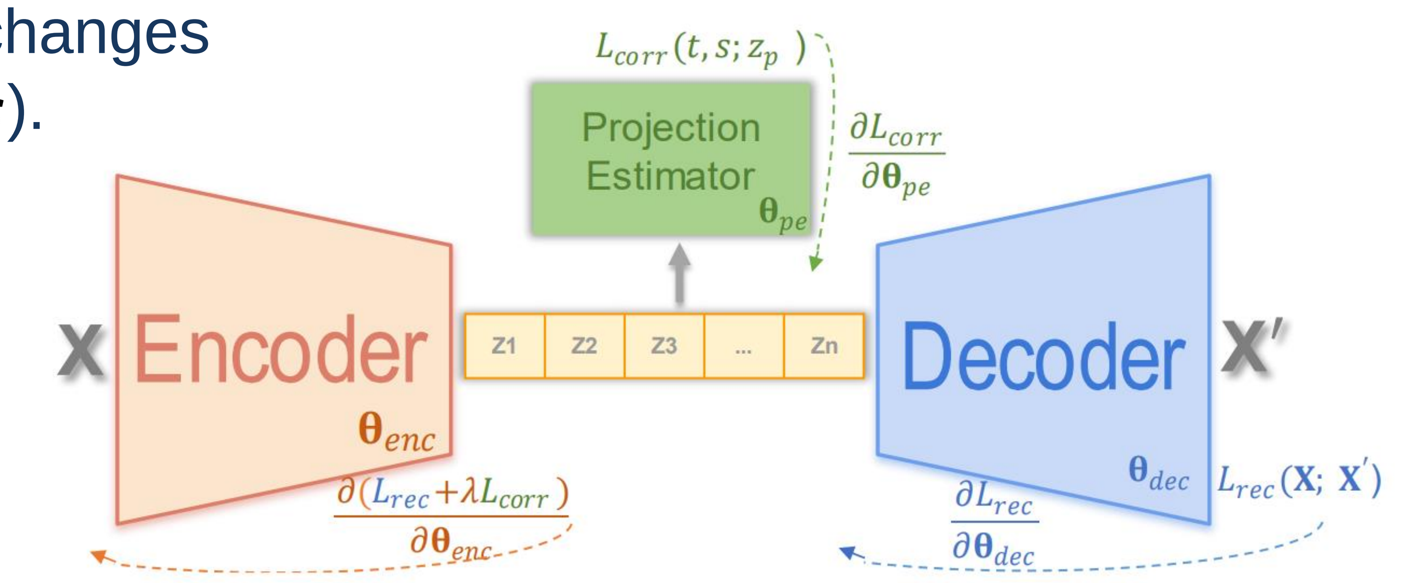
x.liu.1@erasmusmc.nl

Introduction: 3D facial shape analysis was challenged by the high-dimensional complexity of the facial morphology. Recently, geometric deep learning techniques have become state-of-the-art methods for dimensionality of 3D shapes. We thus apply these techniques to facial shape analysis, exploring facial changes related to a target attribute (t) but not being confounded by other biases (s).

Method: We applied the convolution strategy from SpiralNet++ [1], and proposed the projection-wise disentangling approach, which couples a vector direction with the target while being independent to biases (via L_{corr}, Fig. 1b). Here $Corr(.,.)$ is the Pearson correlation coefficient which ranges between $[-1, 1]$



(a)



$$L_{corr}(\mathbf{t}, \mathbf{s}; \mathbf{z_p}) = |Corr(\mathbf{z_p}, \mathbf{s})| - \eta |Corr(\mathbf{z_p}, \mathbf{t})|, \quad L_{rec}(\mathbf{X}; \mathbf{X}') = ||\mathbf{X} - \mathbf{X}'||$$

$$\boldsymbol{\theta}_{enc}, \boldsymbol{\theta}_{dec}, \boldsymbol{\theta}_{pe} \leftarrow argmin \ L_{joint} = L_{rec} + \lambda L_{corr}$$

(b)

Fig. 1: a) The convolution operation on a spiral; b) The proposed projection-wise disentanglement strategy;

A vector direction in the latent space can be represented by $\mathbf{P} = [p_1, p_2, \ldots, p_n]$ as a linear combination of the basis vectors (Fig. 2). Let $\mathbf{D}^{\mathbf{d} \times \mathbf{n}} = [\mathbf{z_1^d}, \mathbf{z_2^d}, \ldots, \mathbf{z_n^d}]$ (d datapoints) be the latent representations of the input data. For each datapoint, $\mathbf{d} = [z_1, z_2, \ldots, z_n]$ is its latent representations, and $z_p = \mathbf{d} * \mathbf{P} / ||\mathbf{P}|| = (p_1 z_1 + p_2 z_2 + \cdots + p_n z_n) / ||\mathbf{P}||$ can be viewed as a scalar projection of $\mathbf{d}$ onto vector $\mathbf{P}$ (Fig. 2). When sampling along $\mathbf{P}$, z_p changes and this change is correlated to t while independent to s.

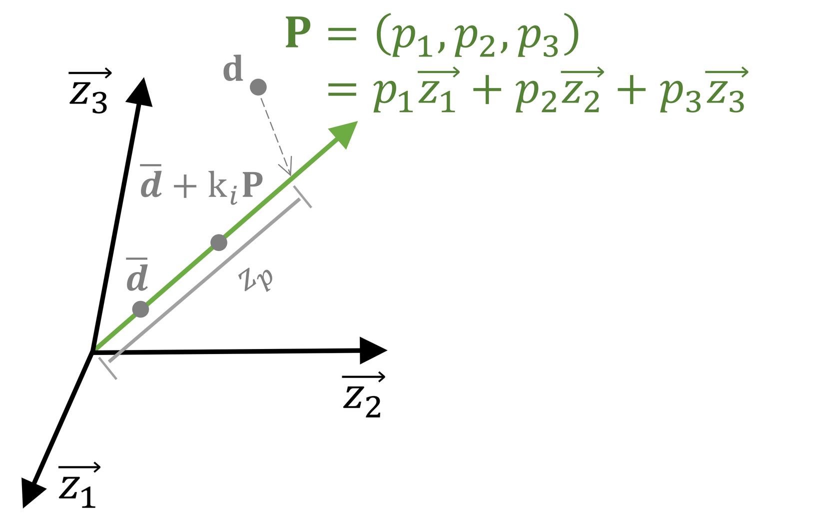


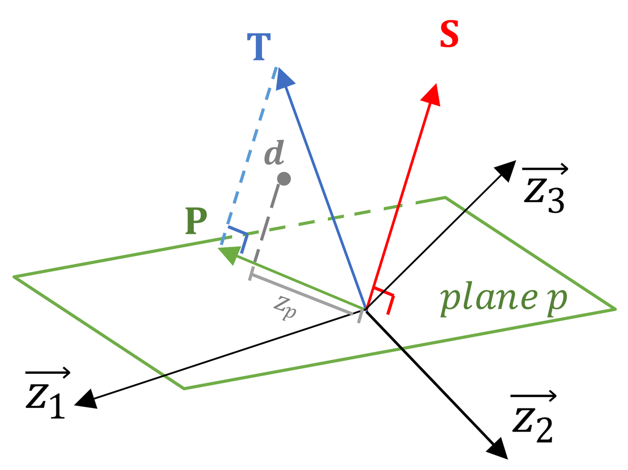
Fig. 2: $\mathbf{P} = [p_1, p_2, \ldots p_n]$ represents a vector in latent space ($n = 3$). $\mathbf{d}$ is the latent representation of a datapoint.

Results:.

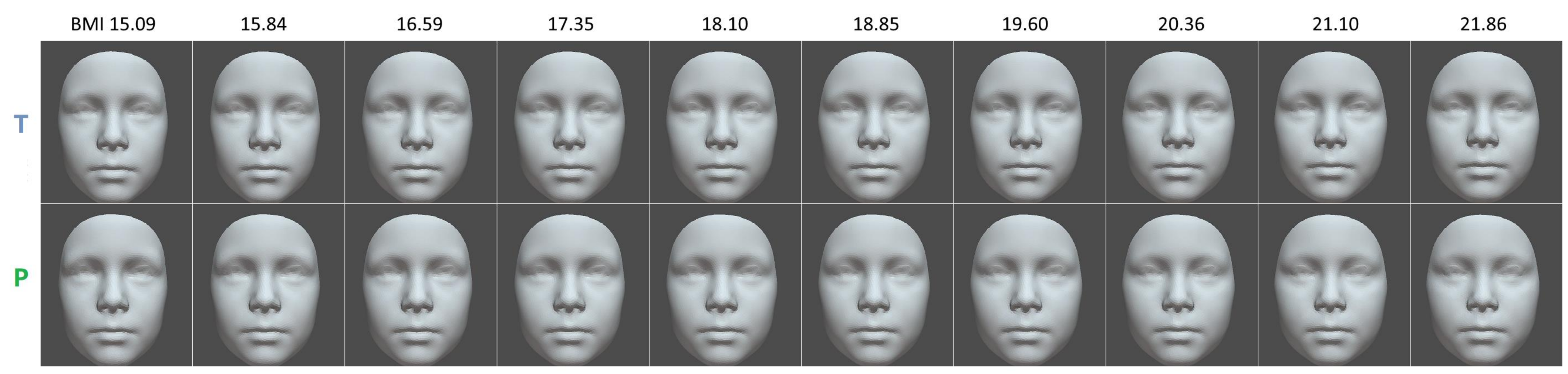
Reconstruction:

Fig. 3: Left: input face; Middle: reconstructed face; Right: reconstruction error.

Disentanglement (Fig 4): We investigated the relation between gender, height and BMI. When analyzing one of the attributes, the other two were considered as biases.
Vector T: a vector direction capture facial changes related to a target attribute, without bias mitigation.
Vector S: a vector direction capture facial changes related to a bias.
Vector P: a vector direction searching by L_{corr} , capturing facial changes related to a target attribute, but being independent to the bias. P can be viewed as the projection of T onto the plane orthogonal with S



(a)



(b)

Fig. 4: a) illustration of vector T, S and P; b) Disentanglement for BMI. 1st row: disentanglement along T; 2nd row: disentanglement along P.

Features visualization:
Fig. 5 provides visualizations of the facial features by computing difference heatmaps between the first and the last frame in Fig 4b. The vector T captured all features related to the target, whereas P captured the features that only related to the target and independent to biases. For gender, the result of P is similar to that of T because gender is nearly unbiased by height and BMI in the dataset.

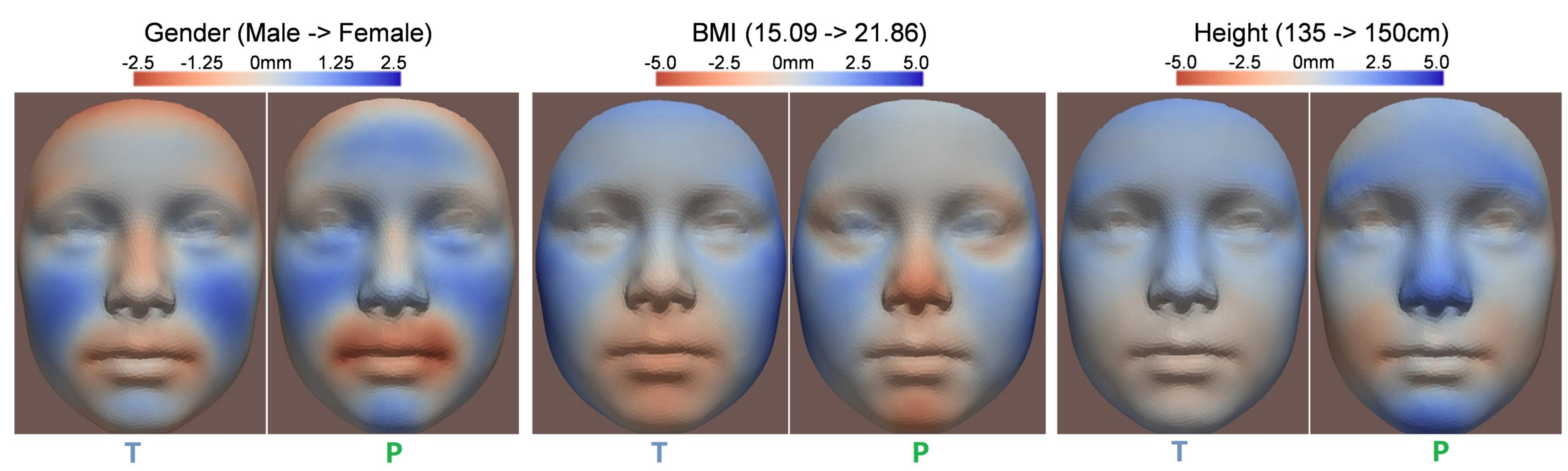


Fig. 5: Visualizations of facial features. Red and blue areas refer to inner and outer facial changes towards the geometric center of the 3D face, respectively.

However, since BMI and height are positively correlated in our dataset, the similar heatmaps for the BMI and height disentanglement in T indicate that it captured common facial features for the two tasks, and thus failed to disentangle the confounding bias. In contrast, P learned a target-specific representation, showing a strong correlation to the target attribute and without being confounded by other (bias) attributes.

[1] Gong S., Chen L., Bronstein M. and Zafeiriou S.: SpiralNet++: A Fast and Highly Efficient Mesh Convolution Operator. In: ICCVW (2019)