# Parametric Design for Human Body Modeling by Wireframe-Assisted Deep Learning

Jida Huang[a], Tsz-Ho Kwok[b,*], Chi Zhou[a]

[a] *Industrial and Systems Engineering, University at Buffalo, SUNY, Buffalo, NY 14260, United States*
[b] *Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, QC H3G 1M8, Canada*

**Abstract**

Statistical learning of human body shape can be used for reconstructing or estimating body shapes from incomplete data, semantic parametric design, modifying images or videos, or simulation. A digital human body is normally represented in a high-dimensional space, and the number of vertices in a mesh is far larger than the number of human bodies in publicly available databases, which results in a model learned by Principle Component Analysis (PCA) can hardly reflect the true variety in human body shapes. While deep learning have been most successful on data with an underlying Euclidean or grid-like structure, the geometric nature of human body is non-Euclidean, it will be very challenging to perform deep learning techniques directly on such non-Euclidean domain. This paper presents a deep neural network (DNN) based hierarchical method for statistical learning of human body by using feature wireframe as one of the layers to separate the whole problem into smaller and more solvable sub-problems. The feature wireframe is a collection of feature curves which are semantically defined on the mesh of human body, and it is consistent to all human bodies. A set of patches can then be generated by clustering the whole mesh surface to separated ones that interpolate the feature wireframe. Since the surface is separated into patches, PCA only needs to be conducted on each patch but not on the whole surface. The spatial relationship between the semantic parameter, the wireframe and the patches are learned by DNN and linear regression respectively. An application of semantic parametric design is used to demonstrate the capability of the method, where the semantic parameters are linked to the feature wireframe instead of the mesh directly. Under this hierarchy, the feature wireframe acts like an agent between semantic parameters and the mesh, and also contains semantic meaning by itself. The proposed method of learning human body statistically with the help of feature wireframe is scalable and has a better quality.

*Keywords:* Feature; wireframe; human body; deep learning; parametric design

## 1. Introduction

Statistical learning of human body is a fundamental problem in many areas and applications such as biometric analysis, generative design, and products customization. Traditional way of acquisition of human body model is using 3D scanner to attain scans of human body and then register the scans together to generate a consistent model. However, this process is time consuming and requires expensive scanning device and reconstruction platform. Therefore, many studies have been done to learn the variations of body shapes [1, 2] and correlate them to their semantic parameters [3] – a parametric design method for human body modeling. With the correlation, the shape of a human body can be generated from a set of given parameters such as body height, chest-girth and waist-girth.

A digital human body is normally represented in a high-dimensional space, e.g., a mesh with $n$ vertices each of which is defined by its three-dimensional (3D) coordinate $(x, y, z)$. Prior to statistical learning, a common practice is to conduct Principal Component Analysis (PCA) on the data to transform the data to a lower-dimensional space such that the complexity is reduced while preserving sufficient variance [3]. PCA can be done by eigenvalue decomposition of a data covariance matrix. However, the number of human bodies ($m$) in publicly available databases is usually far less than the number of vertices, i.e., $m \ll 3n$, and the PCA result only has $m-1$ eigenvectors with non-zero eigenvalues. Therefore, with limited samples, the learning directly on the whole human body can hardly capture the true variety in body shapes. Furthermore, if both $n$ and $m$ are large, it will be very challenging to perform eigenvalue decomposition on such a large matrix ($3n \times 3n$ or $m \times m$).

On the other hand, as a new area in machine learning research, deep learning is becoming an important tool in many applications [4]. Deep learning is successful when dealing with Euclidean data structure with large sample size such as speech and images [5]. However, it is known to suffer from overfitting when the data have high dimension and low sample size [6], such as the database of human body model. Another issue is the long computational time due to the number of training parameters increases dramatically with the dimension of data and the number of

*Corresponding author
*Email address:* tszho.kwok@concordia.ca (Tsz-Ho Kwok)

layers. Despite these challenges, the superior learning capability of deep learning makes it worth to be applied in statistical learning of human body.

To solve the practical issues arisen in the usage of PCA and deep learning, the dimension of data must be reduced. However, many mesh vertices are needed to represent the shape of a human body confidently, and simply downsampling the mesh will lose much shape information. We observe that although the vertices size of a mesh is huge, the vertices are not totally independent to each other. Actually, human bodies are described by a set of feature curves in many garment industries. Based on this observation, we hypothesize that if the learning process can be taken on the feature curves, then the dimension of dataset would be reduced, so that the training through PCA and deep learning will be more efficient and the trained model will be more accurate. To test this hypothesis, this paper develops a wireframe-assisted deep learning method. The feature wireframe is a collection of feature curves which are semantically located on the chest, bust, hip, waist, etc. Instead of analyzing all the mesh vertices directly, we use feature wireframe to separate the whole surface into patches. In this way, PCA does not need to be conducted on the whole surface but just on each patch. The spatial relationship between the patches are preserved through the wireframe. Furthermore, deep learning has a much better performance in this much lower dimensional space. The technical contributions of the paper are summarized as follows:

1. To the best of our knowledge, this is the first time applying feature wireframe as a defined internal layer to guide the statistical learning of human modeling.
2. The feature wireframe divides a whole human model into small patches. Each of the patches has much less shape variation, and PCA can achieve high rate of dimension reduction in the patches, up to 99.72%.
3. With the lower dimensional space, deep learning can be applied to correlate the feature wireframe and semantic parameters, enabling parametric design, in which it is the first time deep learning is applied.

We compare our method with the previous human body learning method [3] which applied PCA on the whole model and used linear regression to learn the correlation. The same human body database (77 subjects) is used, so that the results can be directly compared. To emphasize the usefulness of feature wireframe, we also compare our method with which the DNN replaced by linear regression. Experimental results show that the proposed method with linear regression can already outperform the previous work in terms of reliability, accuracy, and robustness. Together with deep learning, the performance is further enhanced greatly. The proposed method for statistically learning human body with the help of feature wireframe is effective and scalable.

The rest of the paper is organized as follows. Section 2 will briefly review the related works. The technical detail of the methodology for human modeling will be discussed in Section 3, and it is followed by the experimental results in Section4. Section 5 will conclude the paper.

## 2. Related Work

The state of human body learning mainly includes parametric design and shape or pose modeling. This section gives a brief review of these aspects.

### *2.1. Parametric design*

To reduce the complexity of human model and achieve dimension reduction, many studies focus on the parametric design of human body and related product. For example, a feature based parameterization approach of human bodies from the unorganized point cloud is presented in [7]. What's more, Wang et al. [8] investigated a feature-based human model for digital apparel design. Kwok et al. [9] proposed an optimization algorithm for the complexes and the shape of common base domains in cross parameterization for reducing the distortion of the bijective mapping.

Another aspect of research is concentrated on the parametric design of human-related products. For example, an approach for computing planar patterns for compression garments is proposed in [10], while Hasler et al. [11] studied how to estimate the detailed 3D body shape of dressed humans. The approach is based on the space of human shapes learned from a large database of registered body scans. Li et al. [12] proposed a method for fitting a given 3D garment model onto human models of various body shapes and poses. Pons-Moll et al. [13] proposed a method for seamless 4D clothing capture and retargeting. Baek et al. [14] developed parametric human body shape modeling framework for human-centered product design. Au et al. [15] addressed the definition, development and application of garment features with an associative feature approach. Chu et al. [16] proposed a computational framework for personalized design of the eyeglasses frame based on parametric face modeling. Similarly, Huang et al. [17] presented a virtual try-on system based on augmented reality for design personalization of facial accessory products.

### *2.2. Human shape/pose modeling*

Many researchers have investigated the human shape and pose estimation. For example, Anguelov et al. [1] proposed a data-driven method, Shape Completion and Animation for People (SCAPE), for building a human shape model that spans variation in both subject shape and pose. While Hasler et al. [18] presented a learning method for estimating a rigid skeleton for shape and pose. A DeepPose method is proposed for human pose estimation based on Deep Neural Networks in [19]. Shotton et al. [20] proposed a method to predict human pose and the 3D positions of body joints from a single depth image. A biomechanical model of the human body to synthesize realistic swimming

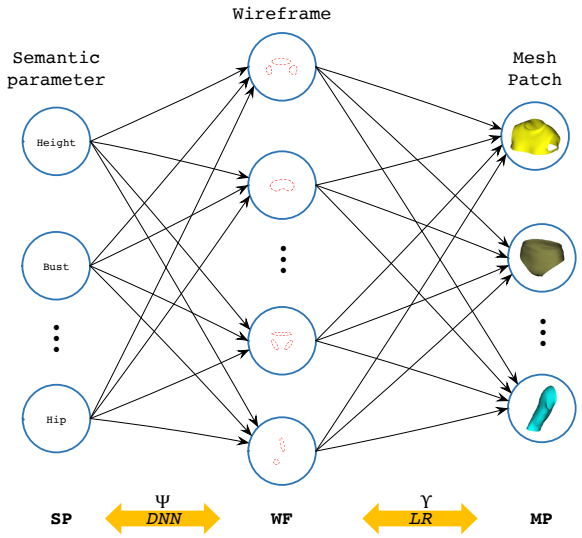



Figure 1: Sketch idea of our parametric design method

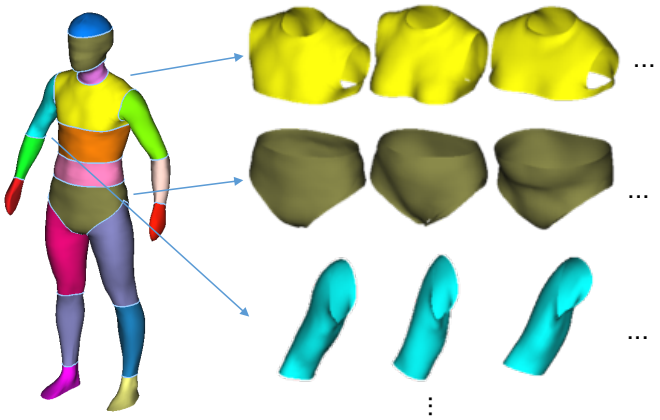


Figure 2: Wireframe definition and patch extraction

animation was proposed in [21]. A retargeting method for human motion to arbitrary 3D mesh models with as little user interaction as possible is presented in [22].

However, most of these works focused on the human pose estimation, few works addressed the human body modeling. In order to generate and animate realistic humans, Hasler et al. [2] proposed a learning based approach which could model muscle deformations as a function of pose. Streuber et al. [23] proposed a method to learn a model of how 3D shape and linguistic descriptions of shape are related. Allen et al. [24] proposed a parametric freeform mesh design to reconstruct human model from range scans. While a system of modeling the effects of muscle, fat, and bone growth with a given input 3D anatomy template was proposed in [25]. Seo et al. [26] proposed an example-based approach which utilizes a modeling synthesizer that learns from preprocessed examples to interpolate new body geometry.

## 3. Methodology

To test our hypothesis in enabling PCA and deep learning on human body data with high dimension and small sample size, this paper aims to develop a wireframe-assisted learning method with the application of parametric design for human body modeling. Instead of correlating semantic parameters and human model directly, we use feature wireframe as an intermediate layer as demonstrated in Fig. 1. The feature wireframe is defined by a set of feature curves based on the anthropometrical rules [7]. This method is widely used as semantics measurements in garment industries. Although it cannot guarantee to capture all the features of a human body, the wireframe contains enough industrial meaning. Specifically, some of the selected feature curves are the girths of head, neck, neck-base, elbow, wrist, bust, under-bust, thigh, knee, and ankle, as well as armhole circumstance. The wireframe is represented and displayed as polylines, but only the points (1197 vertices in the database) are used in the training. Therefore, it is a topology graph that is defined on the mesh of human body, and it is consistent to all human bodies. A set of feature patches can then be generated by clustering the whole mesh surface to separated ones that interpolate the feature wireframe, which is depicted in Fig. 2. Since all the human models in database have the consistent mesh connectivity, the patches can be generated consistently as well for all of the models. Therefore, there are three major layers in the framework for parametric design: semantic parameters ($SP$), feature wireframe ($WF$), and mesh patches ($MP$). Each has the same size of training samples equal to the number of models in the database, and they are used to train the learning model to compute the correlation between the layers. Firstly, we need to find a model that relates the whole wireframe $WF$ to a function of the semantic parameters $SP$ ($\Psi : SP \rightarrow WF$), i.e.,

$$WF \approx \Psi(SP,\ \psi), \tag{1}$$

where $\psi$ is a vector of unknown variables that need to be solved. As the relationship between semantic parameters and the wireframe is highly non-linear, and semantic parameters always inter-affect the wireframe, while deep learning has a good capability to learn such complicated non-linear relationship, hence, a Deep Neural Network (DNN) will be used here for building $\Psi$.

Secondly, assume the mesh is separated by the wireframe into $K$ patches, then the wireframe can be separated into $K$ sets accordingly, each of which is the feature curves that are interpolated by the corresponding patch, i.e., the boundary of the patch. Let the sets of separated wireframes and patches are $wf$ and $mp$, we have

$$wf_k \in WF, \quad mp_k \in MP \quad (k = 1, \ldots, K).$$

Similarly, we need to find a model that relates $mp_k$ to a function of $wf_k$ for each patch ($\Upsilon_k : wf_k \rightarrow mp_k$), i.e.,

$$mp_k \approx \Upsilon_k(wf_k,\ \upsilon_k), \tag{2}$$

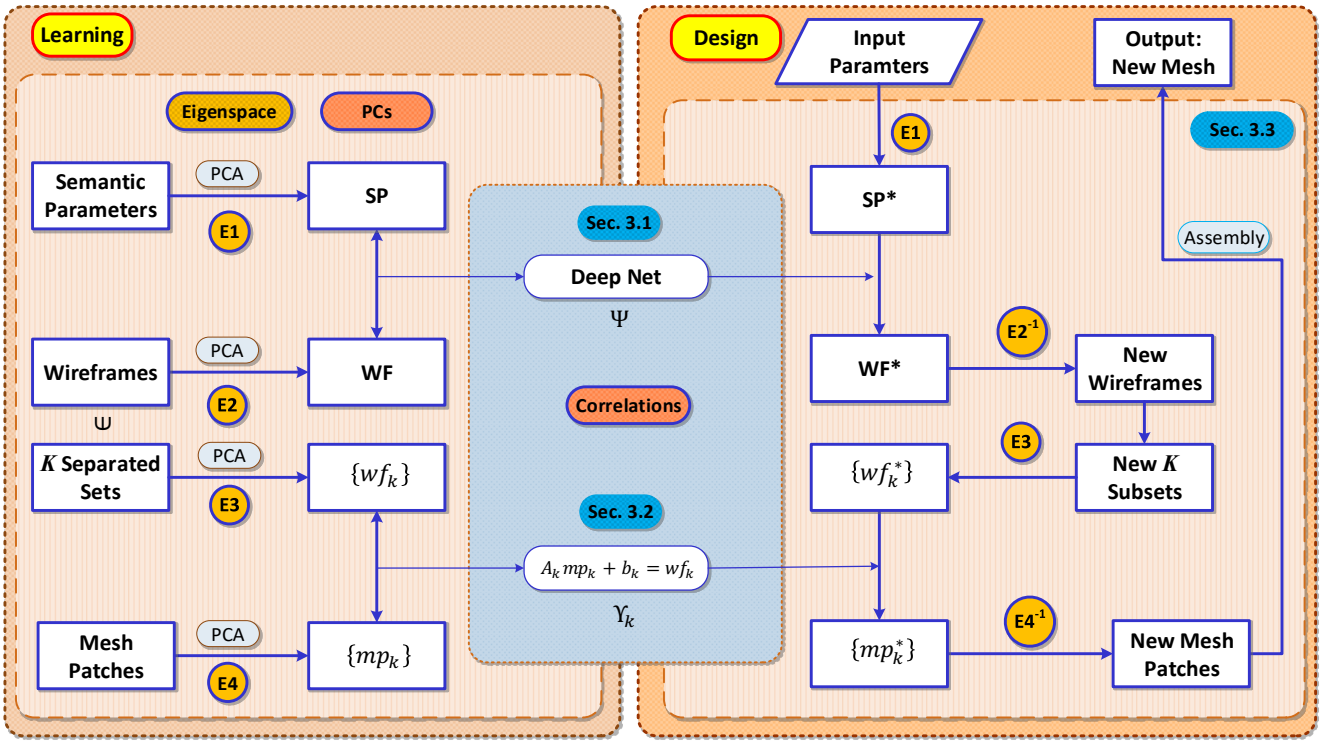



Figure 3: Flowchart of the proposed parametric design method

where $v_k$ is a vector of unknown variables that need to be solved. Thanks to the separation by the wireframe, the models here are close to linear, and we will show that linear regression is good enough to build the function $\Upsilon_k$.

It is worth to note that the original method conducting PCA on all $n$ vertices for $m$ models results in a $3n \times m$ matrix, in which only $m-1$ principal components are meaningful. For $n = 11,000$ and $m = 77$ in the database [3], much information is lost when the matrix is condensed from the dimension of $33,000$ to 76 (434 times) even all principal components are selected. In this new hierarchy, PCA is performed on each patch separately, and each patch has around two hundred vertices (e.g., 198). From 198 to 76 is only 3 times compression, and we will show that just a few principal components are needed to describe the shape confidently because of the low shape variation in each patch. The proposed method can avoid losing feature information of human model and is much faster by solving smaller sub-problems with the help of wireframe.

A more detailed illustration of the proposed methodology is shown in Fig. 3. The proposed method includes two phases: statistical learning phase and parametric design phase. The technical details of learning the correlation functions $\Psi$ through deep neural network and $\Upsilon$ through linear regression are presented in Sections 3.1 and 3.2. They are followed by the parametric design phase in Section 3.3. The detail of PCA is standard and is provided in Appendix A.

### *3.1. Deep Neural Network: Parameter & Wireframe*

The goal of this section is to learn the relationship ($\Psi : SP \rightarrow WF$) between the semantic parameters and the wireframe of human models as shown in the left portion of Figure 1. In the previous work [3], a linear regression is used to correlate the semantic parameters and the mesh. Although the dimension of data is significantly reduced in the wireframe (compared to the mesh), a feature curve may rely on several semantic parameters and their relationship is still non-linear. Furthermore, linear regression fails to separate the influence when the number of semantic parameters $s$ is large [3], and it would over-affect the wireframe generation. Therefore, a non-linear correlation is needed to characterize the relationship between $SP$ and $WF$, and DNN is applied here.

Specifically, the Principal Component Analysis (PCA) is applied to both the semantic parameters and feature wireframe to reduce their data dimension by discarding the last principal components (PCs) with total variance less than 5%, which are unimportant even in the linear space (see Appendix A). Their dimensions are reduced to 11 and 14 on average respectively, which are much smaller than the sample size. Hence, over-fitting in training is less likely to happen here. After that, a DNN is built to learn the relationship between the two sets of PCs: $SP$ and $WF$ (same notation is used for simplicity). The DNN is a feed-forward, artificial neural network with multiple hidden layers between its input and output. The network is called *deep* as many layers are employed [27]. The optimal parameters of the network are obtained by minimizing a

multinomial regression loss function:

$$\ell_{reg}(\psi) = - \sum_{(SP,WF^*(SP))\in\mathcal{T}} \log \Psi_\psi(SP, WF^*(SP)) \quad (3)$$

where $WF^*(SP)$ denotes the ground-truth $WF$ for $SP$, and samples of ground truth are collected: $\mathcal{T} = \{(SP, WF^*(SP))\}$. The network architecture is detailed in the following section.

*3.1.1. Network architecture*

The architecture of deep network consists of different layers:

**Fully connected (FC)** layer is a linearly connected layer to adjust the input and output dimensions. Given a $P$-dimensional input $X^{in} = (x_1^{in}, ..., x_P^{in})$ the fully connected layer produces a $Q$-dimensional output $Y^{out} = (y_1^{out}, ..., y_Q^{out})$ by using a learnable weights $w$,

$$y_q^{out} = \eta(\sum_{p=1}^{P} w_{qp} x_p^{in}); q = 1, ..., Q \quad (4)$$

The output is optionally passed through a non-linear function such as the ReLU [28], $\eta(t) = max\{0, t\}$.

**Dropout**$(\pi)$ layer is a fixed layer to prevent overfitting [29]. The term "dropout" refers to dropping out units (hidden and visible) in a neural network. By dropping a unit out, it means temporarily removing it from the network, along with all its incoming and outgoing connections. The choice of which units to drop is random. It injects binomial noise on each of the computational units of the network. During training stage, an independent and identically distributed binary mask $m_p \sim Binomial(\pi_{drop})$ is generated for each input dimension, each element is 1 with probability $1 - \pi_{drop}$,

$$x_p^{out} = m_p x_p^{in} \quad (5)$$

During testing stage, all possible binary masks have to be integrate over. In practice, the unit is always presented and by using an approximation method: multiplied by the input with the drop probability of the layer:

$$x_p^{out} = \pi_{drop} x_p^{in} \quad (6)$$

Hence, the output at test stage is same as the expected output at training stage.

**Pooling** layer is also a fixed layer which combine the outputs of neuron clusters at one layer into a single neuron in the next layer, it is used to reduce the input dimensions.

**Batch normalization** layer is another fixed layer to reduce the training time of large network [30]. It normalizes each mini-batch during stochastic optimization to have zero mean and unit variance, and then performs a linear transformation of the form:

$$x_p^{out} = \frac{x_p^{in} - \mu}{\sqrt{\sigma^2 + \varepsilon}} \gamma + \beta \quad (7)$$

where $\mu$ and $\sigma^2$ are the mean and the variance of the tanning dataset by using exponential moving average. To avoid numerical errors, a small positive constant $\varepsilon$ is used.

For task of image/speech/computer vision processing which need to extract features from Euclidean data structure, a convolution layer [5] is included in the architecture. In this paper, the input is the PCs of semantic parameters ($SP$) and don't have to extract such features, hence a convolution layer does not needed. The DNN architecture consists different types of layers, acts as a non-linear parametric mapping function $\Psi$ to the output, $WF$, with the set of all learnable parameters $\psi$ of the network.

*3.2. Linear Regression: Wireframe & Mesh Patch*

This section moves the focus to the right part of Fig. 1 and computes the function $\Upsilon$ in Eq.(2) to find the relationship between the feature wireframe and the mesh. The wireframe clusters and separates the human body into a set of mesh patches. Together with the separated patches ($mp_k \in MP$), the wireframe is also separated into different parts ($wf_k \in WF$) correspondingly as the boundary of each patch. The correlation is then built separately on each of the patches, so that the whole problem is divided and solved by a number of sub-problems. Noted that the semantic parameters are correlated with the whole wireframe in the previous section, but a mesh patch is correlated with only part of the wireframe here.

The human models in the database share the same mesh connectivity, and thus the separated patches are consistent among the models too. Let $mp_k^a$ be the $k$-th mesh patch on the $a$-th model, where $a = 1, \ldots, m$, and $wf_k^a$ be the corresponding boundary. Since a mesh patch $k$ among the models (i.e., $mp_k^1, mp_k^2, \ldots, mp_k^m$) can be learned individually now, they are independent of the absolute position on the human body. Therefore, we can first align all the patches from different human models, and then learn the variance of the patches in their own coordinate system, which is very similar to the procedure of Generalized Procrustes analysis (GPA) without the estimation of scaling factor. In this way, the real geometry and shape variances can be learned within a patch rather than the space variance, which was not possible when the training is done on the whole mesh. In short, the alignment is done by finding a rigid transformation that can best-align the patch boundary on each of the human models ($wf_k^2, \ldots, wf_k^m$) to one of them ($wf_k^1$). The rigid transformation can then be used to align the whole patch ($mp_k^a$) to the static one ($mp_k^1$). Noted that the patches are aligned by the transformation that best-align their boundaries instead of the whole patch. This is because the goal here is to eliminate the space variance with the least influence of the shape variance among patches, and more importantly the boundaries are designed in such a way that similar procedure can also be done during the parametric design phase when the mesh patch is not available. Once the wireframe of a new model is synthesized, the patches ($mp_k^a$) can be

Table 1: Dimension reduction of mesh patch and boundary wireframe

| | $wf_k$ | $mp_k$ |
|---|---|---|
| Before PCA | $84 \sim 96$ | $123 \sim 2496$ |
| After PCA | $3 \sim 8$ | $3 \sim 7$ |
| Reduction ratio | $0.85 \sim 0.98$ | $0.97 \sim 0.99$ |

generated and transformed back to the absolute position based on their boundaries ($wf_k^a$).

After the $k$-th patch from all models are aligned, the linear regression similar to the one presented by Chu et al. [3] is used to compute the function $\Upsilon_k$, except that they applied it once to the whole mesh and we apply it $K$ times to each of the patches. Again, the PCA is applied for dimensionality reduction to both the wireframe ($wf_k$) and the mesh patches ($mp_k$). Remarkably, since the meaningful wireframe separates the whole human body into small patches, each patch is relatively simple in geometry complexity. Although the DoF (w.r.t. number of vertices) is still large in a patch (e.g., 2,500), they are highly linear-dependent. This is evidenced by the dimensionality reduction results that the retained numbers of PCs (95% variance) are very similar for both the patch ($mp_k$) and its boundary ($wf_k$) as shown in Table 1. In other words, the wireframe can really capture the high-level spatial relationship between the patches and is a good representation as a intermediate layer to connect the semantic parameters and the mesh. Moreover, a very high reduction ratio (e.g. $99\% : 2,500 \rightarrow 7$) can be achieved for a patch ($mp_k$) and thus the correlation can be built based on the real geometry and shape features.

As the dimensions of $wf_k$ and $mp_k$ after PCA are similar and small, a linear system is good enough to represent the relationship:

$$wp_k = A_k \cdot mf_k + b_k \quad (k = 1, ..., K) \tag{8}$$

where $K$ is the number of patches, $mp_k \in MP$, $wf_k \in WF$, $A_k$ is a relation matrix, $b_k$ is a vector of corresponding residuals, and $\upsilon_k = \{A_k, b_k\}$ in Eq.(2). Depending on the number of PCs retained for $wf_k$ and $mp_k$, this linear system can be over-determined, under-determined, or exactly determined. The QR solver is used to compute the values of $A_k$ and $b_k$. This is done on each of the patches.

### *3.3. Parametric Design*

Once the correlation functions $\Psi$ in Eq.(1) and $\Upsilon_k$ in Eq.(2) are trained and constructed, the learning phase is done, and we can move to the design phase – the right part of Fig.3. In this section, the trained correlations will be used to synthesize new human model with new designated semantic parameters.

Given a new set of semantic parameters, its principal components $SP^*$ can be obtained by the same procedure in Section 3.1. With $SP^*$ and by means of the trained deep net, the new feature wireframe $WF^*$ can be computed by

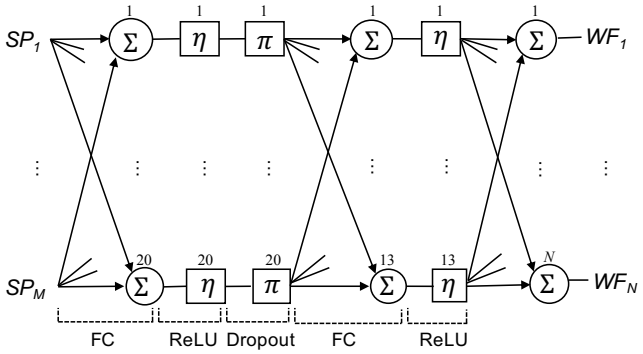



Figure 4: The neural network architecture.

$$WF^* = \Psi(SP^*, \psi). \tag{9}$$

The newly generated wireframe can then be separated into $K$ different sub-sets as the patch boundaries $\{wf_k^*\} \in WF^*$, using the same connectivity in the database. Noted that a boundary alignment is done here as described in Section 3.2. Then, the new mesh patches $\{mp_k^*\}$ can be generated using Eq.(2):

$$mp_k^* = \Upsilon(wf_k^*, \upsilon_k) \quad (k = 1, 2, ..., K) \tag{10}$$

Specifically, it is computed based on the linear regression model in Eq.(8). After all $mp_k^*$ are determined, each of the new mesh patches can be obtained by the back-transformed to the absolute position. The synthesized patches are assembled into an intact human model.

It should be noted that if the learning is simply separated by patches, the reconstructed results will be incompatible, and this is also our motivation to apply wireframe as the middle man to connect the patches. The local wireframes are the subsets of the whole wireframe, in which the correspondences are always defined. As the local wireframes are at the same time the patch boundaries, the patches are always connected and intersect exactly at the wireframes, so the compatibility is guaranteed.

## 4. Experiment

In this paper, we have proposed a new hierarchy method for statistical learning of human models. In order to verify the effectiveness and efficiency of the proposed method, several experiments are conducted in this section to investigate the performance of the proposed methodology compared with the previous methods. The database [3] has 77 subjects of woman model and all of them share the same mesh connectivity. Each model contains $11,072$ mesh vertices and 39 mesh patches separated by wireframe, and there are 1,197 vertices on the feature wireframes. To assess the DNN quality, the first 60 subjects from the database are used for training and the remaining 17 subjects are used for testing. DNN is implemented with Keras in Python, and for the sake of efficiency, the ADAM stochastic optimization algorithm [31]

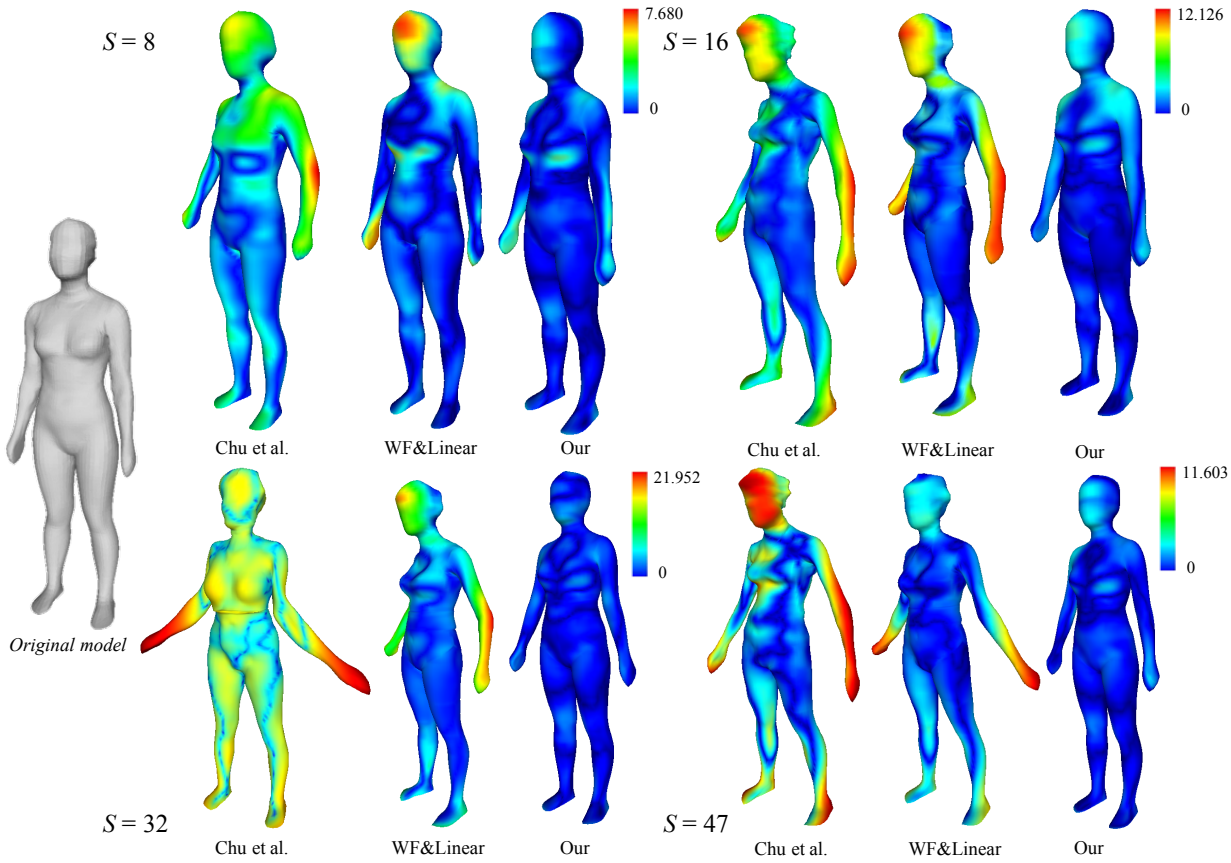



Figure 5: Reconstructing human models with different numbers of semantic parameters ($S$), Color maps display the Hausdorff distance.

is used with a learning rate of $\alpha = 10^{-3}$, $\beta_1 = 0.9$ and $\epsilon = 10^{-8}$. The network architecture has five layers as: FC20+ReLU+Dropout+FC13+ReLU, as shown in Fig.4. The two fully connected layers have 20 and 13 neurons respectively, and the ReLU activation function is used in both layers.

### *4.1. Sensitivity to System Setting*

In the dataset, there are 47 semantic parameters, e.g., body height, neck girth, bust, and waist, collected to describe the human body model. Intuitively, more input semantic parameters would characterize the human model more precisely. However, too many parameters may interaffect the synthesis of human model. It has been seen that the learning system of Chu et al. [3] is very sensitive to the number of input semantic parameters. Being sensitive to system setting leads to a long pre-processing time of calibrating the system parameters, and which will limit its practical uses. Therefore, it is desired to have a learning system that is stable even under different settings.

In this paper, the wireframe is introduced as an interlayer in the learning model, such that the semantic parameters are correlated to an abstract and more meaningful representation of the whole mesh. It is expected that the impact of semantic parameters will be localized and more focused. To evaluate the sensitivity performances of the proposed method and the previous method [3], we conduct an experiment using different semantic parameters to train the learning model, which is then used to synthesize and reconstruct human models. Here, a model is selected randomly from the database, and its parameters are used for the synthesis. The performance can be observed easily in Fig. 5 by comparing the "original model" with reconstructed results of the two methods: "Chu et al" and "Our". It can be seen that the previous method [3] is indeed very sensitive to the number of parameters ($S$). When 8 parameters are used, the reconstructed model is close to original model, but when the number of parameter is larger (e.g., $S = 16$, 32 and 47), the results become unreasonable and the larger the worse. In contrast, our results are more or less similar with different parameter sizes. This experiment shows that our method is less sensitive to the setting of the system, more repeatable, and more practical.

To further verify our contribution of using wireframe as an intermediate layer and applying DNN to model the relationship, an additional comparison is added in Fig. 5 named "WF&Linear", which stands for wireframe is used but only linear regression is applied for both layers (i.e., without DNN). Although it is also affected by the number of parameters used, the results are smoother and closer to the original model for all the tests compared to "Chu et

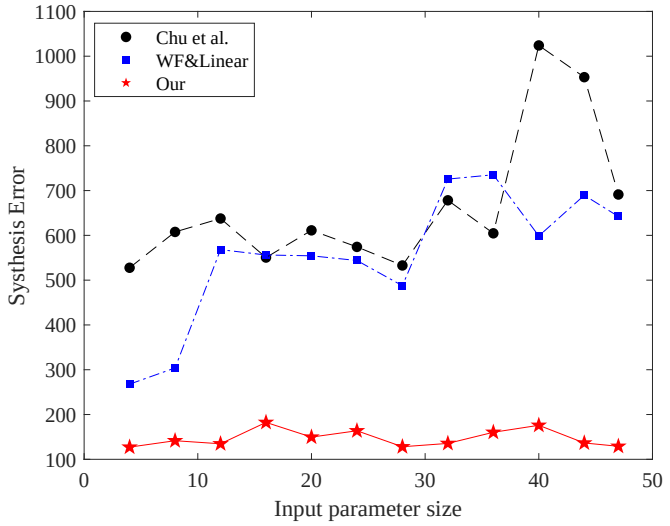



Figure 6: Synthesis error of design methods with various size of input semantic parameters

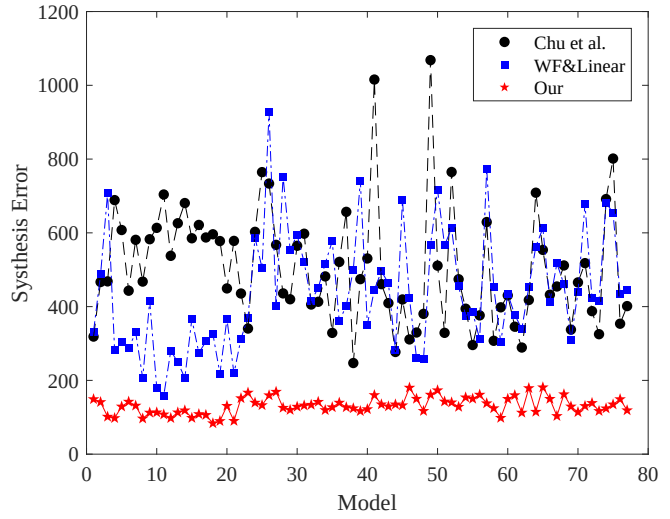



Figure 7: Synthesis error of all 77 models in database with three design methods

al.". This comparison reveals the usefulness of wireframe in stabilizing the learning system. Moreover, the comparison between "WF&Linear" and "Our" demonstrates the usefulness of DNN in characterizing the non-linear correlation between semantic parameters and feature wireframe.

### *4.2. Accuracy of Synthesis*

To quantitatively compare the accuracy of synthesis results with the previous method [3], a $L^2$-norm is defined to measure the difference between two human models $(h, \tilde{h})$:

$$\|h, \tilde{h}\|_2 = \sum_{i=1}^{n} (x_i - \tilde{x}_i)^2, \quad x_i \in h, \tilde{x}_i \in \tilde{h} \qquad (11)$$

where $x_i$ and $\tilde{x}_i$ are the vertices of $h$ and $\tilde{h}$ $(x_i, \tilde{x}_i \in \Re^3)$, and $n$ is the number of vertices ($n = 11,072$ and the unit of database is *cm* in this paper). Given a model $h$ from the database, this experiment is done by synthesizing a new model $\tilde{h}$ with the same semantic parameters of $h$ and comparing the difference between the original model $h$ and the reconstructed model $\tilde{h}$ using Eq.(11). Ideally, if the trained model is perfect, the difference should be zero. Therefore, the smaller the value in difference, the better performance of the method. Firstly, the previous test on sensitivity to parameter size is re-evaluated using this error metric, and the results of synthesis error against parameter size are plotted in Fig. 6. It supports and further verifies the conclusion in Section 4.1 that by using the wireframe as an inter-layer in the learning model, the proposed method can localize the impact of semantic parameters through correlating to an abstract representation of the whole mesh, even when the parameter size increases.

Secondly, a more comprehensive test is done by using all models from the database, and the results are plotted in Fig. 7. It can be seen that "WF&Linear" on average has a smaller value in difference than "Chu et al.", which again indicates the usefulness of wireframe. "Our" gives lowest values in all the tests with a clear level of difference. The results not only verify the use of DNN, but also demonstrate that the proposed method can outperform the previous work significantly. The comparisons are also visualized in Fig. 8, which shows the Hausdorff distance using color map on several models synthesized by three methods. It can be seen that the models reconstructed by "Chu et al." and "WF&Linear' have large differences especially on the head and arms, but our method generates very close results and have a higher accuracy.

### *4.3. Robustness to Extremity*

In this experiment, we test the robustness of methods under some extreme cases in terms of design parameters. The range of each design parameters in the dataset are listed in the table of Fig. 9, and we conduct two experiments based on these values. The first one is to test with extreme values but still within the range, i.e., the smallest value is used for all semantic parameters to synthesize a new model. For example, the range of neck girth parameter in the database is $[29, 44]$, and the input neck girth is set as 29. Reminded that, even the parameters are set as the smallest values in the data ranges, none of the models in the database have exactly the same set of values. The second one is an extrapolation test, i.e., parameters are set out of the data range. For example, the height is set as 185 beyond its range of $[145, 181]$, and similarly for width, hip and inseam. Figure 9 has shown the results generated by the three methods "Chu et al.", "WF&Linear" and "Our" using these two parameter sets, which is also included in its table.

In case 1, the reconstructed model by "Chu et al." [3] is intertwined especially on the hands and body. While "WF&Linear" can generate a more reasonable model but the surface is a bit rough. "Our" method can generate a

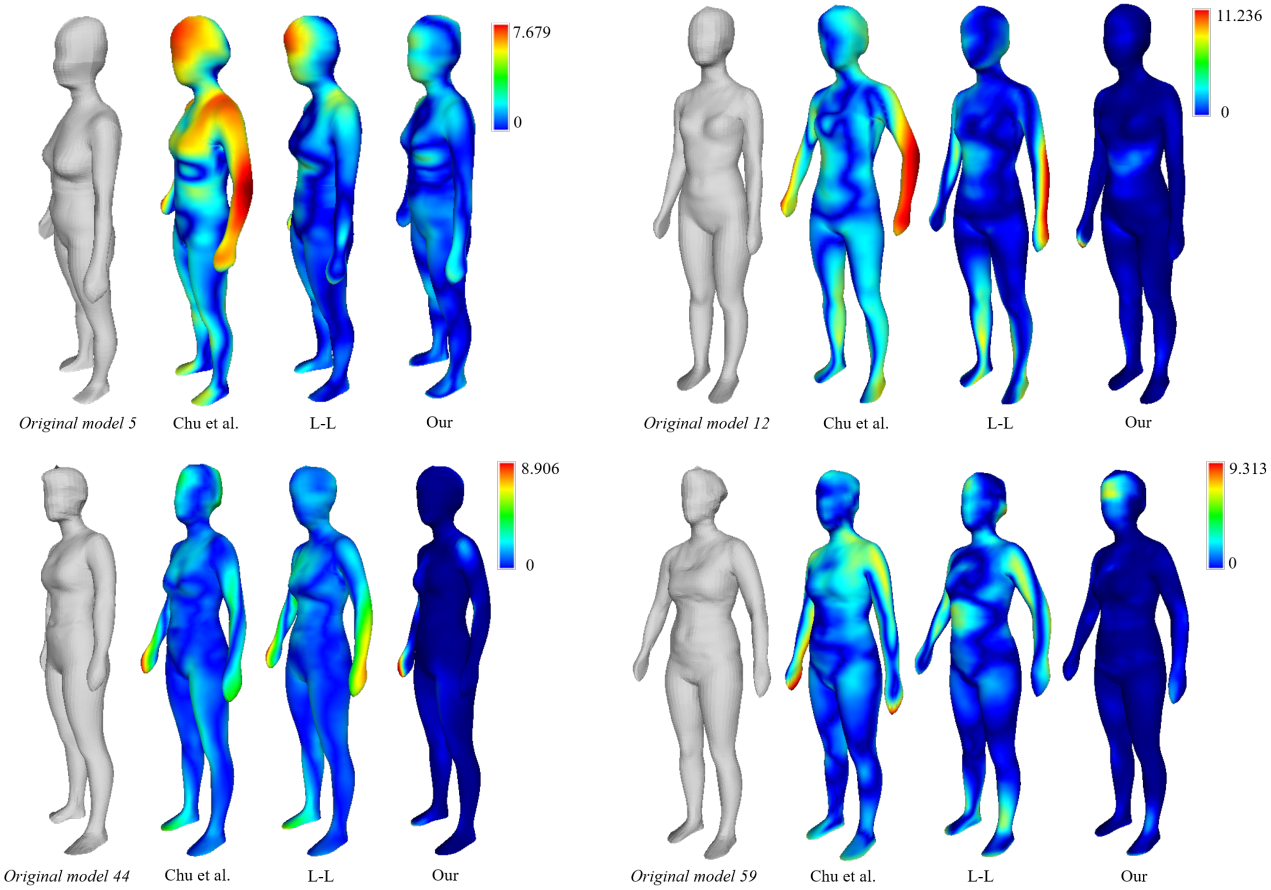



Figure 8: Comparison of Hausdorff distance color map with different design methods for reconstructed models

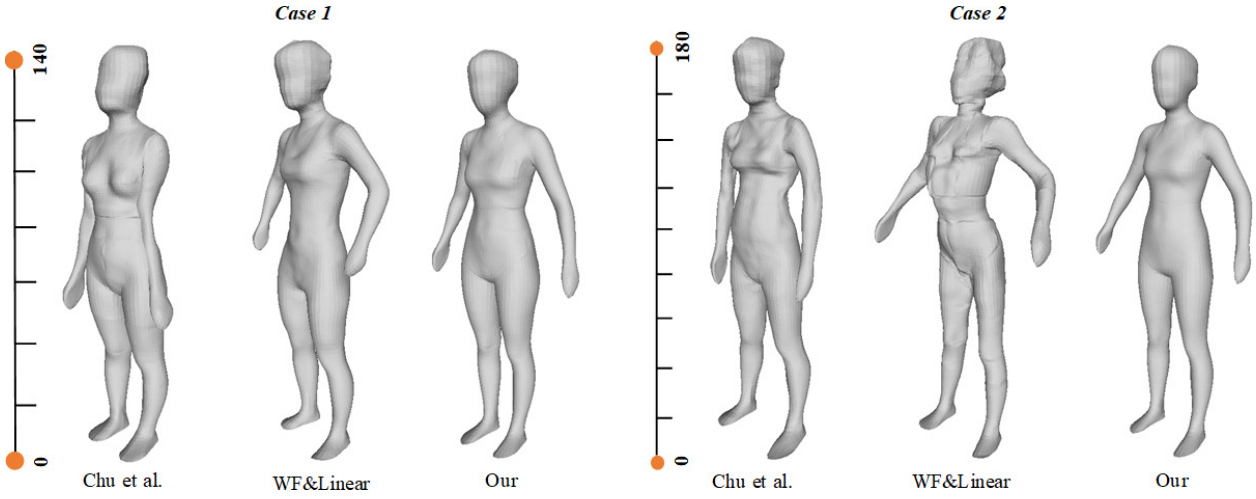
smooth and reasonable model. Quantitatively, the parametric dimensions of the reconstructed models are measured by drawing curves on the digital models and calculating the length of the curves. For example, the input "neck girth" is 29, and the measured values of the models reconstructed by the three methods (Chu et al., WF&Linear, Our) are $(31.5, 31.3, 30.1)$; the input "under bust" is 70, and the results are $(74.8, 71.6, 71.8)$; the input "waist" is 60, and the results are $(62.6, 60.6, 60.0)$. The data show that "WF&Linear" and "Our" can synthesize much more accurate results than "Chu et al." from a set of input parameters that is different from all the models in the database.

In case 2, both the methods of "Chu et al." and "WF&Linear" give very abnormal results with either flattened or stretched shapes. It makes sense because extrapolation is generally very challenging, and Chu et al. [3] used a feasibility check to prevent these cases. Surprisingly, "Our" method can still generate a reasonable model visually, thanks to the stability given by the DNN training with non-linearity. Similarly, the parametric dimensions are measured. For example, the input "width" is 28, and the results are $(33.3, 31.6, 30.6)$; the input "waist" is 80, and the results are $(88.6, 76.3, 78.9)$; the input "inseam" is 85, and the results are $(80.9, 80.4, 84.6)$. The data show that "Our" can outperform both "Chu et al." and "WF&Linear" in terms of extrapolation.

Therefore, the results from both cases reveal that the proposed method has a good robustness on parametric design even for extreme cases and thus more practical for industrial uses.

### *4.4. Application: Generative Design*

Parametric design method have been used to many industrial applications, and digital human body is widely applied in medical, sport, garment design, virtual try-on systems, etc. However, scanning human body is expensive and time consuming, as a result, small-to-medium enterprises may have very limited data. Even a huge amount of data are collected, the design parameters may not all lie in the range of dataset with the evolution of design and requirements. Moreover, to create a new design style that can fit to one or more customer groups, a fast method is needed to generate a comprehensive dataset of human body and explore design possibilities. In such cases, the generative design based on parametric modeling can be applied to facilitate product design and satisfy the requirements of various customer's unique preferences. The challenge in parametric design of human body is that the semantic parameters are usually related to each




| Parameter | Body height | Neck girth | Width | Bust | Under bust | Waist | Hip | Inseam |
|---|---|---|---|---|---|---|---|---|
| Data Range (cm) | [145, 181] | [29, 44] | [29, 63] | [79, 117] | [70, 107] | [60, 115] | [85, 126] | [63, 84] |
| Case 1 | **145** | **29** | **29** | **79** | **70** | **60** | **85** | **63** |
| Case 2 | **185** | 35 | **28** | 82 | 78 | 80 | **82** | **62** |

Figure 9: Comparison of extremity design with different methods

other [32, 33], e.g., height is related to weight, and waist is related to bust.

From the experiments in previous sections we can see that the proposed method has a significant improvement on parametric design of human body design. DNN is used to learn a good representation of the complex correlation among human models. The robust design capability of the proposed method enables its application to high quality generative design. In this section, a generative design experiment is conducted to reveal the generative design capability of the proposed methodology. In the experiment, four design parameters: body height, bust, waist and hip are varied with different values to study the effect on the resulting human model and design capability of the proposed method. Figure 10 shows the generated results with various parameters. It can be seen that the generated human models can reflect the variance of input design values. For example, when the input body height increased from 140$cm$ to 190$cm$ by 10$cm$ for each model, the resulting meshes are also getting taller. Similarly for the bust, when increasing the design values from 78$cm$ to 98$cm$ by increasing 4$cm$ on sequent models, the resulting meshes have a clear increasing effect on the bust area. The similar design effect can be also found with varying the waist and hip values. This indicates the proposed design method is effective and robust.

## 5. Conclusion

In this paper, we presented a wireframe-assisted deep learning method for parametric design of human body modeling. A hierarchy design system consists of building relationship between semantic parameters and feature wireframe by DNN, modeling correlation between feature wireframe and mesh patches with linear regression is developed. Experimental results show that the proposed design method outperforms the previous methods and can capture more feature information of human bodies, less sensitive to system settings, more robust to extreme cases, and more accurate in model reconstruction. The introduced intermediate layer of feature wireframe proved to be useful for human modeling and can be applied to practical applications such as product customization, the DNN based method is robust and could be utilized for generative design in many industrials such as virtual try-on systems. The framework simply follows a previous work to define the wireframe, a future work is to study how the definition of wireframe will affect the learning quality, and find a quantitative way to verify a wireframe.

## Acknowledgement

We acknowledge the support of the Natural Sciences & Engineering Research Council of Canada (NSERC) grant # RGPIN-2017-06707 and National Science Foundation (NSF) # CNS-1547167.

## References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: Shape completion and animation of people, ACM Trans. Graph. 24 (3) (2005) 408–416.

[2] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, H.-P. Seidel, A statistical model of human pose and body shape, Computer Graphics Forum 28 (2) (2009) 337–346.

[3] C.-H. Chu, Y.-T. Tsai, C. C. Wang, T.-H. Kwok, Exemplar-based statistical model for semantic parametric design of human body, Computers in Industry 61 (6) (2010) 541 – 549.

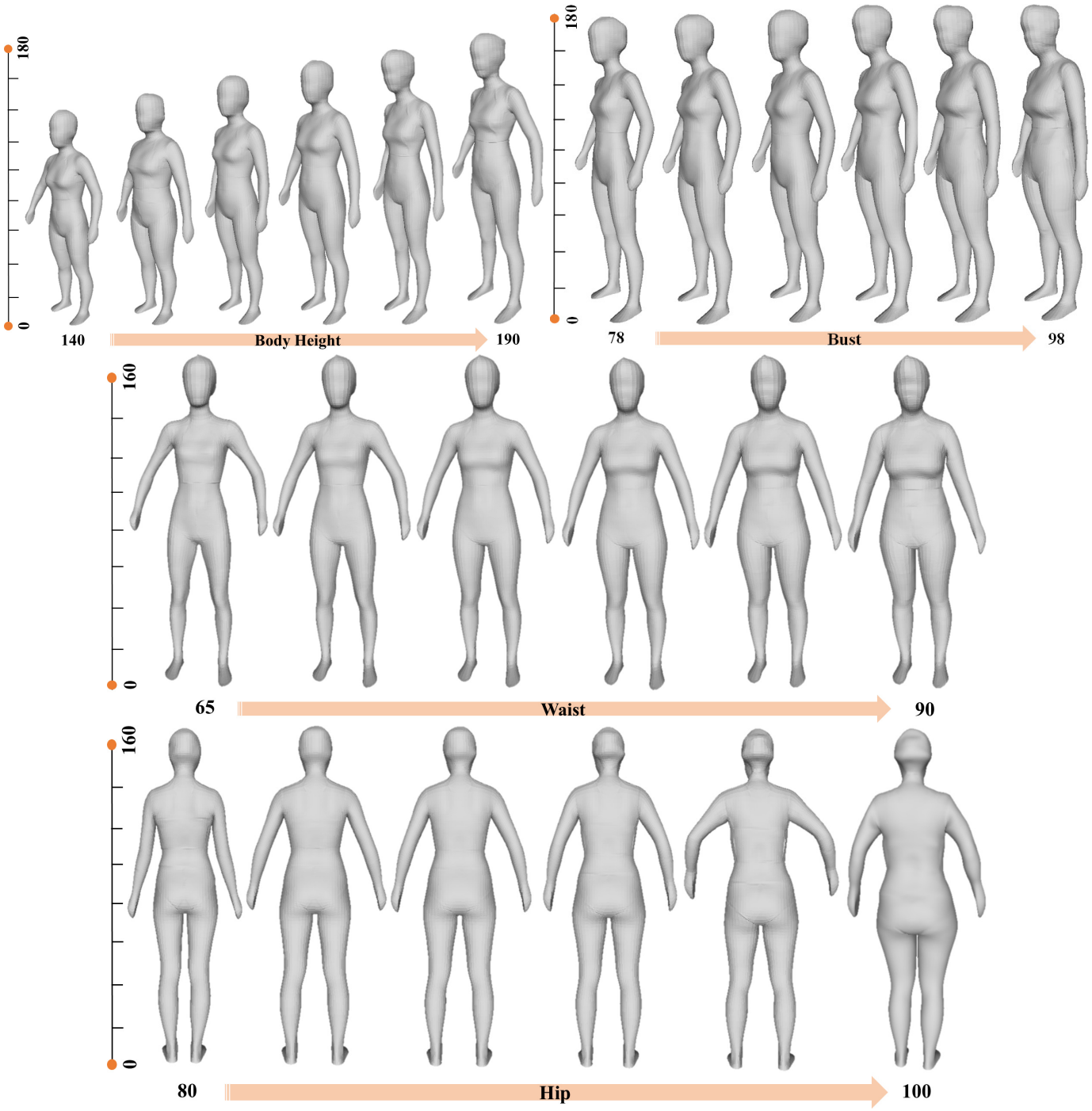



Figure 10: Generative design by varying the values of input semantic parameters

[4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436 EP –.
[5] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: Going beyond euclidean data, IEEE Signal Processing Magazine 34 (4) (2017) 18–42.
[6] B. Liu, Y. Wei, Y. Zhang, Q. Yang, Deep neural networks for high dimension, low sample size data, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 2287–2293. doi:10.24963/ijcai.2017/318.
[7] C. C. Wang, Parameterization and parametric design of mannequins, Computer-Aided Design 37 (1) (2005) 83 – 98.
[8] S. Wang, S. Qin, C. Guan, Feature-based human model for digital apparel design, IEEE Transactions on Automation Science and Engineering 11 (2) (2014) 620–626.
[9] T. H. Kwok, Y. Zhang, C. C. L. Wang, Efficient optimization of common base domains for cross parameterization, IEEE Transactions on Visualization and Computer Graphics 18 (10) (2012) 1678–1692.
[10] C. C. Wang, K. Tang, Pattern computation for compression garment by a physical/geometric approach, Computer-Aided Design 42 (2) (2010) 78 – 86.
[11] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormhlen, H.-P. Seidel, Estimating body shape of dressed humans, Computers and Graphics 33 (3) (2009) 211 – 216.
[12] J. Li, J. Ye, Y. Wang, L. Bai, G. Lu, Fitting 3d garment models onto individual human models, Computers and Graphics 34 (6) (2010) 742 – 755.
[13] G. Pons-Moll, S. Pujades, S. Hu, M. J. Black, Clothcap: Seamless 4d clothing capture and retargeting, ACM Trans. Graph. 36 (4) (2017) 73:1–73:15.
[14] S.-Y. Baek, K. Lee, Parametric human body shape modeling framework for human-centered product design, Computer-Aided Design 44 (1) (2012) 56 – 67.
[15] C. Au, Y.-S. Ma, Garment pattern definition, development and application with associative feature approach, Computers in Industry 61 (6) (2010) 524 – 531.
[16] C.-H. Chu, I.-J. Wang, J.-B. Wang, Y.-P. Luh, 3d parametric human face modeling for personalized product design: Eyeglasses frame design case, Advanced Engineering Informatics 32 (Supplement C) (2017) 202 – 223.
[17] S.-H. Huang, Y.-I. Yang, C.-H. Chu, Human-centric design personalization of 3d glasses frame in markerless augmented reality, Advanced Engineering Informatics 26 (1) (2012) 35 – 45.
[18] N. Hasler, T. Thormählen, B. Rosenhahn, H.-P. Seidel, Learning skeletons for shape and pose, in: Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D '10, ACM, New York, NY, USA, 2010, pp. 23–30.
[19] A. Toshev, C. Szegedy, Deeppose: Human pose estimation via deep neural networks, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[20] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore, Real-time human pose recognition in parts from single depth images, Commun. ACM 56 (1) (2013) 116–124.
[21] W. Si, S.-H. Lee, E. Sifakis, D. Terzopoulos, Realistic biomechanical simulation and control of human swimming, ACM Trans. Graph. 34 (1) (2014) 10:1–10:15.
[22] C. Ufuk, Y. I. O., C. Tolga, Examplebased retargeting of human motion to arbitrary mesh models, Computer Graphics Forum 34 (1) 216–227.
[23] S. Streuber, M. A. Quiros-Ramirez, M. Q. Hill, C. A. Hahn, S. Zuffi, A. O'Toole, M. J. Black, Body talk: Crowdshaping realistic 3d avatars with words, ACM Trans. Graph. 35 (4) (2016) 54:1–54:14.
[24] B. Allen, B. Curless, Z. Popović, The space of human body shapes: Reconstruction and parameterization from range scans, ACM Trans. Graph. 22 (3) (2003) 587–594.
[25] S. Saito, Z.-Y. Zhou, L. Kavan, Computational bodybuilding: Anatomically-based modeling of human bodies, ACM Trans. Graph. 34 (4) (2015) 41:1–41:12.
[26] H. Seo, N. Magnenat-Thalmann, An example-based approach to human body manipulation, Graphical Models 66 (1) (2004) 1 – 23.
[27] D. Boscaini, J. Masci, E. Rodolà, M. Bronstein, Learning shape correspondence with anisotropic convolutional neural networks, in: Advances in Neural Information Processing Systems, 2016, pp. 3189–3197.
[28] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, Omnipress, USA, 2010, pp. 807–814.
[29] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.
[30] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.
[31] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, CoRR abs/1412.6980.
[32] N. Sarris, M. G. Strintzis, 3D modeling and animation: Synthesis and analysis techniques for the human body, IGI Global, 2005.
[33] Y. Zhang, J. Zheng, N. Magnenat-Thalmann, Example-guided anthropometric human body modeling, The Visual Computer 31 (12) (2015) 1615–1631.

## Appendix A. PCA and Dimension Reduction

Demonstrated in Fig. 3 , the correlations between any two dataset are done on their principal components (PCs). The details of PCA are briefly discussed using wireframe here. One of the reasons to apply PCA on the data is for dimensionality reduction, which is a very useful step for processing high-dimensional datasets, while still retaining as much of the variance in the original dataset as possible.

Assume there are $N$ scanned human models in the database, the wireframe on human model can be defined by a set of vertices collected as $H = [h_1\ h_2\ \dots\ h_N]_{3M\times N}$, where $h_i$ is a vector with $M$ three-dimension vertices from the wireframe of the $ith$ human model in the database. Letting:

$$\bar{h} = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{A.1}$$

Here we have $\bar{H} = [h_1 - \bar{h}\ h_2 - \bar{h}\ \dots\ h_N - \bar{h}]$, and its covariance is $V = \bar{H}(\bar{H})^T$, which dimension is $3M \times 3M$. Since $3M \gg N$, we instead compute the transpose of its covariance $W = (\bar{H})^T\bar{H}$. Then by eigenvalue decomposition, we have $Wx = \lambda x$. Then $n$ eigenvectors $[x_j]_{N\times 1}$ could be obtained, with $x_j$, the $jth$ eigenvector of $C$ can be determined: $y_j = \bar{H}x_j$, here $y_j$ is a $3M \times 1$ vector. The normalized eigenvectors $\hat{y} = y_j/\|y_j\|$ $(j = 1, \dots, N)$ are the principal vectors of $\bar{H}$, where each is associated with a variance $\sigma_j$. The vectors are sorted according to $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_N^2$. The largest variance means the corresponding vector $y_j$ has the most dominant effect in the model space. We keep the first $m$ principal components according to the percentage of the total variance explained $\alpha$ by each principal component.

$$\alpha = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_m}{\lambda_1 + \lambda_2 + \cdots + \lambda_m + \cdots + \lambda_N} \geq 0.95 \tag{A.2}$$

Then the wireframe of each human model of all scanned models serving as training data set can be projected onto $m$-dimensional points by:

$$wf_i = \begin{bmatrix} \hat{y}_1^T \\ \hat{y}_2^T \\ \vdots \\ \hat{y}_m^T \end{bmatrix} (h_i - \bar{h}) \ (i = 1, \ldots, N) \qquad \text{(A.3)}$$

Thus, we map $H_{3M \times N}$ into a reduced matrix $WF_{m \times N} = [wf_i]$ $(m \ll 3M)$ spanning the linear space of exemplar patches of human bodies, named as the reduced exemplar matrix. After PCA, a wireframe $h_i$ is projected to a $m$-dimensional point $wf_i$ and its collection is the principal components $WF$.

Similarly, suppose we have $S$ semantic parameters, PCA can be applied to the semantic parameters to obtain $s$ principle components, i.e., $SP = [sp_1 \ sp_2 \ \ldots sp_N]_{s \times N}$ where $sp_i$ is a $s \times 1$ vector.