Research Article

Research on Multi-Precision Fabric Modeling Method Based on Machine Learning

Yanxia. Jin, Zhixu. Chen, Ye. Lu, Jing. Yang, Yabian. Liu, and Zhiru. Shi

School of Big Data, North University of China, Taiyuan 030051, China

Correspondence should be addressed to Yanxia. Jin; jyx@nuc.edu.cn

Received 21 December 2021; Revised 4 March 2022; Accepted 17 March 2022; Published 11 April 2022

Academic Editor: Sikandar Ali

Copyright © 2022 Yanxia. Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the shortcomings of complex calculation and time consumption of fabric physical simulation methods, many single-precision fabric simulation technologies based on machine learning methods have emerged and the amount of calculation is increased for regions with small changes. Aiming at this problem, this paper proposes two multi-precision fabric modeling method based on machine learning. Firstly, the fabric mesh is simulated by the physical method, the initial position of the vertex is calculated, and the deformation of each region of the fabric is measured by Rayleigh quotient curvature, and the multi-precision fabric mesh is constructed. Secondly, the multi-precision fabric graph structure and geometry image are extracted from the multiprecision fabric mesh. Finally, the subgraph convolutional neural network and super-resolution network are trained to model the multi-precision fabric, and we compared the two different multi-precision fabric machine learning modeling methods. Through the experimental verification, in the garment modeling, the garment modeled by the subgraph convolutional network is no longer only dependent on the change of human joints, resulting in a more realistic effect. At the same time, the efficiency of the subgraph convolutional network is 25.3% higher than that of the single-precision garment modeling based on the machine learning method. In the cloth simulation, speed of the super-resolution network is nearly 16 times faster than that of the physical simulation, which supplements the imperfection of insufficient flexibility of the subgraph convolutional neural network modeling.

1. Introduction

Fabric simulation is one of the important research directions in the field of computer animation, and it has many applications in animation movies, games, virtual try-on, and so on. At present, the mainstream fabric simulation generally adopts the spring-mass model [1]. This model obtains a series of particles with regular arrangement through discrete fabric and applies the mathematical and physical laws to the particles to reflect the physical characteristics of the fabric. A large number of particles are needed to ensure the reality. A large number of particles and large-scale solutions increase the computational complexity and resource requirements.

Recently, in the field of fabric simulation, some researchers have linked the idea of machine learning with the optimization of fabric modeling. However, these methods use single-precision mesh for modeling, increasing the amount of computation. In this paper, the number of high-precision meshes is reduced by modeling multi-precision fabric mesh. Using multi-precision meshes can highlight the fold details in the areas where the fabric deforms violently and reduce the amount of calculation.

In this paper, two multi-precision cloth modeling methods based on machine learning are proposed. The first method extracts the multi-precision fabric graph structure based on time and space from the multi-precision fabric mesh. The subgraph convolutional neural network is applied to the multi-precision fabric graph structure, and the fabric vertex feature is updated by aggregating the neighbor node information. In order to obtain a more accurate prediction effect, the multi-precision fabric graph structure constructed in this paper contains not only clothing vertices but also joint nodes. For cloths that are moved by external forces, because only the information of cloth vertices can be aggregated, the predicted results are not good enough. So, the second
method constructs the geometry image from the multi-precision fabric mesh, uses the super-resolution network to predict the high-resolution geometry image of the center from several continuous low-resolution geometry images, and then reconstructs the 3D mesh from the geometry image. In order to better extract the spatiotemporal characteristics of multiframe input, this paper also proposes the spatiotemporal feature progressive fusion module and improves the multiscale feature extraction of ASPP [2]. This paper compares the advantages and disadvantages of the two methods and the existing methods. The two methods in this paper make up for each other’s shortcomings, so that both garment and general free deformation cloth have achieved good simulation results.

2. Related Work

Guan et al. [3] constructed DRAPE model, which divided garment deformation into caused by posture change and caused by body shape, expressed the deformation as a series of nonlinear changes, learned the nonlinear changes through training, and finally integrated the two kinds of garment deformation to obtain the final DRAPE model. Deng et al. [4], based on the research method of Guan et al., extracted 11 joint nodes of human body which affected the garment deformation, extracted the characteristics of human body movement and garment deformation by using BP neural network, and preliminarily found the relationship between human body movement and garment deformation. Shi et al. [5] further defined the character of each joint and used four different machine learning models to learn the relationship between human motion and garment deformation from examples, so as to obtain the deformation degree of each area of garment. Santesteban et al. [6] trained an MLP and an RNN to simulate a garment model by decomposing the garment deformation into static and dynamic folds. The TailorNet model proposed by Patel et al. [7] decomposed the garment deformation into high-frequency and low-frequency components; the low-frequency components were predicted by MLP from the parameters of pose, shape, and style, while the high-frequency components were predicted by the mixture of pose models with specific shape and style. GarNet [8] proposes a network with a two-stream structure. The body stream takes the body represented by the 3D point cloud as input, the clothing stream takes the clothing represented by the 3D triangular mesh as input, and the outputs of the two streams are fed to a set of MLPs. Fusion module is used to model the interaction between clothing and human body. GarNet++ [9] proposes RQ curvature, adds it to the loss term of the original network, and additionally introduces a nearest neighbor merging step to calculate a separate local body feature for each clothing vertex. The combined use of multiple loss terms allows coverage in clothing on the 3D human body to yield a higher level of detail. Jin et al. [10] proposed a pixel-based data-driven clothing framework, which reorganized triangular mesh data into image-based data, given input human pose parameters, trained a CNN to predict cloth images, and finally recovered from the cloth images 3D clothing shape. However, these methods use single-precision mesh for modeling, increasing the amount of computation. In this paper, the number of high-precision meshes is reduced by modeling multi-precision fabric mesh; however, the fabric mesh has irregular topological structure, and regular convolution cannot handle it well. The emergence of neural networks on graph [11] solves this problem well. Graph neural network iterates node state propagation to equilibrium state, applies neural network, and generates output for each node according to node state.

However, the above literature works are all aimed at clothing animation driven by human motion. In these methods, the joint parameters of human motion are the indispensable and important data in the method. For the fabric mesh deformed by external force, such as flag, it does not depend on the joint parameters of human motion. Oh et al. [12] combined the traditional fabric modeling method based on physics with Deep Neural Network (DNN), used DNN model to model fabric in layers, and calculated the finer nodes in layered fabric through DNN model on the rough fabric mesh simulated by physics. Chen et al. [13] proposed a method based on CNN geometry image super-resolution to synthesize fabric resolution, simulated paired high- and low-resolution meshes, converted them into geometry images, and used them as training data. However, single frame input could not guarantee the time consistency between frames in the prediction results. Therefore, for the cloth moving under the action of external force, this paper uses CNN-based image super-resolution method, converts mesh data into image data, and processes it with CNN. In order to obtain the consistent prediction results in time, the network constructed in this paper uses the continuous low-resolution geometry image as the input and outputs the high-resolution geometry image of the center. It supplements the problem of insufficient modeling flexibility of graph convolutional neural network modeling.

3. Method

3.1. Characteristic Curvature. Gaussian curvature is the most common way to measure the curvature of a triangular surface. Meyer et al. [14] defined discrete Gaussian curvature, which can be expressed as

$$K^i_{GC} = \frac{1}{A^i} \left( \frac{2\pi - \sum_{k \in N_i^f} \theta_k}{N_i^f} \right).$$

(1)

In formula (1), $\theta_k$ is the angle of the $K^i$th triangular face, $N_i^f$ is represented as the neighborhood of $F$ adjacent triangular faces of the $i$ vertex, and $A^i$ is the total area of the neighborhood of the triangular face around the $i$ vertex. Gaussian curvature is expressed as the product of two main curvatures. It can well represent the bending degree of the surface in the low deformation clothing region and the medium deformation clothing region. However, in the area of high deformation clothing, it cannot well reflect the deformation trend of folds because if a larger principal curvature is multiplied by a smaller principal curvature, we
will obtain a smaller Gaussian curvature. Therefore, we introduce the characteristic curvature \([9]\), calculate the covariance matrix \(\Sigma_i^k\) for each vertex, and define it as
\[
\sum_i^K = \frac{1}{K} \sum_{j \in N^K_i} (G_j - \bar{G}_i)(G_j - \bar{G}_i)^T,
\]
where \(G_i\) is the 3D position of vertex \(i\), \(G_j\) is the 3D position of vertex \(j\) which shares an edge with vertex \(i\), \(N^K_i\) is a group of vertices which shares an edge with vertex \(i\), and \(\bar{G}_i\) is the average value of all points in \(N^K_i\). Then, we solved the eigenvalue of \(\Sigma_i^k\); the maximum eigenvalue is defined as \(\sigma_{\text{max}}\) and the minimum eigenvalue is defined as \(\sigma_{\text{min}}\). For the low deformation clothing area, the maximum eigenvalue \(\sigma_{\text{max}}^d\) is much larger than the minimum eigenvalue \(\sigma_{\text{min}}^d\). For the high deformation clothing area, the maximum eigenvalue \(\sigma_{\text{max}}^d\) is slightly greater than the minimum eigenvalue \(\sigma_{\text{min}}^d\). Therefore, the eigenvalue distributions of these three different fold regions are different, which makes them easy to distinguish. However, the covariance matrix of low deformation clothing region is unstable in eigenvalue solution. Therefore, we used local neighborhood Rayleigh quotient to replace eigenvalue solution, and it is expressed as
\[
RQ(\Sigma_i^k, G_j) = \frac{G_j^T \Sigma_i^k G_j}{G_j^T G_j}.
\]
Due to the fact that \(\sigma_{\text{min}} \leq RQ(\Sigma_i^k, G_j) \leq \sigma_{\text{max}}\), the minimum and maximum values are used as \(RQ(\Sigma_i^k, G_j)\) curvature estimates in this paper.
\[
\begin{align*}
RQ_{\text{min}}^i & = \min_{j \in \mathcal{N}_i^k} RQ(\Sigma_i^k, G_j), \\
RQ_{\text{max}}^i & = \max_{j \in \mathcal{N}_i^k} RQ(\Sigma_i^k, G_j).
\end{align*}
\]

### 3.2. Multi-Precision Fabric Mesh Division

Although the deformation degree of fabric is complex under the action of external force, without instantaneous large external force and under the same movement trend, the distribution of deformation areas of fabric is similar, and the generation and deformation of fabric wrinkles are relatively stable. In order to effectively and quickly obtain the deformation degree of the fabric mesh region under a certain movement trend, the average deformation degree of fabric mesh vertices under a certain movement trend is sampled and analyzed, and the threshold value of average deformation degree is determined; then the fabric model is divided according to the average deformation threshold.

Intuitively, the greater the change degree of the normal vector of the surface, the more obvious the curvature degree of the surface. Therefore, we divide the fabric area into low, medium, and high deformation fabric area. The low deformation fabric area is described as a gentle area where the normal vector of the surface does not change basically. The medium deformation fabric area is described as a hemispherical area where the normal vector of the surface changes gently. The high deformation fabric area is described as a ridge-shaped area where the normal vector of the surface changes drastically.

The Rayleigh quotient curvature \([9]\) of vertex \(i\) is expressed as
\[
RQ_C^i = \frac{\min_{j \in \mathcal{N}_i^k} (G_j^T \Sigma_i^k G_j / G_j^T G_j)}{\max_{j \in \mathcal{N}_i^k} (G_j^T \Sigma_i^k G_j / G_j^T G_j)},
\]
where \(\Sigma_i^k\) is the covariance matrix and it is defined as
\[
\sum_i^K = \frac{1}{K} \sum_{j \in \mathcal{N}_i^k} (G_j - \bar{G}_i)(G_j - \bar{G}_i)^T.
\]

3.3. Multi-Precision Fabric Graph Structure Construction

The garment mesh produces different degrees of garment fold deformation with human motion. Therefore, we
abstract the multi-precision garment graph structure from the human motion characteristics and the corresponding multi-precision garment mesh structure by studying the relationship between human motion characteristics and multiprecision garment mesh structure deformation.

3.3.1. Feature Extraction of Human Motion. In this paper, SMPL [17] is used as a parameterized human model. The SMPL model represents human body \( M_b(\beta, \theta) \) as a parametric function of posture \( \theta \) and shape \( \beta \),

\[
M_b(\theta) = W(T_b(\theta), J(\theta), \theta, W_b),
\]

\[
T_b(\theta) = B + B_p(\theta).
\] (8)

In linear function \( T_b(\beta, \theta) \), \( B \) represents the vertex data of human skin under T-pose, \( B_p(\theta) \) adds the deformation that depends on posture, \( B(\beta) \) adds the deformation that depends on shape, and \( W \) represents the influence weight of joints node \( J(\beta) \).

3.3.2. Feature Extraction of Garment Deformation. Since the garment model changes with the human body model, we deform each garment vertex according to the vertex function of the SMPL human body model. \( D \) is the displacement between the garment model and \( T_b(\beta, \theta) \). For different garment types, the value of \( D \) is also different.

\[
T_g(\theta, D) = T_b(\theta) + D,
\]

\[
M_g(\theta, D) = W(T_g(\theta, D), J(\beta), \theta, W_g).
\] (9)

When the body mesh deforms, these garment vertices will deform with the body mesh. In this paper, the simulated garment vertex position is written as \( D_g = D + d \), where \( d \) is one corresponding to \( D \) and \( d \) is offset from \( D_g \) to \( D \) garment vertex position. For any \( \theta, D \) is well determined by SMPL’s skinning algorithm, so machine learning only needs to learn the garment vertex offset \( d \) from \( (\theta, d) \), reducing the nonlinear variation.

\[
M_g(\theta, d) = W_d(M_g(\theta, D), d, W_d).
\] (10)

Thus, two kinds of garment models are obtained in this paper; one is rough mesh garment model \( M_g(\beta, \theta, D) \) based on SMPL and the other is fine pleated garment model \( M_g(\beta, \theta, D) \) on \( M_g(\beta, \theta, D) \).

3.3.3. Construction of the Structure of the Spatiotemporal Multi-Precision Graph. We extract 24 joints from the SMPL to construct the space-time joint graph \( G_b = (V_b, E_b) \), where \( V_b \) represents joint nodes and \( E_b \) represents edges. Node set \( V_b = \{V_b^m|t = 1, \ldots, T, m = 1, \ldots, 24\} \), including 24 joint nodes in the SMPL mannequin, where \( m \) represents the number of joint nodes and \( T \) represents the number of frames. The edge set \( E \) contains two subsets, one of which is the spatial edge formed by the connection of two adjacent joint nodes in each frame, expressed as \( E_b^s = \{V_b^m|V_b^{m+1}\} \). The other subset is the time edge formed by connecting the vertex of the same joint node between two adjacent frames, which is represented as \( E_b^t = \{V_b^m|V_b^{m+1+t}\} \).

Since the position of the garment mesh structure changes at every moment based on the position of the last moment, we constructed the spatiotemporal garment graph structure \( G_g = (V_g, E_g) \), where \( V_g \) represents the garment vertex and \( E_g \) represents the edge. Node set \( V_g = \{V_g^{m}|t = 1, \ldots, T, g = 1, \ldots, k\} \), including all garment vertices in the garment model, where \( k \) represents the number of garment vertices and \( T \) represents the number of frames. The edge set \( E_g \) contains two subsets, one of which is the spatial edge formed by the connection of two adjacent garment vertices in each frame, expressed as \( E_g^s = \{V_g^m|V_g^{m+1}\} \). The other subset is the time edge formed by connecting the vertex of the same garment between two adjacent frames, which is represented as \( E_g^t = \{V_g^m|V_g^{m+1+t}\} \).

The spatial and temporal garment graph structure \( G_g = (V_g, E_g) \) is connected with the four joint nodes closest to \( V_g \) in the spatial and temporal joint graph \( G_b = (V_g, E_g) \) to form a multi-precision garment graph structure \( G = (V, E) \). Taking garment vertex \( V^{(t+1)}_g \) as an example, the multi-precision garment graph structure based on space and time is presented, as shown in Figure 1. The red line segment is the time edge, the green line segment is the space edge, the yellow vertex is the garment vertex, and the blue vertex is the joint node.

For the cloth model with external deformation, we only extract the space-time structure of the cloth vertex, similar to the space-time garment graph structure.

3.4. Geometry Images Construction. In this method, geometry images [18] are used as input and output of the model. In order to generate geometry images, the plane parameterization method with fixed boundary is used in this paper. Firstly, the mesh boundary points are uniformly mapped to the 2D square boundary, the parameter coordinates of the internal points are determined by the convex combination of the neighborhood points of one ring, and then the uniform sampling is carried out in this square parameter domain. Under this premise, this paper uses a parameterization method to minimize the geometry stretching [19]. Compared with the traditional conformal and area-preserving parameterization, this method has lower geometry image reconstruction error.

For fabric models such as flag, the initial model is defined by a single mesh with rectangular boundaries, so it does not need to be parameterized. We uniformly sample \( n \times n \) points in its 2D coordinate space. For each sampling point, it is located in a triangle in the 2D space. We find this triangle and calculate the gravity coordinates of the sampling point under this triangle, as shown in Figure 2, where \((x, y)\) represents the sampling point, and \((\alpha, \beta, \gamma)\) is its gravity coordinates. According to the definition of the gravity coordinates, the gravity coordinates can be calculated by using the coordinates of the sampling point and the triangle. When calculating the attributes of the sampling point, the gravity coordinates are used to interpolate the attribute values from the triangle vertex as follows:
\[ V = \alpha V_A + \beta V_B + \gamma V_C, \quad (11) \]

where \( V \) represents the attribute value of the sampling point and \( V_A, V_B, \) and \( V_C \) represent the 3D coordinates of three vertices of the triangle. The 3D space coordinates \((X, Y, Z)\) corresponding to each sampling point converted to color values \((R, G, B)\) by scaling in proportion for visualization. Although the sampling points move along with the fabric, they remain in a fixed position in the 2D space. Therefore, for each frame in the subsequent simulation process, the same weight is used to interpolate the coordinates of the sampling points. For fabric models such as garment, which are defined by irregular mesh, they cannot be well sampled as an image. The mesh has multiple boundaries and is not topologically homeomorphic to the planar 2D rectangular parametric domain. Therefore, for each garment mesh, the mesh is first cut into two meshes with a single boundary, the 3D mesh is mapped to the 2D parametric domain by the parameterized method of minimizing geometry stretching, and the sampling operation is performed as the flag mesh. An example of T-shirt cutting along the joints on both sides, parameterizing and sampling, to generate geometry images is shown in Figure 3. Each garment generates two geometry images on the front and back. In order to construct the dataset, we further downsample each geometry image to generate low-resolution geometry images.

3.5. Subgraph Convolutional Neural Network Algorithm. The core of graph convolutional neural network is to define the aggregation function between nodes. Aggregate function is essentially the neighborhood information of set nodes. In the multi-precision fabric graph structure, the aggregation function is used to recursively update the features of each fabric vertex, which is expressed as the superposition of the features of the neighboring nodes and their own nodes until the stable fabric vertex is reached. Therefore, we present the general framework of graph convolution network by defining aggregation function. The update of fabric vertices will be divided into the two following steps: Firstly, the aggregation function is applied to fabric vertices and their neighbor nodes to obtain the local structural characteristics of fabric vertices. Then the update function is applied to its own and local structural characteristics to obtain the new features of the current fabric vertex. By designing each layer of neural network into the aggregation function and update function mentioned above, each fabric vertex can constantly update itself with its own and neighboring nodes as the source information and then get a new expression depending on the local structure of fabric vertex. However, multi-precision graph structure sacrifices the efficiency of time and memory, because the graph convolutional neural network needs to run the recursive function multiple times on all nodes, which requires storing the intermediate states of all nodes in memory. Therefore, we propose a subgraph convolutional neural network to train multi-precision fabric graph structure.

Based on graph convolutional neural network, the aggregation function of subgraph convolutional neural network is defined as a cyclic recursive function, and each node updates its own expression with neighbor nodes and edges as information sources. In this paper, its hierarchical propagation rule is expressed as follows:

\[ \bar{V}_t = g(V_t, A, p), \]

\[ V_{t+1} = c(\bar{V}_t), \quad (12) \]
where $A$ is the adjacency matrix of the multi-precision fabric graph structure $G = (V, E)$, $V_i = [V_i^{|l = 1}, \ldots, L, n = 1, \ldots, N]$, $l$ represents the number of convolution layers, $N$ represents the number of nodes, $g(\cdot)$ represents the aggregation function, $c(\cdot)$ represents the update function, and $p$ represents the size of the subgraph, which can represent the size of the field of vision of the subgraph convolutional neural network. Each node $V_i^{|l}$ has three features. Searching in $A$, we can obtain adjacent nodes indexes of $V_i^{|l}$, and we will take $V_i^{|l}$ as the target node. According to the number of edges from each node to $V_i^{|l}$, the adjacent nodes of $V_i^{|l}$ are divided into neighbors of order $t$, and the adjacent nodes of $A$ are sorted by order of $1$ to $t$. The middle joint points of each order have a higher priority over the garment vertices. The closer the joint node and garment vertices to the target vertex, the higher the priority it has.

$$
\begin{align*}
    e_{gb} &= \beta^T e_{gb}, \\
    e_{gg} &= \beta^T \alpha e_{gb},
\end{align*}
$$

where $e_{gb}$ represents the importance of the joint node $V_i$, to the garment vertex $V_j$, $e_{gb}$ indicates the importance of the neighbor fabric vertex of the $V_j$ to the fabric vertex $V_j$, and $T$ is the neighbor node of order. Long path attenuation value is set to $\beta(\beta \leq 1)$ relative to the short path. Garment vertex decay value is set to $a(a \leq 1)$ relative to the joint node. For cloth vertices, there is only the $\beta$ parameter and no $a$ parameter. According to the priority of each node, it is stored as a $t$-order tree structure, and the subgraph with size $P$ is sampled in the tree structure. Compared with each sampling node, it is sorted according to the priority of the node, and the tree structure reduces the computational resources of sorting.

On the relationship of time and space, garment vertices use $M_g(\theta, d)$ garment model in time $t$ using $M_g(\beta, \theta, D)$ garment model in $t + 1$ time. On this sequence, we sample $p$ nodes in order. After the sampled nodes are input into the subgraph convolutional neural network as related nodes, the aggregation function $g(\cdot)$ is applied to the feature expression of the garment vertex, and the feature expression of the garment vertex is updated with the aggregation result, until a satisfactory $M_g(\theta, d)$ garment model is generated at the $t + 1$ time.

The selection of the loss function $L$ plays an important role in the training of subgraph convolutional neural networks. $d_p$ is a predictive value and $d_C$ is a true value.

$$
L(d_p, d_C) = \frac{1}{n} \sum_{i=1}^{n} \|d_p - d_C\|^2.
$$

### 3.6. Super-Resolution Network Algorithm

The structure of the super-resolution network model designed is shown in Figure 4. Given $2N + 1$ continuous low-resolution geometry image, the central high-resolution geometry image is predicted. This is an end-to-end network, which is composed of spatiotemporal feature progressive fusion (SFPF), multiscale feature extraction (MSFE), and reconstruction module. To avoid shallow features disappearing during propagation, all MSFE outputs are sent to the end of the network for reconstruction.

#### 3.6.1. SFPF Module

In order to reduce the number of parameters and extract spatiotemporal features at different levels, SFPF structure is proposed, as shown in Figure 5. Firstly, $3 \times 3$ convolutions are used to extract features at $L_0$ level, and then $3 \times 3$ convolutions with step size 2 are used to downsample them with a scaling factor of 2. For each three consecutive features $F_{t-1}$, $F_t$, and $F_{t+1}$, the features are gradually fused from high-level features to low-level features. Specifically, $L$ level features are spliced in the channel dimension. We use $1 \times 1$ convolutions to reduce the number of channels, further merge the time information of the features, and then upsample to $L - 1$ level feature size and fuse with $L - 1$ level feature. Repeat the above operations until $L_0$ level.

#### 3.6.2. MSFE Module

In order to improve ASPP [2], we use small receptive field features in parallel dilated convolution to gradually enrich information captured by large receptive field. As shown in Figure 6, the module consists of dilated convolution, local residual connection, and channel attention. The dilated rate of the dilated convolution is 2, 3, and 4, respectively. The first branch is divided into two groups $y_1$ and $y_2$, the channel dimension after $3 \times 3$ convolutions, $y_1$ splices to the final output, and $y_2$ splices to the next branch, using $1 \times 1$ convolutions to fuse features obtained by the next branch dilated convolution. Repeat the above operation; we use $1 \times 1$ convolutions to fuse different scale feature information, then the features are sent to residual block and attention module.

Since MSFE has treated time information as channel, this paper introduces the channel attention module in CBAM [20]. For a given input feature map $F \in \mathbb{R}^{C \times H \times W}$, the calculation formula is as follows:

$$
F' = M_s(F) \odot F,
$$

$$
M_s(F) = \sigma((MLP(Avg Pool(F))) + MLP(Max Pool(F)))
\begin{align*}
    &= \sigma(W_1(W_c(F_{avg}^c) + W_1(W_c(F_{max}^c)))), \\
    &= \left(W_c \in \mathbb{R}^{C \times r \times C}, W_1 \in \mathbb{R}^{C \times C}ight).
\end{align*}
$$

where $\odot$ represents the element-by-element multiplication, $\sigma$ represents the sigmoid function, $F_{avg}^c$ and $F_{max}^c$ represent the average pooling and maximum pooling of the channel dimension, respectively, and $r$ represents the reduction ratio. We used ReLU function as activation function.

#### 3.6.3. Geometry Image Reconstruction

After the high-resolution geometry image is obtained by network prediction, the adjacent pixels in the horizontal and vertical directions and the positive slope directions of the image are connected to obtain the 3D mesh model [21]. For garment, because the prediction operation is carried out on two cut meshes, respectively, the boundary points of the two high-precision meshes cannot overlap well, which will cause cracks in the reconstruction
mesh. Therefore, the boundary points of the two garment meshes are merged into one mesh, and the coordinates of the new points are taken as the mean values of the two coordinates before merging. In the experiment, we found that the reconstruction mesh can recover a large number of real folds. However, due to the prediction operation, some “noises” are inevitably generated, which is manifested as some unwanted disturbances near the fold-intensive region and the mesh boundary region. It affects the visual effect of reconstruction. Therefore, we use Laplacian smoothing with feature preservation for denoising, as shown in Figure 7.

4. Experiment

The experiment is implemented using Python language in Linux system. All the experiments in this paper are implemented on an Intel Core i7 9700 PC with 8 GB dual-channel memory and NVIDIA GeForce RTX 2070 Super graphics card. In this paper, three basic styles, long sleeves, short sleeves, and trousers, are selected to carry out experiments on male and female models, respectively, to generate instance data of garment animation. Flags, fabric, and ball collisions are used to generate instance data for fabric animation. Each dataset is trained separately, 80% as the training set and 20% as the test set.

4.1. Multi-Precision Fabric Construction. In this experiment, we calculate the Rayleigh quotient curvature of all fabric models, and the initial garment models of men and women were consistent. By specifying the amount of sampling, the sampling results are divided into low, medium, and high deformation regions by threshold. Table 1 shows the number of faces, vertex, sampling times, and facets after multi-precision division in different fabric models.

Figure 8 shows the results of Rayleigh quotient curvature partition for different fabric models. According to the Rayleigh quotient curvature, this experiment sets different division thresholds for different fabric models in the process of fabric division. The threshold values of long sleeves are set to 0.671 and 0.629, short sleeves to 0.714 and 0.686, trousers to 0.732 and 0.656, and flag to 0.137 and 0.128, and collisions between fabric and ball model threshold values are set to 0.173 and 0.167. In this paper, the high deformation fabric area is subdivided by LOOP, the medium deformation fabric area remains unchanged, and the coarsening rate of low deformation fabric area is set to 0.2. Using different subdivision threshold can control the proportion of high-precision mesh to the whole multi-precision mesh and meet the different requirements of simulation effect and efficiency in different application background.

It can be seen from Figure 8 that, for human-driven fabric deformation, the average deformation degree of Rayleigh quotient curvature can effectively measure the deformation degree of the mesh, and the fabric mesh can be divided into low, medium, and high deformation areas by dividing the threshold. By subdividing the high deformation fabric area, coarsening the low deformation fabric area, and keeping the medium deformation fabric area unchanged, the multi-precision fabric mesh dataset is finally constructed correctly. However, for the fabric model with external force
deformation such as flag and fabric ball collision, the average deformation degree of Rayleigh quotient curvature cannot well extract the fabric deformation characteristics.

4.2. Experimental Parameters

4.2.1. Subgraph Convolutional Neural Network. Since the joint nodes connected to the garment vertex selected in this paper are the nearest four joint nodes, the neighbor nodes connected to the joint nodes are also the neighbor nodes of the garment. Therefore, it is enough to sample the second-order neighbor nodes. Figure 9 shows that, under the same number of iterations, the larger the size of the subgraph, the larger the field of view, the larger the number of samples, the smaller the error value, and the more time and resources costed. When exceeding 445, the error value tends to flatten out. For the current machine performance, the maximum value of subgraph size is recommended as 2000. As the number of iterations increases, the influence of the value of subgraph size gradually decreases. In order to take into account performance and efficiency, the value of subgraph size is set as 145, and the number of iterations is set as 68 times.

4.2.2. Super-Resolution Network. The number of input continuous frames is set to 5, and each convolution layer is followed by the Leaky ReLU activation function. The number of multiscale feature extraction blocks is 6 (consistent with the number of ASPPs). Subpixel convolution [22] is used for upsampling, and the loss function is Charbonnier Loss [23]:

$$L = \sqrt{\left(I_{t}^{HR} - I_{t}^{LR}\right)^2 + \varepsilon^2},$$  \hspace{1cm} (16)$$

where $I_{t}^{HR}$ is the predicted pixel value of high-resolution geometry image; $\varepsilon = 10^{-3}$. Batch size is 16. Using Adam optimizer, initial learning rate is $10^{-4}$; at 50k, 100k, 200k, 300k iterations, learning rate decreased to half of the original rate.

4.3. Simulation Results. In this paper, the subgraph convolutional neural network and super-resolution network are trained to predict the fabric mesh, and the prediction results are simulated. In the simulation result graph, the blue is the simulation result of the subgraph convolutional neural network prediction, the yellow is the actual result of the physical simulation, and the pink is the simulation result of the super-resolution network prediction. The superposition part of blue and yellow is subgraph convolutional neural network prediction results and physical simulation results superimposed together. Figures 10(a)~10(f) show the simulation results of three basic garment styles (long sleeves, short sleeves, and trousers) on the male model and the female model, respectively. It can be observed that, in the garment mesh, the subgraph convolutional neural network has greater advantages and more clearly retains the garment wrinkles generated by physical simulation. For the garment wrinkles caused by joint movement near the garment vertex, it has a very realistic effect. For the garment deformation caused by far joint movement, the predicted garment fold appears relatively smooth as marked by the red circle in Figure 10(a). We stack the predicted garment with the actual garment simulated by physics, so as to show the contrast error of the single frame simulation results more vividly.

Figures 11(a)~11(b) show the simulation results of flag and fabric collision, respectively. It can be observed that, in fabric mesh, super-resolution networks produce large-scale
folds more similar to physical simulations on two datasets compared to graph convolutional neural networks. For areas with particularly dense folds, such as the upper right corner of the flag, the results generated by the super-resolution network are relatively smooth, which is not entirely bad. Because the physical simulation produces too many detail folds, it affects the visual quality to some extent. For flat regions and middle dense folds, the results of super-resolution network are basically consistent with the physical simulation.

4.4. Algorithm Comparison. In order to verify the superiority of the method in this paper, the two methods in this paper are compared with TailorNet [7] and ArcSim [24]. The subgraph convolutional neural network method with more advantages in the garment model is compared with TailorNet. The comparison of the average error and the maximum error of garment is shown in Table 2. The error of the method in this paper is slightly lower than that of TailorNet. Experiments show that garment model based on subgraph convolutional neural network not only retains the garment wrinkles based on physical simulation but also is 44 times faster than the physical simulation method. Compared with TailorNet, the average efficiency is improved by 25.3%, and average time consumption per frame is shown in Table 3.

The superiority of subgraph convolutional neural network method is also shown in the simulation of garment with no obvious movement trend. Because the method in TailorNet only depends on SMPL human motion joints, the simulation effect is poor for some garment deformation that is not driven by movement. This paper not only considers the garment deformation caused by joint rotation but also considers the surrounding vertices and space-time characteristics. For garment deformation that is not driven by movement, although our method is not as good as the actual results of physical simulation, compared with literature [7], we retained more obvious wrinkles. As shown in Figure 12(a), yellow garment is the actual garment of physical simulation; in Figure 12(b), blue garment is the garment predicted by our method; in Figure 12(c), red garment is the garment predicted by TailorNet method [7].

The super-resolution network method, which is more advantageous in the fabric model, is compared with ArcSim. Since the low-resolution geometry images in the training and testing are generated by high-resolution geometry images, in comparison, we simulate an additional set of low-precision meshes and then convert the mesh to geometry images. The number of vertices of the mesh is about one-sixteenth of the high-precision mesh. Table 4 lists the average time consumption per frame of ArcSim and our method. This paper counts the three most time-consuming parts of the whole process, namely, low-precision simulation, mesh parameterization, and sampling and generating high-resolution geometry images. For physical simulation, when the number of mesh vertices is quite different, the time-consuming gap between low-precision and high-precision simulations is very large, but the time consumption of this method is mainly concentrated in mesh parameterization. Therefore, for single-boundary regular fabric mesh simulation such as flag, the speed of this method is 14–16 times faster than that of ArcSim. When the number of mesh vertices increases further, the speed increase will be more obvious.

4.5. Ablation Experiment. In the super-resolution network method, in order to verify the effectiveness of the network structure and multiframe input, the comparative experiments are carried out for the two modules of SFPF and MSFE. In the experiment, the MAE (mean absolute error) of 10 geometry images is calculated on the test set as the evaluation index.

The first group of experiments uses the network structure including SFPF modules and removing SFPF modules, and the other experimental environments were the same. The experimental results are shown in Table 5. It can be seen that the MAE of using the SFPF module on all datasets was
Figure 10: Garment simulation results. (a) Men long sleeves. (b) Men short sleeves. (c) Men trousers. (d) Women long sleeves. (e) Women short sleeves. (f) Women trousers.

Table 2: Comparison of average error and maximum error.

| Error type            | Ours     | TailorNet |
|-----------------------|----------|-----------|
| Average error (mm)    | 8.56     | 13.29     |
| Maximum error (mm)    | 10.34    | 14.51     |

Table 3: Comparison of average time consumption per frame.

| Garment model | Way      | Average time spent per frame (ms) |
|---------------|----------|----------------------------------|
| Long sleeves  | Ours     | 1.7                              |
|               | TailorNet| 2.3                              |
|               | Physical | 74.4                             |
| Short sleeves | Ours     | 1.0                              |
|               | TailorNet| 1.4                              |
|               | Physical | 47.3                             |
| Trousers      | Ours     | 0.7                              |
|               | TailorNet| 0.9                              |
|               | Physical | 31.6                             |
lower than removing the SFPF module. This proves that the use of multiframe input time information was helpful for super-resolution reconstruction.

In the second group of experiments, use the network structure including MSFE modules and replace MSFE by ASPP, and the other structures are the same. Table 6 lists the experimental results, and it can be seen that MAE is reduced on all datasets. However, due to the fact that the large number of deformations of the mesh during the collision between cloth and balls is concentrated in specific dense areas, the corresponding geometry images only have large pixel value changes. Therefore, the information obtained by dilated convolution is limited, and the reduction of MAE is small.

5. Conclusion

In order to balance the relationship between efficiency and fidelity in fabric dynamic simulation modeling, this paper uses multi-precision meshes instead of single-precision meshes and proposes two multi-precision fabric simulation methods based on machine learning. The multi-precision fabric mesh preserves the fabric fold details, while reducing the number of mesh vertices and improving the algorithm efficiency. In this paper, the machine learning algorithm is implemented on the multi-precision fabric mesh for the first time, which provides a new idea for improving the efficiency of fabric modeling. The multi-precision fabric modeling method based on subgraph convolutional neural network is...
more suitable for clothing animation driven by human motion. It can be applied to regular grid problems. The multi-precision fabric modeling method based on super-resolution network is more suitable for fabric animation driven by external force. The mesh data is represented as an image for processing, which solves the problem that the first method relies on human joint parameters and cannot be used for general cloth. Experiments in different scenarios show that the fabric simulation speed based on super-resolution network is 14–16 times faster than the physical method, and the multi-precision fabric simulation efficiency based on subgraph convolutional neural network is 25.3% higher than the single-precision machine learning method. However, our method still has limitations. First, a lot of data needs to be processed in the early stages. Second, the image data is sampled from the grid, and there is an error with the original grid during grid reconstruction. Finally, the multi-precision clothing modeling in this paper is mainly on a single shape. Therefore, the future research goals of this paper will be to divide the topology of the fabric mesh, to study the folds of local details, to improve the efficiency of early data processing through classification, and to seek the relationship between multi-precision clothing deformation and human body shape changes. Models are applied on mannequins of different heights, shorts, fats, and thins.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Zhu, X. Jin, and J. Feng, "Overview of clothing animation simulation," Journal of Computer Aided Design and Graphics, vol. 016, no. 5, pp. 613–618, 2004.

[2] L. Wang, "Learning Parallax Attention for Stereo Image Super-Resolution," in Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12242–12251, Long Beach, CA, USA, June 2019.

[3] P. Guan, L. Reiss, and D. A. Hirshberg, "DRAPE: DRessing Any PErson," ACM Transactions on Graphics, vol. 31, no. 4, pp. 1–10, 2012.

[4] Y. Deng, T. Mao, and M. Shi, "Cloth Deformation Prediction Based on Human Motion," in Proceedings of the International Conference on Virtual Reality & Visualization, Hangzhou, China, September 2016.

[5] Y. Shi, D. T. Mao, Y. Deng, and S. Wang, "Study on the correlation between human motion and clothing deformation in clothing animation," Journal of Computer Aided Design and Graphics, vol. 29, 2017.

[6] I. Santesteban, M. A. Otaduy, and D. Casas, "Learning-Based Animation of Clothing for Virtual Try-On," Computer graphics forum: Journal of the European Association for Computer Graphics, vol. 38, 2019.

[7] C. Patel, Z. Liao, and G. Moll, "TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style," in Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, June 2020.

[8] E. Gundogdu, V. Constantin, and A. Seifoddini, "GarNet: A Two-stream Network for Fast and Accurate 3D Cloth Draping," in Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), February 2019.

[9] E. Gundogdu, V. Constantin, S. Parashar, A. Seifoddini, and M. Dang, "GarNet++: Improving Fast and Accurate Static 3D Cloth Draping by Curvature Loss," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44.

[10] N. Jin, Y. Zhu, Z. Geng, and R. Fedkiw, "A Pixel-Based Framework for Data-Driven Clothing," Computer Graphics Forum, vol. 39, no. 8, pp. 135–144, 2020.

[11] F. Scarselli, M. Gori, A. C. Ah Chung Tsoi, M. Hagenbuchner, and G. Monfardini, "The Graph Neural Network Model," IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61–80, 2009.

[12] Y. J. Oh, T. M. Lee, and I. Lee, "Hierarchical Cloth Simulation using Deep Neural Networks," in Proceedings of the Computer Graphics International, New York, NY, USA, June 2018.

[13] L. Chen, J. Ye, L. Jiang, C. Ma, Z. Cheng, and X. Zhang, “Synthesizing cloth wrinkles by CNN-based geometry image super-resolution,” Computer Animations and Virtual Worlds, vol. 29, no. 3-4, Article ID e1810, 2018.

[14] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, "Discrete Differential-Geometry Operators for Triangulated 2-Manifolds,” Mathematics and Visualization, Springer, Berlin, Germany, pp. 35–57, 2003.

[15] G. Michael and P. S. Heckbert, "Surface Simplification using quadric error metrics," in Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, NY, USA, August 1997.

[16] Z. Wang, Research on Key Technologies of Subdivision Surfaces, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2003.

[17] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, J Michael, and S. M. Black, "A skinned multi person linear model," ACM Transactions on Graphics, vol. 34, no. 6, p. 248, 2015.

[18] X. Gu, S. Gortler, and H. Hoppe, "Geometry Images," ACM Transactions on Graphics, vol. 21, no. 3, pp. 355–361, 2002.

[19] S. Yoshizawa, A. Belyaev, and H. P. Seidel, "A fast and simple stretch-minimizing mesh parameterization," in Proceedings Shape Modeling Applications, Genova, Italy, June 2004.

[20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Computer Vision-ECV 2018, Springer, Cham, pp. 3–19, 2018.

[21] J. Xue, Y. Zhang, and D. Kong, "Mesh Parameterizations and Geometry Images," Journal of System Simulation, vol. 24, no. 9, pp. 1780–1784, 2012.

[22] W. Shi, J. Caballero, F. Huszár, J. Totz, A. Aitken, and R. Bishop, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874–1883, Las Vegas, NV, USA, June 2016.

[23] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 624–632, San Juan, PR, USA, June 2017.

[24] R. Narain, A. Samii, and J. F. O’Brien, “Adaptive anisotropic remeshing for cloth simulation,” ACM Transactions on Graphics, vol. 31, no. 6, pp. 1–10, 2012.