Deep Detail Enhancement for Any Garment

Meng Zhang[1] (iD) Tuanfeng Wang[2,3] (iD) Duygu Ceylan[2] (iD) and Niloy J. Mitra[1,2] (iD)

[1]University College London [2]Adobe Research [3]Mihoyo Inc.

(a) coarse simulation (b) our result (c) failed high resolution simulation (d) skinning based deformation (e) our result
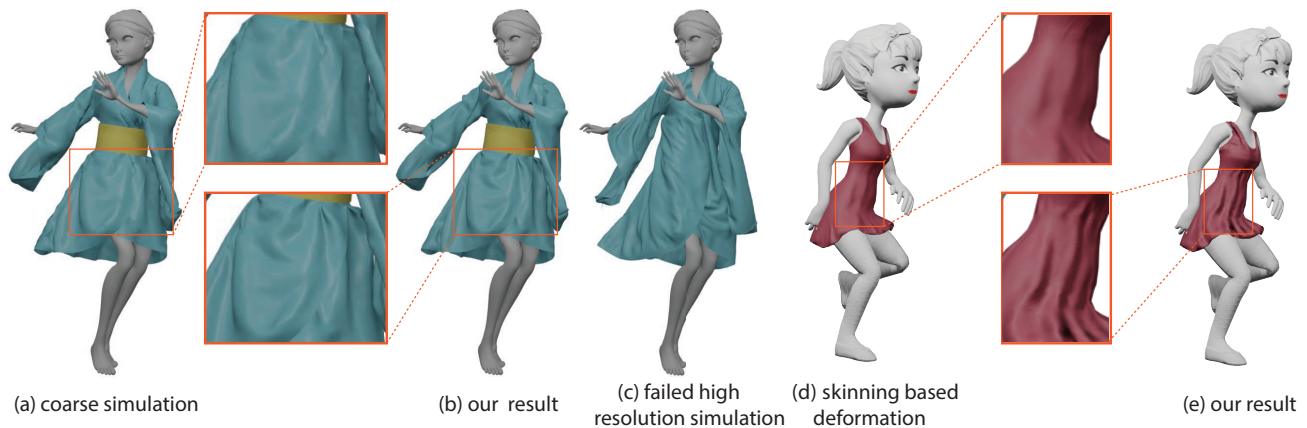
Figure 1: *We present a data-driven approach to synthesize plausible wrinkle details on a coarse garment geometry. Such coarse geometry can be obtained either by running a physically-based simulation on a low resolution version of the garment (a), or by using a skinning-based deformation method (d). Our method can generate fine-scale geometry (b,e) by replacing the expensive simulation process, which, in some cases, are not even currently feasible to setup. For example, high-resolution simulation failed (c) for the blue dress as boundary conditions between the blue dress and the yellow belt were grossly violated. Our method generalizes across garment types, often with different number of patterns, associated parameterization and materials, and undergoing different body motions.*

Abstract

Creating fine garment details requires significant efforts and huge computational resources. In contrast, a coarse shape may be easy to acquire in many scenarios (e.g., via low-resolution physically-based simulation, linear blend skinning driven by skeletal motion, portable scanners). In this paper, we show how to enhance, in a data-driven manner, rich yet plausible details starting from a coarse garment geometry. Once the parameterization of the garment is given, we formulate the task as a style transfer problem over the space of associated normal maps. In order to facilitate generalization across garment types and character motions, we introduce a patch-based formulation, that produces high-resolution details by matching a Gram matrix based style loss, to hallucinate geometric details (i.e., wrinkle density and shape). We extensively evaluate our method on a variety of production scenarios and show that our method is simple, light-weight, efficient, and generalizes across underlying garment types, sewing patterns, and body motion.

CCS Concepts
• *Deep Learning → Motion Generation; Garment Motion Transfer;*

1. Introduction

High fidelity garments are a vital component of many applications including virtual try-on and realistic 3D character design for games, VR, and movies. In real life, garments, as they undergo motion, develop folds and rich details, characteristic of the underlying materials. Naturally, we desire to carry over such realistic details to the digital world as well.

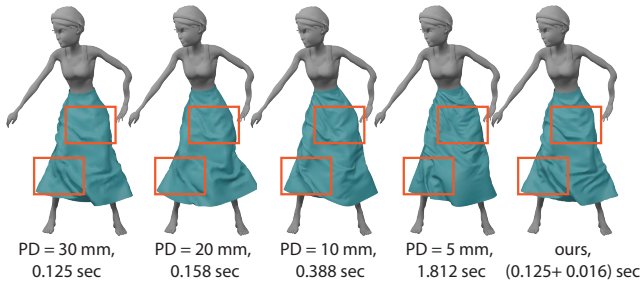Directly capturing such fine geometric details requires profes-

| PD = 30 mm, 0.125 sec | PD = 20 mm, 0.158 sec | PD = 10 mm, 0.388 sec | PD = 5 mm, 1.812 sec | ours, (0.125+ 0.016) sec |

Figure 2: *We show a garment simulated with different mesh resolutions, i.e., varying particle distances (PD), and report the computation time for simulating a single frame. While wrinkle details are captured well when simulating high resolution meshes, i.e., lower PD, computation time significantly increases. Our method captures statistically similar quality details by enhancing a coarse simulation much more efficiently (13x speedup here).*

sional capture setups along with computationally-heavy processing pipeline [BPS*08; CZL*15; PPHB17], which may not be applicable for low-budget scenarios. An alternate option involves physically-based cloth simulation [NSO12; TWL*18; LLK19] to virtually replace the need for physical setups. However, despite many research advances, achieving high quality and robust simulation still has many challenges. The simulation is often sensitive to the choice of parameters and initial conditions and may require a considerable amount of manual work to setup correctly. Even when setup, it remains a computationally expensive process, especially as the resolution of the garment geometry increases (see Figure 2). Further, with increasing complexity of the garment geometry, e.g., garments with many layers or with non-canonical UV layout, it may be difficult to obtain any stable result (see Figure 1(c)). Finally, for each new garment type and/or motion sequences, majority of the setup and simulation steps have to be repeated.

We propose a different approach. As garment undergoes motion, its geometry can be analyzed in two steps. First, the corresponding motion of the body (i.e., the body pose) changes the global shape of the garment, e.g., its boundary and coarse folds. We observe that the global shape to be consistent across different resolutions of a garment (see Figure 2). Such a global geometry can either be simulated efficiently and reliably using a coarsened version of the garment, or by utilizing a skinning based deformation method. Given the coarse geometry, the high frequency wrinkles can be perceived by considering only local neighborhood information. As shown in Figure 2, the highlighted local patches in the coarse garment geometry provide sufficient cues to hallucinate plausible high frequency wrinkles in the same region. Hence, we hypothesize that given the coarse geometry of a garment, the dynamics of the detailed wrinkles can be learned from many such local patches observed from different type of garments undergoing different motions.

In this work, we present a deep learning based method that enhances an input coarse garment geometry by synthesizing realistic and plausible wrinkle details. Inspired by the recent success of image-based artistic style transfer methods [GEB16], we represent the garment geometry as a 2D normal map where we interpret the

coarse normal map as the *content* and fine geometric details as a *style* that can be well captured by the correlation between VGG features captured in the form of Gram matrices, characteristic to different materials. Furthermore, in order to tackle various garment materials within a universal framework, we adopt the conditional instance normalization technique [DSK16] that is proposed in the context of universal style transfer. Finally, we design our detail enhancement network to operate at a patch level enabling generalization across different (i) garment types, (ii) sewing patterns, and (iii) underlying motion sequences. In fact, we show that our network trained on a dataset generated from only a single sheet of cloth by applying random motion sequences (e.g., applying wind force, being hit by a ball) can already reasonably generalize to various garment types such as a skirt, a shirt, or a dress. See Figure 12 for a comparison.

We evaluate our approach on a challenging set of examples demonstrating generalization both across complex garment types (e.g., with layering) as well as a wide range of motion sequences. In addition to enriching low resolution garment simulations with accurate wrinkles, we also demonstrate that our method can enhance coarse garment geometries obtained by skinning methods as often used by computer games, see Figure 1. To our knowledge, we present the first learning-based method that casts garment detail enhancement as a style transfer problem and achieves generalization across unseen garment types, sewing patterns, and body motions all at the same time.

2. Related Work

Realistic and physically based cloth simulation has been extensively studied in the graphics community [CK05; NMK*06; NSO12; LLK19; YZZ*19; TWL*18]. However, due to the associated computational cost and stability concerns, as the complexity of the garments increase, several alternative avenues have been explored. For example, in the industrial settings, skinning based deformation of garments is often favored for its simplicity, robustness, and efficiency. This, however, comes at the expense of losing geometric details. Finally, there has been various efforts to augment coarse simulation output with fine details using various approaches including optimization, data-driven, and learning-based methods.

Optimization-based methods. Müller et al. [MC10] use a constraint-based optimization method to compute a high resolution mesh from a base coarse mesh. Rohmer et al. [RPC*10] uses stretch tensor of the coarse garment animation as a guide to place wrinkles in a post-processing stage. However, they assume the underlying garment motion is smooth which is not always the case with character body motion. Gillette et al. [GPV*15] extends similar ideas to real time animation of cloth.

Data-driven methods. Feng et al. [FYK10] present a deformation transfer system that learns a non-linear mapping from a coarse garment shape to per-vertex displacements that represent fine details. Zurdo et al. [ZBO12] share a similar idea to learn the mapping from a low-resolution simulation of a cloth to high resolution where plausible wrinkles are defined as displacements. Given a set of coarse and mid-scale mesh pairs, Kavan et al. [KGBS11] learn linear upsampling operators. Wang et al. [WHRO10] construct a database of high resolution simulation examples generated by moving each

body joint separately. Given a coarse mesh, wrinkle regions for each joint are synthesized by interpolating between database samples and blended into a final mesh. However, this method is limited to tight fitting clothing. One of the most comprehensive system in this direction is the DRAPE [GRH*12] system that learns a linear subspace model of garment deformation driven by pose variation. Hahn et al. [HTC*14] extend this idea by using adaptive bases to generate richer deformations. Xu et al. [XUC*14] introduce a pose-dependent rigging solution to extract nearby examples of a cloth and blend them in a sensitive-based scheme to generate the final drape.

Learning-based methods. With the recent success of deep learning methods in various imaging and 3D geometry tasks, a recent research direction is to *learn* garment deformations under body motion using neural networks. A popular approach has been to extend parametric human body models with a per-vertex displacement map to capture the deforming garment geometry [AMB*19; BTTP19; PPHB17; JZGF18; TBTP20; JZH*20]. While this is a very efficient representation, it only works well for tight clothes such as t-shirt or pants that are sufficiently close to the body surface. Shen et al. [SLL20] use a GAN-based method to generate garment models by transferring the sewing patterns of different garments to the UV space of the underlying garment. Gundogdu et al. [GCS*19] present GarNet which learns features of garment deformation as a function of body pose and shape at different levels (point-wise, patch-wise, and global features) to reconstruct the detailed draping output. Santesteban et al. [SOC19] learn a garment deformation model via recurrent neural networks that enables real-time virtual try-on. Their approach learns the coarse garment shape based on the body shape and the detailed wrinkles based on the pose dynamics. A similar two-stage strategy is proposed by Patel et al. [PLP20] which decomposes the deformation caused by body pose, shape, and garment style, into high frequency and low-frequency components. Wang et al. [WSFM19] also consider garment deformations due to material properties in their motion-driven network architecture. Others [YFHW18; WCPM18; HDDN19] investigate a subspace technique such as principal component analysis, to reduce the number of variables of the high dimensional garment shape space. While these learning-based methods are efficient and fully differentiable compared to standard physically-based simulation, they are hard to generalize across different garment types and associated sewing patterns. In this work, we seek to combine the best of both worlds, physically-based simulation and learning-based approach. Specifically, we rely on physically-based simulation for fast coarse garment generation, we synthesize fine geometry conditioned on the coarse shape by a neural network. By utilizing a patch-based approach, and training on synthetic data, we demonstrate that our trained model generalizes across different garment types as well as different material configurations, and also extends to coarse geometry obtained using linear blend skinning.

One of the closest learning based methods to our approach is DeepWrinkles [LCT18] which augments a low-resolution normal map of a garment captured by RGB-D sensors with high-frequency details using PatchGAN. Instead of training a discriminator via a generative adversarial setup, our method casts the detail enhancement as a style transfer problem and matches Gram matrix based features. We observe that this avoids oversmoothing details. We use conditional instance normalization to further adapt to different

material specifications. In Section 5, we compare our method to DeepWrinkles and demonstrate superior performance.

Generalizing learning based methods to different garment types with potentially different number of vertices, connectivity, and material properties is an important challenge. In order to address this challenge, the recent work of Ma et al. [MYR*19] adopts a one-hot vector to encode the *pre-defined* garment type and trains a graph convolutional neural network with a generative framework to deform the body mesh into the garment surface. Zhu et al. [ZCJ*20] train an adaptable template for image based garment reconstruction and introduce a large garment dataset with more than 2*K* real 3D garment scans spanning over 10 different categories. This method generates plausible results if an initial template with a reasonable topology is provided. Alldieck et al. [APTM19] use normal and displacement maps based on a fixed body UV parameterization to represent the surface geometry of static garments. The fixed parameterization assumption makes it difficult to generalize the method to new garment designs, especially more complex garments such as multi-layer dresses. We also utilize a normal map representation, however, our patch-based method focuses on local shape variations and is independent of the underlying UV parameterization.

Image-to-image transfer. We draw inspiration from the recent advances in image-to-image transfer methods [LYY*17; FJL*16; HB17; IZZE17] as we represent the garment geometry as a normal map and cast the problem as detail synthesis in this 2D domain. Specifically, we follow the framework proposed by [GEB15; GEB16] that captures the style of an image by the Gram matrices of its VGG features [SZ14]. Similarly, we assume that a coarse normal map can be enhanced with fine scale wrinkles by matching the feature statistics of Gram matrices. Our work can be loosely related to single image super-resolution *(SISR)* where, the aim is to synthesize high-resolution images from blurry counterparts [LTH*17; WLY*15]. In contrast with typical SISR approaches, the criteria adopted in this work is based on fine geometry simulated with high 3D resolution instead of the appearance captured with high 2D resolution.

3. Overview

Our goal is to augment a coarse garment geometry source sequence, $\mathbb{S}$, with fine details such as plausible and realistic wrinkles. We assume that $\mathbb{S}$ is obtained by either a physically-based simulation run on a low-resolution version of the garment or obtained by computationally fast approaches such as skinning based deformation. The output of our pipeline is a fine-detailed surface sequence, $\mathbb{S}^*$, which depicts the source garment with realistic wrinkles in a temporally-consistent manner. The fine wrinkles in $\mathbb{S}^*$ are synthesized either by *sharpening* the original coarse folds in $\mathbb{S}$ or *synthesized* conditioned on nearby coarse folds.

Motivated by the recent advances in image style transfer [GEB16], we cast the 3D wrinkle synthesis task as a *2D detail transfer* problem defined on the normal maps of the garments. Specifically, we present a learning-based method to learn an image-to-image transfer network Φ that synthesizes a detail-rich normal map $\mathbb{N}^*$ corresponding to $\mathbb{S}^*$ given the coarse normal map $\mathbb{N}$ of $\mathbb{S}$ by matching the feature statistics using Gram Matrices (Section 4.1).
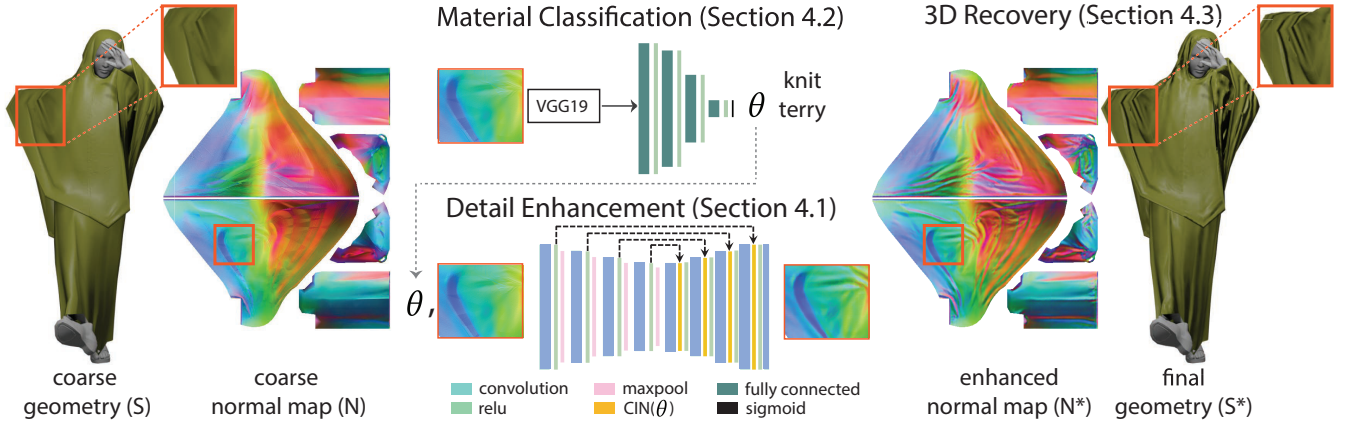
Figure 3: *Given a coarse garment geometry $\mathbb{S}$ represented as a 2D normal map $\mathbb{N}$, we present a method to generate an enhanced normal map $\mathbb{N}^*$ and the corresponding high resolution geometry $\mathbb{S}^*$. At the core of our method is a detail enhancement network that enhances local garment patches in $\mathbb{N}$ conditioned on the material type θ of the garment. We combine such enhanced local patches to generate $\mathbb{N}^*$ which captures fine wrinkle details. We lift $\mathbb{N}^*$ to 3D to generate $\mathbb{S}^*$ by an optimization method that avoids interpenetrations between the garment and the underlying body surface. In case the garment material is not known in advance, we present a material classification network that operates on the local patches cropped from the coarse normal map $\mathbb{N}$.*

However, unlike previous work [LCT18], which focuses on learning garment-specific models using PatchGAN, our goal is to train a *universal detail transfer network* that can generalize over different types of garments (e.g., dress, shirt, skirt) made of different materials (e.g., silk, leather, matte) and undergoing different body motions (e.g., dancing, physical activities). First, in order to ensure generalization across different materials and body motions, we adopt the conditional instance normalization (CIN) approach, introduced by Dumoulin et al. [DSK16] in the context of artistic style transfer. Specifically, given the material type θ of a garment, we shift and scale the learned network parameters to adapt to different materials and motions. Given a sequence of coarse garment geometry, we predict its material type θ via a classifier, and do not require the material specification as part of the input. Second, in order to generalize across different types of garments and different uv-parameterizations utilized to generate the normal maps of the garments, we propose a novel patch-based approach. Specifically, our network works at a patch level by synthesizing details on overlapping patches randomly cropped from the normal map, $\mathbb{N}$, of the garment, $\mathbb{S}$. We merge the resulting detailed patches to obtain the final normal map, $\mathbb{N}^*$, which is lifted to 3D via a gradient-based optimization scheme to yield $\mathbb{S}^*$ (Section 4.3). Our pipeline is illustrated in Figure 3.

4. Algorithm Stages

Given a coarse 3D garment geometry sequence $\mathbb{S}$, for each frame $\mathbb{S}_i \in \mathbb{S}$, our goal is to synthesize $\mathbb{S}_i^*$, which enhances the garment geometry with fine wrinkle details, subject to the underlying body motion and the assigned garment material θ (see Section 4.2). Note that given the nature of garments, any $\mathbb{S}_i$ can be semantically cut into a few developable pieces, i.e., garment patterns. We assume such a parameterization is provided for the first frame $\mathbb{S}_0$ and applicable to the remaining frames in the sequence.

4.1. Detail Enhancement

Using the garment-specific parameterization, we represent each $\mathbb{S}_i$ by a normal map $\mathbb{N}_i$. We treat the detail enhancement problem as generating a normal map $\mathbb{N}_i^*$ conditioned on $\mathbb{N}_i$ and the material characteristics θ. While $\mathbb{N}_i^*$ preserves the content of $\mathbb{N}_i$ (e.g., the density and the position of the wrinkles), the appearance of the details (e.g., shape, granularity) are learned from $\{\mathbb{N}_\theta\}$, a collection of normal maps depicting the detailed geometry of various garments of material θ undergoing different body motions. The detail enhancement step operates at each frame, hence, for simplicity, we ignore the subscript i in the following.

Detail Enhancement Network. We adopt a U-net architecture for our detail enhancement network (see Figure 3). The encoder projects the input $\mathbb{N}$ into a learned latent space. The latent representation $\mathbf{z}$ is then used to construct the output $\mathbb{N}^*$ via the decoder. Although garments made of different materials may share similar coarse geometry, the fine details can vary, conditioned on the material. To adapt our approach to different garment materials and capture fine details, we use θ, a one-hot vector, as input to indicate the type of material $\mathbb{N}$ and $\mathbb{N}^*$ are associated with. We design the decoder using conditional instance normalization (CIN) [DSK16] to adapt for different material types. Specifically, given a material type θ, the activation of a layer i is converted to a normalized activation $\mathbf{z}_i^\theta$ specific to a material type θ according to,

$$\mathbf{z}_i^\theta := \gamma_i^\theta \left(\frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} \right) + \varepsilon_i^\theta,$$

where μ and σ are the mean and standard deviation of $\mathbf{x}_i$, and γ_i^θ and ε_i^θ are material-specific scaling and shifting parameters *learned* for the corresponding layer i.

Loss Function. In general, the loss function for a style transfer task is composed of two parts: (i) content signal, which in our case penalizes the difference between $\mathbb{N}^*$ and $\mathbb{N}$ with respect to
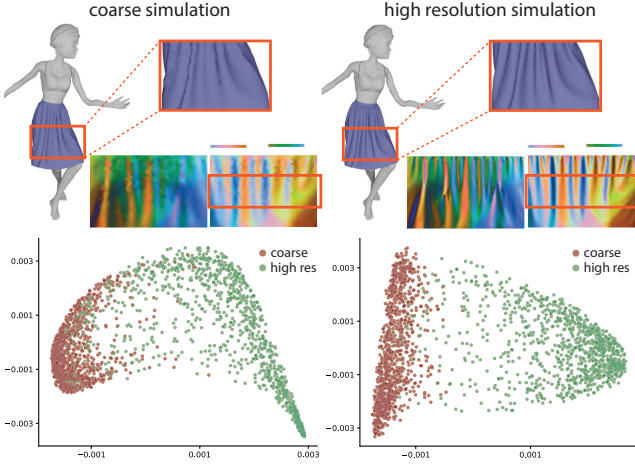
Figure 4: *We pass a given set of normal map patches cropped from coarse and high resolution simulation results through VGG19. Here, we show spectral embedding in $\mathbb{R}^2$ of the feature outputs of candidate layers, and select the layers where the two sets that differ significantly to compute the style loss. We show the feature plots of two such layers. This difference is largely responsible for the drop in visual fidelity going from high resolution to coarse simulation.*

the coarse geometry features; and (ii) style term, which in our case penalizes the perceptual difference between $\mathbb{N}^*$ and $\{\mathbb{N}_\theta\}$, a set of training examples that depict detailed garment geometries of material θ. As indicated by Gatys et al. [GEB16], content and style losses can be formulated using the output of a predefined set of layers, L_C and L_S, of a pre-trained image classification network, e.g., VGG19 [SZ14]. Note that the definition of content and style in our problem is not the same as in a classical image style transfer method. We performed a qualitative study to find the appropriate layers L_C and L_S. Specifically, we pass examples of two sets of normal maps that depict coarse and fine geometry of garments of a particular material through VGG19. For candidate layers, we embed the output features in 2D space using a spectral embedding method and identify the layers where the two sets are well separated (see Figure 4).

Given L_C and L_S, we define content loss between $\mathbb{N}$ and $\mathbb{N}^*$ as

$$L_{content}(\mathbb{N}^*|\mathbb{N}) := \sum_{l \in L_C} [F_l[\mathbb{N}^*] - F_l[\mathbb{N}]]^2/2,$$

where F_l is the VGG features of the selected layer L_C.

We define the style loss based on the correlation of the VGG features across different channels by using *Gram* matrices. Since the input to our network is a normal map that is generated based on a UV parameterization of the garment, it consists of background regions which do not correspond to any UV patch. In order to avoid such empty regions impacting the correlation measure, we adopt a spatial control neural style transfer method by using a UV mask M that masks out the background regions. We capture style by a set of Gram Matrices defined for each L_S layer of VGG:

$$G_l^M(\mathbb{N}) := (B_l^M \circ F_l(\mathbb{N}))^T (B_l^M \circ F_l(\mathbb{N})),$$

where B_l^M is the background mask for layer $l \in L_S$ downsampled accordingly. The style loss is then formulated as:

$$L_{style}(\mathbb{N}^*|\{\mathbb{N}_\theta\}, M) :=$$
$$\sum_{k \in \|\{\mathbb{N}_\theta\}\|} \sum_{l \in \{L_S\}} \frac{1}{4R_l^2} [G_l^M(\mathbb{N}^*) - G_l^M(\mathbb{N}_\theta^k)]^2,$$

where R_l is the number of valid pixels in the down-sampled mask B_l^M.

The final loss function we use to train the network is a weighted sum of the content and style losses:

$$\text{Loss}_\Phi := w_c L_{content}(\mathbb{N}^*|\mathbb{N}) + w_s L_{style}(\mathbb{N}^*|\{\mathbb{N}_\theta\}, M),$$

where w_c and w_s denote the weight of each term respectively. In our tests, we use $w_c = 1$ and $w_s = 10^4$.

Patch-based generalization. Generalizing a CNN based neural network to different types of clothes with unknown UV parameterization is challenging. Particularly, similar 3D garments (e.g., a short and a long skirt) might potentially have very different parameterizations due to different sewing patterns that vary in the number of 2D patterns and their shape. In order to generalize our approach across such varying 2D parameterizations, we adopt a patch-based approach. Specifically, instead of operating with complete normal maps, we use patches cut from the normal maps as input and output of our network. While we use randomly cut patches during training, at test time we use overlapping patches sampled regularly on the image plane. We resize each input patch to ensure the absolute size of one pixel remains unchanged. The output patches that are enhanced by our network are then combined to generate the final detailed normal map $\mathbb{N}^*$ by averaging the RGB values in the overlapping regions of the output patches.

4.2. Material Classification

We introduce a patch-based material classification method to predict the material properties in such cases. Specifically, we train a patch-based material-classification network, Π, which is composed of
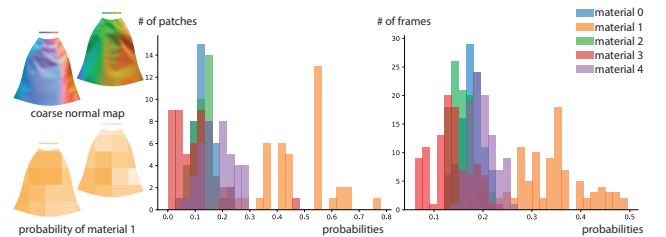

Figure 5: *For an input coarse normal map, we crop 42 patches as the input of our material classification network. For each material type in our dataset, We plot the histogram of patches based on the probability of them belonging to that material. As shown, material 1 has the highest distribution among the five materials. We also color code the patches based on their probability of belonging to material 1, where darker color means a higher probability. When we evaluate our voting strategy over a sequence of 100 frames, material 1 gets the highest number of votes for each frame.*

4 fully-connected layers as shown in Figure 3. The input to this network is the flattened feature vector of VGG calculated from a patch cropped from $\mathbb{N}$. The dimension of the last layer is the number of the materials captured in our dataset. The last layer is followed with a *Sigmoid function* and a normalization layer to output the probability of the input patch belonging to each material. During training, we use *Cross-Entropy* loss between the predicted probabilities and the ground truth material type.

Humans perceive the characteristics of a material when they observe a garment in motion. Similarly, we observe that performing material classification on a single patch is not sufficiently accurate [YLL17]. Hence, for a given input garment sequence, we choose randomly cropped patches and run material classification on each of them sampled over time. We use a voting strategy to decide the final material type from all such individual predictions as shown in Figure 5. We also experimented with providing concatenated sequential data as input to our material classification network and obtained similar results. As opposed to providing concatenated sequential input, our voting strategy enables sampling as many patches as desired over time without increasing the network capacity. We trained our material classification network on synthetically generated coarse simulations where we have access to ground truth material properties. At test time, we apply the network on both coarse simulations and inputs that are generated by skinning. For skinning data, since there is no available physical material parameter, the output of our network has lower confidence. Nevertheless, the automatic classification frees the user from having to guess the underlying material parameters.

4.3. 3D Recovery

Once the enhanced normal map $\mathbb{N}^*$ is calculated, we finally lift the details into 3D. Since coarse geometry $\mathbb{S}$ is known, one common solution in the CG industry is to bake the normalmap onto the original coarse surface during rendering. The skinning based examples shown in Figure 1 and 8 are rendered using this approach. On the other hand, real 3D results are useful in many follow up applications, such as animation, texture painting, etc. We adopt a two-stage approach to recover the detailed 3D geometry for such applications.

Normal map guided deformation. We perform upsampling over the surface of the coarse geometry $\mathbb{S}$ and deform the sampled vertices P based on the enhanced normal map $\mathbb{N}^*$. Our method uses an iterative gradient optimization scheme to find the deformed vertex positions p which minimizes the following energy:

$$\underset{\{P\}}{\arg\min} \sum_{p \in \{P\}} \sum_{q \in R_p^1} \|n_p \cdot \frac{q-p}{\|q-p\|}\|^2 + \eta \sum_{p \in \{P\}} \|\Delta p\|^2$$
$$+ \omega \sum_{p \in \{P\}} \|p_0 - p\|^2,$$

where n_p is the normal of vertex p as denoted on $\mathbb{N}^*$, R_p^1 is the one-ring neighborhood of p, Δ is the Laplacian operator, and p_0 is the initial position of p obtained by upsampling the coarse geometry, and η and ω are the weights of the last two regularization terms. While the Laplacian term acts as a smoothness term to avoid sharp

deformations, the last term penalizes the final shape from deviating from the original geometry (e.g., by shifting or scaling).

Avoiding interpenetration. The above deformation approach may cause interpenetrations between the garment and the underlying body surface. Hence, we adopt a post-processing step to resolve such penetrations as introduced by Guan et al. [GRH*12]. Specifically, we first find the set of garment vertices $P_C \in P$, which are inside the body mesh. We project each such vertex $p \in P_C$ to the nearest location, p^*, on the body surface. We update the vertex positions p such that the following objective function is minimized:

$$\underset{\{p\}}{\arg\min} \sum_{p \in \{P_C\}} \|p^* - p\|^2 + \phi \sum_{p \in \{P\}} \|\nabla p\|^2.$$

In our experiments, we use $\eta = 5 \times 10^3$, $\omega = 8$, and $\phi = 5 \times 10^3$. The objective functions are optimized by a standard gradient descent algorithm.

4.4. Data Generation

Our goal is to generalize our method to a large variety of garment types and motion sequences. To ensure such a generalization capability, we generate our training data by performing physically-based simulation over 3 different garment-motion pairs (see Figure 6). These are:

1. a piece of hanging sheet interacting with 2 balls and blowing wind (901 frames at 30 fps);
2. a long skirt over a virtual character performing Samba dance (988 frames at 30 fps);
3. a pleated skirt over a virtual character performing modern dance (850 frames at 30 fps).

For the last two cases, the underlying virtual character is created by Adobe Fuse (https://www.adobe.com/products/fuse.html) and the motion sequences are obtained from Mixamo (https://www.mixamo.com/). We use Marvellous Designer (https://www.marvelousdesigner.com) to drape the garments over the 3D character body and perform the physically-based simulation. We simulate each garment with different material samples (silk chamuse, denim lightweight, knit terry, wool melton, silk chiffon) and in two resolution settings. To obtain the coarse geometry, we run the simulation with a particle distance of 30*mm*. This produces meshes of $4440, 2287, 1756$ vertices for the aforementioned 3 types of garments respectively. To generate the fine scale geometry, we run the simulation with a particle distance of 10*mm* which results in meshes with $39330, 18950, 14433$ vertices respectively. The normal map of the 3D garment surface is then calculated. For each example, we generate the corresponding normal map using a given uv parameterization for each garment type. We scale the normal maps of different examples to ensure a single pixel corresponds to the same unit in 3D and randomly cut patches from the normal maps and use each pair of non-empty coarse and high resolution patches as input and target of our network. Finally, we also augment our dataset by rotating and reflecting the patches.

We observe that, sometimes due to instabilities in the physical simulation, simulation results with different particle distances may reveal geometric variations, e.g., in the position and the shape of the wrinkles. This results in input and output patch pairs that are
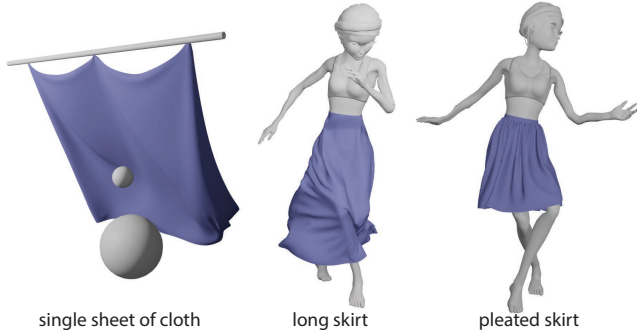
single sheet of cloth   long skirt   pleated skirt

Figure 6: *Our base training data is simulated from a single sheet of cloth interacting with two balls and blowing wind. We also simulate a long skirt and a pleated skirt under different body motions. For each training example, we simulate both a coarse and a fine resolution mesh and obtain the corresponding normal map pairs.*

not well aligned. To avoid this problem, we instead downsample the normal maps of the high resolution mesh to generate the coarse normal maps that will be provided as input to our network. During testing, the inputs to our network are obtained from the coarse simulation results. Our experiments demonstrate that training with the downsampled normal maps while predicting with the coarse normal maps as input at test is a successful strategy.

Network architecture. Our network broadly consists of an encoder and a decoder. There are 4 layers in the encoder to project the input patch image to a compact latent space, by down-sampling from 128×128 to 8×8. In each layer, there are blocks in the order of 2D convolution, Relu, and maxpool. The decoder is used to up-sample the patch tensor from 8×8 to the original size, mainly using 2D trans-convolutions, and a 2D convolution in the last layer. Because different materials share one network, we adopt CIN layer to scale and shift the latent code to the specified material related space, after every 2D trans-convolution. Finally, we use skip connection, as we regard our problem as constructing the residual for the fine image based on the coarse one. In our experiments, we use the Adam optimizer, with a learning rate of $1e-4$ and parameters $\beta_1 = \beta_2 = 0.9$. Project code and data will be released.

5. Results

We evaluate our method on the product space of a variety of garments undergoing various body motions. We present qualitative and quantitative evaluation, comparison with baseline methods, and also ablation study.

5.1. Dataset

As input, we test on coarse garment geometries obtained by either running a physically-based simulation on a low resolution mesh of the garment, as well as garments deformed via a skeleton rig using *linear blend skinning* (LBS). Similar to the dataset generation, we use motion sequences from Mixamo and utilize Marvelous Designer(MD) as our physical simulator. As seen in the inset figure, although the resolutions of the LBS meshes are much coarser

compared to our training data, we scale the normal maps to ensure the pixels from different normal maps are representing the same absolute area over the 3D surface. Our method generalizes well to these cases as well. For skinning-based examples, we test our method on a character from Mixamo, as well as a character from the Microsoft Rocketbox dataset (https://github.com/microsoft/Microsoft-Rocketbox). We show qualitative examples in Figures 1 and 8.

5.2. Evaluation

Qualitative evaluation. The examples demonstrate the generalization capability of our method to *unseen garments and motions* during training time. Even for garment types that are significantly different from our training data, e.g., the multi-layer skirt in Figures 1, 7, and the hijab in Figure 8, our network synthesizes plausible geometric details an order of magnitude faster than running a physically based simulation with a high resolution mesh (see Figure 2).

coarse input in training stage   coarse input in testing

When evaluating our method on skinning-based rigged characters, since no garment material information is provided, we use our material classification method to first predict plausible material parameters based on the input normal maps. We then utilize this material information to synthesize vivid wrinkles that enhances the input garment geometry. Empirically, even when material specification was available, we found training end-to-end with the material classification network along with CIN resulted in faster convergence [HB17].
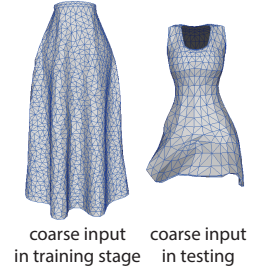
Material properties of a garment plays an important role in the resulting folds and wrinkles. Our method captures such deformations by taking the material parameters as an additional input. In Figure 9, we show examples of garments made from different materials undergoing the same body motion. Our method is able to generate plausible details in all cases that well respect the input coarse geometries.

Quantitative evaluation. In order to evaluate our approach quantitatively, when ground truth data is available, we define an *improvement score* given an input patch, $(\mathbb{N})$, an output patch, $(\mathbb{N}^*)$, and the corresponding ground truth patch, $(\hat{\mathbb{N}})$, as follows:

$$\text{improvement score} := 100 \left| \frac{L_{style}(\mathbb{N}) - L_{style}(\mathbb{N}^*)}{L_{style}(\mathbb{N}) - L_{style}(\hat{\mathbb{N}})} \right|.$$

A higher improvement score, 100 being perfect, indicates that the output patch is closer to the ground truth in terms of the styles of the wrinkles. Table 1 reports the mean and standard deviation of the improvement score for three different garment types made from two different materials and undergoing two different motion types. While one garment and motion type has been seen during training (dress A, motion 1), the remaining garment types and the motion sequence are unseen. Not surprisingly, the method performs well ($\sim 98\%$) on seen garment types and motions. The main result is the generalization to
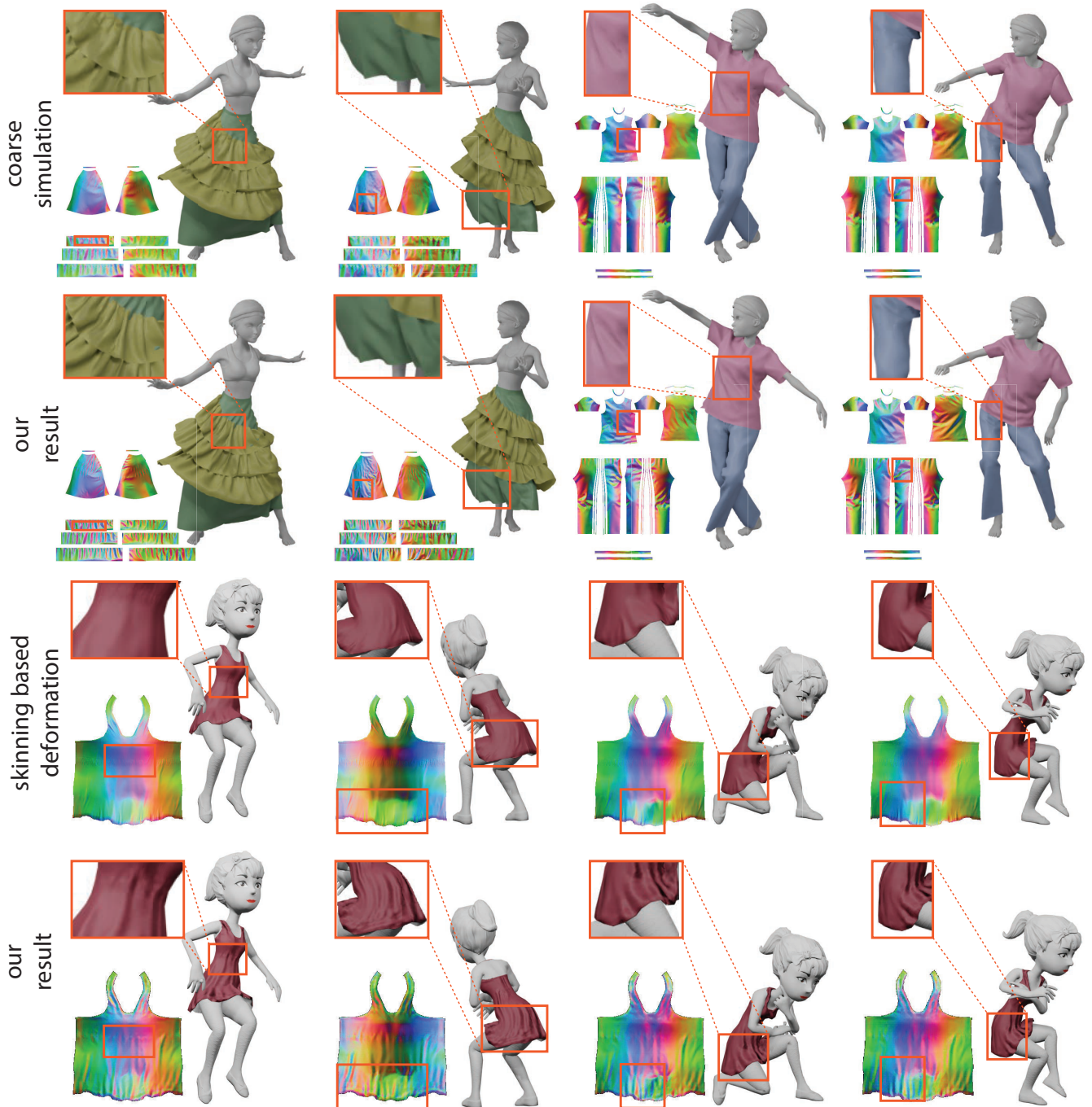
Figure 7: *We evaluate our method both on coarse garment simulations and garments deformed by linear blend skinning. Our method can generalize across different garment types and motions. The materials for the simulated garments are as follows: t-shirt (knit terry), pants (wool melton), skirt (silk chamuse), skirt laces (silk chiffon). For the skinning based example, our material classification network predicts it as silk chiffon. Please see supplementary video.*

unseen data and achieves high improvement scores regardless with only slight degradation in terms of performance ($\sim 91 - 96\%$).

When ground truth is not available, we evaluate the performance

from a statistics perspective. Given a sequence of animated garments, we randomly crop patches from each frame and generate the corresponding samples of Gram matrices of style features. We collect such samples from three sources: (a) patches from the coarse input;
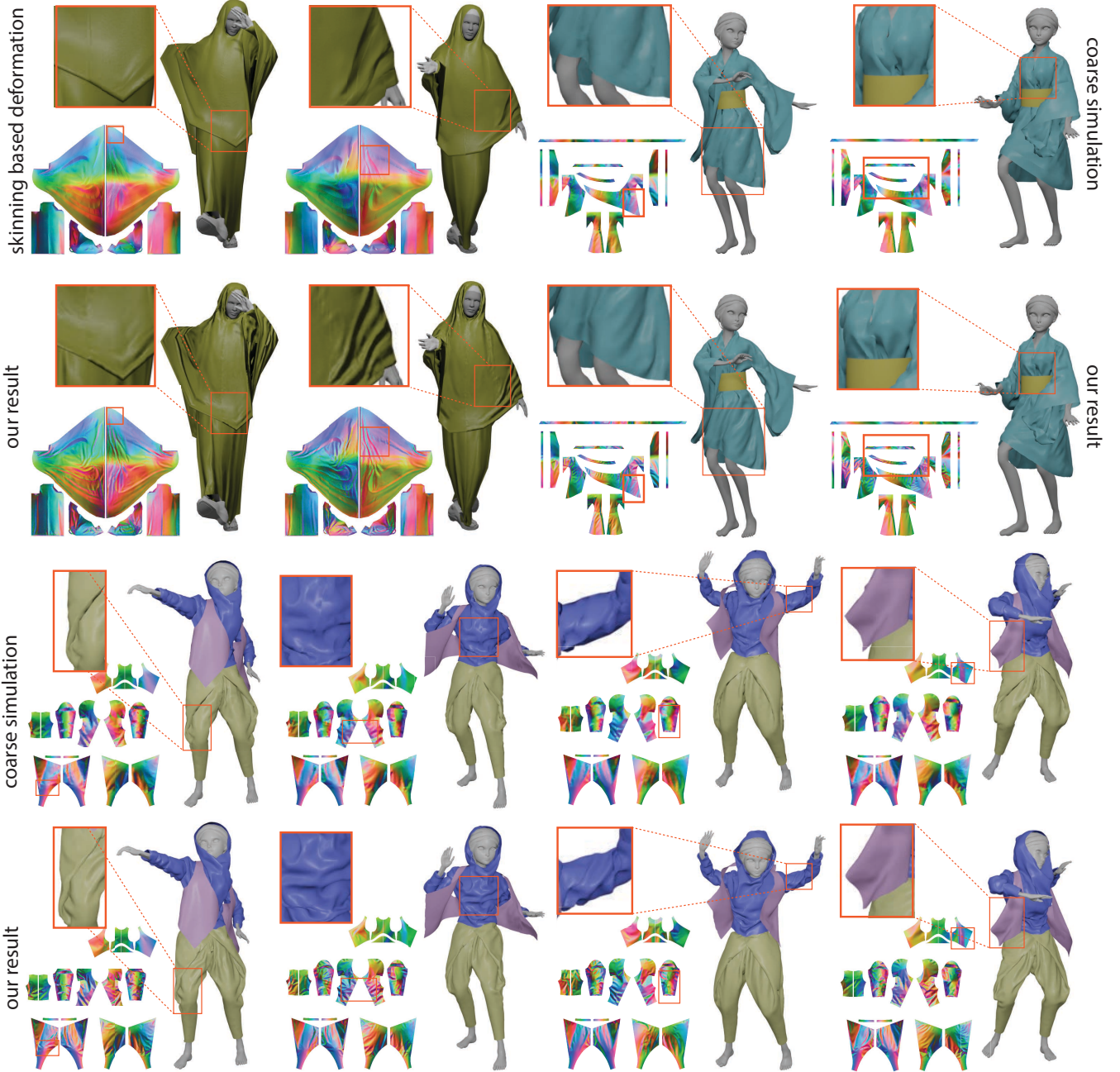
Figure 8: *We evaluate our method both on coarse garment simulations and garments deformed by linear blend skinning. The materials for the simulated garments are as follows: hanfu (silk chamuse), hood (denim lightweight), pants (knit terry), vest (wool melton). For the skinning based example, our material classification network predicts it as knit terry. Please see supplementary video.*

(b) patches generated by our method; and (c) patches from high resolution simulation (not for corresponding frames). We embed the patches using spectral embedding from $\mathbb{R}^{262144} \to \mathbb{R}^2$, and measure the distances in the distribution of the resultant point sets (denotes as P_a, P_b, and P_c, respectively) using Chamfer distance (CD) as $C_1 := \mathrm{CD}(P_a, P_c)$ and $C_2 := \mathrm{CD}(P_b, P_c)$, and improvement score defined as $\mathrm{DR} := 100|C_1 - C_2|/C_1$ in Table 2. As shown in the table,

the Chamfer distance between our result and the ground truth is significantly lower than the distance between the coarse input and the ground truth. This indicates that our method not only enhances the details of individual frames, but also improves overall plausibility for a motion sequence.

Continuity in time. While we do use time information in the ma-

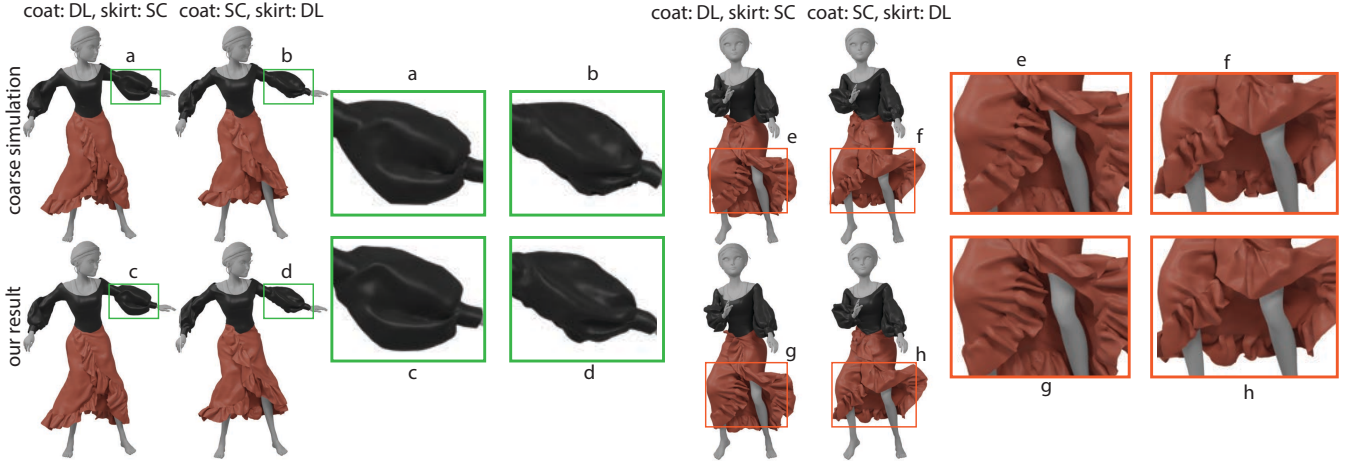coat: DL, skirt: SC   coat: SC, skirt: DL   coat: DL, skirt: SC   coat: SC, skirt: DL

Figure 9: *We demonstrate how our method works on garments made of different materials and undergoing the same motion. Note how the coarse inputs vary with material change and the synthesized details that align well with the input. The two different materials used are denim light (DL) and silk chiffon (SC).*

terial classification, the main detail enhancement step works at frame level. In earlier versions of our network, we experimented with time-dependent networks, with GRUs and LSTMs. While the results were better on training sets, we found the networks did not generalize across unseen data. In contrast, the proposed per-frame processing still leverages continuity data in the coarse geometry sequence and produces time-continuous details (please see supplementary video), with superior generalization behavior. We quantitatively evaluate our performance over continuous body motion, see Figure 6 (right).

Table 1: *We report the mean and standard deviation of improvement scores on three dresses made of two different materials and undergoing two different motion sequences. While Dress A and Motion 1 was seen during training, the remaining dresses and the motion sequence are unseen. Both motions are sequences are composed of 200 frames.*

	Std / Mean	Dress A	Dress B	Dress C
Silk chamuse	Motion 1	0.87 / 98.41	1.01 / 96.22	3.02 / 91.87
	Motion 2	1.53 / 97.55	0.92 / 95.39	4.56 / 91.25
Denim light	Motion 1	0.57 / 98.35	0.92 / 95.28	2.03 / 92.45
	Motion 2	1.08 / 97.39	1.42 / 95.25	3.25 / 92.44

Table 2: *We report the Chamfer distance between (i) the coarse input and the ground truth and (ii) our result and the ground truth. We calculate a matching score {$DR := 100|C_1 - C_2|/C_1$} to show the relative improvement. Please refer to the text for details. Note that Dress A and Motion 1 are seen during training while the remaining dresses and the motion sequence are unseen. Dress C is tight, as indicated by the low Chamfer distance to start, and therefore leaves little scope to improve in terms of detail enhancement.*

C_2 \ DR C_1		Dress A	Dress B	Dress C
Silk chamuse	Motion 1	0.09 \ 94.81 / 1.77	0.09 \ 74.32 / 0.36	0.12 \ 14.95 / 0.15
	Motion 2	0.11 \ 92.25 / 1.39	0.12 \ 89.20 / 1.09	0.10 \ 10.56 / 0.11
Denim light	Motion 1	0.01 \ 92.54 / 1.31	0.13 \ 85.86 / 0.89	0.12 \ 6.36 / 0.13
	Motion 2	0.12 \ 92.33 / 1.50	0.11 \ 88.23 / 0.93	0.09 \ 7.70 / 0.10

5.3. Ablation Study

In order to ensure coarse and high resolution patches are aligned during training, we generate the input to our network by downsampling the high resolution normal maps. We validate this choice, by visualizing the 2D embedding of the output features of some of the layers of VGG19 that we choose as style layers in Figure 11. As shown in the figure, the distribution of the features of the downsampled high resolution normal maps are close to the actual coarse normal maps. Further, given the downsampled normal maps as input, our method can successfully generate results that share similar distribution as the ground truth.
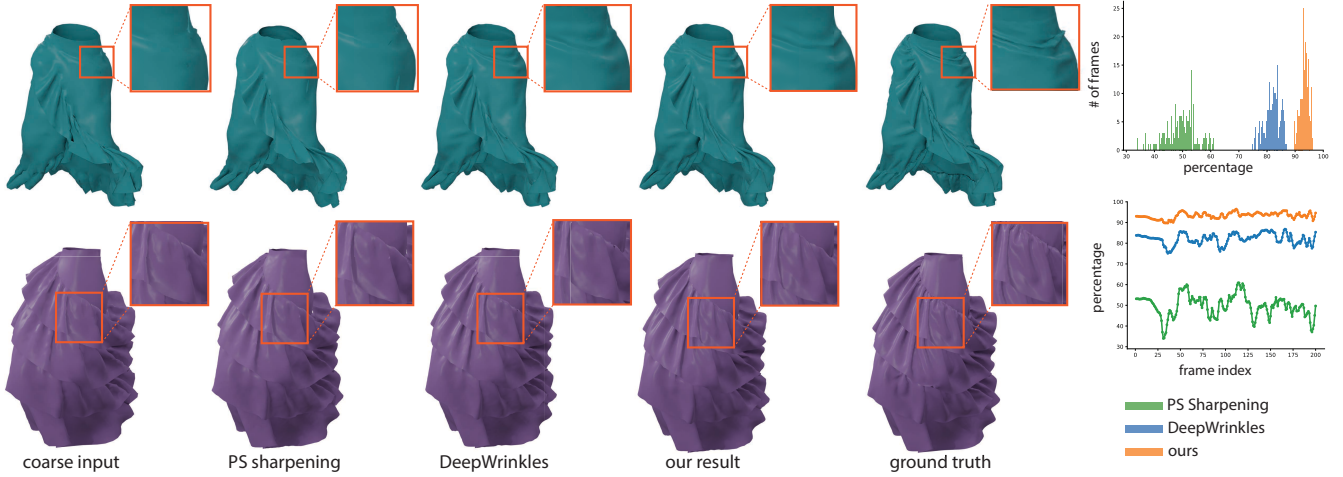
| coarse input | PS sharpening | DeepWrinkles | our result | ground truth |

Figure 10: *Given a coarse garment geometry, we compare our results with image based methods, i.e., sharpening feature in Photoshop as well as our implementation of the DeepWrinkles [LCT18] method. Our results are closest to the target ground truth visually. We also outperform the alternative approaches in terms of the improvement score, that we show the example scores corresponding to the case of green skirt.*
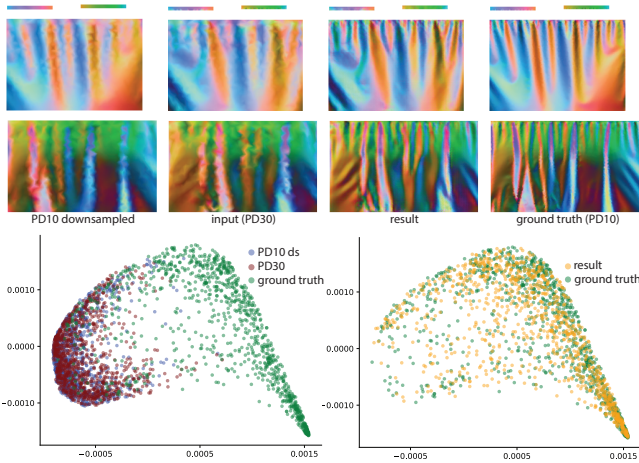

Figure 11: *Given an input normal map obtained by running a coarse simulation with a particle distance of 30 mm (PD30), our network produces a result that is close to the ground truth (high resolution simulation run with a particle distance of 10 mm) in terms of Gram matrix distribution of the output features of VGG19 layers that we use as style layers. We note that when training our network, as input we use downsampled high resolution normal maps (PD10 ds) which have similar feature distribution as the coarse simulation normal maps (PD30) used during test time.*

We also evaluate the performance of our method trained with different portions of our dataset. As shown in Figure 12, even when trained with a single sheet of cloth, our method generalizes reasonably well to different garment types. The quality of the results improve as the complete dataset is utilized.

5.4. Comparison

In Figure 10, we compare our method to other alternative approaches. Specifically, we compare to an image-based enhancement method, namely *image sharpening*, featured in Adobe Photoshop [Ado20]. We observe that PS sharpening can only sharpen or sometimes over-sharpen the original coarse folds, but hardly hallucinate the wrinkle details by synthesizing conditioned on the nearby coarse folds.

We further compare our method to the state-of-the-art neural network-based approach, e.g., DeepWrinkles [LCT18]. We implement their method and train both their generative model and our network with a portion of our dataset, specifically with training data obtained from the long skirt (as shown in Figure 6). We show that our method synthesizes sharper details for different garment types, e.g., the irregular lace skirt and the multi-layer dress as shown in Figure 10. We also report the improvement score across the whole
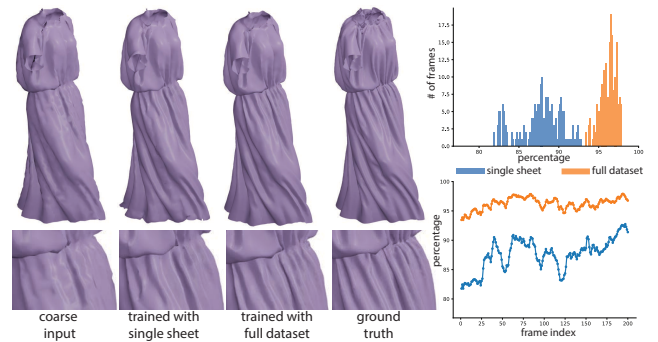

| coarse input | trained with single sheet | trained with full dataset | ground truth |

Figure 12: *We show results of our method when trained with (i) data generated from a single sheet only and (ii) all the three datasets (see Figure 6).*

motion sequence of each method demonstrating the superiority of our approach quantitatively.

6. Conclusions and Future Work

We have presented a deep learning based approach to synthesize plausible wrinkle details on coarse garment geometries. Our approach draws inspiration from the recent success of image artistic style methods and casts the problem as detail transfer across the normal maps of the garments. We have shown that our method generalizes across different garment types, sewing patterns, and motion sequences. We have also trained and tested our method with different material parameters under a universal framework. We have demonstrated results on inputs obtained from running physical simulation on low-resolution garment geometries as well as garments deformed by skinning.

Limitations and Future work. While our method shows impressive generalization capability, it has several limitations that can be addressed in future work. First of all, in order to generalize the detail enhancement network across different materials, we provide the material parameters as input. Hence, at run time, the network can be tested only on material types that have been seen during training. Generalizing our method to unseen materials, e.g., via learning an implicit material basis, is an interesting future direction. The current bottleneck is getting diverse simulations by sampling a variety of realistic materials due to challenges in getting robust and realistic simulations, in reasonable time, using commercial simulation frameworks. In our current implementation, our method is trained with regularly cropped 2D patches on the normal maps. Cropping geodesic patches over 3D surface instead to minimize averaging errors across overlapping regions and integrating the post-processing step into the network is an exciting direction. Furthermore, we assume that the coarse and high-resolution garments share the same set of garment pieces. In a possible garment design scenario, the designer might add some accessories to an existing garment, e.g., adding a belt or a layer to a skirt. It would be interesting to hallucinate the wrinkle details after such edits using the coarse simulation of the base garment only.

Acknowledgements

References

[Ado20] ADOBE. *Adobe Photoshop CC*. Version 19.1.2. May 18, 2020. URL: https://www.adobe.com/products/photoshop.html 11.

[AMB*19] ALLDIECK, THIEMO, MAGNOR, MARCUS, BHATNAGAR, BHARAT LAL, et al. "Learning to reconstruct people in clothing from a single RGB camera". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, 1175–1186 3.

[APTM19] ALLDIECK, THIEMO, PONS-MOLL, GERARD, THEOBALT, CHRISTIAN, and MAGNOR, MARCUS. "Tex2shape: Detailed full human body geometry from a single image". *Proceedings of the IEEE International Conference on Computer Vision*. 2019, 2293–2303 3.

[BPS*08] BRADLEY, DEREK, POPA, TIBERIU, SHEFFER, ALLA, et al. "Markerless garment capture". *ACM SIGGRAPH 2008 papers*. 2008, 1–9 2.

[BTTP19] BHATNAGAR, BHARAT LAL, TIWARI, GARVITA, THEOBALT, CHRISTIAN, and PONS-MOLL, GERARD. "Multi-garment net: Learning to dress 3d people from images". *Proceedings of the IEEE International Conference on Computer Vision*. 2019, 5420–5430 3.

[CK05] CHOI, KWANG-JIN and KO, HYEONG-SEOK. "Research Problems in Clothing Simulation". *Comput. Aided Des.* 37.6 (May 2005), 585–592 2.

[CZL*15] CHEN, XIAOWU, ZHOU, BIN, LU, FEI-XIANG, et al. "Garment modeling with a depth camera." *ACM Trans. Graph.* 34.6 (2015), 203–1 2.

[DSK16] DUMOULIN, VINCENT, SHLENS, JONATHON, and KUDLUR, MANJUNATH. "A learned representation for artistic style". *arXiv preprint arXiv:1610.07629* (2016), 2, 4.

[FJL*16] FIŠER, JAKUB, JAMRIŠKA, ONDŘEJ, LUKÁČ, MICHAL, et al. "StyLit: illumination-guided example-based stylization of 3D renderings". *ACM Transactions on Graphics (TOG)* 35.4 (2016), 1–11 3.

[FYK10] FENG, WEI-WEN, YU, YIZHOU, and KIM, BYUNG-UCK. "A Deformation Transformer for Real-Time Cloth Animation". *ACM Trans. Graph.* 29.4 (July 2010), 2.

[GCS*19] GUNDOGDU, ERHAN, CONSTANTIN, VICTOR, SEIFODDINI, AMROLLAH, et al. "Garnet: A two-stream network for fast and accurate 3d cloth draping". *Proceedings of the IEEE International Conference on Computer Vision*. 2019, 8739–8748 3.

[GEB15] GATYS, LEON, ECKER, ALEXANDER S, and BETHGE, MATTHIAS. "Texture synthesis using convolutional neural networks". *Advances in neural information processing systems*. 2015, 262–270 3.

[GEB16] GATYS, LEON A, ECKER, ALEXANDER S, and BETHGE, MATTHIAS. "Image style transfer using convolutional neural networks". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 2414–2423 2, 3, 5.

[GPV*15] GILLETTE, RUSSELL, PETERS, CRAIG, VINING, NICHOLAS, et al. "Real-Time Dynamic Wrinkling of Coarse Animated Cloth". *SCA*. SCA '15. 2015 2.

[GRH*12] GUAN, PENG, REISS, LORETTA, HIRSHBERG, DAVID A, et al. "Drape: Dressing any person". *ACM Transactions on Graphics (TOG)* 31.4 (2012), 1–10 3, 6.

[HB17] HUANG, XUN and BELONGIE, SERGE. "Arbitrary style transfer in real-time with adaptive instance normalization". *Proceedings of the IEEE International Conference on Computer Vision*. 2017, 1501–1510 3, 7.

[HDDN19] HOLDEN, DANIEL, DUONG, BANG CHI, DATTA, SAYANTAN, and NOWROUZEZAHRAI, DEREK. "Subspace neural physics: fast data-driven interactive simulation". *Proceedings of the 18th annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 2019, 1–12 3.

[HTC*14] HAHN, FABIAN, THOMASZEWSKI, BERNHARD, COROS, STELIAN, et al. "Subspace clothing simulation using adaptive bases". *ACM Transactions on Graphics (TOG)* 33.4 (2014), 1–9 3.

[IZZE17] ISOLA, PHILLIP, ZHU, JUN-YAN, ZHOU, TINGHUI, and EFROS, ALEXEI A. "Image-to-image translation with conditional adversarial networks". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, 1125–1134 3.

[JZGF18] JIN, NING, ZHU, YILIN, GENG, ZHENGLIN, and FEDKIW, RONALD. "A pixel-based framework for data-driven clothing". *arXiv preprint arXiv:1812.01677* (2018) 3.

[JZH*20] JIANG, BOYI, ZHANG, JUYONG, HONG, YANG, et al. "BCNet: Learning Body and Cloth Shape from A Single Image". *European Conference on Computer Vision*. Springer. 2020 3.

[KGBS11] KAVAN, LADISLAV, GERSZEWSKI, DAN, BARGTEIL, ADAM W., and SLOAN, PETER-PIKE. "Physics-Inspired Upsampling for Cloth Simulation in Games". *ACM Trans. Graph.* 30.4 (July 2011) 2.

[LCT18] LAHNER, ZORAH, CREMERS, DANIEL, and TUNG, TONY. "Deepwrinkles: Accurate and realistic clothing modeling". *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, 667–684 3, 4, 11.

[LLK19] LIANG, JUNBANG, LIN, MING, and KOLTUN, VLADLEN. "Differentiable Cloth Simulation for Inverse Problems". *Advances in Neural Information Processing Systems*. 2019, 771–780 2.

[LTH*17] LEDIG, CHRISTIAN, THEIS, LUCAS, HUSZÁR, FERENC, et al. "Photo-realistic single image super-resolution using a generative adversarial network". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, 4681–4690 3.

[LYY*17] LIAO, JING, YAO, YUAN, YUAN, LU, et al. "Visual attribute transfer through deep image analogy". *arXiv preprint arXiv:1705.01088* (2017) 3.

[MC10] MÜLLER, MATTHIAS and CHENTANEZ, NUTTAPONG. "Wrinkle Meshes". *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '10. Madrid, Spain: Eurographics Association, 2010, 85–92 2.

[MYR*19] MA, QIANLI, YANG, JINLONG, RANJAN, ANURAG, et al. "Learning to Dress 3D People in Generative Clothing". *arXiv preprint arXiv:1907.13615* (2019) 3.

[NMK*06] NEALEN, ANDREW, MÜLLER, MATTHIAS, KEISER, RICHARD, et al. "Physically Based Deformable Models in Computer Graphics". *Computer Graphics Forum* 25.4 (2006), 809–836 2.

[NSO12] NARAIN, RAHUL, SAMII, ARMIN, and O'BRIEN, JAMES F. "Adaptive anisotropic remeshing for cloth simulation". *ACM transactions on graphics (TOG)* 31.6 (2012), 1–10 2.

[PLP20] PATEL, CHAITANYA, LIAO, ZHOUYINGCHENG, and PONS-MOLL, GERARD. "The virtual tailor: Predicting clothing in 3d as a function of human pose, shape and garment style". *arXiv preprint arXiv:2003.04583* (2020) 3.

[PPHB17] PONS-MOLL, GERARD, PUJADES, SERGI, HU, SONNY, and BLACK, MICHAEL J. "ClothCap: Seamless 4D clothing capture and retargeting". *ACM Transactions on Graphics (TOG)* 36.4 (2017), 1–15 2, 3.

[RPC*10] ROHMER, DAMIEN, POPA, TIBERIU, CANI, MARIE-PAULE, et al. "Animation Wrinkling: Augmenting Coarse Cloth Simulations with Realistic-Looking Wrinkles". *ACM Trans. Graph.* 29.6 (Dec. 2010) 2.

[SLL20] SHEN, YU, LIANG, JUNBANG, and LIN, MING C. "GAN-based Garment Generation Using Sewing Pattern Images". *Proceedings of the European Conference on Computer Vision (ECCV)*. Vol. 1. 2. 2020, 3 3.

[SOC19] SANTESTEBAN, IGOR, OTADUY, MIGUEL A, and CASAS, DAN. "Learning-Based Animation of Clothing for Virtual Try-On". *Computer Graphics Forum*. Vol. 38. 2. Wiley Online Library. 2019, 355–366 3.

[SZ14] SIMONYAN, KAREN and ZISSERMAN, ANDREW. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556* (2014) 3, 5.

[TBTP20] TIWARI, GARVITA, BHATNAGAR, BHARAT LAL, TUNG, TONY, and PONS-MOLL, GERARD. "SIZER: A Dataset and Model for Parsing 3D Clothing and Learning Size Sensitive 3D Clothing". *European Conference on Computer Vision (ECCV)*. Springer. Aug. 2020 3.

[TWL*18] TANG, MIN, WANG, TONGTONG, LIU, ZHONGYUAN, et al. "I-cloth: incremental collision handling for GPU-based interactive cloth simulation". *ACM Transactions on Graphics (TOG)* 37.6 (2018), 1–10 2.

[WCPM18] WANG, TUANFENG Y, CEYLAN, DUYGU, POPOVIC, JOVAN, and MITRA, NILOY J. "Learning a shared shape space for multimodal garment design". *arXiv preprint arXiv:1806.11335* (2018) 3.

[WHRO10] WANG, HUAMIN, HECHT, FLORIAN, RAMAMOORTHI, RAVI, and O'BRIEN, JAMES F. "Example-Based Wrinkle Synthesis for Clothing Animation". *ACM Trans. Graph.* 29.4 (July 2010) 2.

[WLY*15] WANG, ZHAOWEN, LIU, DING, YANG, JIANCHAO, et al. "Deep networks for image super-resolution with sparse prior". *Proceedings of the IEEE international conference on computer vision*. 2015, 370–378 3.

[WSFM19] WANG, TUANFENG Y, SHAO, TIANJIA, FU, KAI, and MITRA, NILOY J. "Learning an intrinsic garment space for interactive authoring of garment animation". *ACM Transactions on Graphics (TOG)* 38.6 (2019), 1–12 3.

[XUC*14] XU, WEIWEI, UMETANI, NOBUYUKI, CHAO, QIANWEN, et al. "Sensitivity-optimized rigging for example-based real-time clothing synthesis." *ACM Trans. Graph.* 33.4 (2014), 107–1 3.

[YFHW18] YANG, JINLONG, FRANCO, JEAN-SÉBASTIEN, HÉTROY-WHEELER, FRANCK, and WUHRER, STEFANIE. "Analyzing clothing layer deformation statistics of 3d human motions". *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, 237–253 3.

[YLL17] YANG, SHAN, LIANG, JUNBANG, and LIN, MING C. "Learning-based cloth material recovery from video". *Proceedings of the IEEE International Conference on Computer Vision*. 2017, 4383–4393 6.

[YZZ*19] YU, TAO, ZHENG, ZERONG, ZHONG, YUAN, et al. "Simulcap: Single-view human performance capture with cloth simulation". *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2019, 5499–5509 2.

[ZBO12] ZURDO, JAVIER S, BRITO, JUAN P, and OTADUY, MIGUEL A. "Animating wrinkles by example on non-skinned cloth". *IEEE Transactions on Visualization and Computer Graphics* 19.1 (2012), 149–158 2.

[ZCJ*20] ZHU, HEMING, CAO, YU, JIN, HANG, et al. "Deep Fashion3D: A Dataset and Benchmark for 3D Garment Reconstruction from Single Images". *arXiv preprint arXiv:2003.12753* (2020) 3.