SCANimate: Weakly Supervised Learning of Skinned Clothed Avatar Networks
SCANimate: Skinned Clothing Avatar Networks 的弱監督學習

Shunsuke Saito, Jinlong Yang, Qianli Ma, Michael J. Black

Max Planck Institute for Intelligent Systems, Tubingen, Germany
ETH Zurich

CVPR 2021 Best Paper Candidate
https://arxiv.org/abs/2104.03313
https://scanimate.is.tue.mpg.de/

Abstract 摘要

We present SCANimate, an end-to-end trainable framework that takes raw 3D scans of a clothed human and turns them into an animatable avatar.
我們展示了 SCANimate，這是一個端到端的可訓練框架，它對一個穿著衣服的人進行原始 3D 掃描並將它們變成一個可動畫的化身。

These avatars are driven by pose parameters and have realistic clothing that moves and deforms naturally.
這些化身由姿勢參數驅動，並擁有可自然移動和變形的逼真服裝。

SCANimate does not rely on a customized mesh template or surface mesh registration.
SCANimate 不依賴於自定義網格模板或表面網格註冊。

We observe that fitting a parametric 3D body model, like SMPL, to a clothed human scan is tractable while surface registration of the body topology to the scan is often not, because clothing can deviate significantly from the body shape.
我們觀察到，將參數化 3D 身體模型（如 SMPL）擬合到穿著衣服的人體掃描是容易處理的，而身體拓撲結構與掃描的表面配准通常則不然，因為衣服可能會顯著偏離身體形狀。

We also observe that articulated transformations are invertible, resulting in geometric cycle-consistency in the posed and unposed shapes.
我們還觀察到鉸接變換是可逆的，從而導致擺姿勢和未擺姿勢的形狀的幾何循環一致性。

These observations lead us to a weakly supervised learning method that aligns scans into a canonical pose by disentangling articulated deformations without templatebased surface registration.
這些觀察使我們找到了一種弱監督學習方法，該方法通過在沒有基於模板的表面配準的情況下解開鉸接變形來將掃描對齊到規範姿勢。

Furthermore, to complete missing regions in the aligned scans while modeling posedependent deformations, we introduce a locally pose-aware implicit function that learns to complete and model geometry with learned pose correctives.
此外，為了在對姿態相關變形建模的同時完成對齊掃描中的缺失區域，我們引入了一個局部姿態感知隱式函數，該函數學習使用學習的姿態校正來完成和建模幾何。

In contrast to commonly used global pose embeddings, our local pose conditioning significantly reduces long-range spurious correlations and improves generalization to unseen poses, especially when training data is limited. Our method can be applied to poseaware appearance modeling to generate a fully textured avatar.
與常用的全局姿勢嵌入相比，我們的局部姿勢調節顯著降低了遠程虛假相關性，並提高了對未知姿勢的泛化能力，尤其是在訓練數據有限的情況下。 我們的方法可以應用於姿態感知外觀建模以生成完全紋理化的頭像。

We demonstrate our approach on various clothing types with different amounts of training data, outperforming existing solutions and other variants in terms of fidelity and generality in every setting.
我們在具有不同訓練數據量的各種服裝類型上展示了我們的方法，在每種設置的保真度和通用性方面均優於現有解決方案和其他變體。

The code is available at https://scanimate.is.tue.mpg.de.

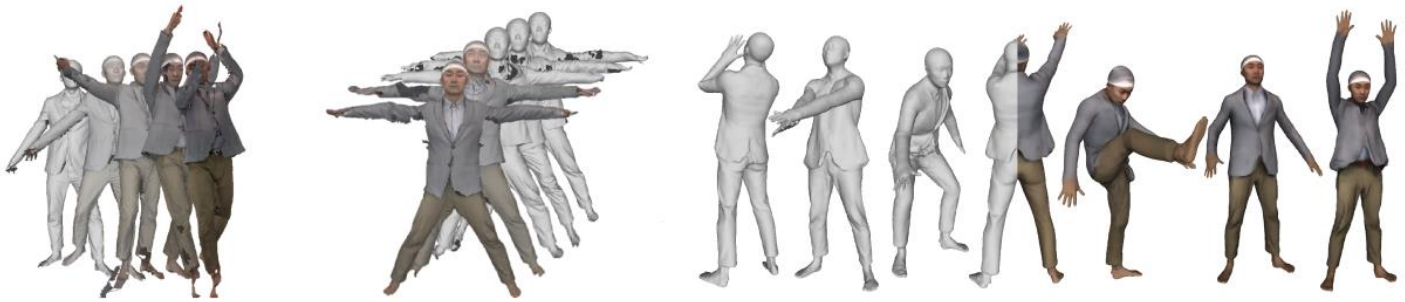

Figure 1: **SCANimate.** Given a set of raw scans with multiple poses containing self-intersections, holes, and noise (left), SCANimate automatically aligns all scans to a canonical pose (middle) and learns a Scanimat, a fully animatable avatar that produces pose-dependent deformations and texture without garment-specific templates or mesh registration (right).

Figure 1: SCANimate.

Given a set of raw scans with multiple poses containing self-intersections, holes, and noise (left), SCANimate automatically aligns all scans to a canonical pose (middle) and learns a Scanimat, a fully animatable avatar that produces pose-dependent deformations and texture without garment-specific templates or mesh registration (right).
給定一組包含自相交、孔洞和噪聲的多個姿勢的原始掃描（左），SCANimate 自動將所有掃描與規範姿勢對齊（中）並學習 Scanimat，一個完全可動畫化的化身，產生依賴於姿勢的變形和 沒有服裝特定模板或網格註冊的紋理（右）。

1. Introduction 前言

Parametric models of 3D human bodies are widely used for the analysis and synthesis of human shape, pose, and motion.
3D 人體參數模型廣泛用於分析和合成人體形狀、姿勢和運動。

While existing models typically represent “minimally clothed” bodies [4, 26, 43, 53, 67], many applications require realistically clothed bodies.
雖然現有模型通常代表“最少衣服”的身體 [4, 26, 43, 53, 67]，但許多應用程序需要逼真的衣服。

Our goal is to make it easy to produce a realistic 3D avatar of a clothed person that can be reposed and animated as easily as existing models like SMPL [43].
我們的目標是使製作一個穿著衣服的人的逼真 3D 化身變得容易，該化身可以像 SMPL [43] 等現有模型一樣輕鬆地放置和動畫化。

In particular, the model must support clothing that moves and deforms naturally, with detailed 3D wrinkles, and the rendering of realistically textured images.
特別是，模型必須支持自然移動和變形的衣服，具有詳細的 3D 皺紋，以及逼真紋理圖像的渲染。

To that end, we introduce SCANimate (Skinned Clothed Avatar Networks for animation), which creates high-quality animatable clothed humans, called Scanimats, from raw 3D scans.
為此，我們引入了 SCANimate（用於動畫的皮膚服裝化身網絡），它通過原始 3D 掃描創建高質量的可動畫服裝人，稱為 Scanimat。

SCANimate has the following properties:
SCANimate 具有以下屬性：

(1) we learn an articulated clothed human model directly from raw scans, completely eliminating the need for surface registration of a custom template or synthetic clothing simulation data,
(1) 我們直接從原始掃描中學習一個鉸接式服裝人體模型，完全不需要自定義模板或合成服裝模擬數據的表面配準，

(2) our parametric model retains the complex and detailed deformations of clothing present in the original scans such as wrinkles and sliding effects of garments with arbitrary topology,
(2) 我們的參數化模型保留了原始掃描中服裝的複雜和詳細的變形，例如具有任意拓撲結構的服裝的皺紋和滑動效果，

(3) a Scanimat can be animated directly using SMPL pose parameters, and
(3) Scanimat 可以直接使用 SMPL 姿勢參數進行動畫處理，並且

(4) our approach predicts pose-dependent clothing deformations based on local pose parameters, providing

generalization to unseen poses.
(4) 我們的方法基於局部姿勢參數預測依賴於姿勢的服裝變形，提供對看不見的姿勢的概括。

Recent data-driven approaches have shown promise for learning parametric models of clothed humans from real-world observations [39, 46, 55, 57].
最近的數據驅動方法顯示了從現實世界觀察中學習穿衣人類參數模型的前景 [39, 46, 55, 57]。

However, these approaches typically limit the supported clothing types and topology because they require accurate surface registration of a common template mesh to 3D training scans [39, 46, 57].
然而，這些方法通常會限制支持的服裝類型和拓撲，因為它們需要將通用模板網格準確的表面配準到 3D 訓練掃描 [39, 46, 57]。

Concurrent work by Ma et al. [45] learns clothing deformation without surface registration, yet it is unclear if the method works on raw scans with noise and holes.
Ma et al. [45] 的並行工作在沒有表面配準的情況下學習服裝變形，但目前尚不清楚該方法是否適用於帶有噪聲和孔洞的原始掃描。

Learning from real-world observations is essentially challenging because raw 3D scans are un-ordered point clouds with missing data, changing topology, multiple clothing layers, and sliding motions between the body and garments.
從現實世界的觀察中學習本質上是具有挑戰性的，因為原始 3D 掃描是無序的點雲，具有缺失數據、不斷變化的拓撲、多個服裝層以及身體和服裝之間的滑動運動。

Although one can learn from synthetic data generated by physics-based clothing simulation [23, 25, 55], the results are less realistic, the data preparation is time consuming and non-trivial to scale to the real-world clothing.
雖然可以從基於物理的服裝模擬生成的合成數據中學習 [23, 25, 55]，但結果不太現實，數據準備非常耗時且難以擴展到真實世界的服裝。

To address these issues, SCANimate learns directly from raw scans of people in clothing.
為了解決這些問題，SCANimate 直接從穿著衣服的人的原始掃描中學習。

Body scanning is becoming common, and scans can be obtained from a variety of devices.
身體掃描正變得越來越普遍，並且可以從各種設備獲得掃描。

Scans contain high-frequency details, capture varied clothing topology, and are inherently realistic.
掃描包含高頻細節，捕捉各種服裝拓撲結構，並且本質上是逼真的。

To make learning from scans possible, we make several contributions: canonicalization, implicit skinning fields, cycle consistency, and implicit shape learning.
為了使從掃描中學習成為可能，我們做出了一些貢獻：規範化、隱式蒙皮場、循環一致性和隱式形狀

學習。

Canonicalization and Implicit Skinning Fields. 規範化和隱式蒙皮字段。

The first step involves transforming the raw scans to a common pose so we can learn to model pose-dependent surface deformations (e.g. bulging, stretching, wrinkling, and sliding), i.e. pose "correctives".
第一步涉及將原始掃描轉換為常見姿勢，以便我們可以學習建模依賴於姿勢的表面變形（例如凸出、拉伸、起皺和滑動），即姿勢"矯正"。

But we are not seeking a traditional "registration" of the scans to a common mesh topology, since this is, in general, not feasible with clothed bodies.
但我們並不尋求傳統的掃描"註冊"到常見的網格拓撲，因為這通常不適用於穿著衣服的身體。

Instead, we learn continuous functions of 3D space that allow us to transform posed scans to a canonical pose and back again.
相反，我們學習了 3D 空間的連續函數，這使我們能夠將姿勢掃描轉換為規範姿勢，然後再返回。

The key idea is to build this on linear blend skinning (LBS), which traditionally defines weights on the surface of a mesh that encode how much each vertex is influenced by the rotation of a body joint.
關鍵思想是將其構建在線性混合蒙皮 (LBS) 上，該方法傳統上定義了網格表面的權重，用於編碼每個頂點受身體關節旋轉影響的程度。

To deal with raw scans of unknown topology, we extend this notion by defining skinning weights implicitly everywhere in 3D space.
為了處理未知拓撲的原始掃描，我們通過在 3D 空間中隱式地定義蒙皮權重來擴展這個概念。

Specifically, given a 3D location x, we regress a continuous vector function g represented by a neural network, $g(x):R3 \rightarrow RJ$, which defines the skinning weights.
具體來說，給定 3D 位置 x，我們回歸由神經網絡表示的連續向量函數 g，$g(x):R3 \rightarrow RJ$ ，它定義了蒙皮權重。

An inverse LBS function uses the regressed skinning weights to "undo" the pose of the body and transforms the points into the canonical space.
逆 LBS 函數使用回歸的蒙皮權重來"撤消"身體的姿勢並將點轉換為規範空間。

As this representation makes no assumptions about the topology or resolution of input scans, we can canonicalize arbitrary non-watertight meshes.
由於此表示不對輸入掃描的拓撲或分辨率進行假設，因此我們可以規範化任意非水密網格。

Furthermore, we can easily generate animations of the parametric clothed avatar by applying forward LBS to the clothed body in the canonical pose with the learned pose correctives.

此外，我們可以通過將前向 LBS 應用到具有學習姿勢校正的規範姿勢中的衣服身體，輕鬆生成參數化衣服化身的動畫。

Cycle Consistency. 循環一致性。

Despite the desirable properties of canonicalization, learning the skinning function is ill-posed since we do not have ground truth training data that specifies the weights.
儘管規範化具有理想的特性，但學習蒙皮函數是不適定的，因為我們沒有指定權重的真實訓練數據。

To address this, we exploit two key observations.
為了解決這個問題，我們利用了兩個關鍵觀察結果。

First, as demonstrated in previous work [27, 70, 73], fitting a parametric human body model such as SMPL [43] to 3D scans is more tractable than surface registration.
首先，如之前的工作 [27、70、73] 所示，將參數人體模型（例如 SMPL [43]）擬合到 3D 掃描比表面配準更容易處理。

We leverage SMPL's skinning weights, which are defined only on the body surface, to regularize our more general skinning function.
首先，如之前的工作 [27、70、73] 所示，將參數人體模型（例如 SMPL [43]）擬合到 3D 掃描比表面配準更容易處理。

Second, the transformations between the posed space and the canonical space should be cycle-consistent.
其次，姿勢空間和規範空間之間的轉換應該是循環一致的。

Namely, inverse LBS and forward LBS together should form an identity mapping as illustrated in Fig. 3, which provides a self-supervision signal for training the skinning function.
即，逆 LBS 和前向 LBS 一起應該形成一個身份映射，如圖 3 所示，它為訓練蒙皮函數提供自監督信號。

After training the skinning function, we obtain the canonicalized scans (all in the same pose).
在訓練蒙皮函數後，我們獲得了規範化的掃描（所有在相同的姿勢）。

Learning Implicit Pose Correctives. 學習隱式姿勢矯正。

Given the canonicalized scans, we learn a model that captures the pose-dependent deformations.
鑑於規範化掃描，我們學習了一個模型來捕捉姿態相關的變形。

However a problem remains:
但是仍然存在一個問題：

the original raw scans often contain holes, and so do the canonicalized scans.
原始原始掃描通常包含漏洞，規範化掃描也是如此。

To deal with this and with the arbitrary topology of clothing, we use an implicit surface representation [13, 47, 54].
為了解決這個問題以及服裝的任意拓撲結構，我們使用隱式表面表示 [13, 47, 54]。

As multiple canonicalized scans will miss different regions, with this approach, they complement each other, while retaining details present in the original inputs.
由於多個規範化掃描會遺漏不同的區域，因此通過這種方法，它們可以相互補充，同時保留原始輸入中存在的細節。

Furthermore, unlike traditional approaches [39, 46, 55, 71], where pose-dependent deformations are conditioned on entire pose parameters, we spatially filter out irrelevant pose features from the input conditions by leveraging the learned skinning weights.
此外，與傳統方法 [39, 46, 55, 71] 不同，與姿勢相關的變形以整個姿勢參數為條件，我們通過利用學習到的蒙皮權重從輸入條件中空間過濾掉不相關的姿勢特徵。

In this way, we effectively prune long-range spurious correlations between garment deformations and body joints, achieving plausible pose correctives for unseen poses even from a small number of training scans.
通過這種方式，我們有效地修剪了服裝變形和身體關節之間的長程虛假相關性，即使從少量的訓練掃描中也能實現對看不見的姿勢的合理姿勢校正。

The resulting learned Scanimat can be easily reposed and animated with SMPL pose parameters.
由此產生的學習 Scanimat 可以輕鬆地使用 SMPL 姿勢參數進行調整和動畫處理。

In summary, our main contributions are
總之，我們的主要貢獻是

(1) the first endto-end trainable framework to build a high-quality parametric clothed human model from raw scans,
(1) 第一個端到端可訓練框架，用於從原始掃描構建高質量的參數化服裝人體模型，

(2) a novel weaklysupervised formulation with geometric cycle-consistency that disentangles articulated deformations from the local pose correctives without requiring ground-truth training data,
and
(2) 一種具有幾何週期一致性的新型弱監督公式，可在不需要地面實況訓練數據的情況下將關節變形與局部姿態校正分開，
和

(3) a locally pose-aware implicit surface representation that models pose-dependent clothing deformation and

generalizes to unseen poses.
(3) 一種局部姿勢感知的隱式表面表示，它對依賴於姿勢的服裝變形進行建模並泛化到看不見的姿勢。

Our results show that SCANimate is superior to existing solutions in terms of generality and accuracy.
我們的結果表明，SCANimate 在通用性和準確性方面優於現有解決方案。

Furthermore, we perform an extensive study to evaluate the technical contributions that are critical for success.
此外，我們進行了廣泛的研究，以評估對成功至關重要的技術貢獻。

The code and example Scanimats can be found at https://scanimate.is.tue.mpg.de.

2. Related Work 相關工作

Parametric Models for Human Bodies and Clothing. 人體和服裝的參數模型。

Parametric body models [4, 26, 43, 53, 67] learn statistical body shape variations and pose-dependent shape correctives that capture non-linear body deformation and compensate for linear blend skinning artifacts [3, 30, 37, 38, 41].
參數身體模型 [4, 26, 43, 53, 67] 學習統計的身體形狀變化和姿勢相關的形狀校正，以捕捉非線性身體變形並補償線性混合蒙皮偽影 [3, 30, 37, 38, 41] .

While these approaches achieve high-fidelity and intuitive control of human body shape and pose, they only focus on bodies without clothing.
雖然這些方法實現了對人體形狀和姿勢的高保真和直觀控制，但它們只關注沒有衣服的身體。

Similar ideas have been extended to model clothed bodies by introducing additional garment layers [16, 17, 23, 25, 34, 39, 68] or adding displacements or transformations to the base human body mesh [1, 2, 46, 51, 71].
通過引入額外的服裝層 [16、17、23、25、34、39、68] 或向基礎人體網格 [1、2、46、51、71]。

These parametric clothed human models decompose garment deformations into articulated deformations and local deformations such that pose correctives only focus on nonrigid local deformations.
這些參數化的穿著人體模型將服裝變形分解為關節變形和局部變形，這樣姿勢矯正只關注非剛性局部變形。

Thus, it is essential to obtain the inverse skinning transformation [55] by using the surface registration of a well-defined template [39, 43, 46, 71, 73] or using synthetic simulation data [15, 23, 25, 55].
因此，必須通過使用定義明確的模板的表面配準 [39, 43, 46, 71, 73] 或使用合成模擬數據 [15, 23, 25, 55] 來獲得逆皮膚變換 [55] .

However, these requirements limit the applicability of the approaches to fairly simple clothing, with a fixed

topology, and without complex interactions between garments and the body.
然而，這些要求限制了這些方法對相當簡單的服裝的適用性，具有固定的拓撲結構，並且服裝和身體之間沒有復雜的相互作用。

In contrast, our work uses a weakly supervised approach to build a parametric clothed human model from raw scans without the requirement of a template and surface registration.
相比之下，我們的工作使用弱監督方法從原始掃描構建參數化服裝人體模型，而無需模板和表面配準。

We canonicalize posed scans and learn an implicit surface with arbitrary topology [22] conditioned on pose parameters by leveraging a fitted human body model to the scan data [5, 8, 70, 72, 73].
我們將姿勢掃描規範化，並通過將擬合的人體模型用於掃描數據 [5, 8, 70, 72, 73]，以姿勢參數為條件學習具有任意拓撲結構的隱式表面 [22]。

Moon et al. [50] similarly propose a weakly supervised method for learning a finegrained hand model from scan data by deforming a fitted base hand model [59]; the approach is non-trivial to extend to human clothing with varying topology.
Moon et al. [50] 同樣提出了一種弱監督方法，通過對擬合的基礎手部模型進行變形，從掃描數據中學習細粒度手部模型 [59]； 這種方法很容易擴展到具有不同拓撲結構的人類服裝。

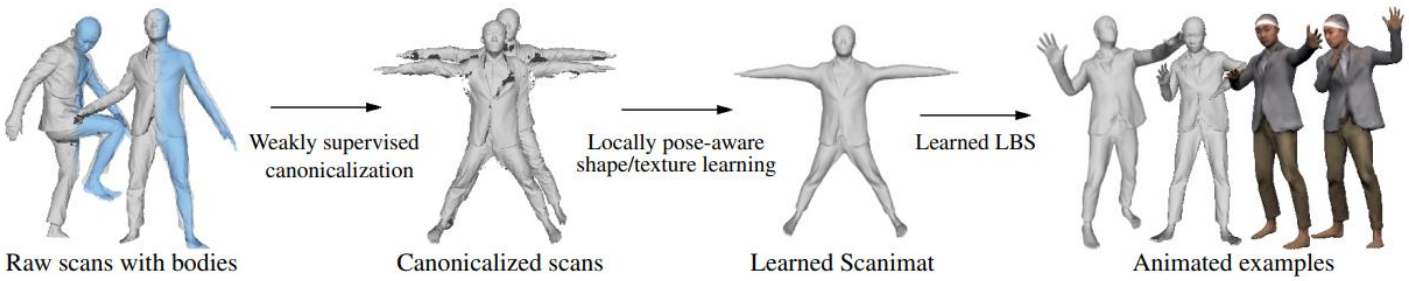

Figure 2: **Overview.** SCANimate learns a pose-aware parametric clothed human model directly from raw scans in a weakly supervised manner. The resulting Scanimats can be animated with SMPL pose parameters, producing realistic pose-dependent deformations and texture.

Figure 2: Overview. 圖 2：概述。

SCANimate learns a pose-aware parametric clothed human model directly from raw scans in a weakly supervised manner.
SCANimate 以弱監督的方式直接從原始掃描中學習姿勢感知參數化服裝人體模型。

The resulting Scanimats can be animated with SMPL pose parameters, producing realistic pose-dependent deformations and texture.
生成的 Scanimat 可以使用 SMPL 姿勢參數進行動畫處理，從而產生逼真的姿勢相關變形和紋理。

The most related work to ours is Neural Articulated Shape Approximation (NASA) [18], where the composition of occupancy networks [13, 47] articulated by the fitted SMPL model are directly learned from posed scans in the same spirit as structured implicit functions [20, 21].

與我們最相關的工作是神經鉸接形狀近似 (NASA) [18]，其中由擬合的 SMPL 模型連接的佔用網絡 [13, 47] 的組成直接從構成的掃描中學習，其精神與結構化隱函數相同[20, 21]。

Concurrent work, LEAP [48], extends a similar framework to a multisubject setting.
並行工作 LEAP [48] 將類似的框架擴展到多學科設置。

Through an extensive study in Sec. 4.1, we find that the compositional implicit functions proposed in [18] are more prone to artifacts and less generalizable to unseen poses than our LBS-based formulation.
通過在 Sec. 4.1 中的廣泛研究，我們發現 [18] 中提出的組合隱式函數更容易出現偽影，並且比我們基於 LBS 的公式更不容易泛化到看不見的姿勢。

Pose Canonicalization via Inverse LBS. 通過反向 LBS 進行姿勢規範化。

The key to successful canonicalization is learning transformations in the form of skinning weights in a continuous space.
成功規範化的關鍵是以連續空間中的皮膚權重的形式學習轉換。

Learning skinning weights for varied topologies has become possible using neural networks with graph convolutions [7, 42, 69].
使用具有圖卷積的神經網絡可以學習各種拓撲的蒙皮權重 [7, 42, 69]。

Given a neutral-posed template, these networks predict skinning weights together with a skeleton [69] or posedependent deformations [7].
給定一個中性姿勢模板，這些網絡與骨架 [69] 或姿勢相關變形 [7] 一起預測蒙皮權重。

While they predict skinning weights on a neutral-posed template in a fully supervised manner, our problem requires learning skinning weights, not only on the surface mesh, but in both the canonical and posed space without ground-truth skinning weights.
雖然他們以完全監督的方式預測中性姿勢模板上的蒙皮權重，但我們的問題需要學習蒙皮權重，不僅在表面網格上，而且在沒有真實蒙皮權重的規範和姿勢空間中。

Extending LBS skinning weights from an underlying body model to the continuous space is used in the data preparation step of ARCH [29] and LoopReg [9].
ARCH [29] 和 LoopReg [9] 的數據準備步驟中使用了將 LBS 蒙皮權重從底層身體模型擴展到連續空間。

However, in these approaches, the skinning weights are uniquely determined by the underlining body and not learnable.
但是，在這些方法中，蒙皮權重由下劃線主體唯一確定，並且不可學習。

We argue, and experimentally demonstrate, in Sec. 4.1 that jointly learning skinning weights leads to visually

pleasing canonicalization while maximizing the reproducibility of input scans by the reconstructed parametric model.
我們在 Sec. 4.1 中論證並通過實驗證明，聯合學習蒙皮權重會導致視覺上令人愉悅的規範化，同時通過重建的參數模型最大化輸入掃描的可重複性。

Inspired by recent unsupervised methods using cycle consistency [12, 76], we leverage geometric cycle consistency between the canonical space and posed space to learn skinning weights in a weakly supervised manner without requiring any ground-truth training data.
受最近使用循環一致性 [12, 76] 的無監督方法的啟發，我們利用規範空間和姿勢空間之間的幾何循環一致性以弱監督的方式學習蒙皮權重，而無需任何地面實況訓練數據。

Concurrent work, FTP [66], proposes a similar idea but is limited to body modeling; instead, we extend the traditional LBS to the entire 3D space and enable clothing surface modeling.
並行工作 FTP [66] 提出了類似的想法，但僅限於身體建模； 相反，我們將傳統的 LBS 擴展到整個 3D 空間並啟用服裝表面建模。

Reconstructing Clothed Humans. 重建穿衣人。

Reconstructing humans from depth maps [14, 65, 72], images [11, 32, 36], or video [33, 35] is also extensively studied. While many works focus on the minimally clothed human body [11, 24, 32, 36, 40], recent approaches show promise in reconstructing clothed human models from RGB inputs using the SMPL mesh with displacements [1, 2, 75], external garment layers [10, 31], depth maps [19, 63], voxels [64, 74], or implicit functions [29, 60, 61].
從深度圖 [14、65、72]、圖像 [11、32、36] 或視頻 [33、35] 重建人類也被廣泛研究。 雖然許多作品都集中在穿著最少的人體 [11, 24, 32, 36, 40]，但最近的方法顯示出使用具有位移的 SMPL 網格從 RGB 輸入重建穿著人體模型的前景 [1, 2, 75]，外衣 層 [10、31]、深度圖 [19、63]、體素 [64、74] 或隱函數 [29、60、61]。

However, these approaches do not learn, or infer, pose-dependent deformation of garments, and simply apply articulated deformations to the reconstructed shapes.
然而，這些方法不會學習或推斷服裝的姿勢相關變形，而是簡單地將鉸接變形應用於重建的形狀。

This results in unrealistic pose-dependent deformations that lack garment specific wrinkles. Our work differs by focusing on learning pose-dependent clothing deformation from scans.
這會導致不切實際的姿勢相關變形，缺少服裝特定的皺紋。 我們的工作不同之處在於專注於從掃描中學習與姿勢相關的服裝變形。

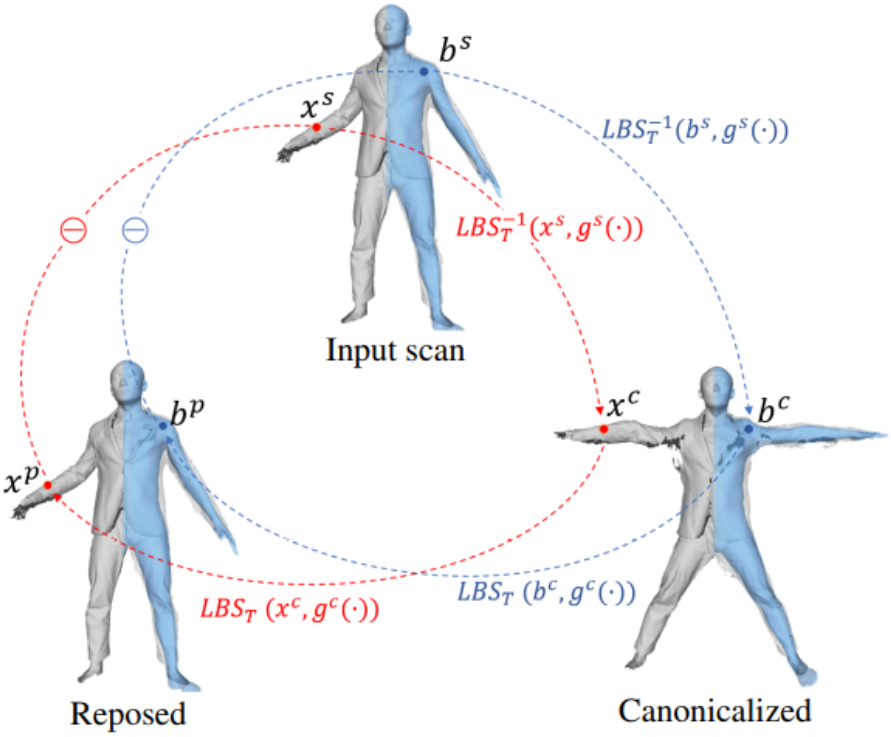

Figure 3: **Canonicalization with cycle consistency.** The geometric cycle consistency loss, with the guidance from the underlining body model, leads to successful canonicalization.

Figure 3: Canonicalization with cycle consistency.
圖 3：具有循環一致性的規範化。

The geometric cycle consistency loss, with the guidance from the underlining body model, leads to successful canonicalization.
在下劃線身體模型的指導下，幾何循環一致性損失導致成功的規範化。

3. Method 方法

Figure 2 shows an overview of our pipeline.
圖 2 顯示了我們管道的概覽。

The input is a set of raw 3D scans of a person in clothing, together with fitted minimally clothed body models.
輸入是一組穿著衣服的人的原始 3D 掃描，以及合身的最少衣服的身體模型。

Here we use the SMPL model [43] fit to the scans to obtain body joints and blend skinning weights, which we exploit in learning.
在這裡，我們使用適合掃描的 SMPL 模型 [43] 來獲取身體關節並混合蒙皮權重，我們在學習中利用這些權重。

Given the input, we first learn bidirectional transformations between the posed space and canonical space by predicting skinning weights as a function of space coordinates (Sec. 3.1).

給定輸入，我們首先通過將蒙皮權重預測為空間坐標的函數來學習姿勢空間和規範空間之間的雙向變換（第 3.1 節）。

To address the lack of ground truth correspondence of the scan data, we leverage geometric cycle consistency to learn continuous skinning functions.
為了解決掃描數據缺乏基本真實對應的問題，我們利用幾何循環一致性來學習連續蒙皮函數。

The raw scans are canonicalized with the learned bidirectional transformations.
原始掃描通過學習到的雙向變換進行規範化。

We further learn a locally pose-aware signed distance function, parameterized by a neural network, from canonicalized scans using implicit geometric regularization [22] (Sec. 3.2).
我們進一步從使用隱式幾何正則化 [22]（第 3.2 節）的規範化掃描中學習由神經網絡參數化的局部姿勢感知有符號距離函數。

For implementation details, including hyper parameters and network architectures, see Appendix A.
有關實現細節，包括超參數和網絡架構，請參見附錄 A。

3.1. Canonicalization 規範化

Instead of a traditional skinning scheme that assigns a skinning weight vector w ∈ RJ , where J is the number of joints, to each point on a surface, we extend the notion of skinning using a continuous function:
與為表面上的每個點分配蒙皮權重向量 w ∈ RJ 的傳統蒙皮方案不同，我們使用連續函數擴展蒙皮的概念：

we train a model that takes any point in the space as input and outputs its skinning weight vector w. Figure 3 illustrates the principles.
我們訓練一個模型，該模型將空間中的任何點作為輸入並輸出其蒙皮權重向量 w。 圖 3 說明了這些原理。

We specifically focus on points from two surfaces, the clothing surface X and the body surface B.
我們特別關注來自兩個表面的點，衣服表面 X 和身體表面 B。

To be more specific about the canonicalization step, let us first define the posed space and the canonical space.
為了更具體地了解規範化步驟，讓我們首先定義姿勢空間和規範空間。

The posed space is defined for each scan, and the canonical space is shared across all the scans.
為每次掃描定義姿勢空間，並且所有掃描共享規範空間。

Let Xsi = {xs ∈ R3} be vertices on the original scan in the posed space, where i is the frame index of the scans,

and Xci= {xc ∈ R3} be vertices on the unposed scans in the canonical space, which are not known.
設 Xsi = {xs ∈ R3} 是在正態空間中原始掃描上的頂點，其中 i 是掃描的幀索引，而 Xci= {xc ∈ R3} 是規範空間中未定式掃描上的頂點，它們是 未知。

Now we seek a mapping function that aligns the posed scans in a canonical pose.
現在我們尋找一個映射函數，以規範姿勢對齊姿勢掃描。

While the mapping function can be arbitrarily defined, we observe that this can be formulated as a composition of the known rigid transformations of body joints, Ti = {Ti j ∈ SE(3), j = 1, … , J}, which come from the fitted SMPL model.
雖然映射函數可以任意定義，但我們觀察到，這可以表示為已知的身體關節剛性變換的組合，Ti = {Ti j ∈ SE(3), j = 1, ... , J}，它們來自 來自擬合的 SMPL 模型。

More specifically, given a set of blending weights w, we define linear blend skinning (LBS) and inverse linear blend skinning (LBS-1) functions as follows:
更具體地說，給定一組混合權重 w，我們定義線性混合蒙皮 (LBS) 和逆線性混合蒙皮 (LBS-1) 函數如下：

$$\begin{aligned} \boldsymbol{X}_i^p &= LBS_{\mathbf{T}_i}(\boldsymbol{X}_i^c, \boldsymbol{w}(\boldsymbol{X}_i^c)) = (\sum w_j \mathbf{T}_{i,j}) \boldsymbol{X}_i^c \\ \boldsymbol{X}_i^c &= LBS_{\mathbf{T}_i}^{-1}(\boldsymbol{X}_i^s, \boldsymbol{w}(\boldsymbol{X}_i^s)) = (\sum w_j \mathbf{T}_{i,j})^{-1} \boldsymbol{X}_i^s, \end{aligned} \quad (1)$$

where Xpi = {xp ∈ R3} are the vertices of the reposed scans and ideally should have the same value as Xsi.
其中 Xpi = {xp ∈ R3} 是放置掃描的頂點，理想情況下應該與 Xsi 具有相同的值。

The LBS function maps arbitrary points in the canonical space to the posed space represented by Ti and the inverse LBS function maps points in the posed space to the canonical space.
LBS 函數將規範空間中的任意點映射到由 Ti 表示的姿勢空間，逆 LBS 函數將姿勢空間中的點映射到規範空間。

In other words, the equations above show that given skinning weights w on vertices, we can not only apply any pose to the canonicalized shapes as in a traditional character animation pipeline [41], but also transform back the posed shapes into the canonical space.
換句話說，上面的等式表明，給定頂點上的蒙皮權重 w，我們不僅可以像傳統角色動畫管道 [41] 那樣將任何姿勢應用於規範化的形狀，而且還可以將構成的形狀轉換回規範空間。

Implicit Skinning Fields. 隱式蒙皮字段。

In contrast to traditional applications, where the skinning weights for each point are predefined, either by artists or by automatic methods [6, 29, 71], skinning weights on the raw scan data are not known a priori.
與傳統應用相比，每個點的蒙皮權重是由藝術家或通過自動方法預定義的 [6, 29, 71]，原始掃描數據的

蒙皮權重不是先驗的。

Fortunately, we can learn them in a weakly supervised manner, such that all the scans can be decomposed into articulated deformations and non-rigid deformations.
幸運的是，我們可以以弱監督的方式學習它們，這樣所有的掃描都可以分解為鉸接變形和非剛性變形。

To this end, we introduce two neural networks called the forward skinning net and the inverse skinning net:
為此，我們引入了兩個神經網絡，稱為正向蒙皮網絡和逆向蒙皮網絡：

$$\begin{aligned} \boldsymbol{w}(\boldsymbol{x}_i^c) &= g_{\Theta_1}^c(\boldsymbol{x}_i^c) : \mathbb{R}^3 \rightarrow \mathbb{R}^J \\ \boldsymbol{w}(\boldsymbol{x}_i^s) &= g_{\Theta_2}^s(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s) : \mathbb{R}^3 \times \mathbb{R}^{\mathcal{Z}_s} \rightarrow \mathbb{R}^J, \end{aligned} \tag{2}$$

where zsi represents a latent embedding, and $\Theta 1$ and $\Theta 2$ are the learnable parameters of the multilayer perceptrons (MLP), which we omit below for notational brevity.
其中 zsi 表示潛在嵌入，$\Theta 1$ 和 $\Theta 2$ 是多層感知器 (MLP) 的可學習參數，為了符號簡潔，我們在下面省略了這些參數。

The forward skinning net predicts LBS skinning weights of queried 3D locations in the canonical space.
前向蒙皮網絡預測規範空間中查詢的 3D 位置的 LBS 蒙皮權重。

Similarly, the inverse skinning net predicts skinning weights in the posed space of each training scan.
類似地，逆皮膚網絡預測每次訓練掃描的姿勢空間中的皮膚權重。

Notably, this continuous representation is advantageous over other alternatives including fully connected networks and graph convolutional networks [42, 46, 58] as it does not depend on a fixed number of vertices or predefined topology.
值得注意的是，這種連續表示優於其他替代方案，包括全連接網絡和圖卷積網絡 [42,46,58]，因為它不依賴於固定數量的頂點或預定義的拓撲。

Empirically we observe that jointly learning zsi in an auto-decoding fashion [54] leads to superior performance compared to taking pose parameters as input; see Appendix B for discussion.
根據經驗，我們觀察到，與將姿勢參數作為輸入相比，以自動解碼[54]方式聯合學習 zsi 會帶來更好的性能； 參見附錄 B 進行討論。

By combining Eq. 1 and 2, we can compute the mappings between the canonical and posed spaces via:
通過結合 Eq. 1 和 Eq. 2，我們可以通過以下方式計算規範空間和姿勢空間之間的映射：

$$\begin{aligned} \boldsymbol{x}_i^p &= LBS_{\mathbf{T}_i}(\boldsymbol{x}_i^c, g^c(\boldsymbol{x}_i^c)) \\ \boldsymbol{x}_i^c &= LBS_{\mathbf{T}_i}^{-1}(\boldsymbol{x}_i^s, g^s(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s)). \end{aligned} \tag{3}$$

Note that these functions are differentiable.
請注意，這些函數是可微的。

Learning Skinning. 學習剝皮。

To successfully train gc(·) and gs(·) without ground truth weights on the scans, we leverage two key observations:
為了在掃描時成功訓練 gc(·) 和 gs(·)，而無需在掃描上使用真實權重，我們利用了兩個關鍵觀察結果：

(1) the regions close to the human body model are highly correlated with the nearest body parts where ground-truth skinning weights are available;
(1) 接近人體模型的區域與最近的可以得到 ground-truth 蒙皮權重的身體部位高度相關；

(2) any points in the posed space should be mapped back to the same points after reapplying LBS to the canonicalized points.
(2) 在將 LBS 重新應用於規範化點後，應將姿勢空間中的任何點映射回相同的點。

To utilize (1), we use the underlying SMPL body model's LBS skinning weights as guidance for the canonical and posed space.
為了利用（1），我們使用底層 SMPL 身體模型的 LBS 蒙皮權重作為規範和姿勢空間的指導。

More specifically, gc(·) and gs(·) at points on the scans are loosely guided by the nearest neighbor point on the body model and its SMPL skinning weights, propagating skinning weights from body models to the input scans.
更具體地說，掃描點上的 gc(·) 和 gs(·) 由身體模型上的最近鄰點及其 SMPL 蒙皮權重鬆散地引導，將蒙皮權重從身體模型傳播到輸入掃描。

Most importantly, observation (2) plays a central role in the success of the weakly supervised learning.
最重要的是，觀察 (2) 在弱監督學習的成功中起著核心作用。

It allows us to formulate cycle consistency constraints, updating both gc(·) and gs(·) such that wrongly associated skinning weights that break the cycle consistency are highly penalized.
它允許我們制定循環一致性約束，更新 gc(·) 和 gs(·)，這樣錯誤關聯的破壞循環一致性的蒙皮權重就會受到高度懲罰。

Our evaluation in Sec. 4.1 shows that the cycle consistency constraints are critical to decompose articulated deformations.
我們在 Sec. 4.1 中的評估表明，循環一致性約束對於分解鉸接變形至關重要。

Note that the jointly learned gc(·) is used to learn and animate the pose-aware clothed human model (see Sec. 3.2).
請注意，聯合學習的 gc(·) 用於學習和動畫姿勢感知的衣服人體模型（參見第 3.2 節）。

Our final objective function is defined as:
我們的最終目標函數定義為：

$$E_{cano}(\Theta_1, \Theta_2, \{\boldsymbol{z}_i^s\}) = \sum_i (\lambda_B E_B + \lambda_S E_S + E_C + E_R), \quad (4)$$

where EB and ES are body-guided loss functions, EC is based on cycle consistency, and ER is a regularization term.
其中 EB 和 ES 是身體引導的損失函數，EC 基於循環一致性，而 ER 是正則化項。

EB ensures gc(·) and gc(·) predict SMPL skinning weights on the body surface by
EB 確保 gc(·) 和 gc(·) 通過以下方式預測體表上的 SMPL 蒙皮權重

$$E_B = \sum_{\boldsymbol{b}_i^c \in \mathbf{B}_i^c} \|g^c(\boldsymbol{b}_i^c) - \boldsymbol{w}'(\boldsymbol{b}_i^c)\| + \sum_{\boldsymbol{b}_i^s \in \mathbf{B}_i^s} \|g^s(\boldsymbol{b}_i^s, \boldsymbol{z}_i^s) - \boldsymbol{w}'(\boldsymbol{b}_i^s)\|, \quad (5)$$

where Bci = {bci ∈ R3} and Bsi = {bsi ∈ R3} are vertices on the canonical and posed body surfaces, and w' (·) are the SMPL LBS weights.
其中 Bci = {bci ∈ R3} 和 Bsi = {bsi ∈ R3} 是規範和姿勢體表面上的頂點，w' (·) 是 SMPL LBS 權重。

See Appendix A for details to obtain Bci .
有關獲取 Bci 的詳細信息，請參閱附錄 A。

Similarly, ES is the regression loss between the predicted weights and the LBS weights on the nearest neighbor body vertex:
同樣，ES 是預測權重與最近鄰體頂點上的 LBS 權重之間的回歸損失：

$$E_S = \sum_{\boldsymbol{x}_i^s \in \mathbf{X}_i^s} (\|g^s(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s) - \boldsymbol{w}'(\operatorname{argmin}_{\boldsymbol{b}_i^s \in \mathbf{B}_i^s} d(\boldsymbol{x}_i^s, \boldsymbol{b}_i^s))\| + \|g^c(\boldsymbol{x}_i^c) - \boldsymbol{w}'(\operatorname{argmin}_{\boldsymbol{b}_i^s \in \mathbf{B}_i^s} d(\boldsymbol{x}_i^s, \boldsymbol{b}_i^s))\|). \quad (6)$$

Note that this nearest neighbor assignment is also used in [29] for training data preparation.
請注意，此最近鄰分配也用於 [29] 中的訓練數據準備。

However, in Sec. 4.1, we show that this alone is prone to inaccurate assignments, causing severe artifacts.
然而，在 Sec. 4.1 中，我們表明僅此一項就容易導致分配不准確，從而導致嚴重的偽影。

We facilitate cycle consistency with two terms.
我們通過兩個術語促進循環一致性。

EC' directly constrains the consistency of skinning weights between the canonical space and the posed space, and EC00 facilitates cycle consistency on the vertices of the posed meshes as follows:
EC' 直接約束規範空間和姿勢空間之間蒙皮權重的一致性，EC00 促進了姿勢網格頂點的循環一致性，如下所示：

$$E_C = \lambda_{C'} E_{C'} + \lambda_{C''} E_{C''} \tag{7}$$

$$E_{C'} = \sum_{\boldsymbol{x}_i^s \in \mathbf{X}_i^s} \|g^s(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s) - g^c(\boldsymbol{x}_i^c)\| \tag{8}$$

$$E_{C''} = \sum_{\boldsymbol{x}_i^s \in \mathbf{X}_i^s} \|\boldsymbol{x}_i^p - \boldsymbol{x}_i^s\|. \tag{9}$$

Notice that cycle consistency can hold only if we start from the posed space since points in the canonical space can be mapped to the same location in case of self-intersection.
請注意，只有當我們從姿勢空間開始時，循環一致性才能保持，因為在自相交的情況下，規範空間中的點可以映射到相同的位置。

Lastly, our regularization term consists of a sparsity constraint ESp, a smoothness term ESm, and a statistical regularization on the latent code EZ as follows:
最後，我們的正則化項由稀疏約束 Esp、平滑項 ESm 和潛在代碼 EZ 上的統計正則化組成，如下所示：

$$E_R = \lambda_{Sp} E_{Sp} + \lambda_{Sm} E_{Sm} + \lambda_Z E_Z, \tag{10}$$

$$E_{Sp} = \sum_{\boldsymbol{x}_i^s} |g^s(\boldsymbol{x}_i^s, \boldsymbol{z}_i^s)|^\beta \quad \beta = 0.8, \tag{11}$$

$$E_{Sm} = \sum_{\boldsymbol{e} \in \mathbf{E}/\mathbf{C}} \|g^s(\boldsymbol{e}_1, \boldsymbol{z}_i^s) - g^s(\boldsymbol{e}_2, \boldsymbol{z}_i^s)\|, \tag{12}$$

$$E_Z = \|\boldsymbol{z}_i^p\|_2^2, \tag{13}$$

where e = (e1, e2), E is the set of edges on the triangulated scans and we mask out concave regions C so that

skinning weights are not propagated across merged body parts due to self-intersection (See Appendix A for details.).
其中 e = (e1, e2)，E 是三角掃描上的邊集，我們屏蔽了凹面區域 C，這樣由於自相交，蒙皮權重不會在合併的身體部位之間傳播（詳見附錄 A）。

After training, we canonicalize all the scans by applying the inverse LBS transform (Eq. 3) to all vertices on the scans.
訓練後，我們通過將逆 LBS 變換（Eq. 3）應用於掃描上的所有頂點來規範化所有掃描。

By eliminating triangles with large distortion (see Appendix A for details), we obtain the canonical scans used to learn a pose-aware parametric clothed human model.
通過消除大失真的三角形（詳見附錄 A），我們獲得了用於學習姿勢感知參數化服裝人體模型的規範掃描。

3.2. Locally Pose-aware Implicit Shape Learning 局部姿態感知隱式形狀學習

Given the canonicalized partial scans together with the learned skinning weights, we learn a parametric clothed human model with pose-aware deformations.
給定規範化的部分掃描以及學習的蒙皮權重，我們學習了具有姿勢感知變形的參數化服裝人體模型。

To this end, we base our shape representation on an implicit surface representation [13, 47, 54] as it supports arbitrary topology with fine details.
為此，我們將形狀表示基於隱式表面表示 [13, 47, 54]，因為它支持具有精細細節的任意拓撲。

However, real scans have holes and such partial scans cause difficulty obtaining ground truth occupancy labels since the meshes are not water-tight.
然而，真實的掃描有漏洞，並且由於網格不防水，因此這種部分掃描導致難以獲得地面真實佔用標籤。

To handle partial scans as input, we learn a signed distance function f(x) based on a multilayer perceptron (for brevity, we omit the network parameters), using implicit geometric regularization (IGR) [22] by minimizing the following objective function:
為了將部分掃描作為輸入處理，我們學習了一個基於多層感知器的有符號距離函數 f(x)（為簡潔起見，我們省略了網絡參數 ），使用隱式幾何正則化 (IGR) [22] 通過最小化以下目標函數：

$$E_{shape}(\Phi) = \sum_i \left(E_{LS} + \lambda_{igr} E_{IGR} + \lambda_o E_O\right) \quad (14)$$

$$E_{LS} = \sum_{\boldsymbol{x} \in \mathbf{X}'^c_i} \left(|f(\boldsymbol{x})| + \|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) - \boldsymbol{n}(\boldsymbol{x})\|\right), \quad (15)$$

$$E_{IGR} = \mathbb{E}_{\boldsymbol{x}} \left(\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x})\| - 1\right)^2, \quad (16)$$

$$E_O = \mathbb{E}_{\boldsymbol{x}} \left(\exp(-\alpha \cdot |f(\boldsymbol{x})|)\right) \quad \alpha \gg 1, \quad (17)$$

where ELS ensures the zero level-set of the predicted SDF lies on the given points with its surface normal aligned with that of the input scans, n(x).
其中 ELS 確保預測 SDF 的零水平集位於給定點上，其表面法線與輸入掃描的表面法線 n(x) 對齊。

EIGR is the Eikonal regularization term that regularizes the function f to satisfy the Eikonal equation || ∇xf(·)|| = 1.
EIGR 是 Eikonal 正則化項，它對函數 f 進行正則化以滿足 Eikonal 方程 || ∇xf(·)|| = 1。

EO regularizes off-surface SDF values from being close to the level-set surface as in [62].
EO 正則化表面 SDF 值，使其遠離水平集表面，如 [62] 中所示。

Remarkably, this formulation does not require ground truth signed distance for non-surface points and naturally fills in the missing regions by leveraging the inductive bias of multilayer perceptrons as shown in [22].
值得注意的是，該公式不需要非表麵點的真實符號距離，並通過利用多層感知器的歸納偏差自然地填充缺失區域，如 [22] 所示。

To learn pose-dependent deformations of clothing, we could condition the function f with the pose features θ∈RJx4 (we use quaternions as in [53]).
為了學習服裝的姿勢相關變形，我們可以用姿勢特徵 θ∈RJx4 來調節函數 f（我們使用 [53] 中的四元數）。

However, the straightforward approach of concatenating the pose features with Cartesian coordinates, namely f(x, θ), suffers from overfitting due to the limited pose variations in the training data and spurious correlations between joints.
然而，將姿勢特徵與笛卡爾坐標連接的直接方法，即 f(x, θ)，由於訓練數據中姿勢變化有限和關節之間的虛假相關性而遭受過度擬合。

Since the relationship between body joints and clothing deformation tends to be non-local [68], it is also important to limit the influence of irrelevant joints to reduce spurious correlations [53].
由於身體關節和服裝變形之間的關係往往是非局部的 [68]，因此限制無關關節的影響以減少虛假相關性也很重要 [53]。

Thus, we need an attention mechanism to associate spatial locations with only the relevant pose features.
因此，我們需要一種注意力機制來將空間位置僅與相關的姿勢特徵相關聯。

To this end, we modify the function f:
為此，我們修改函數 f：

$$f(\boldsymbol{x}, (W \cdot g^c(\boldsymbol{x})) \circ \theta), W \in \mathbb{R}^{J \times J}, \qquad (18)$$

where gc(·) is the skinning network learned in Sec. 3.1, W is the weight map that converts skinning weights into pose attention weights, and denotes element-wise product.
其中 gc(·) 是在 Sec. 3.1 中學習的蒙皮網絡，W 是將蒙皮權重轉換為姿勢注意權重的權重圖，並表示逐元素乘積。

Specifically, if we want a 3D point that is skinned to the nth joint with non-zero skinning weights to pay attention to the mth joint, Wm,n and Wn,m are set to 1, otherwise, they are set to 0.
具體來說，如果我們想要一個以非零蒙皮權重剝皮到第 n 個關節的 3D 點注意第 m 個關節，則 Wm,n 和 Wn,m 設置為 1，否則設置為 0。

The weight map is essential because the movement of one joint will be propagated to regions associated with neighboring body joints (e.g. raising the shoulders lifts up an entire T-shirt).
權重圖是必不可少的，因為一個關節的運動將傳播到與相鄰身體關節相關的區域（例如，抬高肩膀會抬起整件 T 卹）。

In this paper, we set Wn,m = 1 when nth joint is within 4-ring neighbors of mth joint in the kinematic tree.
在本文中，當第 n 個關節在運動樹中第 m 個關節的 4 環鄰居內時，我們設置 Wn,m = 1。

By reducing spurious correlations, our formulation significantly reduces over-fitting artifacts given a set of unseen poses, demonstrating better generalization ability even with a small number of input scans (see Sec. 4.1).
通過減少虛假相關性，我們的公式顯著減少了給定一組看不見的姿勢的過度擬合偽影，即使使用少量輸入掃描也表現出更好的泛化能力（參見第 4.1 節）。

4. Experimental Results 實驗結果

Dataset and Metric. 數據集和指標。

For evaluation and comparison with baseline methods, we use the CAPE dataset [46], which includes raw 3D scan sequences and SMPL model fits.
為了評估和與基線方法進行比較，我們使用 CAPE 數據集 [46]，其中包括原始 3D 掃描序列和 SMPL 模型擬合。

We evaluate generalization to unseen poses with both pose interpolation (denoted as Int. in tables) and extrapolation tasks (denoted as Ex. in tables).
我們使用姿勢插值（在表中表示為 Int.）和外推任務（在表中表示為 Ex.）來評估對未知姿勢的泛化。

The motion sequences are randomly split into training (80%) and test (20%) sets, where the test sequences are used to evaluate extrapolation.
運動序列被隨機分成訓練 (80%) 和測試 (20%) 集，其中測試序列用於評估外推。

For the training sequences, we choose every 10th frame starting from the first frame as training scans and every 10th frame with 5 frame strides from the training sequences for the interpolation evaluation.
對於訓練序列，我們選擇從第一幀開始的每 10 幀作為訓練掃描，每 10 幀從訓練序列中選擇 5 幀步幅進行插值評估。

We perform Marching Cubes [44] to the predicted implicit surface in canonical space as in Eq 18 and then pose it by forward LBS in Eq. 1 to get the resulting meshes.
我們在標準空間中對預測的隱式表面執行行進立方體 [44]，如方程 18 所示，然後通過方程 18 中的前向 LBS 構成它。 1 獲取生成的網格。

For quantitative evaluation, we use scanto-mesh distance Ds2m (cm) and surface normal consistency Dn, where a nearest neighbor vertex on the resulting meshes is used to compute the average L2 norm.
對於定量評估，我們使用掃描到網格距離 Ds2m (cm) 和表面法線一致性 Dn，其中結果網格上的最近鄰頂點用於計算平均 L2 範數。

In addition, we conduct a perceptual study to assess the plausibility score, P, of generated garment shapes and deformations.
此外，我們進行了一項感知研究，以評估生成的服裝形狀和變形的合理性分數 P。

Workers on Amazon Mechanical Turk (AMT) are given a pair of side-by-side images or videos showing a rendered result from our approach and another approach;
Amazon Mechanical Turk (AMT) 上的工作人員會收到一對並排的圖像或影像，顯示我們的方法和另一種方法的渲染結果；

the left-right order of the results is randomized.
結果的左右順序是隨機的。

The task is to choose the result with the most realistic clothing.
任務是選擇具有最逼真服裝的結果。

We continue this N times and compute the probability of the other approach being favored P = M/N, where M is how many times the users chose the other method over ours.

我們繼續這 N 次併計算另一種方法被青睞的概率 P = M/N，其中 M 是用戶選擇另一種方法而不是我們的次數。

In other words, we set our approach as baseline with a constant score P = 0.5; for other approaches, if P < 0.5, ours achieves higher fidelity.
換句話說，我們將我們的方法設置為基線，恆定分數 P = 0.5； 對於其他方法，如果 P < 0.5，我們的實現了更高的保真度。

The perceptual score for image and video pairs is denoted as Pi and Pv, respectively. While Pi focuses on the plausibility of static clothing, Pv reveals the temporal consistency and realism of pose-dependent clothing deformations.
圖像和視頻對的感知分數分別表示為 Pi 和 Pv。 Pi 側重於靜態服裝的合理性，而 Pv 則揭示了與姿勢相關的服裝變形的時間一致性和真實性。

Note that we provide only the perceptual scores for the extrapolation task as numerical evaluation is difficult due to the stochasticity of clothing deformations.
請注意，我們僅提供外推任務的感知分數，因為由於服裝變形的隨機性，數值評估很困難。

4.1. Evaluation 評估

Canonicalization. 規範化。

The goal of canonicalization is to disentangle articulated deformations from other non-rigid deformations for effective shape learning.
規範化的目標是將鉸接變形與其他非剛性變形分開，以實現有效的形狀學習。

We choose two baseline approaches to replace our canonicalization module.
我們選擇兩種基線方法來替換我們的規範化模塊。

The first, as used by [29], copies skinning weights on the clothed scans from the nearest neighbor body vertex.
第一個，如 [29] 所用，從最近的相鄰身體頂點複製衣服掃描的蒙皮權重。

The other approach is based on weighted correspondences by interpolating skinning weights from the k nearest neighbors (we use k = 6) in the spirit of [71].
另一種方法是根據 [71] 的精神，通過從 k nearest neighbors（我們使用 k = 6）插入蒙皮權重來基於加權對應。

This reduces the impact of a wrong clothing-body association that limits the performance of single nearest neighbor assignment.
這減少了限制單個最近鄰分配性能的錯誤服裝-身體關聯的影響。

Figure 4 shows that the two baseline methods break the cycle consistency with wrong associations of the skinning weights, resulting in noticeable artifacts.
圖 4 顯示了兩種基線方法由於蒙皮權重的錯誤關聯而破壞了循環一致性，從而導致了明顯的偽影。

The inaccurate canonicalization results are propagated to the parametric model learning, substantially degrading the quality of reconstructed avatars as shown in Tab. 1.
不准確的規範化結果被傳播到參數模型學習，大大降低了重建化身的質量，如表 1 所示。 1.

Our approach with cycle consistency successfully normalizes the input scans into a canonical pose while retaining coherent geometric details, enabling the parametric modeling of clothed avatars.
我們的循環一致性方法成功地將輸入掃描歸一化為規範姿勢，同時保留了連貫的幾何細節，從而實現了服裝化身的參數化建模。

Locally Pose-aware Shape Learning. 局部姿勢感知形狀學習。

We evaluate our local pose representation using the learned skinning weights for pose-dependent shape learning and compare against commonly used global pose conditioning [18, 39, 46, 55, 71].
我們使用學習的蒙皮權重來評估我們的局部姿勢表示，用於姿勢相關的形狀學習，並與常用的全局姿勢調節進行比較 [18, 39, 46, 55, 71]。

To this end, we replace the second input of Eq. 18 with the global pose parameter, θ , as a baseline.
為此，我們將 Eq. 18 的第二個輸入替換為全局姿態參數 θ 作為基線。

To assess the generalization ability, both models are trained on 100%, 50%, 10% and 5% of the original training set.
為了評估泛化能力，兩個模型都在原始訓練集的 100%、50%、10% 和 5% 上進行訓練。

Table 2 shows that our local pose conditioning achieves better reconstruction accuracy and fidelity for both interpolation and extrapolation.
表 2 表明，我們的局部姿態調節在插值和外插方面實現了更好的重建精度和保真度。

Note that the performance of global pose conditioning drastically degrades when the training data is reduced to less than 10%, suffering from severe overfitting.
請注意，當訓練數據減少到 10% 以下時，全局姿態調節的性能會急劇下降，遭受嚴重的過度擬合。

In contrast, our approach keeps roughly equivalent reconstruction accuracy even when only 5% of the original training data is used, exhibiting few noticeable artifacts (see Fig. 5).
相比之下，即使僅使用 5% 的原始訓練數據，我們的方法也保持大致相當的重建精度，幾乎沒有明顯的偽影（見圖 5）。

Comparison with SoTA. 與 SoTA 的比較。

We compare the proposed method with two state-of-the-art methods that also learn an articulated parametric human model with pose correctives from real world scans [18, 46].
我們將所提出的方法與兩種最先進的方法進行比較，這兩種方法還可以學習帶有來自現實世界掃描的姿勢校正的鉸接參數人體模型 [18, 46]。

CAPE [46] learns posedependent deformations on a fixed mesh topology using graph convolutions [58], but requires surface registration for training.
CAPE [46] 使用圖卷積 [58] 在固定網格拓撲上學習姿態相關變形，但需要表面配准進行訓練。

NASA [18], on the other hand, can be learned without registration but needs to determine occupancy values.
另一方面，NASA [18] 無需註冊即可學習，但需要確定佔用值。

We train both methods using registered CAPE data.
我們使用註冊的 CAPE 數據訓練這兩種方法。

Table 3 shows that our approach achieves superior reconstruction accuracy and perceptual realism, while Fig. 6 illustrates limitations of the prior methods.
表 3 顯示我們的方法實現了卓越的重建精度和感知真實感，而圖 6 說明了先前方法的局限性。

As CAPE relies on a template mesh with a fixed topology, the reconstructions are not only less detailed but also fail to capture topological changes such as the lifting up of the jacket.
由於 CAPE 依賴於具有固定拓撲結構的模板網格，因此重建不僅不夠詳細，而且無法捕捉拓撲變化，例如夾克的抬起。

While NASA can model pose-dependent shapes using articulated implicit functions, discontinuites and ghosting artifacts are visible, as the implicit functions of each body part are learned independently, which limits generalization to unseen poses.
雖然 NASA 可以使用鉸接的隱函數對依賴於姿勢的形狀進行建模，但由於每個身體部位的隱函數是獨立學習的，因此可以看到不連續和重影偽影，這限制了對看不見的姿勢的泛化。

Table 1: Quantitative comparison of canonicalization. As the perceptual score is pair-wise and compared against ours, we put 0.5 for the proposed approach throughout the tables. D_{s2m} is in centimeters throughout the tables.

		Ours	NN [29]	KNN
Int.	$D_{s2m} \downarrow$	**0.570**	1.25	1.25
	$D_n \downarrow$	**0.253**	0.301	0.299
	$P_i \uparrow$	**0.5**	0.374	0.396
	$P_v \uparrow$	**0.5**	0.435	0.431
Ex.	$P_i \uparrow$	**0.5**	0.262	0.312
	$P_v \uparrow$	**0.5**	0.392	0.449

Table 1: Quantitative comparison of canonicalization.
表 1：規範化的定量比較。

As the perceptual score is pair-wise and compared against ours, we put 0.5 for the proposed approach throughout the tables.
由於感知分數是成對的並與我們的比較，我們在整個表格中為所提出的方法設置了 0.5。

Ds2m is in centimeters throughout the tables.
Ds2m 在整個表格中以厘米為單位。

Table 2: Quantitative evaluation of the importance of locality in the pose conditioning on different sizes of training data.

Train size (%)		100	50	10	5
Local pose conditioning (Ours)					
Int.	$D_{s2m} \downarrow$	**0.570**	0.663	0.699	0.732
	$D_n \downarrow$	**0.253**	0.253	0.261	0.268
	$P_i \uparrow$	**0.5**	0.476	0.466	0.398
	$P_v \uparrow$	**0.5**	0.453	0.435	0.425
Ex.	$P_i \uparrow$	**0.5**	0.429	0.359	0.359
	$P_v \uparrow$	**0.5**	0.408	0.408	0.343
Global pose conditioning					
Int.	$D_{s2m} \downarrow$	0.768	0.786	1.54	2.38
	$D_n \downarrow$	0.253	0.256	0.293	0.354
	$P_i \uparrow$	0.424	0.393	0.350	0.252
	$P_v \uparrow$	0.468	0.457	0.363	0.301
Ex.	$P_i \uparrow$	0.417	0.401	0.291	0.192
	$P_v \uparrow$	0.436	0.382	0.311	0.203

Table 2: Quantitative evaluation of the importance of locality in the pose conditioning on different sizes of training data.
表 2：對不同規模訓練數據的姿勢調節中局部重要性的定量評估。

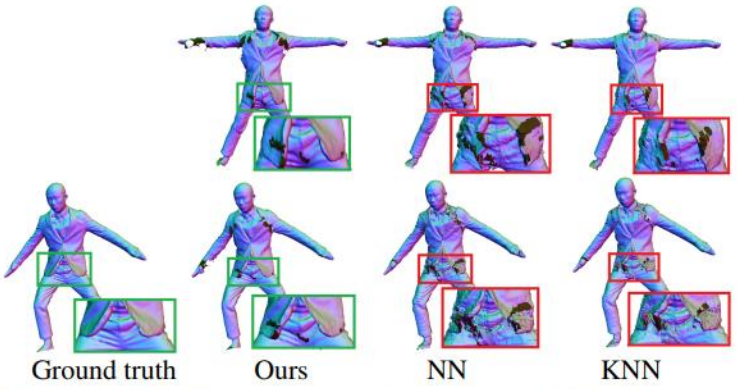

Figure 4: **Qualitative comparison on canonicalization.** Top: canonicalization results. Bottom: reposed canonicalization results. Compared with our method, the baseline methods suffer from severe artifacts.

Figure 4: Qualitative comparison on canonicalization.
圖 4：規範化的定性比較。

Top: canonicalization results.
頂部：規範化結果。

Bottom: reposed canonicalization results.
底部：保留的規範化結果。

Compared with our method, the baseline methods suffer from severe artifacts.
與我們的方法相比，基線方法存在嚴重的偽影。

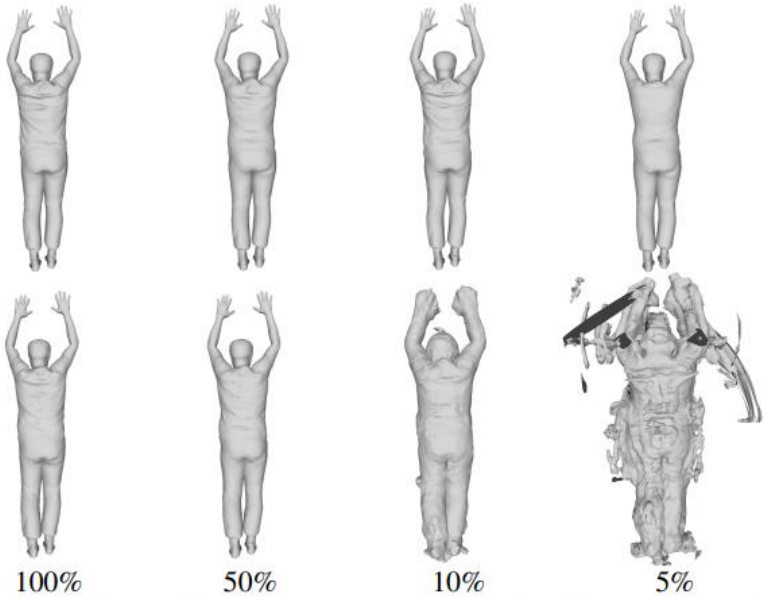

Figure 5: **Evaluation of pose encoding with different sizes of training data.** Top row: our local pose encoding. Bottom row: global pose encoding. While the global pose encoding suffers from severe overfitting artifacts, our local pose encoding generalizes well even if data size is severely limited.

Figure 5: Evaluation of pose encoding with different sizes of training data. Top row: our local pose encoding.
圖 5：使用不同大小的訓練數據評估姿勢編碼。 第一行：我們的局部姿勢編碼。

Bottom row: global pose encoding. While the global pose encoding suffers from severe overfitting artifacts, our local pose encoding generalizes well even if data size is severely limited.
底行：全局姿態編碼。 雖然全局姿態編碼遭受嚴重的過度擬合偽影，但即使數據大小受到嚴重限制，我們的局部姿態編碼也能很好地泛化。

In contrast, our approach can produce highly detailed and globally coherent pose-dependent deformations without template-registration.
相比之下，我們的方法可以在沒有模板註冊的情況下產生高度詳細和全局連貫的姿態相關變形。

Learning a Fully Textured Avatar. 學習全紋理化身。

We extend our poseaware shape modeling to appearance modeling by predicting texture fields [52, 60]; see Appendix A for details.
我們通過預測紋理字段將姿態感知形狀建模擴展到外觀建模 [52, 60]； 有關詳細信息，請參閱附錄 A。

Figure 7 shows that high-resolution texture can be modeled without 2D texture mapping, which illustrates another advantage of eliminating the template-mesh requirement.
圖 7 顯示可以在沒有 2D 紋理映射的情況下對高分辨率紋理進行建模，這說明了消除模板網格要求的另一個優勢。

Table 3: Comparison with the state-of-the-art pose-aware shape modeling methods.

		Ours	CAPE [46]	NASA [18]
Int.	$D_{s2m} \downarrow$	**0.570**	0.970	1.12
	$D_n \downarrow$	**0.253**	0.308	0.289
	$P_i \uparrow$	**0.5**	0.268	0.432
	$P_v \uparrow$	**0.5**	0.455	0.457
Ex.	$P_i \uparrow$	**0.5**	0.214	0.343
	$P_v \uparrow$	**0.5**	0.422	0.395

Table 3: Comparison with the state-of-the-art pose-aware shape modeling methods.
表 3：與最先進的姿勢感知形狀建模方法的比較。

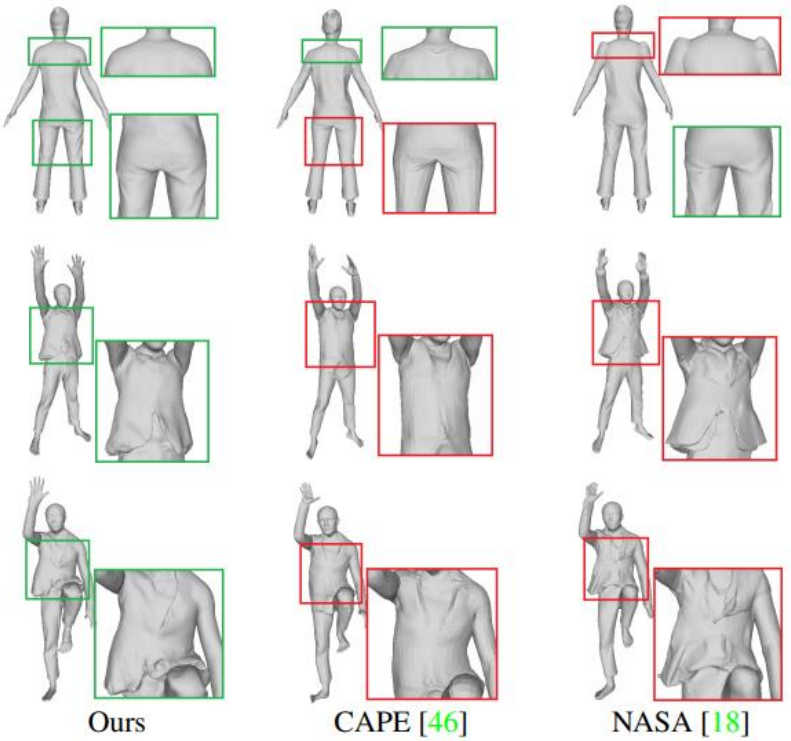

Figure 6: **Comparison with the SoTA methods.** We show qualitative results on the extrapolation task, illustrating the advantages of our method as well as the limitations of the existing approaches.

Figure 6: Comparison with the SoTA methods.
圖 6：與 SoTA 方法的比較。

We show qualitative results on the extrapolation task, illustrating the advantages of our method as well as the limitations of the existing approaches.
我們展示了外推任務的定性結果，說明了我們方法的優點以及現有方法的局限性。

Figure 7: **Textured Scanimats.** Our method can be extended to texture modeling, allowing us to automatically build a Scanimat with high-resolution realistic texture.

Figure 7: Textured Scanimats.
圖 7：紋理掃描圖。

Our method can be extended to texture modeling, allowing us to automatically build a Scanimat with high-resolution realistic texture.
我們的方法可以擴展到紋理建模，使我們能夠自動構建具有高分辨率逼真紋理的 Scanimat。

5. Discussion and Future Work 討論和未來工作

We introduced SCANimate, a fully automatic framework to create high-quality avatars (Scanimats), with realistic clothing deformations, driven by pose parameters, that are directly learned from raw 3D scans.
我們引入了 SCANimate，這是一個全自動框架，用於創建高質量的化身 (Scanimats)，具有逼真的服裝變形，由姿勢參數驅動，直接從原始 3D 掃描中學習。

Our experiments show that decomposing articulated deformations from scanned data is now possible in a weakly supervised manner by combining body-guided supervision with cycle-consistency regularization.
我們的實驗表明，通過將身體引導的監督與循環一致性正則化相結合，現在可以以弱監督的方式從掃描數據中分解鉸接變形。

Previously, the difficulty of accurate and coherent surface registration limited the field from analysing and modeling complex clothing deformations involving multiple garments from real-world observations.
以前，準確和連貫的表面配準的困難限制了該領域分析和建模涉及來自真實世界觀察的多件服裝的複

雜服裝變形。

Our approach enables, for the first time, learning of physically plausible clothing deformations from raw scans, unlocking the possibility of realistic avatar learning from data.
我們的方法第一次能夠從原始掃描中學習物理上合理的服裝變形，從而開啟了從數據中學習真實化身的可能性。

Limitations and Future Work. The current representation works well for clothing that is topologically similar to the body.
局限性和未來的工作。 當前的表示適用於拓撲類似於身體的服裝。

The method may fail for clothing, like skirts, that deviates significantly from the body; see Appendix B for an example.
該方法可能不適用於與身體顯著偏離的衣服，例如裙子； 有關示例，請參見附錄 B。

Clothing wrinkles tend to be stochastic; that is, for a specific pose, they may differ depending on the preceding sequence of poses.
服裝皺紋往往是隨機的； 也就是說，對於特定的姿勢，它們可能會根據前面的姿勢序列而有所不同。

The current model, however, is deterministic. Future work should factor the surface texture into albedo, shape, and lighting enabling more realistic relighting of Scanimats.
然而，當前的模型是確定性的。 未來的工作應該將表面紋理考慮到反照率、形狀和照明中，從而實現更逼真的 Scanimat 重新照明。

Additionally, an adversarial texture loss [28] could improve visual quality.
此外，對抗性紋理損失 [28] 可以提高視覺質量。

Here we model a person in a single garment.
在這裡，我們為穿著單件衣服的人建模。

Learning a generative model with clothing variety should be possible but will require training data of varied clothing in varied poses.
學習具有多種服裝的生成模型應該是可能的，但需要不同姿勢的不同服裝的訓練數據。

Most exciting is the idea of fitting Scanimats to, or even learning them from, images or videos.
最令人興奮的是將 Scanimat 擬合到圖像或視頻中，甚至從圖像或視頻中學習它們的想法。

Finally, extending this approach to model hand articulation and facial expressions should be possible using expressive body models like SMPL-X [56].
最後，應該可以使用像 SMPL-X [56] 等富有表現力的身體模型將這種方法擴展到手部關節和麵部表情的模型。