On Self-Contact and Human Pose
關於自我接觸和人體姿勢

Lea Müller, Ahmed A. A. Osman, Siyu Tang, Chun-Hao P. Huang, Michael J. Black

Max Planck Institute for Intelligent Systems, Tubingen

ETH Zurich

{lea.mueller, ahmed.osman, stang, paul.huang, black}@tuebingen.mpg.de

https://arxiv.org/abs/2104.03176
https://tuch.is.tue.mpg.de/

Abstract 摘要

People touch their face 23 times an hour, they cross their arms and legs, put their hands on their hips, etc.
人們每小時摸臉 23 次，他們交叉雙臂和雙腿，把手放在臀部等。

While many images of people contain some form of selfcontact, current 3D human pose and shape (HPS) regression methods typically fail to estimate this contact.
雖然許多人的圖像包含某種形式的自我接觸，但當前的 3D 人體姿勢和形狀 (HPS) 回歸方法通常無法估計這種接觸。

To address this, we develop new datasets and methods that significantly improve human pose estimation with self-contact.
為了解決這個問題，我們開發了新的數據集和方法，通過自我接觸顯著改善人體姿勢估計。

First, we create a dataset of 3D Contact Poses (3DCP) containing SMPL-X bodies fit to 3D scans as well as poses from AMASS, which we refine to ensure good contact.
首先，我們創建了一個 3D 接觸姿勢 (3DCP) 數據集，其中包含適合 3D 掃描的 SMPL-X 身體以及來自 AMASS 的姿勢，我們對其進行改進以確保良好的接觸。

Second, we leverage this to create the Mimic-The-Pose (MTP) dataset of images, collected via Amazon Mechanical Turk, containing people mimicking the 3DCP poses with selfcontact.
其次，我們利用它來創建通過 Amazon Mechanical Turk 收集的 Mimic-The-Pose (MTP) 圖像數據集，其中包含通過自我接觸模仿 3DCP 姿勢的人。

Third, we develop a novel HPS optimization method, SMPLify-XMC, that includes contact constraints and uses

the known 3DCP body pose during fitting to create near ground-truth poses for MTP images.
第三，我們開發了一種新穎的 HPS 優化方法 SMPLify-XMC，它包括接觸約束並在擬合期間使用已知的 3DCP 身體姿勢為 MTP 圖像創建接近真實的姿勢。

Fourth, for more image variety, we label a dataset of in-the-wild images with Discrete Self-Contact (DSC) information and use another new optimization method, SMPLify-DC, that exploits discrete contacts during pose optimization.
第四，為了獲得更多的圖像多樣性，我們使用離散自接觸 (DSC) 信息標記野外圖像數據集，並使用另一種新的優化方法 SMPLify-DC，該方法在姿勢優化過程中利用離散接觸。

Finally, we use our datasets during SPIN training to learn a new 3D human pose regressor, called TUCH (Towards Understanding Contact in Humans).
最後，我們在 SPIN 訓練期間使用我們的數據集來學習一個新的 3D 人體姿勢回歸器，稱為 TUCH（Towards Understanding Contact in Humans）。

We show that the new selfcontact training data significantly improves 3D human pose estimates on withheld test data and existing datasets like 3DPW.
我們表明，新的自我接觸訓練數據顯著改善了對保留測試數據和現有數據集（如 3DPW）的 3D 人體姿勢估計。

Not only does our method improve results for selfcontact poses, but it also improves accuracy for non-contact poses.
我們的方法不僅改善了自接觸姿勢的結果，而且還提高了非接觸姿勢的準確性。

The code and data are available for research purposes at https://tuch.is.tue.mpg.de.
程式碼和數據可用於研究目的，網址為 https://touch.is.the.mpg.de。

1. Introduction 前言

Self-contact takes many forms.
自我接觸有多種形式。

We touch our bodies both consciously and unconsciously [25].
我們有意識地和無意識地觸摸我們的身體 [25]。

For the major limbs, contact can provide physical support, whereas we touch our faces in ways that convey our emotional state.
對於主要肢體，接觸可以提供身體支持，而我們以表達情緒狀態的方式觸摸我們的臉。

We perform self-grooming, we have nervous gestures, and we communicate with each other through combined face and hand motions (e.g. “shh”).

我們進行自我修飾，我們有緊張的手勢，我們通過面部和手部動作（例如“噓”）相互交流。

We may wring our hands when worried, cross our arms when defensive, or put our hands behind our head when confident.
我們可能會在擔心時擰手，在防禦時交叉雙臂，或者在自信時把手放在腦後。

A Google search for “sitting person” or “thinking pose” for example, will return images, the majority of which, contain self-contact.
例如，Google search “坐著的人”或“思考姿勢”將返回圖像，其中大部分包含自我接觸。

Although self-contact is ubiquitous in human behavior, it is rarely explicitly studied in computer vision.
儘管自我接觸在人類行為中無處不在，但在計算機視覺中卻很少被明確研究。

For our purposes, self-contact comprises “self touch” (where the hands touch the body) and contact between other body parts (e.g. crossed legs).
就我們的目的而言，自我接觸包括“自我接觸”（手接觸身體）和身體其他部位之間的接觸（例如交叉雙腿）。

We ignore body parts that are frequently in contact (e.g. at the crotch or armpits) and focus on contact that is communicative or functional.
我們忽略經常接觸的身體部位（例如在胯部或腋窩處），而專注於交流或功能性的接觸。

Our goal is to estimate 3D human pose and shape (HPS) accurately for any pose.
我們的目標是準確估計任何姿勢的 3D 人體姿勢和形狀 (HPS)。

When self-contact is present, the estimated pose should reflect the true 3D contact.
當存在自接觸時，估計的姿勢應反映真實的 3D 接觸。

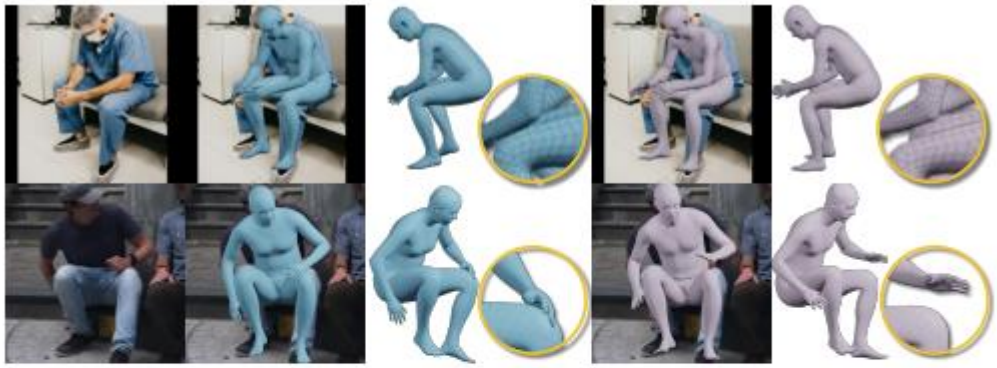

Figure 1. The first column shows images containing self-contact. In blue (left), results of TUCH, compared to SPIN results in violet (right). When rendered from the camera view, the estimated pose may look fine (column two vs. four). However, when rotated, it is clear that training TUCH with self-contact information improves 3D pose estimation (column three vs. five).

Figure 1. The first column shows images containing self-contact.
圖 1. 第一列顯示包含自我接觸的圖像。

In blue (left), results of TUCH, compared to SPIN results in violet (right).
藍色（左）是 TUCH 的結果，而 SPIN 的結果是紫色（右）。

When rendered from the camera view, the estimated pose may look fine (column two vs. four).
從相機視圖渲染時，估計的姿勢可能看起來不錯（第二列與第四列）。

However, when rotated, it is clear that training TUCH with self-contact information improves 3D pose estimation (column three vs. five).
然而，當旋轉時，很明顯用自接觸信息訓練 TUCH 可以改善 3D 姿態估計（第三列與第五列）。

Unfortunately, existing methods that compute 3D bodies from images perform poorly on images with self-contact; see Fig. 1.
不幸的是，現有的從圖像計算 3D 物體的方法在具有自接觸的圖像上表現不佳。 見圖 1。

Body parts that should be touching generally are not.
應該接觸的身體部位通常不是。

Recovering human meshes from images typically involves either learning a regressor from pixels to 3D pose and shape [20, 24], or fitting a 3D model to image features using an optimization method [4, 36, 47, 48].
從圖像中恢復人體網格通常涉及學習從像素到 3D 姿勢和形狀的回歸量 [20, 24]，或使用優化方法將 3D 模型擬合到圖像特徵 [4, 36, 47, 48]。

The learning approaches rely on labeled training data.
學習方法依賴於標記的訓練數據。

Unfortunately, current 2D datasets typically contain labeled keypoints or segmentation masks but do not provide any information about 3D contact.
不幸的是，當前的 2D 數據集通常包含標記的關鍵點或分割掩碼，但不提供有關 3D 接觸的任何資訊。

Similarly, existing 3D datasets typically avoid capturing scenarios with self-contact because it complicates mesh processing.
同樣，現有的 3D 數據集通常會避免使用自接觸捕獲場景，因為它會使網格處理複雜化。

What is missing is a dataset with in-the-wild images and reliable data about 3D self-contact.
缺少的是包含野外圖像和有關 3D 自我接觸的可靠數據的數據集。

To address this limitation, we introduce three new datasets that focus on self-contact at different levels of

detail.
為了解決這個限制，我們引入了三個新的數據集，專注於不同細節層次的自我接觸。

Additionally, we introduce two new optimizationbased methods that fit 3D bodies to images with contact information.
此外，我們引入了兩種新的基於優化的方法，將 3D 身體與帶有聯繫資訊的圖像相匹配。

We leverage these to estimate pseudo groundtruth 3D poses with self-contact.
我們利用這些來估計具有自接觸的偽地面實況 3D 姿勢。

To make reasoning about contact between body parts, the hands, and the face possible, we represent pose and shape with the SMPL-X [36] body model, which realistically captures the body surface details, including the hands and face.
為了使有關身體部位、手和臉之間接觸的推理成為可能，我們用 SMPL-X [36] 身體模型表示姿勢和形狀，該模型真實地捕捉了身體表面的細節，包括手和臉。

Our new datasets then let us train neural networks to regress 3D HPS from images of people with self-contact more accurately than state-ofthe- art methods.
然後，我們的新數據集讓我們訓練神經網絡，以比最先進的方法更準確地從具有自我接觸的人的圖像中回歸 3D HPS。

To begin, we first construct a 3D Contact Pose (3DCP) dataset of 3D meshes where body parts are in contact.
首先，我們首先構建身體部位接觸的 3D 網格的 3D 接觸姿勢 (3DCP) 數據集。

We do so using two methods.
我們使用兩種方法來做到這一點。

First, we use high-quality 3D scans of subjects performing self-contact poses.
首先，我們使用高質量 3D 掃描對象執行自接觸姿勢。

We extend previous mesh registration methods to cope with selfcontact and register the SMPL-X mesh to the scans.
我們擴展了先前的網格配準方法以應對自接觸並將 SMPL-X 網格配準到掃描。

To gain more variety of poses, we search the AMASS dataset [30] for poses with self-contact or “near” self-contact.
為了獲得更多種類的姿勢，我們在 AMASS 數據集 [30] 中搜索具有自我接觸或“接近”自我接觸的姿勢。

We then optimize these poses to bring nearby parts into full contact while resolving interpenetration.
然後我們優化這些姿勢，使附近的零件完全接觸，同時解決相互滲透。

This provides a dataset of valid, realistic, self-contact poses in SMPL-X format.
這提供了 SMPL-X 格式的有效、真實、自接觸姿勢的數據集。

Second, we use these poses to collect a novel dataset of images with near ground-truth 3D pose.
其次，我們使用這些姿勢來收集具有接近真實 3D 姿勢的新圖像數據集。

To do so, we show rendered 3DCP meshes to workers on Amazon Mechanical Turk (AMT).
為此，我們向 Amazon Mechanical Turk (AMT) 上的工作人員展示了渲染的 3DCP 網格。

Their task is to Mimic The Pose (MTP) as accurately as possible, including the contacts, and submit a photograph.
他們的任務是盡可能準確地模仿姿勢 (MTP)，包括聯繫人，並提交照片 。

We then use the “true” pose as a strong prior and optimize the pose in the image by extending SMPLify-X [36] to enforce contact.
然後我們使用“真實”姿勢作為強先驗，並通過擴展 SMPLify-X [36] 來優化圖像中的姿勢以加強接觸。

A key observation is that, if we know about self-contact (even approximately), this greatly reduces pose ambiguity by removing degrees of freedom.
一個關鍵的觀察是，如果我們知道自接觸（甚至是近似的），這會通過消除自由度來大大減少姿勢歧義。

Thus, knowing contact makes the estimation of 3D human pose from 2D images more accurate.
因此，知道接觸使得從 2D 圖像中估計 3D 人體姿勢更加準確。

The resulting method, SMPLify-XMC (for SMPLify-X with Mimicked Contact), produces high-quality 3D reference poses and body shapes in correspondence with the images.
由此產生的方法 SMPLify-XMC（用於 SMPLify-X with Mimicked Contact）產生與圖像對應的高質量 3D 參考姿勢和身體形狀。

Third, to gain even more image variety, we take images from three public datasets [16, 17, 29] and have them labeled with discrete body-part contacts.
第三，為了獲得更多的圖像多樣性，我們從三個公共數據集 [16、17、29] 中獲取圖像，並用離散的身體部位接觸標記它們。

This results in the Discrete Self-Contact (DSC) dataset.
這導致離散自接觸 (DSC) 數據集。

To enable this, we define a partitioning of the body into regions that can be in contact.

為了實現這一點，我們將身體劃分為可以接觸的區域。

Given labeled discrete contacts, we extend SMPLify to optimize body shape using image features and the discrete contact labels.
給定標記的離散接觸，我們擴展 SMPLify 以使用圖像特徵和離散接觸標籤優化身體形狀。

We call this method SMPLify-DC, for SMPLify with Discrete Self-Contact.
我們稱這種方法為 SMPLify-DC，用於具有離散自接觸的 SMPLify。

Given the MTP and DSC datasets, we finetune a recent about 3D contact.
鑑於 MTP 和 DSC 數據集，我們對最近的 3D 接觸進行了微調。

Similarly, existing 3D datasets typically avoid capturing scenarios with self-contact because it complicates mesh processing.
同樣，現有的 3D 數據集通常會避免使用自接觸捕獲場景，因為它會使網格處理複雜化。

What is missing is a dataset with in-the-wild images and reliable data about 3D self-contact.
缺少的是包含野外圖像和有關 3D 自我接觸的可靠數據的數據集。

To address this limitation, we introduce three new datasets that focus on self-contact at different levels of detail.
為了解決這個限制，我們引入了三個新的數據集，專注於不同細節層次的自我接觸。

Additionally, we introduce two new optimization-based methods that fit 3D bodies to images with contact information.
此外，我們引入了兩種新的基於優化的方法，將 3D 身體與帶有聯繫資訊的圖像相匹配。

We leverage these to estimate pseudo groundtruth 3D poses with self-contact.
我們利用這些來估計具有自接觸的偽地面實況 3D 姿勢。

To make reasoning about contact between body parts, the hands, and the face possible, we represent pose and shape with the SMPL-X [36] body model, which realistically captures the body surface details, including the hands and face.
為了使有關身體部位、手和臉之間接觸的推理成為可能，我們用 SMPL-X [36] 身體模型表示姿勢和形狀，該模型真實地捕捉了身體表面的細節，包括手和臉。

Our new datasets then let us train neural networks to regress 3D HPS from images of people with self-contact more accurately than state-ofthe- art methods.
然後，我們的新數據集讓我們訓練神經網絡，以比最先進的方法更準確地從具有自我接觸的人的圖像中回歸 3D HPS。

To begin, we first construct a 3D Contact Pose (3DCP) dataset of 3D meshes where body parts are in contact.
首先，我們首先構建身體部位接觸的 3D 網格的 3D 接觸姿勢 (3DCP) 數據集。

We do so using two methods.
我們使用兩種方法來做到這一點。

First, we use high-quality 3D scans of subjects performing self-contact poses.
首先，我們使用高質量 3D 掃描對象執行自接觸姿勢。

We extend previous mesh registration methods to cope with selfcontact and register the SMPL-X mesh to the scans.
我們擴展了先前的網格配準方法以應對自接觸並將 SMPL-X 網格配準到掃描。

To gain more variety of poses, we search the AMASS dataset [30] for poses with self-contact or “near” self-contact.
為了獲得更多種類的姿勢，我們在 AMASS 數據集 [30] 中搜索具有自我接觸或“接近”自我接觸的姿勢。

We then optimize these poses to bring nearby parts into full contact while resolving interpenetration.
然後我們優化這些姿勢，使附近的零件完全接觸，同時解決相互滲透。

This provides a dataset of valid, realistic, self-contact poses in SMPL-X format.
這提供了 SMPL-X 格式的有效、真實、自接觸姿勢的數據集。

Second, we use these poses to collect a novel dataset of images with near ground-truth 3D pose.
其次，我們使用這些姿勢來收集具有接近真實 3D 姿勢的新圖像數據集。

To do so, we show rendered 3DCP meshes to workers on Amazon Mechanical Turk (AMT).
為此，我們向 Amazon Mechanical Turk (AMT) 上的工作人員展示了渲染的 3DCP 網格。

Their task is to Mimic The Pose (MTP) as accurately as possible, including the contacts, and submit a photograph.
他們的任務是盡可能準確地模仿姿勢 (MTP)，包括聯繫人，並提交照片 。

We then use the “true” pose as a strong prior and optimize the pose in the image by extending SMPLify- X [36] to enforce contact.
然後，我們使用“真實”姿勢作為強先驗，並通過擴展 SMPLify-X [36] 來強制接觸來優化圖像中的姿勢。

A key observation is that, if we know about self-contact (even approximately), this greatly reduces pose ambiguity by removing degrees of freedom.

一個關鍵的觀察是，如果我們知道自接觸（甚至是近似的），這會通過消除自由度來大大減少姿勢歧義。

Thus, knowing contact makes the estimation of 3D human pose from 2D images more accurate.
因此，知道接觸使得從 2D 圖像中估計 3D 人體姿勢更加準確。

The resulting method, SMPLify-XMC (for SMPLify-X with Mimicked Contact), produces high-quality 3D reference poses and body shapes in correspondence with the images.
由此產生的方法 SMPLify-XMC（用於 SMPLify-X with Mimicked Contact）產生與圖像對應的高質量 3D 參考姿勢和身體形狀。

Third, to gain even more image variety, we take images from three public datasets [16, 17, 29] and have them labeled with discrete body-part contacts.
第三，為了獲得更多的圖像多樣性，我們從三個公共數據集 [16、17、29] 中獲取圖像，並用離散的身體部位接觸標記它們。

This results in the Discrete Self-Contact (DSC) dataset.
這導致離散自接觸 (DSC) 數據集。

To enable this, we define a partitioning of the body into regions that can be in contact.
為了實現這一點，我們將身體劃分為可以接觸的區域。

Given labeled discrete contacts, we extend SMPLify to optimize body shape using image features and the discrete contact labels.
給定標記的離散接觸，我們擴展 SMPLify 以使用圖像特徵和離散接觸標籤優化身體形狀。

We call this method SMPLify-DC, for SMPLify with Discrete Self-Contact.
我們稱這種方法為 SMPLify-DC，用於具有離散自接觸的 SMPLify。

Given the MTP and DSC datasets, we finetune a recent HPS regression network, SPIN [24].
鑑於 MTP 和 DSC 數據集，我們對最近的 HPS 回歸網絡 SPIN [24] 進行了微調。

When we have 3D reference poses, i.e. for MTP images, we use these as though they were ground truth and do not optimize them in SPIN.
當我們有 3D 參考姿勢時，即對於 MTP 圖像，我們使用這些姿勢就好像它們是真實情況一樣，並且不會在 SPIN 中對其進行優化。

When discrete contact annotations are available, i.e. for DSC images, we use SMPLify-DC to optimize the fit in the SPIN training loop.
當離散接觸註釋可用時，即對於 DSC 圖像，我們使用 SMPLify-DC 來優化 SPIN 訓練循環中的擬合。

Fine-tuning SPIN on MTP and DSC significantly improves accuracy of the regressed poses when there is contact (evaluated on 3DPW [45]).
當有接觸時，在 MTP 和 DSC 上微調 SPIN 顯著提高了回歸姿勢的準確性（在 3DPW [45] 上評估）。

Surprisingly, the results on non-self-contact poses also improve, suggesting that
令人驚訝的是，非自我接觸姿勢的結果也有所改善，這表明

(1) gathering accurate 3D poses for in-thewild images is beneficial, and
(1) 為野外圖像收集準確的 3D 姿勢是有益的，並且

(2) that self-contact can provide valuable constraints that simplify pose estimation.
(2) 自接觸可以提供簡化姿勢估計的有價值的約束。

We call our regression method TUCH (Towards Understanding Contact in Humans).
我們稱我們的回歸方法為 TUCH（Towards Understanding Contact in Humans）。

Figure 1 illustrates the effect of exploiting self-contact in 3D HPS estimation.
圖 1 說明了在 3D HPS 估計中利用自接觸的效果。

By training with self-contact, TUCH significantly improves the physical plausibility.
通過自我接觸訓練，TUCH 顯著提高了物理可信度。

In summary, the key contributions of this paper are:
總之，本文的主要貢獻是：

(1) We introduce TUCH, the first HPS regressor for self-contact poses, trained end-to-end.
(1) 我們介紹了 TUCH，第一個用於自接觸姿勢的 HPS 回歸器，經過端到端訓練。

(2) We create a novel dataset of 3D human meshes with realistic contact (3DCP).
(2) 我們創建了具有真實接觸 (3DCP) 的 3D 人體網格的新數據集。

(3)We define a “Mimic The Pose” MTP task and a new optimization method to create a novel dataset of in-the-wild images with accurate 3D reference data.
(3) 我們定義了“模仿姿勢” MTP 任務和新的優化方法，以創建具有準確 3D 參考數據的野外圖像的新數據集。

(4) We create a large dataset of images with reference poses that use discrete contact labels.
(4) 我們創建了一個包含使用離散接觸標籤的參考姿勢的大型圖像數據集。

(5) We show in experiments that taking self-contact information into account improves pose estimation in two ways (data and losses), and in turn achieves state-of-the-art results on 3D pose estimation benchmarks.

(5) 我們在實驗中表明，考慮自接觸信息可以通過兩種方式（數據和損失）改進姿勢估計，進而在 3D 姿勢估計基准上實現最先進的結果。

(6) The data and code are available for research purposes.
(6) 數據和代碼可用於研究目的。

2. Related Work 相關工作

3D pose estimation with contact. 接觸的 3D 姿態估計。

Despite rapid progress in 3D human pose estimation [19, 20, 24, 33, 36, 42, 47], and despite the role that self-contact plays in our daily lives, only a handful of previous works discuss selfcontact.
儘管 3D 人體姿勢估計取得了快速進展 [19, 20, 24, 33, 36, 42, 47]，儘管自我接觸在我們的日常生活中扮演著重要角色，但之前只有少數作品討論過自我接觸。

Information about contact can benefit 3D HPS estimation in many ways, usually by providing additional physical constraints to prevent undesirable solutions such as interpenetration between limbs.
有關接觸的信息可以在許多方面有益於 3D HPS 估計，通常是通過提供額外的物理約束來防止不受歡迎的解決方案，例如四肢之間的相互滲透。

Body contact. 身體接觸。

Lee and Chen [26] approximate the human body as a set of line segments and avoid collisions between the limbs and torso.
Lee and Chen [26] 將人體近似為一組線段，並避免四肢和軀幹之間的碰撞。

Similar ideas are adopted in [3, 10] where line segments are replaced with cylinders.
在 [3, 10] 中採用了類似的想法，其中線段被圓柱體替換。

Yin et al. [50] build a pose prior to penalize deep interpenetration detected by the Open Dynamics Engine [43].
Yin et al. [50] 在懲罰 Open Dynamics Engine [43] 檢測到的深度互穿之前構建一個姿勢。

While efficient, these stickman-like representations are far from realistic.
雖然有效，但這些類似火柴人的表示遠非現實。

Using a full 3D body mesh representation, Pavlakos et al. [36] take advantage of physical limits and resolve interpenetration of body parts by adding an interpenetration loss.
Pavlakos et al. [36] 使用完整的 3D 身體網格表示，利用物理限制並通過添加互穿損失來解決身體部位的互穿問題。

When estimating multiple people from an image, Zanfir et al. [51] use a volume occupancy exclusion loss to

prevent penetration.
從圖像中估計多人時，Zanfir et al. [51]使用體積佔用排除損失來防止滲透。

Still, other work has exploited textual and ordinal descriptions of body pose [37, 38].
儘管如此，其他工作還是利用了身體姿勢的文本和順序描述 [37, 38]。

This includes constraints like “Right hand above the hips”.
這包括諸如“右手在臀部上方”之類的約束。

These methods, however, do not consider self-contact.
然而，這些方法不考慮自接觸。

Most similar to us is the work of Fieraru et al. [8], which utilizes discrete contact annotations between people.
與我們最相似的是 Fieraru et al. [8] 的工作，它利用人與人之間的離散接觸註釋。

They introduce contact signatures between people based on coarse body parts.
他們根據粗糙的身體部位在人與人之間引入接觸特徵。

This is similar to how we collect the DSC dataset.
這類似於我們收集 DSC 數據集的方式。

Contemporaneous with our work, Fieraru et al. [9] extend this to self-contact with a 2-stage approach.
這類似於我們收集 DSC 數據集的方式。

They train a network to predict “self-contact signatures”, which are used for optimization-based 3D pose estimation.
他們訓練一個網絡來預測“自接觸簽名”，這些簽名用於基於優化的 3D 姿態估計。

In contrast, TUCH is trained end-to-end to regress body pose with contact information.
相比之下，TUCH 接受端到端的訓練，以使用接觸資訊回歸身體姿勢。

World contact. 世界聯繫方式。

Multiple methods use the 3D scene to help estimate the human pose.
多種方法使用 3D 場景來幫助估計人體姿勢。

Physical constraints can come from the ground plane [46, 51], an object [13, 21, 23], or contextual scene information [11, 49].
物理約束可以來自地平面 [46, 51]、對象 [13, 21, 23] 或上下文場景資訊 [11, 49]。

Li et al. [27] use a DNN to detect 2D contact points between objects and selected body joints.

Li et al. [27] 使用 DNN 來檢測對象和選定身體關節之間的 2D 接觸點。

Narasimhaswamy et al. [34] categorize hand contacts into self, person-person, and object contacts and aim to detect them from in-the-wild images.
Narasimhaswamy et al. [34] 將手部接觸分為自我接觸、人與人接觸和物體接觸，旨在從野外圖像中檢測它們。

Their dataset does not provide reference 3D poses or shape.
他們的數據集不提供參考 3D 姿勢或形狀。

All the above works make a similar observation: human pose estimation is not a stand-alone task;
以上所有工作都做出了類似的觀察：人體姿態估計不是一項獨立的任務；

considering additional physical contact constraints improves the results.
考慮額外的物理接觸約束可以改善結果。

We go beyond prior work by addressing self-contact and showing how training with self-contact data improves pose estimation overall.
我們通過解決自我接觸並展示使用自我接觸數據進行訓練如何改善整體姿勢估計來超越先前的工作。

3D body datasets. 3D 身體數據集。

While there are many datasets of 3D human scans, most of these have people standing in an “A” or “T” pose to explicitly minimize self-contact [40].
雖然有許多 3D 人體掃描數據集，但其中大多數都讓人們以“A”或“T”姿勢站立以明確減少自我接觸 [40]。

Even when the body is scanned in varied poses, these poses are designed to avoid self-contact [2, 6, 7, 39].
即使以各種姿勢掃描身體，這些姿勢也旨在避免自我接觸 [2, 6, 7, 39]。

For example, the FAUST dataset has a few examples of self-contact and the authors identify these as the major cause of error for scan processing methods [5].
例如，FAUST 數據集有一些自我接觸的例子，作者將這些確定為掃描處理方法錯誤的主要原因 [5]。

Recently, the AMASS [30] dataset unifies 15 different optical marker-based motion capture (mocap) datasets within a common 3D body parameterization, offering around 170k meshes with SMPL-H [41] topology.
最近，AMASS [30] 數據集在一個常見的 3D 身體參數化中統一了 15 個不同的基於光學標記的運動捕捉（mocap）數據集，提供了大約 17 萬個具有 SMPL-H [41] 拓撲的網格。

Since mocap markers are sparse and often do not cover the hands, such datasets typically do not explicitly capture self-contact.

由於 mocap 標記稀疏且通常不覆蓋手部，因此此類數據集通常不會明確捕獲自我接觸。

As illustrated in Table 1, none of these datasets explicitly addresses self-contact.
如表 1 所示，這些數據集都沒有明確解決自我接觸問題。

Pose mimicking. Our Mimic-The-Pose dataset uses the idea that people can replicate a pose that they are shown.
姿勢模仿。 我們的 Mimic-The-Pose 數據集使用人們可以複製他們展示的姿勢的想法。

Several previous works have explored this idea in different contexts.
之前的幾部作品已經在不同的背景下探索了這個想法。

Taylor et al. [44] crowd-source images of people in the same pose by imitation.
Taylor et al. [44] 通過模仿人群來源的相同姿勢的人的圖像。

While they do not know the true 3D pose, they are able to train a network to match images of people in similar poses.
雖然他們不知道真正的 3D 姿勢，但他們能夠訓練網絡來匹配具有相似姿勢的人的圖像。

Marinoiu et al. [31] motion capture subjects reenacting a 3D pose from a 2D image.
Marinoiu et al. [31] 運動捕捉對像從 2D 圖像中重現 3D 姿勢。

They found that subjects replicated 3D poses with a mean joint error of around 100mm.
他們發現受試者複製了 3D 姿勢，平均關節誤差約為 100 毫米。

This is on par with existing 3D pose regression methods, pointing to people's ability to approximately recreate viewed poses.
這與現有的 3D 姿勢回歸方法相當，表明人們能夠近似地重新建立查看過的姿勢。

Fieraru et al. [9] ask subjects to reproduce contact from an image in a lab setting.
Fieraru et al. [9] 要求受試者在實驗室環境中從圖像中再現接觸。

They manually annotate the contact, whereas our MTP task is done in people's homes and SMPLify-XMC is used to automatically optimize the pose and contact.
他們手動註釋接觸，而我們的 MTP 任務是在人們家中完成的，而 SMPLify-XMC 用於自動優化姿勢和接觸。

Name	Meshes	Meshes with self-contact
3DCP Scan (ours)	190	188
3D BodyTex [1]	400	3
SCAPE [2]	70	0
Hasler et al. [12]	520	0
FAUST [5]	100/ 400	20/ 140

Table 1. Existing 3D human mesh datasets with the number of poses and the number of contact poses identified by visual inspection. 3DCP Scan is the scan subset of 3DCP (see Section 4). FAUST (train/test) includes scans with self-contact, i.e. 20 in the training and 140 in the test set. However, in FAUST the variety is low as each subject is scanned in the same 10/20 poses, whereas in 3DCP Scan each subject does different poses.

Table 1. Existing 3D human mesh datasets with the number of poses and the number of contact poses identified by visual inspection.
表 1. 現有的 3D 人體網格數據集，包含通過目視檢查識別的姿勢數和接觸姿勢數。

3DCP Scan is the scan subset of 3DCP (see Section 4).
3DCP 掃描是 3DCP 的掃描子集（見第 4 節）。

FAUST (train/test) includes scans with self-contact, i.e. 20 in the training and 140 in the test set.
FAUST（訓練/測試）包括具有自我接觸的掃描，即訓練中的 20 次和測試集中的 140 次。

However, in FAUST the variety is low as each subject is scanned in the same 10/20 poses, whereas in 3DCP Scan each subject does different poses.
然而，在 FAUST 中，多樣性較低，因為每個對像都以相同的 10/20 姿勢進行掃描，而在 3DCP 掃描中，每個對像都有不同的姿勢。

3. Self-Contact 自我聯繫

An intuitive definition of contact between two meshes, e.g. a human and an object, is based on intersecting triangles.
兩個網格之間接觸的直觀定義，例如 一個人和一個物體，是基於相交的三角形。

Self-contact, however, must be formulated to exclude common, but not functional, triangle intersections, e.g. at the crotch or armpits.
然而，必須制定自接觸以排除常見但不是功能性的三角形交叉點，例如 在胯部或腋窩處。

Intuitively, vertices are in self-contact if they are close in Euclidean distance (near zero) but distant in geodesic

distance, i.e. far away on the body surface.
直觀地說，如果頂點在歐幾里德距離上很近（接近零）但在測地距離上很遠，即遠離體表，則它們處於自接觸狀態。

Definition 3.1.

Given a mesh M with vertices MV , we define two vertices vi and vj ∈ MV to be in self-contact, if (i) ||vi − vj|| < teucl , and (ii) geo(vi, vj) > tgeo, where teucl and tgeo are predefined thresholds and geo(vi, vj) denotes the geodesic distance between vi and vj .

定義 3.1。 給定一個頂點為 MV 的網格 M，我們定義兩個頂點 vi 和 vj 2 MV 自接觸，如果 (i) kvi vjk < teucl ，以及 (ii) geo(vi; vj) > tgeo，其中 teucl 和 tgeo 是預定義的閾值，geo(vi; vj) 表示 vi 和 vj 之間的測地線距離。

We use shape-independent geodesic distances precomputed on the neutral, mean-shaped SMPL and SMPL-X models.
我們使用在中性、平均形狀 SMPL 和 SMPL-X 模型上預先計算的與形狀無關的測地距離。

Following this definition, we denote the set of vertex pairs in self-contact as MC := {(vi, vj)|vi, vj ∈ MV and vi, vj satisfy Definition 3.1}.
按照這個定義，我們將自接觸中的頂點對集表示為 MC := {(vi, vj)|vi, vj ∈ MV 和 vi, vj 滿足定義 3.1}。

M is a self-contact mesh when |MC|> 0.
當|MC|> 0 時，M 是自接觸網格。

We further define an operator U(·) that returns a set of unique vertices in MC, and an operator fg(·) that takes vi as input and returns the Euclidean distance to the nearest vj that is far enough in the geodesic sense.
我們進一步定義了一個運算符 U(·)，它返回 MC 中的一組唯一頂點，以及一個運算符 fg(·)，它將 vi 作為輸入並返回到最近的在測地線意義上足夠遠的 vj 的歐幾里德距離。

Formally, U(MC) = {v0, v1, v2, … , vn}, where ∀vi ∈ U(MC) , ∃vj 2 U(MC), such that (vi, vj) ∈ MC.
形式上，U(MC) = {v0, v1, v2, … , vn}，其中∀vi ∈ U(MC) , ∃vj 2 U(MC)，使得(vi, vj) ∈ MC。

$$\text{Formally, } \mathcal{U}(M_C) = \{v_0, v_1, v_2, \ldots, v_n\}, \text{ where } \forall v_i \in \mathcal{U}(M_C) \ , \exists v_j \in \mathcal{U}(M_C), \text{ such that } (v_i, v_j) \in M_C.$$

fg(vi) := minvj ∈MG(vi) ||vi - vj||, where MG(vi) := {vj |geo(vi, vj) > tgeo}.

$$f_g(v_i) := \min_{v_j \in M_G(v_i)} \|v_i - v_j\|, \text{ where } M_G(v_i) := \{v_j | geo(v_i, v_j) > t_{geo}\}.$$

We further cluster self-contact meshes into distinct types.

我們進一步將自接觸網格劃分為不同的類型。

To that end, we define self-contact signatures S ∈ {0, 1}KxK; see [9] for a similar definition.
為此，我們定義了自接觸簽名 S ∈ {0, 1}KxK; 類似的定義見[9]。

We first segment the vertices of a mesh into K regions Rk, where Rk ∩ Rl = ∅; for k ≠ l and ∪Kk=1 Rk = MV .
我們首先將網格的頂點分割為 K 個區域 Rk，其中 Rk \ Rl = ; 對於 k 6= l 和 SKk=1 Rk = MV 。

We use fine signatures to cluster self-contact meshes from AMASS (see Sup. Mat.) and rough signatures (see Fig. 18) for human annotation.
我們使用精細簽名來聚類來自 AMASS（參見 Sup. Mat.）的自接觸網格和用於人類註釋的粗略簽名（參見圖 18）。

Definition 3.2.
定義 3.2。

Two regions Rk and Rl are in contact if ∃(vi, vj) ∈ MC, such that vi ∈ Rk and vj ∈ Rl holds.
如果∃(vi, vj) ∈ MC，則兩個區域 Rk 和 Rl 接觸，使得 vi ∈ Rk 和 vj ∈ Rl 成立。

If Rk and Rl are in contact, Skl = Slk = 1.
如果 Rk 和 Rl 接觸，則 Skl = Slk = 1。

MS denotes the contact signature for mesh M.
MS 表示網格 M 的接觸簽名。

To detect self-contact, we need to be able to quickly compute the distance between two points on the body surface.
為了檢測自接觸，我們需要能夠快速計算身體表面兩點之間的距離。

Vertex-to-vertex distance is a poor approximation of this due to the varying density of vertices across the body.
由於整個身體的頂點密度不同，頂點到頂點距離是一個很差的近似值。

Consequently, we introduce HD SMPL-X and HD SMPL to efficiently approximate surface-to-surface distance.
因此，我們引入了 HD SMPL-X 和 HD SMPL 來有效地近似表面到表面的距離。

For this, we uniformly, and densely, sample points, MP ∈ RP3 with P = 20,000 on the body.
為此，我們在身體上均勻且密集地採樣點 MP ∈ RP3，P = 20,000。

A sparse linear regressor P regresses MP from the mesh vertices MV , MP = PMV .
稀疏線性回歸器 P 從網格頂點 MV ，MP = PMV 回歸 MP。

The geodesic distance geoHD(p1, p2) between p1 ∈ MP and p2 ∈ MP is approximated via geo(m, n), where m = arg minv∈MV ||v - p1|| and n = arg minv∈MV ||v - p2||.
p1 ∈ MP 和 p2 ∈ MP 之間的測地距離 geoHD(p1, p2) 通過 geo(m, n) 近似，其中 m = arg minv∈MV ||v - p1|| 並且 n = arg minv∈MV ||v - p2||。

In practice, we use mesh surface points only when contact is present by following a three-step procedure as illustrated in Fig. 2.
在實踐中，我們僅在存在接觸時通過遵循圖 2 所示的三步程序使用網格表麵點。

First, we use Definition 3.1 to detect vertices in contact, MC.
首先，我們使用定義 3.1 來檢測接觸的頂點 MC。

Then we select all points in MP lying on faces that contain vertices in MC, denoted as MD.
然後我們選擇 MP 中所有位於包含 MC 頂點的面上的點，表示為 MD。

Last, for pi ∈ MD we find the closest mesh surface point minpj∈MD ||pi – pj||, such that geoHD(pi, pj) > tgeo.
最後，對於 pi ∈ MD，我們找到最近的網格表面點 (mesh surface point) minpj∈MD ||pi – pj||，使得 geoHD(pi, pj) > tgeo。

With HD(X) : X ⊂ MV→MD⊂MP we denote the function that maps from a set of mesh vertices to a set of mesh surface points.
HD(X) : X ⊂ MV→MD⊂MP 我們表示從一組網格頂點映射到一組網格表麵點的函數。

As the number of points, P, increases, the point-to-point distance approximates the surface-to-surface distance.
隨著點數 P 的增加，點到點的距離接近表面到表面的距離。

Figure 2. Visualization of the function HD(X), that maps from mesh vertices to mesh surface points.
圖 2. 函數 HD(X) 的可視化，它從網格頂點映射到網格表面點 (mesh surface point)。

Figure 2. Visualization of the function $HD(X)$, that maps from mesh vertices to mesh surface points. First, a SMPL-X mesh with vertices in contact highlighted. Second, in yellow, all faces containing a vertex in contact are selected. Then, all points lying on a face containing a vertex in contact are selected from M_P, denoted as M_D. M_P is a fixed set of mesh surface points that are regressed from mesh vertices. Note that in image one and two the finger vertices are denser than the arm and chest vertices, in contrast to the more uniform density in images three and four.

First, a SMPL-X mesh with vertices in contact highlighted.
首先，突出顯示頂點接觸的 SMPL-X 網格。

Second, in yellow, all faces containing a vertex in contact are selected.
其次，在黃色中，所有包含接觸頂點的面都被選中。

Then, all points lying on a face containing a vertex in contact are selected from MP , denoted as MD.
然後，從 MP 中選擇包含接觸頂點的面上的所有點，記為 MD。

MP is a fixed set of mesh surface points that are regressed from mesh vertices.
MP 是一組固定的網格表面點 (mesh surface point)，從網格頂點回歸。

Note that in image one and two the finger vertices are denser than the arm and chest vertices, in contrast to the more uniform density in images three and four.
請注意，在圖像 1 和圖像 2 中，手指頂點比手臂和胸部頂點更密集，與圖像 3 和 4 中更均勻的密度形成對比。

Figure 3. Self-contact optimization. Column 1 and 2: a pose selected from AMASS with near self-contact (between the fingertips and the foot) and interpenetration (thumb and foot). Column 3 and 4: after self-contact optimization, all fingers are in contact with the foot and interpenetration is reduced.

Figure 3. Self-contact optimization. Column 1 and 2: a pose selected from AMASS with near self-contact (between the fingertips and the foot) and interpenetration (thumb and foot).
圖 3. 自接觸優化。 第 1 列和第 2 列：從 AMASS 中選擇的具有近乎自接觸（指尖和足部之間）和相互滲透（拇指和足部）的姿勢。

Column 3 and 4: after self-contact optimization, all fingers are in contact with the foot and interpenetration is reduced.
第三列和第四列：自接觸優化後，所有手指都與足部接觸，減少了相互滲透。

4. Self-Contact Datasets 自接觸數據集

Our goal is to create datasets of in-the-wild images paired with 3D human meshes as pseudo ground truth.
我們的目標是創建與 3D 人體網格配對的野外圖像數據集作為偽地面實況。

Unlike traditional pipelines that collect images first and then annotate them with pose and shape parameters [18, 45], we take the opposite approach.
與首先收集圖像然後用姿勢和形狀參數對其進行註釋的傳統管道不同 [18, 45]，我們採用相反的方法。

We first curate meshes with self-contact and then pair them with images through a novel pose mimicking and fitting procedure.
我們首先通過自我接觸來管理網格，然後通過一種新穎的姿勢模仿和擬合程序將它們與圖像配對。

We use SMPL-X to create the 3DCP and MTP dataset to better fit contacts between hands and bodies.
我們使用 SMPL-X 來創建 3DCP 和 MTP 數據集，以更好地適應手和身體之間的接觸。

However, to fine-tune SPIN [24], we convert MTP data to SMPL topology, and use SMPLify-DC when optimizing with discrete contact.
然而，為了微調 SPIN [24]，我們將 MTP 數據轉換為 SMPL 拓撲，並在使用離散接觸進行優化時使用 SMPLify-DC。

4.1. 3D Contact Pose (3DCP) Meshes - 3D 接觸姿勢 (3DCP) 網格

We create 3D human meshes with self-contact in two ways: with 3D scans and with motion capture data.
我們通過兩種方式創建具有自我接觸的 3D 人體網格：使用 3D 掃描和使用運動捕捉數據。

3DCP Scan.

We scan 6 subjects (3 males, 3 females) in self-contact poses.
我們以自我接觸姿勢掃描 6 名受試者（3 名男性，3 名女性）。

We then register the SMPL-X mesh topology to the raw scans.
然後我們將 SMPL-X 網格拓撲註冊到原始掃描。

These registrations are obtained using Co-Registration [14], which iteratively deforms the SMPL-X template mesh V to minimize the point-to-plane distance between the scan points S ∈ R^(Nx3), where N is the number of scan points and the template points V ∈ R^(10375x3).
這些配準是使用 Co-Registration [14] 獲得的，它反複變形 SMPL-X 模板網格 V 以最小化掃描點 S ∈ R^(Nx3) 之間的點到平面距離，其中 N 是掃描次數 點和模板點 V ∈ R^(10375x3)。

However, registering poses with self-contact is challenging.
然而，通過自我接觸來註冊姿勢是具有挑戰性的。

When body parts are in close proximity, the standard process can result in interpenetration.
當身體部位非常接近時，標準流程可能會導致相互滲透。

To address this, we add a self-contact-preserving energy term to the objective function.
為了解決這個問題，我們在目標函數中添加了一個自接觸保持能量項。

If two vertices vi and vj are in contact according to Definition 3.1, we minimize the point-to-plane distance between triangles including vi and the triangular planes including vj .
如果根據定義 3.1 兩個頂點 vi 和 vj 接觸，我們最小化包括 vi 在內的三角形和包括 vj 在內的三角形平面之間的點到平面距離。

This term ensures that body parts that are in contact remain in contact; see Sup. Mat. for details.
該術語確保接觸的身體部位保持接觸； see Sup. Mat. for details.

3DCP Mocap.

While mocap datasets are usually not explicitly designed to capture self-contact, it does occur during motion capture.

雖然動作捕捉數據集通常沒有明確設計用於捕捉自我接觸，但它確實發生在動作捕捉期間。

We therefore search the AMASS dataset for poses that satisfy our self-contact definition.
因此，我們在 AMASS 數據集中搜索滿足我們自接觸定義的姿勢。

We find that some of the selected meshes from AMASS contain small amounts of self-penetration or near contact.
我們發現 AMASS 的一些選定網格包含少量自滲透或接近接觸。

Thus, we perform self-contact optimization to fix this while encouraging contact, as shown in Fig. 3; see Sup. Mat. for details.
因此，我們執行自接觸優化來解決這個問題，同時鼓勵接觸，如圖 3 所示； see Sup. Mat. for details.

4.2. Mimic-The-Pose(MTP) Data - 模仿姿勢（MTP）數據

To collect in-the-wild images with near ground-truth 3D human meshes, we propose a novel two-step process (see Fig. 4).
為了使用近乎真實的 3D 人體網格收集野外圖像，我們提出了一種新穎的兩步過程（見圖 4）。

First, using meshes from 3DCP as examples, workers on AMT are asked to mimic the pose as accurately as possible while someone takes their photo showing the full body (the mimicked pose).
首先，使用來自 3DCP 的網格作為示例，要求 AMT 上的工作人員盡可能準確地模仿姿勢，同時有人拍攝顯示全身的照片（模仿姿勢）。

Mimicking poses may be challenging for people when only a single image of the pose is presented [31].
當僅呈現姿勢的單個圖像時，模仿姿勢對人們來說可能具有挑戰性[31]。

Thus, we render each 3DCP mesh from three different views with the contact regions highlighted (the presented pose).
因此，我們從三個不同的視圖渲染每個 3DCP 網格，突出顯示接觸區域（呈現的姿勢）。

We allot 3 hours time for ten poses.
我們為十個姿勢分配了 3 小時的時間。

Participants also provide their height and weight.
參與者還提供了他們的身高和體重。

All participants gave informed consent for the capture and the use of their imagery.
所有參與者都知情同意拍攝和使用他們的圖像。

Please see Sup. Mat. for details.

所有參與者都知情同意拍攝和使用他們的圖像。

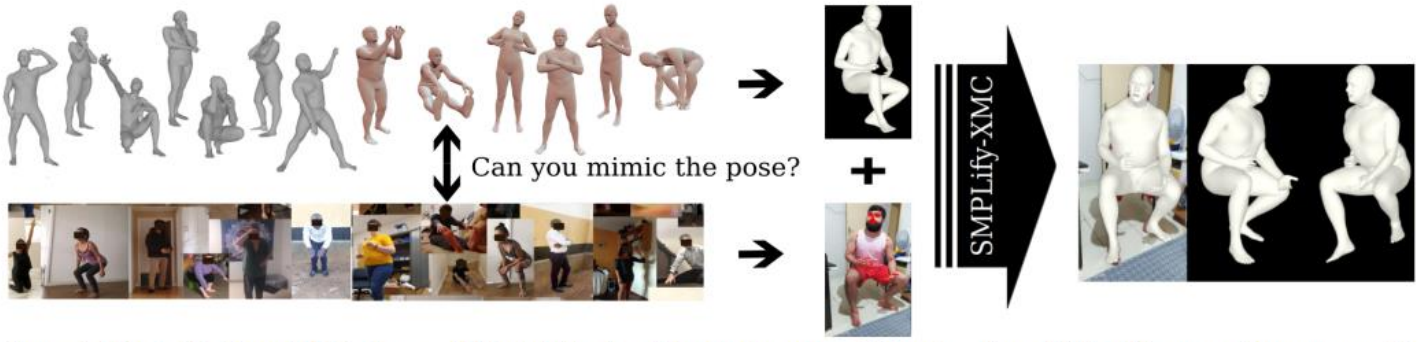

Figure 4. Mimic-The-Pose (MTP) dataset. MTP is built via: (1) collecting many 3D meshes that exhibit self-contact. In grey, new 3D scans in self-contact poses, in brown self-contact poses optimized from AMASS mocap data. (2) collecting images in the wild, by asking workers on AMT to mimic poses and contacts. (3) the presented meshes are refined via SMPLify-XMC to match the image features.

Figure 4. Mimic-The-Pose (MTP) dataset. MTP is built via: (1) collecting many 3D meshes that exhibit self-contact.

圖 4. Mimic-The-Pose (MTP) 數據集。 MTP 是通過以下方式構建的：(1) 收集許多表現出自接觸的 3D 網格。

In grey, new 3D scans in self-contact poses, in brown self-contact poses optimized from AMASS mocap data.

在灰色中，新的 3D 掃描自接觸姿勢，棕色自接觸姿勢從 AMASS 運動捕捉數據優化。

(2) collecting images in the wild, by asking workers on AMT to mimic poses and contacts.

(2) 在野外收集圖像，要求 AMT 上的工作人員模仿姿勢和接觸。

(3) the presented meshes are refined via SMPLify-XMC to match the image features.

(3) 呈現的網格通過 SMPLify-XMC 細化以匹配圖像特徵。

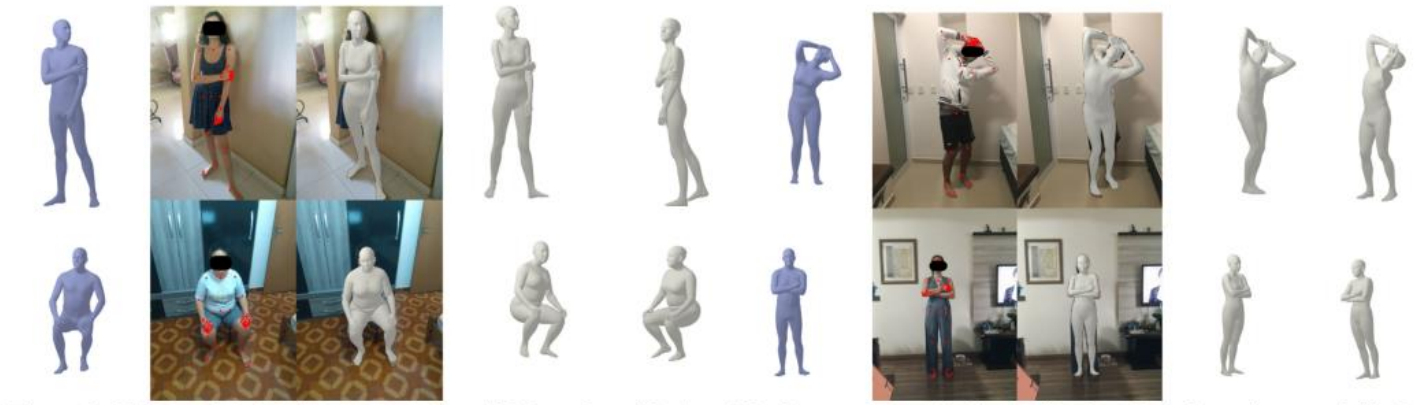

Figure 5. MTP results. Meshes presented to AMT workers (blue) and the images they submitted with OpenPose keypoints overlaid. In grey, the pseudo ground-truth meshes computed by SMPLify-XMC.

Figure 5. MTP results.

Meshes presented to AMT workers (blue) and the images they submitted with OpenPose keypoints overlaid.

呈現給 AMT 工作人員（藍色）的網格以及他們提交的帶有 OpenPose 關鍵點的圖像。

In grey, the pseudo ground-truth meshes computed by SMPLify-XMC.
灰色部分是由 SMPLify-XMC 計算的偽地面真實網格。

SMPLify-XMC.

The second step applies a novel optimization method to estimate the pose in the image, given a strong prior from the presented pose.
第二步應用一種新的優化方法來估計圖像中的姿勢，給定來自所呈現姿勢的強先驗。

The presented pose θ~, shape β~ , and gender is not mimicked perfectly.
所呈現的姿勢 θ~、形狀 β~ 和性別沒有被完美模仿。

To obtain pseudo ground-truth pose and shape, we adapt SMPLify- X [36], a multi-stage optimization method, that fits SMPLX pose θ, shape β, and expression ψ to image features starting from the mean pose and shape.
為了獲得偽地面真實姿勢和形狀，我們採用 SMPLify-X [36]，一種多階段優化方法，從平均姿勢和形狀開始，將 SMPLX 姿勢 θ、形狀 β 和表達式 ϕ 擬合到圖像特徵。

We make use of the presented pose θ~ in three ways:
我們以三種方式利用所呈現的姿勢 θ~：

first, to initialize the optimization and solve for global orientation and camera;
首先，初始化優化並求解全局方向和相機；

second, it serves as a pose prior; and third its contact is used to keep relevant body parts close to each other.
其次，它作為先驗姿勢； 第三，它的接觸用於保持相關的身體部位彼此靠近。

We refer to this new optimization method as SMPLify-XMC.
我們將這種新的優化方法稱為 SMPLify-XMC。

In the first stage, we optimize body shape β, camera Π (rotation, translations, and focal length), and body global orientation θg, while the pose θ is initialized as θ~ and stays constant; see Sup. Mat. for a description of the first stage.
在第一階段，我們優化了身體形狀 β、相機Π（旋轉、平移和焦距）和身體全局方向 θg，而姿勢 θ 被初始化為 θ~ 並保持不變； 見 Sup. Mat. 第一階段的描述。

In the second and third stage, we jointly optimize θ, β, and to minimize L(θ, β, Π).
在第二和第三階段，我們聯合優化 θ, β ，並最小化 L(θ, β, Π)。

$$\mathcal{L}(\theta, \beta, \Pi) = E_J + \lambda_{m_h} E_{m_h} + \lambda_{\tilde{\theta}} \mathcal{L}_{\tilde{\theta}} + \lambda_M \mathcal{L}_M + \lambda_{\tilde{C}} \mathcal{L}_{\tilde{C}} + \lambda_S \mathcal{L}_S. \quad (1)$$

EJ denotes the same re-projection loss as specified in [36]1.
EJ 表示與 [36]1 中指定的相同的重新投影損失。

We use the standard SMPLify-X priors for left and right hand Emh.
我們對左右手 Emh 使用標準的 SMPLify-X 先驗。

While the pose prior in [36] penalizes deviation from the mean pose, here, L θ ~ is an L2-Loss that penalizes deviation from the presented pose.
雖然 [36] 中的先驗姿勢會懲罰與平均姿勢的偏差，但在這裡，L θ ~ 是懲罰與所呈現姿勢的偏差的 L2-Loss。

The measurements loss LM takes ground-truth height and weight into account; see Sup. Mat. for details.
測量損失 LM 考慮了真實高度和重量； see Sup. Mat. for details

The term L~C acts on ~MC, the vertices in self-contact on the presented mesh.
術語 L~C 作用於~MC，即所呈現網格上自接觸的頂點。

To ensure the desired self-contact, one could seek to minimize the distances between vertices in contact, e.g. ||vi – vj||, (vi, vj) ∈ ~MC.
為了確保所需的自接觸，人們可以尋求最小化接觸頂點之間的距離，例如 ||vi – vj||, (vi, vj) ∈ ~MC。

However, with this approach, we observe slight mesh distortions, when presented and mimicked contact are different.
然而，使用這種方法，我們觀察到輕微的網格扭曲，當呈現和模擬的接觸不同時。

Instead, we use a term that encourages every vertex in ~MC to be in contact.
相反，我們使用一個術語來鼓勵~MC 中的每個頂點都接觸。

More formally,
更正式地說，

$$\mathcal{L}_{\tilde{C}} = \frac{1}{|\mathcal{U}(\tilde{M}_C)|} \sum_{v_i \in \mathcal{U}(\tilde{M}_C)} \tanh(f_g(v_i)). \quad (2)$$

####
1

We denote loss terms defined in prior work as E while ours as L.
我們將先前工作中定義的損失項表示為 E，而我們的為 L。

The third stage actives LS for fine-grained self-contact optimization, which resolves interpenetration while encouraging contact.
第三階段激活 LS 進行細粒度的自接觸優化，在鼓勵接觸的同時解決相互滲透。

The objective is LS = λ C LC + λ P LP + λ ALA.
目標是 LS = λ C LC + λ P LP + λ ALA。

Vertices in contact are pulled together via a contact term LC;
接觸的頂點通過接觸項 LC 被拉到一起；

vertices inside the mesh are pushed to the surface via a pushing term LP, and LA aligns the surface normals of two vertices in contact.
網格內的頂點通過推送項 LP 被推送到表面，並且 LA 對齊兩個接觸頂點的表面法線。

To compute these terms, we must first find which vertices are inside, MI ⊂ MV , or in contact, MC ⊂ MV .
為了計算這些項，我們必須首先找到哪些頂點在裡面， MI ⊂ MV ，或接觸， MC ⊂ MV 。

MC is computed following Definition 3.1 with tgeo = 30cm and teucl = 2cm.
MC 按照定義 3.1 計算，tgeo = 30cm and teucl = 2cm。

The set of inside vertices MI is detected by generalized winding numbers [15].
內部頂點集 MI 由廣義繞組數檢測[15]。

SMPL-X is not a closed mesh and thus complicating the test for penetration.
SMPL-X 不是封閉的網格，因此使滲透測試複雜化。

Consequently, we close it by adding a vertex at the back of the mouth.
因此，我們通過在嘴巴後部添加一個頂點來關閉它。

In addition, neighboring parts of SMPL and SMPL-X often intersect, e.g. torso and upper arms.
因此，我們通過在嘴巴後部添加一個頂點來關閉它。

We identify such common self-intersections and filter them out fromMI .
我們識別出這些常見的自相交並將它們從 MI 中過濾掉。

See Sup. Mat. for details.

To capture fine-grained contact, we map the union of inside and contact vertices onto the HD SMPL-X surface,

i.e. MD = HD(MI∪MC), which is further segmented into an inside MDI and outside MDCI subsets by testing for intersections.
為了捕捉細粒度接觸，我們將內部頂點和接觸頂點的並集映射到 HD SMPL-X 表面上，即 MD = HD(MI∪MC)，通過測試相交將其進一步分割為內部 MDI 和外部 MDCI 子集 .

The self-contact objectives are defined as
自接觸目標定義為

$$\mathcal{L}_C = \sum_{p_i \in M_{D_I^0}} \alpha_1 \tanh(\frac{f_g(p_i)}{\alpha_2})^2,$$
$$\mathcal{L}_P = \sum_{p_i \in M_{D_I}} \beta_1 \tanh(\frac{f_g(p_i)}{\beta_2})^2,$$
$$\mathcal{L}_A = \sum_{(p_i,p_j) \in M_{D_C}} 1 + \langle N(p_i), N(p_j) \rangle.$$

fg denotes the function that finds the closest point pj ∈ MD.
fg 表示找到最近點 pj ∈ MD 的函數。

MDC is the subset of vertices in contact in MD.
MDC 是 MD 中接觸的頂點的子集。

We use $\alpha 1 = \alpha 2 = 0.005$, $\beta 1 = 1.0$, and $\beta 2 = 0.04$ and visualize the contact and pushing functions in the Sup. Mat.
我們使用 $\alpha 1 = \alpha 2 = 0.005$、$\beta 1 = 1.0$ 和 $\beta 2 = 0.04$ 並可視化 Sup. Mat. 中的接觸和推動函數。

Fig. 5 shows examples of our pseudo ground-truth meshes.
圖 5 顯示了我們的偽地面實況網格的示例。

4.3. Discrete SelfContact 離散自接觸

(DSC) Data Images in the wild collected for human pose estimation normally come with 2D keypoint annotations, body segmentation, or bounding boxes.
(DSC) 為人體姿態估計而收集的野外數據圖像通常帶有 2D 關鍵點註釋、身體分割或邊界框。

Such annotations lack 3D information.
此類註釋缺少 3D 信息。

Discrete self-contact annotation, however, provides useful 3D information about pose.
然而，離散自接觸註釋提供了關於姿勢的有用 3D 信息。

We use K = 24 regions and label their pairwise contact for three publicly available datasets, namely Leeds

Sports Pose (LSP), Leeds Sports Pose Extended (LSPet), and DeepFashion (DF).
我們使用 K = 24 個區域並為三個公開可用的數據集標記它們的成對接觸，即 Leeds Sports Pose (LSP)、Leeds Sports Pose Extended (LSPet) 和 DeepFashion (DF)。

An example annotation is visualized in Fig. 18.
圖 18 顯示了一個示例註釋。

Of course, such labels are noisy because it can be difficult to accurately determine contact from an image.
當然，這樣的標籤是有噪聲的，因為很難從圖像中準確地確定接觸。

See Sup. Mat. for details.

4.4. Summary of the Collected Data 收集的數據摘要

Our 3DCP human mesh dataset consists of 190 meshes containing self-contact from 6 subjects, 159 SMPL-X bodies fit to commercial scans from AGORA [35], and 1304 self-contact optimized meshes from mocap data.
我們的 3DCP 人體網格數據集由 190 個網格組成，其中包含來自 6 個主體的自接觸、159 個 SMPL-X 身體適合來自 AGORA [35] 的商業掃描，以及 1304 個來自運動捕捉數據的自接觸優化網格。

From these 1653 poses, we collect 3731 mimicked pose images from 148 unique subjects (52 female; 96 male) for MTP and fit pseudo ground-truth SMPL-X parameters.
從這 1653 個姿勢中，我們從 148 個獨特的受試者（52 個女性；96 個男性）中收集了 3731 個模擬姿勢圖像，用於 MTP 並擬合偽地面實況 SMPL-X 參數。

MTP is diverse in body shapes and ethnicities.
MTP 的體型和種族各不相同。

Our DSC dataset provides annotations for 30K images.
我們的 DSC 數據集為 30K 圖像提供註釋。

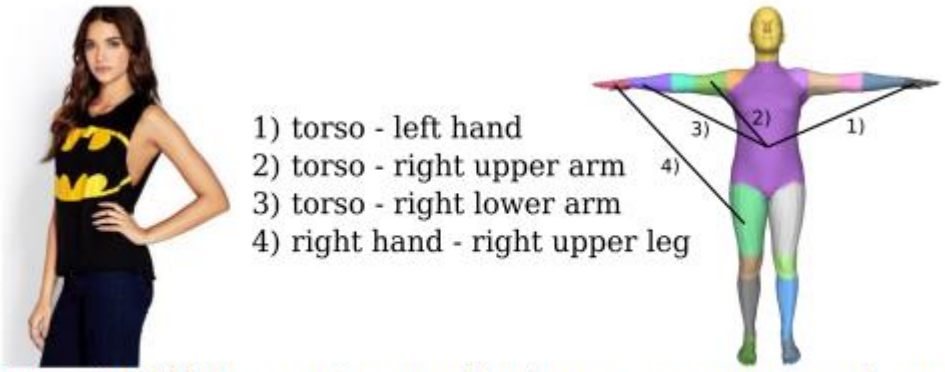

Figure 6. DSC dataset. Image with discrete contact annotation on the left. Right: DSC signature with $K = 24$ regions.

Figure 6. DSC dataset. Image with discrete contact annotation on the left. Right: DSC signature with K = 24 regions.
圖 6. DSC 數據集。 左側帶有離散接觸註釋的圖像。 右圖：K = 24 個區域的 DSC 特徵。

5. TUCH

Finally, we train a regression network that has the same design as SPIN [24].
最後，我們訓練了一個與 SPIN [24] 具有相同設計的回歸網絡。

At each training iteration, the current regressor estimates the pose, shape, and camera parameters of the SMPL model for an input image.
在每次訓練迭代中，當前回歸器都會為輸入圖像估計 SMPL 模型的姿態、形狀和相機參數。

Using groundtruth 2D keypoints, an optimizer refines the estimated pose and shape, which are used, in turn, to supervise the regressor.
使用 groundtruth 2D 關鍵點，優化器優化估計的姿勢和形狀，進而用於監督回歸器。

We follow this regression-optimization scheme for DSC data, where we have no 3D ground truth.
我們對 DSC 數據遵循此回歸優化方案，其中我們沒有 3D 地面實況。

To this end, we adapt the in-the-loop SMPLify routine to account for discrete self-contact labels, which we term SMPLify-DC.
為此，我們採用在環 SMPLify 例程來解釋離散的自接觸標籤，我們將其稱為 SMPLify-DC。

For MTP images, we use the pseudo ground truth from SMPLify-XMC as direct supervision with no optimization involved.
對於 MTP 圖像，我們使用來自 SMPLify-XMC 的偽地面實況作為直接監督，不涉及優化。

We explain the losses of each routine below.
我們在下面解釋每個例程的損失。

Regressor. 回歸器。

Similar to SPIN, the regressor of TUCH predicts pose, shape, and camera, with the loss function:
與 SPIN 類似，TOUCH 的回歸預測姿勢、形狀和相機，損失函數為：

$$L_R = E_J + \lambda_\theta E_\theta + \lambda_\beta E_\beta + \lambda_C \mathcal{L}_C + \lambda_P \mathcal{L}_P. \quad (3)$$

EJ denotes the joint re-projection loss.

EJ 表示聯合重投影損失。

LP and LC are self-contact loss terms used in LS in SMPLify-XMC, where LP penalizes mesh intersections and LC encourages contact.
LP 和 LC 是 SMPLify-XMC 中 LS 中使用的自接觸損耗項，其中 LP 懲罰網格交叉點，而 LC 鼓勵接觸。

Further, Eθ and Eβ are L2-Losses that penalize deviation from the pseudo ground-truth pose and shape.
此外，Eθ 和 Eβ 是 L2 損失，懲罰與偽地面真實姿勢和形狀的偏差。

Optimizer.
優化器。

We develop SMPLify-DC to fit pose opt, shape opt, and camera opt to DSC data, taking groundtruth keypoints and contact as constraints.
我們開發了 SMPLify-DC 來擬合姿勢選擇、形狀選擇和相機選擇到 DSC 數據，將真實關鍵點和接觸作為約束。

Typically, in human mesh optimization methods the camera is fit first, then the model parameters follow.
通常，在人體網格優化方法中，首先擬合相機，然後是模型參數。

However, we find that this can distort body shape when encouraging contact.
然而，我們發現這在鼓勵接觸時會扭曲體型。

Therefore, we optimize shape and camera translation first, using the same camera fitting loss as in [24].
因此，我們首先使用與 [24] 中相同的相機擬合損失來優化形狀和相機平移。

After that, body pose and global orientation are optimized under the objective
之後，在目標下優化身體姿勢和全局方向

$$L_O(\theta) = E_J + \lambda_\theta E_\theta + \lambda_C \mathcal{L}_C + \lambda_P \mathcal{L}_P + \lambda_D \mathcal{L}_\mathcal{D}. \quad (4)$$

The discrete contact loss, LD, penalizes the minimum distance between regions in contact.
離散接觸損耗 LD 會影響接觸區域之間的最小距離。

Formally, given a contact signature S where Sij = Sji = 1 if two regions Ri and Rj are annotated to be in contact, we define LD.
形式上，如果兩個區域 Ri 和 Rj 被註釋為接觸，則給定接觸簽名 S，其中 Sij = Sji = 1，我們定義 LD。

$$\mathcal{L}_{\mathcal{D}} = \sum_{i=1}^{K} \sum_{j=i+1}^{K} \mathbf{S}_{ij} \min_{v \in R_i, u \in R_j} ||v - u||^2.$$

Given the optimized pose θ opt, shape β opt, and camera Π opt, we compute the re-projection error and the minimum distance between the regions in contact.
給定優化的姿勢選擇、形狀選擇和相機選擇，我們計算重投影誤差和接觸區域之間的最小距離。

When the re-projection error improves, and more regions with contact annotations are closer than before, we keep the optimized pose as the current best fit.
當重新投影誤差改善，並且更多具有接觸註釋的區域比以前更接近時，我們將優化的姿勢保持為當前的最佳擬合。

When no ground truth is available, the current best fits are used to train the regressor.
當沒有可用的基本事實時，當前的最佳擬合用於訓練回歸器。

We make three observations:
我們提出三點意見：

(1) The optimizer is often able to fix incorrect poses estimated by the regressor because it considers the ground-truth keypoints and contact (see Fig. 7).
(1) 優化器通常能夠修復回歸器估計的不正確姿勢，因為它考慮了真實關鍵點和接觸（見圖 7）。

(2) Discrete contact labels bring overall improvement by helping resolve depth ambiguity (see Fig. 8).
(2) 離散接觸標籤通過幫助解決深度歧義帶來整體改進（見圖 8）。

(3) Since we have mixed data in each mini-batch, the direct supervision of MTP data improves the regressor, which benefits SMPLify-DC by providing better initial estimates.
(3) 由於我們在每個小批量中都有混合數據，MTP 數據的直接監督改進了回歸器，這通過提供更好的初始估計使 SMPLify-DC 受益。

Implementation details. 實施細則。

We initialize our regression network with SPIN weights [24].
我們用 SPIN 權重初始化我們的回歸網絡 [24]。

For SMPLify-DC, we run 10 iterations per stage and do not use the HD operator to speed up the optimization process.
對於 SMPLify-DC，我們每個階段運行 10 次迭代，並且不使用 HD 算子來加速優化過程。

For the 2D re-projection loss, we use ground-truth keypoints when available and, for MTP and DF images,

OpenPose detections weighted by confidence.
對於 2D 重投影損失，我們使用可用的地面實況關鍵點，對於 MTP 和 DF 圖像，使用置信度加權的 OpenPose 檢測。

From DSC data we only use images where the full body is visible and ignore annotated region pairs that are connected in the DSC segmentation (see Sup. Mat.).
從 DSC 數據中，我們只使用全身可見的圖像，而忽略在 DSC 分割中連接的帶註釋的區域對（參見 Sup. Mat.）。

Figure 7. Initial wrong contact (left) from the regressor is fixed by SMPLify-DC after 5 (middle) and 10 (right) iterations.

Figure 7. Initial wrong contact (left) from the regressor is fixed by SMPLify-DC after 5 (middle) and 10 (right) iterations.
圖 7. 來自回歸器的初始錯誤接觸（左）在 5（中）和 10（右）次迭代後由 SMPLify-DC 修復。

6. Evaluation . 評估

We evaluate TUCH on the following three datasets: 3DPW [45], MPI-INF-3DHP [32], and 3DCP Scan.
我們在以下三個數據集上評估 TUCH：**3DPW [45]、MPI-INF-3DHP [32]** 和 **3DCP Scan**。

This latter dataset consists of RGB images taken during the 3DCP Scan scanning process.
後一個數據集由在 3DCP 掃描掃描過程中拍攝的 RGB 圖像組成。

While TUCH has never seen these images or subjects, the contact poses were mimicked in creation of MTP, which is used in training.
雖然 TUCH 從未見過這些圖像或主題，但在創建用於訓練的 MTP 時模仿了接觸姿勢。

We use standard evaluation metrics for 3D pose, namely Mean Per-Joint Position Error (MPJPE) and the Procrustesaligned version (PA-MPJPE), and Mean Vertex-to-Vertex Error (MV2VE) for shape and contact.
我們對 3D 姿勢使用標準評估指標，即平均每關節位置誤差 (MPJPE) 和 Procrustesaligned 版本 (PA-MPJPE)，以及用於形狀和接觸的平均頂點到頂點誤差 (MV2VE)。

Tables 2 and 3 summarize the results of TUCH on 3DPW and 3DCP Scan.
表 2 和表 3 總結了 TUCH 在 3DPW 和 3DCP 掃描上的結果。

Interestingly, TUCH is more accurate than SPIN on 3DPW.

表 2 和表 3 總結了 TUCH 在 3DPW 和 3DCP 掃描上的結果。

See Sup. Mat. for results of fine-tuning EFT.
See Sup. Mat. 用於微調 EFT 的結果。

We further evaluate our results w.r.t. contact.
我們進一步評估我們的結果 w.r.t. 接觸。

To this end, we divide the 3DPW test set into subsets, namely for tgeo = 50cm: self-contact (teucl < 1cm), no self-contact (teucl > 5cm), and unclear (1cm < teucl < 5cm).
為此，我們將 3DPW 測試集劃分為子集，即對於 tgeo = 50cm：自接觸（teucl < 1cm）、無自接觸（teucl > 5cm）和不清楚（1cm < teucl < 5cm）。

For 3DPW we obtain 8752 self-contact, 16752 no self-contact, and 9491 unclear poses.
對於 3DPW，我們獲得了 8752 個自接觸、16752 個無自接觸和 9491 個不清楚的姿勢。

Table 4 shows a clear improvement on poses with contact and unclear poses compared to a smaller improvement on poses without contact.
表 4 顯示了與無接觸姿勢的較小改進相比，有接觸姿勢和不清楚姿勢的明顯改進。

To further understand the improvement of TUCH over SPIN, we break down the improved MPJPE in 3DPW selfcontact into the pairwise body-part contact labels defined in the DSC dataset.
為了進一步了解 TUCH 相對於 SPIN 的改進，我們將 3DPW 自接觸中改進的 MPJPE 分解為 DSC 數據集中定義的成對身體部位接觸標籤。

Specifically, for each contact pair, we search all poses in 3DPW self-contact that have this particular self-contact.
具體來說，對於每個接觸對，我們搜索 3DPW 自接觸中具有這種特定自接觸的所有姿勢。

We find a clear improvement for a large number of contacts between two body parts, frequently between arms and torso, or e.g. left hand and right elbow, which is common in arms-crossed poses (see Fig. 9).
我們發現兩個身體部位之間的大量接觸有明顯的改善，經常在手臂和軀幹之間，或者例如 左手和右手肘，這在雙臂交叉姿勢中很常見（見圖 9）。

	MPJPE		PA-MPJPE	
	3DPW	MI	3DPW	MI
SPIN [24]	96.9	105.2	59.2	**67.5**
EFT [18]	-	-	**54.2**	68.0
TUCH	**84.9**	**101.2**	55.5	68.6

Table 2. Evaluation on 3DPW and MPI-INF-3DHP (MI). Bold numbers indicate the best result; units are *mm*. We report the EFT result denoted in their publication when 3DPW was not part of the training data. Please note that SPIN is trained on MI, but we do not include MI in the fine-tuning set. MI contains mostly indoor lab sequences (100% train, 75% test), while DSC and MTP contain only in-the-wild images. This domain gap likely explains the decreased performance in PA-MPJPE.

Table 2. Evaluation on 3DPW and MPI-INF-3DHP (MI).
表 2. 對 3DPW 和 MPI-INF-3DHP (MI) 的評估。

Bold numbers indicate the best result; units are mm.
粗體數字表示最佳結果； 單位是毫米。

We report the EFT result denoted in their publication when 3DPW was not part of the training data.
當 3DPW 不是訓練數據的一部分時，我們報告了在他們的出版物中指出的 EFT 結果。

Please note that SPIN is trained on MI, but we do not include MI in the fine-tuning set.
請注意，SPIN 是在 MI 上訓練的，但我們不將 MI 包括在微調集中。

MI contains mostly indoor lab sequences (100% train, 75% test), while DSC and MTP contain only in-the-wild images.
MI 主要包含室內實驗室序列（100% 訓練，75% 測試），而 DSC 和 MTP 僅包含野外圖像。

This domain gap likely explains the decreased performance in PA-MPJPE.
這種域差距可能解釋了 PA-MPJPE 性能下降的原因。

	MPJPE	PA-MPJPE	MV2VE
SPIN [24]	79.7	50.6	95.7
EFT [18]	71.4	48.3	83.9
TUCH	**69.5**	**42.5**	**81.5**

Table 3. Evaluation on 3DCP Scan. Numbers are in *mm*. Note that in contrast to TUCH, this version of SPIN did not see poses in the MTP dataset during training. Please see Table 5 and the corresponding text for an ablation study.

Table 3. Evaluation on 3DCP Scan.
表 3. 3DCP 掃描評估。

Numbers are in mm. Note that in contrast to TUCH, this version of SPIN did not see poses in the MTP dataset during training.
數字以毫米為單位。 請注意，與 TUCH 相比，此版本的 SPIN 在訓練期間未在 MTP 數據集中看到姿勢。

Please see Table 5 and the corresponding text for an ablation study.
有關消融研究，請參閱表 5 和相應文本。

	MPJPE				PA-MPJPE			
	contact	no contact	unclear	total	contact	no contact	unclear	total
SPIN	100.2	95.5	96.7	96.9	59.1	61.7	55.7	59.2
TUCH	**85.1**	**86.6**	**81.9**	**84.9**	**54.1**	**58.6**	**51.2**	**55.5**

Table 4. Evaluation of TUCH for contact classes in 3DPW. Numbers are in *mm*. See text.

Table 4. Evaluation of TUCH for contact classes in 3DPW.
表 4. TUCH 對 3DPW 中接觸類的評估。

Numbers are in mm. See text.
數字以毫米為單位。 見正文。

TUCH incorporates self-contact in various ways: annotations of training data, in-the-loop fitting, and in the regression loss.
TUCH 以多種方式結合了自接觸：訓練數據的註釋、在環擬合和回歸損失。

We evaluate the impact of each in Table 5.
我們評估了表 5 中每個的影響。

S+ is SPIN but it sees MTP+DSC images in fine-tuning and runs standard in-the-loop SMPLify with no contact information.
S+ 是 SPIN，但它可以在微調中看到 MTP+DSC 圖像，並在沒有聯繫資訊的情況下運行標準的在環 SMPLify。

S++ is S+ but uses pseudo ground truth computed with SMPLify-XMC on MTP images;
S++ 是 S+，但在 MTP 圖像上使用 SMPLify-XMC 計算的偽地面實況；

thus self-contact is used to generate the data but nowhere else.
因此，自我接觸用於生成數據，而不是其他地方。

S+ vs. SPIN suggests that, while poses in 3DCP Scan appear in MTP, just seeing similar poses for training and testing does not yield improvement.

S+ vs. SPIN 表明，雖然 3DCP 掃描中的姿勢出現在 MTP 中，但僅看到類似的訓練和測試姿勢並不會產生改進。

S+ vs. TUCH is a fair comparison as both see the same images during training.
S+ 與 TUCH 是一個公平的比較，因為兩者在訓練期間看到的圖像相同。

The improved results of TUCH confirm the benefit of using self-contact.
TUCH 的改進結果證實了使用自接觸的好處。

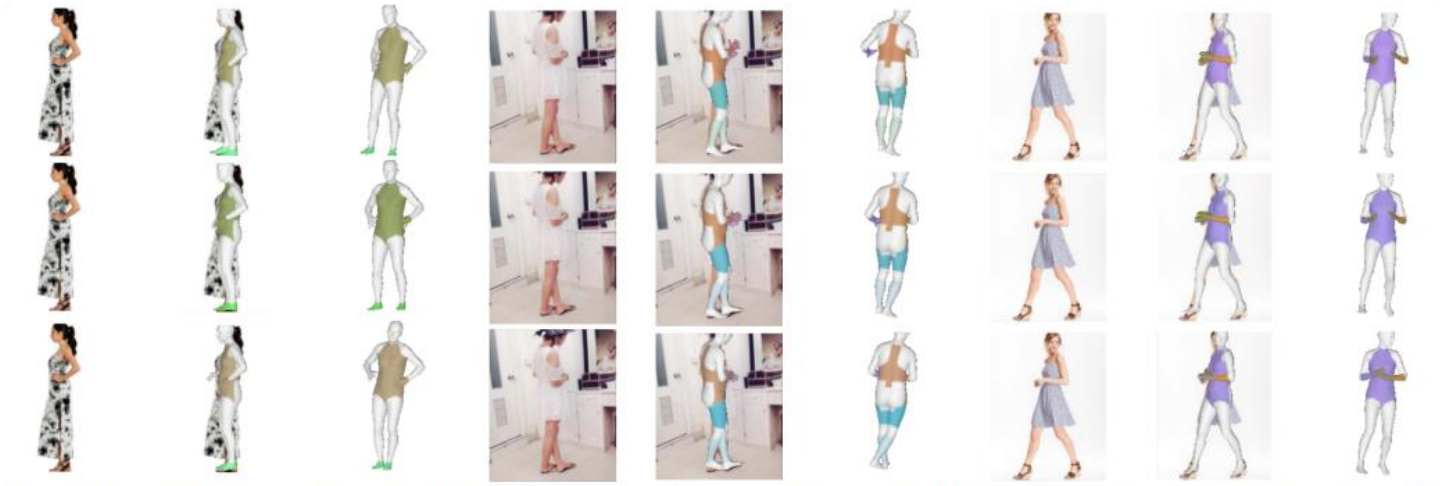

Figure 8. Impact of discrete self-contact labels in human pose estimation. Body parts labeled in contact are shown in the same color. First row shows an initial SPIN estimate, second row the SMPLify fit, third row the SMPLify-DC fit after 20 iterations.

Figure 8. Impact of discrete self-contact labels in human pose estimation.
圖 8. 離散自接觸標籤對人體姿勢估計的影響。

Body parts labeled in contact are shown in the same color.
標記為接觸的身體部位以相同的顏色顯示。

First row shows an initial SPIN estimate, second row the SMPLify fit, third row the SMPLify-DC fit after 20 iterations.
第一行顯示初始 SPIN 估計，第二行顯示 SMPLify 擬合，第三行顯示 20 次迭代後的 SMPLify-DC 擬合。

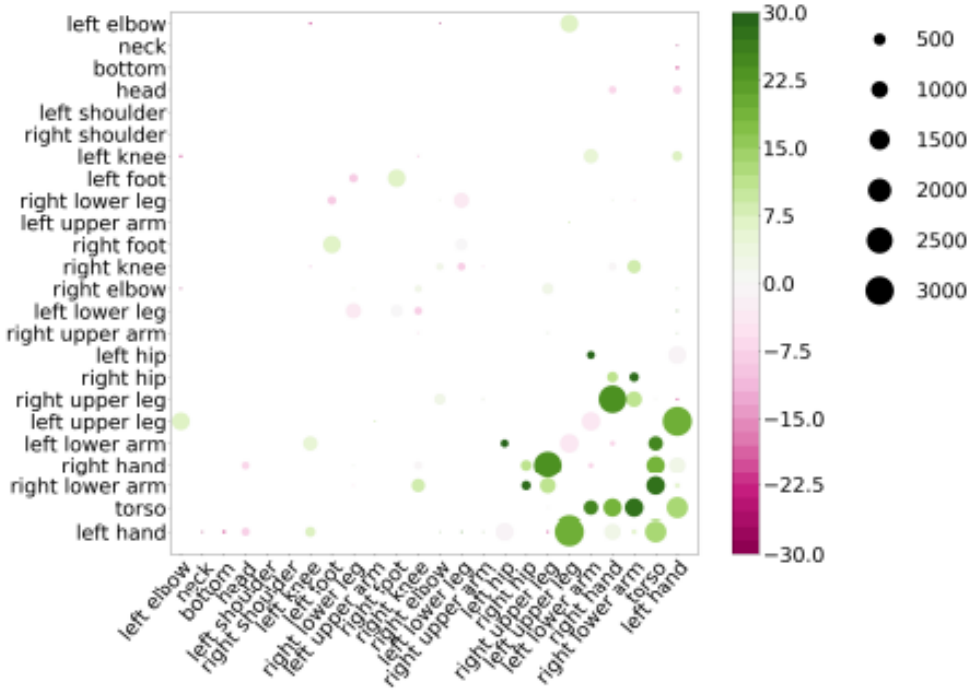

Figure 9. Average MPJPE difference (SPIN - TUCH), evaluated on the *self-contact* subset of 3DPW. The axes show labels for the DSC regions. Green indicates that TUCH has a lower error than SPIN on average across all poses with the corresponding regions in contact. The circle size represents the number of images per region. Regions with small circle sizes are less common.

Figure 9. Average MPJPE difference (SPIN - TUCH), evaluated on the self-contact subset of 3DPW.
圖 9. 在 3DPW 的自接觸子集上評估的平均 MPJPE 差異 (SPIN - TUCH)。

The axes show labels for the DSC regions.
軸顯示 DSC 區域的標籤。
Green indicates that TUCH has a lower error than SPIN on average across all poses with the corresponding regions in contact.
綠色表示 TUCH 在所有姿勢中的平均誤差低於 SPIN，並且相應區域接觸。

The circle size represents the number of images per region.
圓圈大小表示每個區域的圖像數量。

Regions with small circle sizes are less common.
小圓圈大小的區域不太常見。

	SPIN	S+	S++	TUCH
3DPW	96.9/ 59.2	96.1/ 61.4	85.0/ 56.3	**84.9/ 55.5**
3DCP Scan	82.2/ 52.1	86.9/ 52.3	**74.8**/ 45.7	75.2/ **45.4**
MI	105.2/ 67.5	105.8/ 69.4	103.1/ 69.0	**101.2/ 68.6**

Table 5. MPJPE/PA-MPJPE (mm) to examine the impact of data and algorithm on 3DPW, 3DCP Scan, and MPI-INF-3DHP (MI).

Table 5. MPJPE/PA-MPJPE (mm) to examine the impact of data and algorithm on 3DPW, 3DCP Scan, and MPI-INF-3DHP (MI).
表 5. MPJPE/PA-MPJPE (mm) 用於檢查數據和演算法對 3DPW、3DCP 掃描和 MPI-INF-3DHP (MI) 的影響。

7. Conclusion 結論

In this work, we address the problem of HPS estimation when self-contact is present.
在這項工作中，我們解決了存在自接觸時 HPS 估計的問題。

Self-contact is a natural, com-mon occurrence in everyday life, but SOTA methods fail to estimate it.
自我接觸是日常生活中自然而常見的現象，但 SOTA 方法無法估計它。

One reason for this is that no datasets pairing images in the wild and 3D reference poses exist.
原因之一是不存在在野外和 3D 參考姿勢中配對圖像的數據集。

To address this problem we introduce a new way of collecting data:
為了解決這個問題，我們引入了一種新的數據收集方式：

we ask humans to mimic presented 3D poses.
我們要求人類模仿呈現的 3D 姿勢。

Then we use our new SMPLify-XMC method to fit pseudo ground-truth 3D meshes to the mimicked images, using the presented pose and self-contact to constrain the optimization.
然後我們使用我們新的 SMPLify-XMC 方法將偽真實 3D 網格擬合到模擬圖像，使用呈現的姿勢和自接觸來約束優化。

We use the new MTP data along with discrete self-contact annotations to train TUCH; the first end-to-end HPS regressor that also handles poses with self-contact.
我們使用新的 MTP 數據以及離散的自接觸註釋來訓練 TUCH； 第一個端到端的 HPS 回歸器，它也可以通過自接觸處理姿勢。

TUCH uses MTP data as if it was ground truth, while the discrete, DSC, data is exploited during SPIN training

via SMPLify-DC.
TUCH 使用 MTP 數據，就好像它是地面實況一樣，而離散的 DSC 數據在 SPIN 訓練期間通過 SMPLify-DC 被利用。

Overall, incorporating contact improves accuracy on standard benchmarks like 3DPW, remarkably, not only for poses with selfcontact, but also for poses without self-contact.
總體而言，結合接觸提高了標準基準（如 3DPW）的準確性，不僅適用於具有自接觸的姿勢，而且還適用於沒有自接觸的姿勢。