PoseAug: A Differentiable Pose Augmentation Framework for 3D Human Pose Estimation

PoseAug：用於 3D 人體姿勢估計的可微姿勢增強框架

Kehong Gong, Jianfeng Zhang, Jiashi Feng

National University of Singapore

CVPR 2021 Best Paper Candidate

https://arxiv.org/abs/2105.02465

https://github.com/jfzhang95/PoseAug

Abstract 摘要

Existing 3D human pose estimators suffer poor generalization performance to new datasets, largely due to the limited diversity of 2D-3D pose pairs in the training data.

現有的 3D 人體姿態估計器對新數據集的泛化性能較差，這主要是由於訓練數據中 2D-3D 姿態對的多樣性有限。

To address this problem, we present PoseAug, a new autoaugmentation framework that learns to augment the available training poses towards a greater diversity and thus improve generalization of the trained 2D-to-3D pose estimator.

為了解決這個問題，我們提出了 PoseAug，這是一種新的自動增強框架，它可以學習將可用的訓練姿勢增加到更大的多樣性，從而提高經過訓練的 2D 到 3D 姿勢估計器的泛化。

Specifically, PoseAug introduces a novel pose augmentor that learns to adjust various geometry factors (e.g., posture, body size, view point and position) of a pose through differentiable operations.

具體來說，PoseAug 引入了一種新穎的姿勢增強器，它可以通過可微分操作來學習調整姿勢的各種幾何因素（例如，姿勢、身體大小、視點和位置）。

With such differentiable capacity, the augmentor can be jointly optimized with the 3D pose estimator and take the estimation error as feedback to generate more diverse and harder poses in an online manner.

有了這種可區分的容量，增強器可以與 3D 姿態估計器聯合優化，並將估計誤差作為反饋，以在線方式生成更多樣和更難的姿態。

Moreover, PoseAug introduces a novel part-aware Kinematic Chain Space for evaluating local joint-angle plausibility and develops a discriminative module accordingly to ensure the plausibility of the augmented poses.

此外，PoseAug 引入了一種新穎的部分感知運動鏈空間來評估局部關節角度的合理性，並相應地開發

了一個判別模塊以確保增強姿勢的合理性。

These elaborate designs enable PoseAug to generate more diverse yet plausible poses than existing offline augmentation methods, and thus yield better generalization of the pose estimator.
這些精心設計的設計使 PoseAug 能夠生成比現有離線增強方法更多樣化但更合理的姿勢，從而更好地泛化姿勢估計器。

PoseAug is generic and easy to be applied to various 3D pose estimators.
PoseAug 是通用的，易於應用於各種 3D 姿勢估計器。

Extensive experiments demonstrate that PoseAug brings clear improvements on both intra-scenario and cross-scenario datasets.
大量實驗表明，PoseAug 在場景內和跨場景數據集上都帶來了明顯的改進。

Notably, it achieves 88.6% 3D PCK on MPI-INF-3DHP under cross-dataset evaluation setup, improving upon the previous best data augmentation based method [22] by 9.1%.
值得注意的是，它在跨數據集評估設置下在 MPI-INF-3DHP 上實現了 88.6% 的 3D PCK，比之前基於數據增強的最佳方法 [22] 提高了 9.1%。

Code can be found at: https://github.com/jfzhang95/PoseAug.

1. Introduction 前言

3D human pose estimation aims to estimate 3D body joints in images or videos.
3D 人體姿態估計旨在估計圖像或視頻中的 3D 身體關節。

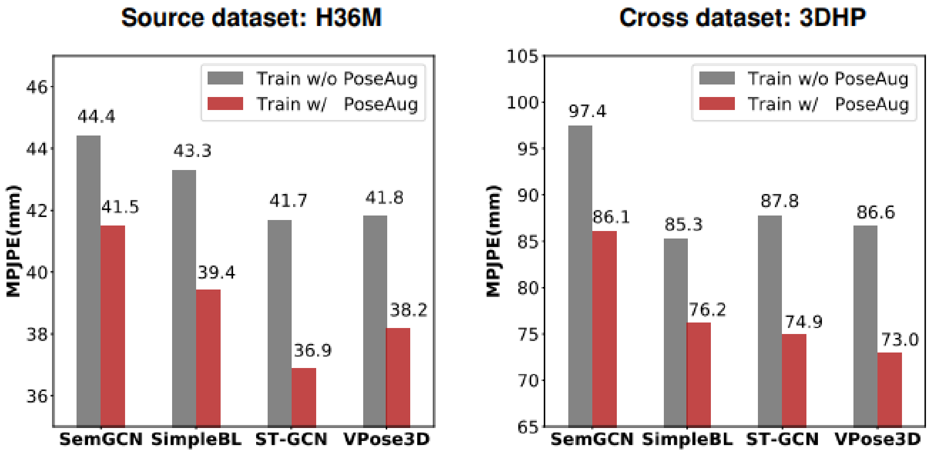
It is a fundamental task with broad applications in action recognition [47, 39], human-robot interaction [11], human tracking [29], etc.
這是一項基本任務，在動作識別 [47, 39]、人機交互 [11]、人體跟踪 [29] 等方面具有廣泛的應用。

This task is typically solved using learning-based methods [26, 52, 3, 32] with ground truth annotations that are collected in the laboratorial environments [16].
該任務通常使用基於學習的方法 [26, 52, 3, 32] 和在實驗室環境 [16] 中收集的地面實況註釋來解決。

Despite their success in indoor scenarios, these methods are hardly generalizable to cross-scenario datasets (e.g., an in-the-wild dataset).
儘管它們在室內場景中取得了成功，但這些方法很難推廣到跨場景數據集（例如，野外數據集）。

We argue that their poor generalization is mainly due to the limited diversity of training data, such as limited variations in human posture, body size, camera view point and position.
我們認為，它們的泛化能力較差主要是由於訓練數據的多樣性有限，例如人體姿勢、身體大小、相機

視點和位置的變化有限。



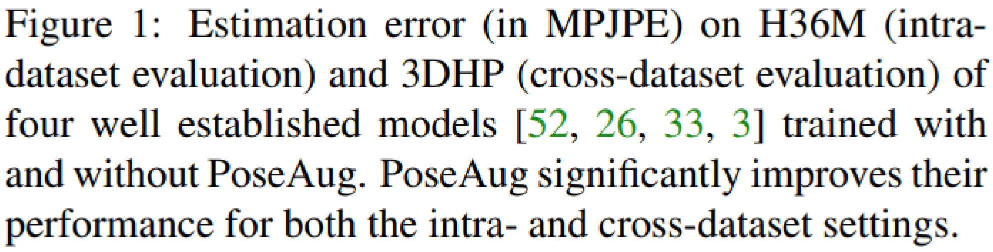


Figure 1: Estimation error (in MPJPE) on H36M (intra-dataset evaluation) and 3DHP (cross-dataset evaluation) of four well established models [52, 26, 33, 3] trained with and without PoseAug. PoseAug significantly improves their performance for both the intra- and cross-dataset settings.

Figure 1: Estimation error (in MPJPE) on H36M (intradataset evaluation) and 3DHP (cross-dataset evaluation) of four well established models [52, 26, 33, 3] trained with and without PoseAug.
圖 1：使用和不使用 PoseAug 訓練的四個完善模型 [52、26、33、3] 在 H36M（數據集內評估）和 3DHP（跨數據集評估）上的估計誤差（在 MPJPE 中）。

PoseAug significantly improves their performance for both the intra- and cross-dataset settings.
PoseAug 顯著提高了數據集內和跨數據集設置的性能。

Recent works explore data augmentation to improve the training data diversity and enhance the generalization of their trained models.
最近的工作探索了數據增強，以提高訓練數據的多樣性並增強其訓練模型的泛化能力。

They either generate data through image composition [37, 29, 28] and synthesis [5, 42], or directly generate 2D-3D pose pairs from the available training data by applying pre-defined transformations [22].
他們要么通過圖像合成 [37, 29, 28] 和合成 [5, 42] 來生成數據，要么通過應用預定義的轉換 [22] 從可用的訓練數據直接生成 2D-3D 姿勢對。

However, all of these works regard data augmentation and model training as two separate phases, and conduct data augmentation in an offline manner without interaction with model training.
然而，所有這些工作都將數據增強和模型訓練視為兩個獨立的階段，並在不與模型訓練交互的情況下以離線方式進行數據增強。

Consequently, they tend to generate ineffective augmented data that are too easy for model training, leading

to marginal boost to the model generalization.
因此，它們往往會生成對模型訓練來說太容易的無效增強數據，從而導致模型泛化的邊際提升。

Moreover, these methods heavily rely on pre-defined rules such as joint angle limitations [1] and kinematics constraints [37] for data augmentation, which limit the diversity of the generated data and make the resulting model hardly generalize to more challenging in-the-wild scenes.
此外，這些方法在很大程度上依賴於預定義的規則，例如用於數據增強的關節角度限制 [1] 和運動學約束 [37]，這限制了生成數據的多樣性，並使生成的模型難以泛化到更具挑戰性的 - 狂野的場景。

To improve the diversity of augmented data, we propose PoseAug, a novel auto-augmentation framework for 3D human pose estimation.
為了提高增強數據的多樣性，我們提出了 PoseAug，一種用於 3D 人體姿態估計的新型自動增強框架。

Instead of conducting data augmentation and network training separately, PoseAug jointly optimizes the augmentation process with network training endto-end in an online manner.
PoseAug 不是分別進行數據增強和網絡訓練，而是通過在線方式端到端地與網絡訓練共同優化增強過程。

Our main insight is that the feedback from the training process can be used as effective guidance signals to adapt and improve the data augmentation.
我們的主要見解是來自訓練過程的反饋可以用作有效的指導信號來適應和改進數據增強。

Specifically, PoseAug exploits a differentiable augmentation module (the ‘augmentor’) implemented by a neural network to directly augment 2D-3D pose pairs in the training data.
具體來說，PoseAug 利用由神經網絡實現的可微增強模塊（“增強器”）來直接增強訓練數據中的 2D-3D 姿勢對。

Considering the potential domain shift with respective to geometry in pose pairs (e.g., postures, view points) [36, 22, 50], the augmentor learns to perform three types of augmentation operations to respectively control 1) the skeleton joint angle, 2) the body size, and 3) the view point and human position.
考慮到相對於姿勢對中的幾何體的潛在域偏移（例如，姿勢、視點）[36、22、50]，增強器學習執行三種類型的增強操作以分別控制 1）骨架關節角度，2） 身體大小，以及 3) 視點和人體位置。

In this way, the augmentor is able to produce augmented poses with more diverse geometric features and thus relieves the diversity limitation issue.
通過這種方式，增強器能夠產生具有更多樣化幾何特徵的增強姿態，從而緩解多樣性限制問題。

With its differentiable capacity, the augmentor can be optimized together with the pose estimator end-to-end via an error feedback strategy.
憑藉其可微分的能力，增強器可以通過錯誤反饋策略與姿態估計器一起進行端到端的優化。

Concretely, by taking increasing training loss of the estimator as the learning target, the augmentor can learn to enrich the input pose pairs via enlarging data variations and difficulties; in turn, through combating such increasing difficulties, the pose estimator can become increasingly more powerful during the training process.
具體而言，以增加估計器的訓練損失為學習目標，增強器可以學習通過擴大數據變化和難度來豐富輸入姿勢對； 反過來，通過克服這些日益增加的困難，姿勢估計器在訓練過程中可以變得越來越強大。

To ensure the plausibility of the augmented poses, we use a pose discriminator module to guide the augmentation, to avoid generating implausible joint angles [1], unreasonable positions or view points that may hamper model training.
為了確保增強姿勢的合理性，我們使用姿勢鑑別器模塊來指導增強，以避免產生不合理的關節角度 [1]、不合理的位置或可能妨礙模型訓練的觀點。

In particular, the module consists of a 3D pose discriminator for enhancing the joint angle plausibility and a 2D pose discriminator for guiding the body size, view point and position plausibility.
特別是，該模塊由一個用於增強關節角度可信度的 3D 姿勢鑑別器和一個用於指導身體尺寸、視點和位置可信度的 2D 姿勢鑑別器組成。

The 3D pose discriminator adopts the Kinematic Chain Space (KCS) [44] representation and extends it into a part-aware KCS for local-wise supervision.
3D 姿勢鑑別器採用運動鏈空間 (KCS) [44] 表示並將其擴展為局部感知 KCS，以進行局部監督。

More concretely, it splits skeleton joints into several parts and focuses on joint angles in each part separately instead of the whole body pose, which yields greater flexibility of the augmented poses.
更具體地說，它將骨骼關節分成幾個部分，並分別關注每個部分的關節角度，而不是整個身體姿勢，從而產生更大的增強姿勢的靈活性。

By jointly training the pose augmentor, estimator and discriminator in an end-to-end manner (Fig. 2), PoseAug can largely improve the training data diversity, and thus boost model performance on both source and more challenging cross-scenario datasets.
通過以端到端的方式聯合訓練姿態增強器、估計器和鑑別器（圖 2），PoseAug 可以在很大程度上提高訓練數據的多樣性，從而提高模型在源數據集和更具挑戰性的跨場景數據集上的性能。

Our PoseAug framework is flexible regarding the choice of the 3D human pose estimator.
我們的 PoseAug 框架可以靈活地選擇 3D 人體姿勢估計器。

This is demonstrated by the clear improvements made with PoseAug on four representative 3D pose estimation models [52, 26, 33, 3] over both source (H36M) [16] and cross-scenario (3DHP) [29] datasets (Fig. 1).
PoseAug 在四個代表性 3D 姿態估計模型 [52, 26, 33, 3] 上對源 (H36M) [16] 和跨場景 (3DHP) [29] 數據集的明顯改進證明了這一點（圖 1） ）。

Remarkably, it brings more than 13.1% average improvement w.r.t. MPJPE for all models on 3DHP.
值得注意的是，它帶來了超過 13.1% 的平均改進 w.r.t. MPJPE 適用於 3DHP 上的所有模型。

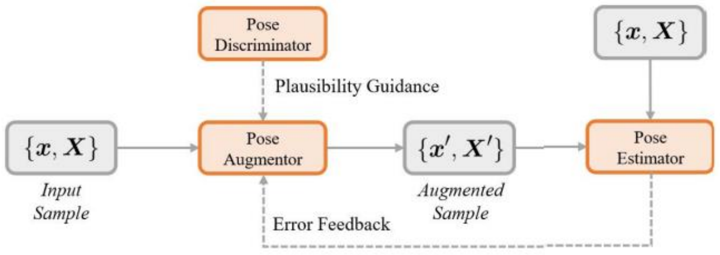
Moreover, it achieves 88.6% 3D PCK on 3DHP under cross-dataset evaluation setup, improving upon the previous best data augmentation based method [22] by 9.1%.
此外，它在跨數據集評估設置下在 3DHP 上實現了 88.6% 的 3D PCK，比之前最好的基於數據增強的方法 [22] 提高了 9.1%。

Our contributions are three-fold. 我們的貢獻是三方面的。

1) To the best of our knowledge, we are the first to investigate differentiable data augmentation on 3D human pose estimation.
1) 據我們所知，我們是第一個研究 3D 人體姿勢估計的可微數據增強的人。

2) We propose a differentiable pose augmentor, together with the error feedback design, which generates diverse and realistic 2D-3D pose pairs for training the 3D pose estimator, and largely enhances the model's generalization ability.
2）我們提出了一個可微的姿態增強器，連同誤差反饋設計，它生成多樣化和逼真的 2D-3D 姿態對來訓練 3D 姿態估計器，並在很大程度上增強了模型的泛化能力。

3) We propose a new part-aware 3D discriminator, which enlarges the feasible region of augmented poses via local-wise supervision, ensuring both data plausibility and diversity.
3）我們提出了一種新的部分感知 3D 鑑別器，它通過局部監督擴大了增強姿勢的可行區域，確保了數據的合理性和多樣性。


Pose Discriminator
Plausibility Guidance
$\{\boldsymbol{x}, \boldsymbol{X}\}$
$\{\boldsymbol{x}, \boldsymbol{X}\}$
Pose Augmentor
$\{\boldsymbol{x}', \boldsymbol{X}'\}$
Pose Estimator
Input Sample
Augmented Sample
Error Feedback

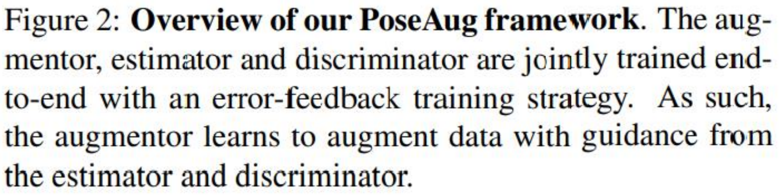


Figure 2: **Overview of our PoseAug framework**. The augmentor, estimator and discriminator are jointly trained end-to-end with an error-feedback training strategy. As such, the augmentor learns to augment data with guidance from the estimator and discriminator.

Figure 2: Overview of our PoseAug framework.
圖 2：我們的 PoseAug 框架概述。

The augmentor, estimator and discriminator are jointly trained endto-end with an error-feedback training

strategy.
增強器、估計器和鑑別器使用錯誤反饋訓練策略進行端到端的聯合訓練。

As such, the augmentor learns to augment data with guidance from the estimator and discriminator.
因此，增強器學習在估計器和鑑別器的指導下增強數據。

2. Related Work 相關工作

3D human pose estimation Recent progress of 3D human pose estimation is largely driven by the deployment of various deep neural network models [41, 26, 12, 52, 31, 3, 38, 53].
3D 人體姿態估計 3D 人體姿態估計的最新進展主要是由各種深度神經網絡模型的部署推動的 [41, 26, 12, 52, 31, 3, 38, 53]。

However, they all highly rely on well-annotated data for fully-supervised model training and hardly generalize to the new scenarios that present unseen patterns in the training dataset, such as new camera views and subject poses.
然而，它們都高度依賴於良好註釋的數據進行全監督模型訓練，並且很難推廣到在訓練數據集中呈現出看不見的模式的新場景，例如新的相機視圖和主體姿勢。

Thus some recent works explore to leverage external information to improve their generalization ability.
因此，最近的一些工作探索利用外部信息來提高其泛化能力。

For example, some methods [54, 48, 9, 44, 14, 45, 6, 33, 20] utilize 2D pose data collected in the wild for model training, e.g., through exploring kinematics priors for regularization or post-processing [54, 9, 33], and adversarial training [48, 44].
例如，一些方法 [54, 48, 9, 44, 14, 45, 6, 33, 20] 利用在野外收集的 2D 姿態數據進行模型訓練，例如，通過探索運動學先驗進行正則化或後處理 [54 , 9, 33] 和對抗性訓練 [48, 44]。

More recently, geometry-based self-supervised learning [36, 10, 4, 19, 34, 23, 35] has been used to train models with unlabeled data.
最近，基於幾何的自監督學習 [36, 10, 4, 19, 34, 23, 35] 已被用於訓練具有未標記數據的模型。

Though effective, applying these methods is largely constrained by the availability of suitable external datasets.
儘管有效，但應用這些方法在很大程度上受限於合適的外部數據集的可用性。

Instead of focusing on complex network architectures and learning schemes, we explore a learnable pose augmentation framework to enrich the 3D pose data at hand directly.
我們不是專注於複雜的網絡架構和學習方案，而是探索可學習的姿勢增強框架來直接豐富手頭的 3D 姿勢數據。

Specifically, the proposed frame work can generate 2D-3D pose pairs with both diversity and plausibility for training pose estimation models.
具體來說，所提出的框架可以生成具有多樣性和合理性的 2D-3D 姿勢對，用於訓練姿勢估計模型。

In addition, our framework is generic and can adapt to those methods to further improve their performance.
此外，我們的框架是通用的，可以適應這些方法以進一步提高其性能。

Data augmentation on 3D human poses 3D 人體姿勢的數據增強

Data augmentation is widely used to alleviate the bottleneck of training data diversity and improve model generalization ability.
數據增強被廣泛用於緩解訓練數據多樣性的瓶頸，提高模型泛化能力。

Some works augment data by stitching image patches [37, 29, 51], and some generate new data with graphics engines [5, 42].
一些作品通過拼接圖像塊來增加數據 [37, 29, 51]，一些作品使用圖形引擎 [5, 42] 生成新數據。

More recently, Li et al., [22] directly augment 2D-3D pose pairs through randomly applying partial skeleton recombination and joint angle perturbation on source datasets.
最近，Li et al., [22] 通過在源數據集上隨機應用部分骨架重組和關節角度擾動來直接增加 2D-3D 姿勢對。

To ensure data plausibility, several constraints are imposed, including joint angle limitation [1] and fixed augmentation range on view point and human position.
為了確保數據的合理性，施加了一些約束，包括關節角度限制 [1] 和視點和人體位置的固定增強範圍。

Despite the good results on source data, these pre-defined rules limit the data diversity expansion and harm the model applicability to more challenging in-the-wild scenarios.
儘管在源數據上取得了良好的結果，但這些預定義的規則限制了數據多樣性的擴展，並損害了模型對更具挑戰性的野外場景的適用性。

Unlike all these methods, we make the first attempt to explore learnable data augmentation on 3D human pose estimation, which is shown effective for improving model generalization ability.
與所有這些方法不同，我們首次嘗試探索 3D 人體姿態估計的可學習數據增強，這對於提高模型泛化能力是有效的。

3. Method 方法

3.1. Problem Definition 問題定義

Let x ∈ R2xJ denote 2D spatial coordinates of J keypoints of the human in the image, and X ∈ R3xJ denote the corresponding 3D joint position in the camera coordinate system.
設 x∈R2xJ 表示圖像中人體 J 個關鍵點的 2D 空間坐標，X∈R3xJ 表示相機坐標系中對應的 3D 關節位置。

We aim to obtain a 3D pose estimator P : x → X to recover the 3D pose information from the input 2D pose.
我們的目標是獲得一個 3D 姿態估計器 P : x → X 以從輸入的 2D 姿態中恢復 3D 姿態資訊。

Conventionally, the estimator P, with parameters , is trained on a well-annotated source dataset (e.g., well-controlled indoor environment [16]) by solving the following optimization problem:
通常，帶有參數 的估計器 P 通過解決以下優化問題在註釋良好的源數據集（例如，控制良好的室內環境 [16]）上進行訓練：

$$\min_{\theta} \mathcal{L}_{\mathcal{P}}(\mathcal{P}_{\theta}, \mathcal{X}) = \mathcal{L}_{\mathcal{P}}(\mathcal{P}_{\theta}(\boldsymbol{x}), \boldsymbol{X}), \qquad (1)$$

where X = {x,X} denotes paired 2D-3D poses from the source training dataset, and the loss function LP is typically defined as mean square errors (MSE) between predicted and ground truth 3D poses.
其中 X = {x,X} 表示來自源訓練數據集的成對 2D-3D 姿勢，損失函數 LP 通常定義為預測和真實 3D 姿勢之間的均方誤差 (MSE)。

However, it is often observed that the pose estimator P trained on such an indoor dataset can hardly generalize to a new dataset (e.g., in-the-wild scenario) which features more diverse poses, body sizes, view points or human positions [14, 50, 46].
然而，經常觀察到，在這樣的室內數據集上訓練的姿勢估計器 P 很難推廣到一個新的數據集（例如，野外場景），該數據集具有更多樣化的姿勢、身體尺寸、視點或人體位置 [14] , 50, 46]。

To improve generalization ability of the model, we propose to design a pose augmentor A : X → X’, to augment the training pose pair X into a more diverse one X’ = {x’,X’} for training the model P:
為了提高模型的泛化能力，我們建議設計一個姿勢增強器 A : X → X'，將訓練姿勢對 X 增強為更多樣化的 X' = {x',X'} 用於訓練模型 P：

$$\min_{\theta} \mathcal{L}_{\mathcal{P}}(\mathcal{P}_{\theta}, \mathcal{A}(\mathcal{X})). \qquad (2)$$

There are several strategies to construct the augmentor in an offline manner, e.g., random [5, 29, 42] or evolution-based augmentations [22].
有幾種以離線方式構建增強器的策略，例如隨機 [5, 29, 42] 或基於進化的增強 [22]。

Differently, we propose to implement the augmentor A via a neural network with parameters θ A and train it jointly with the estimator in an online manner, such that the pose estimator loss can be fully exploited as a

surrogate for the augmentation diversity and effectively guide the augmentor learning.
不同的是，我們建議通過具有參數 A 的神經網絡實現增強器 A 並以在線方式與估計器聯合訓練它，這樣姿勢估計器損失就可以被充分利用作為增強多樣性的替代物並有效地指導增強器 學習。

In particular, the augmentor is trained to generate harder augmented samples that could increase the training loss of the current pose estimator:
特別是，增強器經過訓練以生成更難的增強樣本，這可能會增加當前姿勢估計器的訓練損失：

$$\min_{\theta} \max_{\theta_A} \mathcal{L}_{\mathcal{P}}(\mathcal{P}_{\theta}, \mathcal{A}_{\theta_A}(\mathcal{X})). \qquad (3)$$

3.2. PoseAug Formulation 公式

Our proposed framework aims to generate diverse training data, with proper difficulties for the pose estimator, to improve model generalization performance.
我們提出的框架旨在生成不同的訓練數據，對姿勢估計器具有適當的難度，以提高模型泛化性能。

Two challenges thus need to be tackled: how to make the augmented data diverse and beneficial for model training; and how to make them natural and realistic.
因此需要解決兩個挑戰：如何使增強數據多樣化並有利於模型訓練； 以及如何使它們自然逼真。

To address them, we propose two novel ideas in training the augmentor.
為了解決這些問題，我們提出了兩個訓練增強器的新想法。

Error feedback learning for online pose augmentation 在線姿態增強的錯誤反饋學習

Instead of performing random pose augmentation in an offline manner [37, 5, 22], the proposed pose augmentator A deploys a differentiable design which enables online joint-training with the pose estimator P.
所提出的姿勢增強器 A 不是以離線方式執行隨機姿勢增強 [37, 5, 22]，而是部署了一種可微分設計，該設計能夠與姿勢估計器 P 進行在線聯合訓練。

Using the training error from the pose estimator P as feedback (see Eqn. (3)), the pose augmentor A learns to generate poses that are most suitable for the current pose estimator—the augmented poses present proper difficulties and diversity due to online augmentation, thus maximally benefiting generalization of the trained 3D pose estimation model.
使用姿勢估計器 P 的訓練誤差作為反饋（參見方程（3）)，姿勢增強器 A 學習生成最適合當前姿勢估計器的姿勢——由於在線增強，增強的姿勢呈現出適當的難度和多樣性 ，從而最大限度地有利於訓練的 3D 姿態估計模型的泛化。

Discriminative learning for plausible pose augmentation 合理姿勢增強的判別式學習

Purely pursuing error-maximized augmentations may result in implausible training poses that violate the biomechanical structure of human body and may hurt model performance.
純粹追求誤差最大化的增強可能會導致違反人體生物力學結構的難以置信的訓練姿勢，並可能損害模型性能。

Previous augmentation methods [37, 5, 22] mostly rely on pre-defined rules for ensuring plausibility (e.g., joint angle constraint [1]), which however would severely limit the diversity of generated poses.
以前的增強方法 [37, 5, 22] 主要依賴於預先定義的規則來確保合理性（例如，關節角度約束 [1]），但是這會嚴重限制生成姿勢的多樣性。

For example, some harder yet plausible poses may fail to pass their rulebased plausibility check [22] and will not be adopted for model training.
例如，一些更難但合理的姿勢可能無法通過其基於規則的合理性檢查 [22] 並且不會被用於模型訓練。

To address this issue, we deploy a pose discriminator module over the local relation of body joints [44] to assist training the augmentor, thus ensuring the plausibility of augmented poses without sacrificing the diversity.
為了解決這個問題，我們在身體關節的局部關係上部署了一個姿勢鑑別器模塊 [44] 來幫助訓練增強器，從而在不犧牲多樣性的情況下確保增強姿勢的合理性。

3.3. Architecture 架構

Fig. 2 summarizes our PoseAug architecture design.
圖 2 總結了我們的 PoseAug 架構設計。

It includes
這包括

1) a pose augmentor that augments the input pose pair {x,X} to an augmented one {x’,X’} for pose estimator P training;
1) 一個姿態增強器，將輸入姿態對 {x,X} 增強為增強的 {x',X'}，用於姿態估計器 P 訓練；

2) a pose discriminator module with two discriminators in 3D and 2D spaces, to ensure the plausibility of the augmented data; and
2) 一個姿勢鑑別器模塊，在 3D 和 2D 空間中具有兩個鑑別器，以確保增強數據的合理性； 和

3) a 3D pose estimator, that provides pose estimation error feedback.
3) 3D 姿態估計器，提供姿態估計誤差反饋。

Augmentor

Given a 3D pose X ∈ R3xJ , the augmentor first obtains its bone vector B ∈ R3x(J-1) via a hierarchical transformation1 B = H(X) [44, 22], which can be further decomposed into a bone direction vector ^B (representing the joint angle) and a bone length vector ||B|| (representing the body size).
1) 一個姿態增強器，將輸入姿態對 {x,X} 增強為增強的 {x',X'}，用於姿態估計器 P 訓練；

Then the augmentor applies multi-layer perceptron (MLP) for feature extraction from the input 3D pose X.
然後增強器應用多層感知器（MLP）從輸入的 3D 姿態 X 中提取特徵。

Additionally, a noise vector based on Gaussian distribution is concatenated with X in the feature extraction process to incur sufficient randomness for enhancing the feature diversity.
此外，基於高斯分佈的噪聲向量在特徵提取過程中與 X 連接，以產生足夠的隨機性以增強特徵多樣性。

The extracted features are then used for regressing three operation parameters (ba, bl and (R, t)) to change the joint angles, body size, as well as view point and position as illustrated in Fig. 3.
然後將提取的特徵用於回歸三個操作參數（ba、bl 和（R，t））以改變關節角度、身體大小以及視點和位置，如圖 3 所示。

Among these parameters,
在這些參數中，

1) γ ba ∈ R3x(J-1) is the bone angle residual vector that is used for adjusting the Bone Angle (BA) as follows:
1) γba ∈ R3x(J-1) 是骨骼角度殘差向量，用於調整骨骼角度 (BA) 如下：

$$1)\ \gamma_{\boldsymbol{ba}} \in \mathbb{R}^{3\times(J-1)}$$

$$\hat{\boldsymbol{B}}' = \hat{\boldsymbol{B}} + \gamma_{\boldsymbol{ba}}, \qquad \text{(BA operation).} \qquad (4)$$

Specifically, BA operation will rotate the input bone direction vector ^B by γ ba, generating a new bone direction vector ^B’.
具體來說，BA 操作會將輸入的骨骼方向向量 ^B 旋轉 γ ba，生成一個新的骨骼方向向量 ^B'。

2) γ bl∈ R1(J-1) represents the bone length ratio vector that is used for adjusting the Bone Length (BL):
2) γbl∈R1(J-1)表示用於調整骨骼長度(BL)的骨骼長度比率向量：

$$\|\boldsymbol{B}'\| = \|\boldsymbol{B}\| \times (1+\gamma_{\boldsymbol{bl}}), \qquad \text{(BL operation).} \qquad (5)$$

BL operation modifies the input bone length vector ||B|| by γ bl to adjust the body size.
BL 操作修改輸入骨骼長度向量 ||B|| 通過 γ bl 來調整 body 大小。

Notably, to ensure biomechanical symmetry, the left and right body parts share the same parameters.
值得注意的是，為了確保生物力學對稱性，左右身體部位共享相同的參數。

3) R ∈ R3x3 and t ∈ R3x1 denote the rotation and translation parameters respectively for Rigid Transformation (RT) operation to control pose view point and position:
3) R ∈ R3x3 和 t ∈ R3x1 分別表示剛性變換 (RT) 操作的旋轉和平移參數，以控制姿勢視點和位置：

$$\boldsymbol{X}' = \boldsymbol{R}[\mathcal{H}^{-1}(\boldsymbol{B}')] + \boldsymbol{t}, \qquad \text{(RT operation)}, \qquad (6)$$

where B’ = ||B’|| x ^B’ is the augmented bone vector from the above BA and BL operations.
其中 B' = ||B'|| x ^B' 是來自上述 BA 和 BL 操作的增強骨骼向量。

H-1 is the inverse hierarchical conversion to transform B0 back to a 3D pose [44, 22].
H-1 是將 B0 轉換回 3D 姿勢的逆層次轉換 [44, 22]。

By applying these operations, the augmentor can generate the augmented 3D pose X’ with more challenging pose, body size, view point and position from the original 3D pose X (Fig. 3).
通過應用這些操作，增強器可以從原始 3D 姿勢 X（圖 3）生成具有更具挑戰性的姿勢、身體尺寸、視點和位置的增強 3D 姿勢 X'。

The augmented pose is then re-projected to 2D with x’ =Π(X’), where Π : R3→ R2 denotes perspective projection [15] via the camera parameters from the original data.
然後使用 x' =Π(X') 將增強姿態重新投影到 2D，其中 Π : R3→ R2 表示通過來自原始數據的相機參數的透視投影 [15]。

The augmented 2D-3D pair {x’,X’} is then used for further training the pose estimator.
然後使用增強的 2D-3D 對 {x',X'} 進一步訓練姿勢估計器。

####
1
The hierarchical transformation converts the J joints of X into J -1 column vectors of B, each of which represents a line segment connecting two adjacent joints.
層次變換將 X 的 J 個關節轉換成 B 的 J -1 個列向量，每一個代表連接兩個相鄰關節的線段。

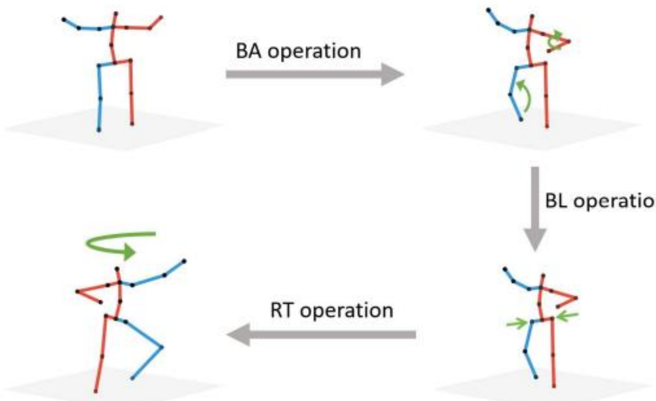


Figure 3: **Augmentation operations with PoseAug.** A source 3D pose is augmented by modifying its posture (via BA operation), body size (via BL operation) and view point and position (via RT operation).

Figure 3: Augmentation operations with PoseAug. A source 3D pose is augmented by modifying its posture (via BA operation), body size (via BL operation) and view point and position (via RT operation).
圖 3：使用 PoseAug 進行增強操作。 通過修改姿勢（通過 BA 操作）、身體尺寸（通過 BL 操作）以及視點和位置（通過 RT 操作）來增強源 3D 姿勢。

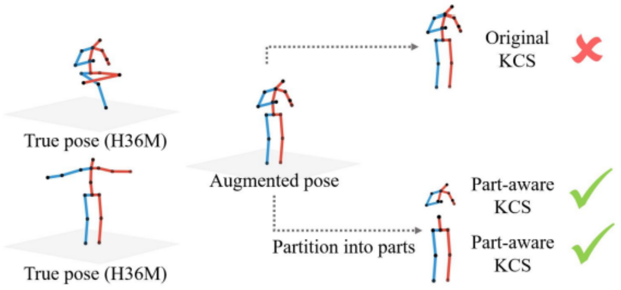


Figure 4: **Illustrations of the difference between original and part-aware KCS based discriminator.** Given a novel and valid augmented pose, the original KCS based discriminator would wrongly classify it as fake as it does not appear in source data (H36M), while the part-aware KCS based discriminator would recognize is as real and approve it, since it inspects local joint relations. It can be seen the part-aware KCS based discriminator can help the augmentor generate more diverse and plausible pose augmentation.

Figure 4: Illustrations of the difference between original and part-aware KCS based discriminator.
圖 4：原始和基於部分感知 KCS 的鑑別器之間差異的說明。

Given a novel and valid augmented pose, the original KCS based discriminator would wrongly classify it as fake as it does not appear in source data (H36M), while the part-aware KCS based discriminator would recognize is

as real and approve it, since it inspects local joint relations.
給定一個新穎且有效的增強姿勢，原始的基於 KCS 的鑑別器會錯誤地將其分類為假的，因為它沒有出現在源數據 (H36M) 中，而基於部分感知的 KCS 鑑別器會識別它是真實的並批准它，因為它 檢查地方聯合關係。

It can be seen the part-aware KCS based discriminator can help the augmentor generate more diverse and plausible pose augmentation.
可以看出，基於部分感知 KCS 的鑑別器可以幫助增強器生成更多樣化和更合理的姿勢增強。

Discriminator 鑑別器

Due to lacking priors in the augmentation procedure, the augmented poses may present implausible joint angles that violate the bio-mechanical structure [1], or unreasonable positions and view points.
由於增強過程中缺乏先驗，增強的姿勢可能會出現不合理的關節角度，違反生物力學結構 [1]，或不合理的位置和觀點。

Though such poses are indeed harder cases for the estimator, training on them would not benefit the model generalization ability.
雖然這樣的姿勢對於估計器來說確實是更難的情況，但對它們進行訓練不會有利於模型的泛化能力。

To ensure the plausibility of the augmented poses, we introduce a pose discriminator module to guide the augmentation.
為了確保增強姿勢的合理性，我們引入了一個姿勢鑑別器模塊來指導增強。

Specifically, the module consists of a 3D pose discriminator D3d for evaluating the joint angle plausibility and a 2D discriminator D2d for evaluating the body size, viewpoint and position plausibility.
具體來說，該模塊由用於評估關節角度可信度的 3D 姿勢鑑別器 D3d 和用於評估身體尺寸、視點和位置可信度的 2D 鑑別器 D2d 組成。

The key to the 3D pose discriminator design is to ensure the pose plausibility without sacrificing the diversity.
3D 姿勢鑑別器設計的關鍵是在不犧牲多樣性的情況下確保姿勢的合理性。

Inspired by the Kinematic Chain Space (KCS) [44], we design a part-aware KCS as input to the discriminator.
受運動鏈空間 (KCS) [44] 的啟發，我們設計了一個部分感知 KCS 作為鑑別器的輸入。

Instead of taking the whole body pose into consideration as in the original KCS, our part-aware KCS only focuses on local joint angle and thus enlarges the feasible region of the augmented pose, ensuring both plausibility and diversity (Fig. 4).
我們的部分感知 KCS 沒有像原始 KCS 那樣考慮整個身體姿勢，而是只關注局部關節角度，從而擴大了增強姿勢的可行區域，確保了合理性和多樣性（圖 4）。

Specifically, to compute the part-aware KCS of an input pose, either X or its augmentation X’, we convert the pose to its bone direction vector ^B as above and separate it into 5 parts (torso and left/right arm/leg) [1], denoted as ^Bi, i =1,…,5; respectively.
具體來說，為了計算輸入姿勢X或其增強X'的部分感知KCS，我們將姿勢轉換為其骨骼方向向量^B，並將其分成5個部分（軀乾和左/右臂/腿) [1], 記為^Bi, i =1,⋯,5; 分別。

We then calculate the following local joint angle matrix KCSilocal for the i-th part:
然後，我們為第i部分計算以下局部關節角度矩陣KCSilocal：

$$KCS^{i}_{local} = \hat{\boldsymbol{B}}_i^{\top} \hat{\boldsymbol{B}}_i, \tag{7}$$

which encapsulates the inter joint angle information within the i-th part.
它將關節間角度資訊封裝在第i部分中。

Based on the above local KCS representation, a 3D pose discriminator D3d is constructed which takes the KCSilocal as input and is trained for distinguishing the original and augmented 3D poses.
基於上述局部KCS表示，構建了一個3D姿態鑑別器D3d，它以KCSilocal作為輸入，並被訓練以區分原始和增強的3D姿態。

Besides the 3D discriminator, we also introduce a 2D discriminator to guide the augmentor to generate real body size, view points and positions.
除了3D鑑別器，我們還引入了2D鑑別器來引導增強器生成真實的身體尺寸、視點和位置。

As the 2D poses contain information such as view point (rotation), position (translation), and body size (bone length), the 2D discriminator can learn such information through adversarial training and guide the pose augmentor in generating realistic rotation R, translation t, and bone length ratio γ bl.
由於2D姿勢包含諸如視點（旋轉）、位置（平移）和身體大小（骨骼長度）等信息，2D鑑別器可以通過對抗性訓練學習這些資訊，並指導姿勢增強器生成逼真的旋轉R、平移t, 和骨骼長度比γbl。

Estimator
估算器

The pose estimator P estimates 3D poses from 2D poses.
姿態估計器P從2D姿態估計3D姿態。

We use the original and augmented 2D-3D pose pair {x,X} and {x’,X’} to train the pose estimator.
我們使用原始和增強的2D-3D姿勢對{x,X}和{x',X'}來訓練姿勢估計器。

The pose estimator contains a feature extractor to capture internal features from 2D poses, and a regression

layer to estimate the corresponding 3D poses.
姿勢估計器包含一個特徵提取器，用於從 2D 姿勢中捕獲內部特徵，以及一個回歸層來估計相應的 3D 姿勢。

Moreover, any existing effective estimator can be implemented in our PoseAug framework.
此外，任何現有的有效估計器都可以在我們的 PoseAug 框架中實現。

In Sec. 4.3, we conduct experiments to check robustness of PoseAug with different estimators, and the results show PoseAug can bring noticeable improvements on both source and cross-scenario datasets for all models.
在 4.3，我們用不同的估計器進行實驗來檢查 PoseAug 的魯棒性，結果表明 PoseAug 可以為所有模型的源數據集和跨場景數據集帶來顯著的改進。

3.4. Training Loss 訓練損失

Pose estimation loss 姿態估計損失

We adopt the mean squared errors (MSE) of the ground truth (GT) X and predicted poses ~X as the pose estimation loss, which is formulated as LP
我們採用 ground truth (GT) X 的均方誤差（MSE）和預測姿態~X 作為姿態估計損失，公式為 LP

$$\mathcal{L}_{\mathcal{P}} = \|\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_2^2. \tag{8}$$

We train the pose estimator using LP with both original and augmented pose pairs jointly, which can significantly boost performance for the challenging in-the-wild scenes.
我們使用 LP 與原始和增強的姿勢對聯合訓練姿勢估計器，這可以顯著提高具有挑戰性的野外場景的性能。

Pose augmentation loss 姿態增強損失

To facilitate model training, augmented data should harder than the original one, i.e.,LP(X’) > LP(X), but not too hard to hurt the training process.
為了促進模型訓練，增強數據應該比原始數據更難，即 LP(X’) > LP(X)，但不要太難傷害訓練過程。

A simple way to design the loss function is to let the difference between the pose estimation loss on augmented and original data within a proper range.
設計損失函數的一個簡單方法是讓增強數據和原始數據上的姿態估計損失之間的差異在一個合適的範圍內。

Inspired by [25, 21], we implement a controllable feedback loss as Lfb
受 [25, 21] 的啟發，我們實現了一個可控的反饋損失作為 Lfb

$$\mathcal{L}_{fb} = |1.0 - \exp[\mathcal{L}_{\mathcal{P}}(\boldsymbol{X}') - \beta\mathcal{L}_{\mathcal{P}}(\boldsymbol{X})]|, \qquad (9)$$

where β > 1 controls the difficulty level for the generated poses, making the value of LP(X’) stay within a certain range w.r.t. LP(X).
其中 β > 1 控制生成姿勢的難度級別，使 LP(X') 的值保持在一定範圍內 w.r.t. LP(X)。

During training, as the pose estimator becomes increasingly more powerful, we accordingly increase β value to generate more challenging augmentation data for training it.
在訓練期間，隨著姿態估計器變得越來越強大，我們相應地增加 β 值以生成更具挑戰性的增強數據來訓練它。

Additionally, to prevent extremely hard cases from causing training collapse, we introduce a rectified L2 loss for regularizing the augmentation parameters γ ba and γ bl:
此外，為了防止極端困難的情況導致訓練崩潰，我們引入了修正的 L2 損失來正則化增強參數 γ ba 和 γ bl：

$$\mathcal{L}_{reg}(\boldsymbol{\gamma}) = \begin{cases} 0, & \text{if } \bar{\boldsymbol{\gamma}} < threshold, \\ \|\boldsymbol{\gamma}\|^2, & \text{otherwise}, \end{cases} \qquad (10)$$

where γ denotes γ ba and γ bl, and - γ denotes the mean value over all of its elements.
其中 γ 表示 γ ba 和 γ bl，- γ 表示其所有元素的平均值。

Combining Eqn. (9) and Eqn. (10), the overall augmentation loss LA is formulated as LA
結合 Eqn. (9) 和 Eqn. (10)，整體增強損失 LA 被公式化為 LA

$$\mathcal{L}_{\mathcal{A}} = \mathcal{L}_{fb} + \mathcal{L}_{reg}. \qquad (11)$$

Pose discrimination loss 姿勢辨別損失

For the discrimination loss LD, we adopt the LS-GAN loss [25] for both 3D and 2D spaces:
對於鑑別損失 LD，我們對 3D 和 2D 空間均採用 LS-GAN 損失 [25]：

$$\begin{aligned} \mathcal{L}_{\mathcal{D}} = {} & \mathbb{E}[(D_{3d}(\boldsymbol{X}) - 1)^2] + \mathbb{E}[D_{3d}(\boldsymbol{X}')^2] \\ & + \mathbb{E}[(D_{2d}(\boldsymbol{x}) - 1)^2] + \mathbb{E}[D_{2d}(\boldsymbol{x}')^2], \end{aligned} \qquad (12)$$

where {x,X} and {x’,X’} denote the original (real) and the augmented (fake) pose pairs, respectively.

其中 {x,X} 和 {x',X'} 分別表示原始（真實）和增強（假）姿勢對。

End-to-end training strategy 端到端的培訓策略

With the differentiable design, the pose augmentor, discriminator and estimator can be jointly trained end-to-end.
通過可微分設計，可以對姿勢增強器、鑑別器和估計器進行端到端的聯合訓練。

We update them alternatively by minimizing losses Eqn. (11), Eqn. (12) and Eqn. (8).
我們通過最小化損失方程來交替更新它們。 Eqn. (11)、 Eqn. (12)、Eqn. (8)).

In addition, we first pre-train the pose estimator P before training the whole framework end-to-end, which ensures stable training and produces better performance.
此外，在端到端訓練整個框架之前，我們首先預訓練姿勢估計器 P，以確保穩定的訓練並產生更好的性能。

4. Experiments 實驗

We study four questions in experiments.
我們在實驗中研究了四個問題。
1) Is PoseAug able to improve performance of 3D pose estimator for both intra-dataset and cross-dataset scenarios?
1) PoseAug 是否能夠提高數據集內和跨數據集場景的 3D 姿態估計器的性能？

2) Is PoseAug effective at enhancing diversity of training data?
2) PoseAug 在增強訓練數據的多樣性方面是否有效？

3) Is PoseAug consistently effective for different pose estimators and cases with limited training data?
3) PoseAug 對於不同的姿勢估計器和訓練數據有限的情況是否始終有效？

4) How does each component of PoseAug take effect?
4）PoseAug 的各個組件是如何生效的？

We experiment on H36M, 3DHP and 3DPW.
我們在 H36M、3DHP 和 3DPW 上進行了實驗。

Throughout the experiments, unless otherwise stated we adopt single-frame version of VPose [33] as pose estimator.
在整個實驗過程中，除非另有說明，否則我們採用 VPose [33] 的單幀版本作為姿態估計器。

4.1. Datasets 數據集

Human3.6M (H36M) [16] Following previous works [26, 52], we train our model on subjects S1, 5, 6, 7, 8 of H36M and evaluate on subjects S9 and S11.
Human3.6M (H36M) [16] 根據之前的工作 [26, 52]，我們在 H36M 的科目 S1、5、6、7、8 上訓練我們的模型，並對科目 S9 和 S11 進行評估。

We use two evaluation metrics: Mean Per Joint Position Error (MPJPE) in millimeters and MPJPE over aligned predictions with GT 3D poses by a rigid transformation (PA-MPJPE).
我們使用兩個評估指標：以毫米為單位的平均每關節位置誤差 (MPJPE) 和通過剛性變換 (PA-MPJPE) 與 GT 3D 姿勢對齊預測的 MPJPE。

MPI-INF-3DHP (3DHP) [29] It is a large 3D pose dataset with 1.3 million frames, presenting more diverse motions than H36M.
MPI-INF-3DHP (3DHP) [29] 它是一個包含 130 萬幀的大型 3D 姿勢數據集，呈現出比 H36M 更多樣化的運動。

We use its test set to evaluate the model’s generalization ability to unseen environments, using metrics of MPJPE, Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC).
我們使用其測試集來評估模型對未知環境的泛化能力，使用 MPJPE、正確關鍵點百分比 (PCK) 和曲線下面積 (AUC) 等指標。

3DPW [43] It is an in-the-wild dataset with more complicated motions and scenes.
3DPW [43] 它是一個具有更複雜運動和場景的野外數據集。

To verify generalization of the proposed method to challenging in-the-wild scenarios, we use its test set for evaluation with PA-MPJPE as metric.
為了驗證所提出的方法對具有挑戰性的野外場景的泛化，我們使用其測試集以 PA-MPJPE 作為度量進行評估。

MPII [2] and LSP [17]
They are in-the-wild datasets with only 2D body joint annotations and used for qualitatively evaluating model generalization for unseen poses.
MPII [2] 和 **LSP [17]**
它們是只有 2D 身體關節註釋的野外數據集，用於定性評估模型泛化對看不見的姿勢。

4.2. Results 結果

Results on H36M

We compare PoseAug with state-of-theart methods [52, 38, 33, 30, 22] on H36M.
我們在 H36M 上將 PoseAug 與最先進的方法 [52, 38, 33, 30, 22] 進行了比較。

Similar to [22], we use 2D poses from HR-Net [40] as inputs.
與 [22] 類似，我們使用來自 HR-Net [40] 的 2D 姿勢作為輸入。

As shown in Table 1, our method outperforms SOTA methods [52, 38, 33, 30] by a large margin, indicating its effectiveness.
如表 1 所示，我們的方法大大優於 SOTA 方法 [52, 38, 33, 30]，表明其有效性。

Notably, compared with the previous best augmentation method [22], our PoseAug achieves lower MPJPE even though it uses external bone length data for data augmentation and nearly 3more data than ours for model training.
值得注意的是，與之前的最佳增強方法 [22] 相比，我們的 PoseAug 實現了更低的 MPJPE，即使它使用外部骨骼長度數據進行數據增強，並且比我們的模型訓練多出近 3 個數據。

This clearly verifies advantages of PoseAug’s online augmentation scheme—it can generate more diverse and informative data that better benefit model training.
這清楚地驗證了 PoseAug 在線增強方案的優勢——它可以生成更多樣化和信息量更大的數據，更好地有利於模型訓練。

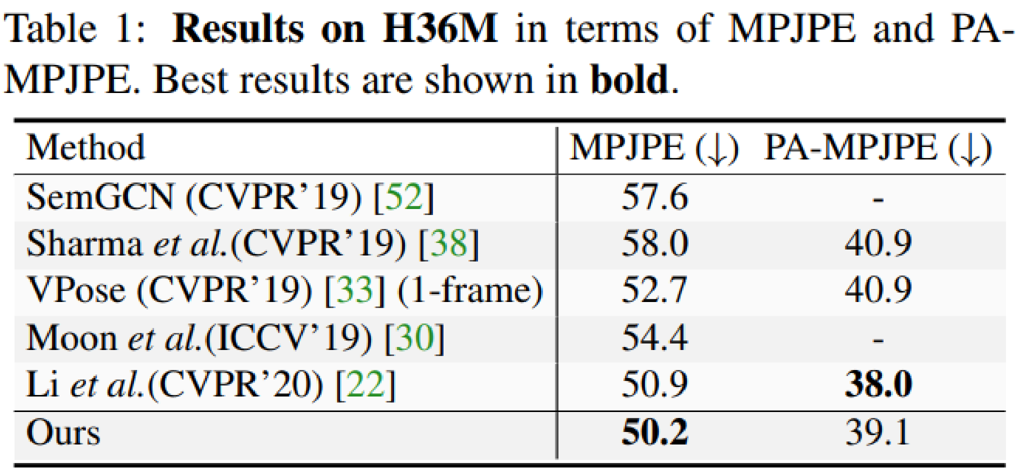


Table 1: **Results on H36M** in terms of MPJPE and PA-MPJPE. Best results are shown in **bold**.

Method	MPJPE (↓)	PA-MPJPE (↓)
SemGCN (CVPR’19) [52]	57.6	-
Sharma *et al.*(CVPR’19) [38]	58.0	40.9
VPose (CVPR’19) [33] (1-frame)	52.7	40.9
Moon *et al.*(ICCV’19) [30]	54.4	-
Li *et al.*(CVPR’20) [22]	50.9	**38.0**
Ours	**50.2**	39.1

Table 1: Results on H36M in terms of MPJPE and PAMPJPE.
表 1：H36M 在 MPJPE 和 PAMPJPE 方面的結果。

Best results are shown in bold.
最佳結果以粗體顯示。

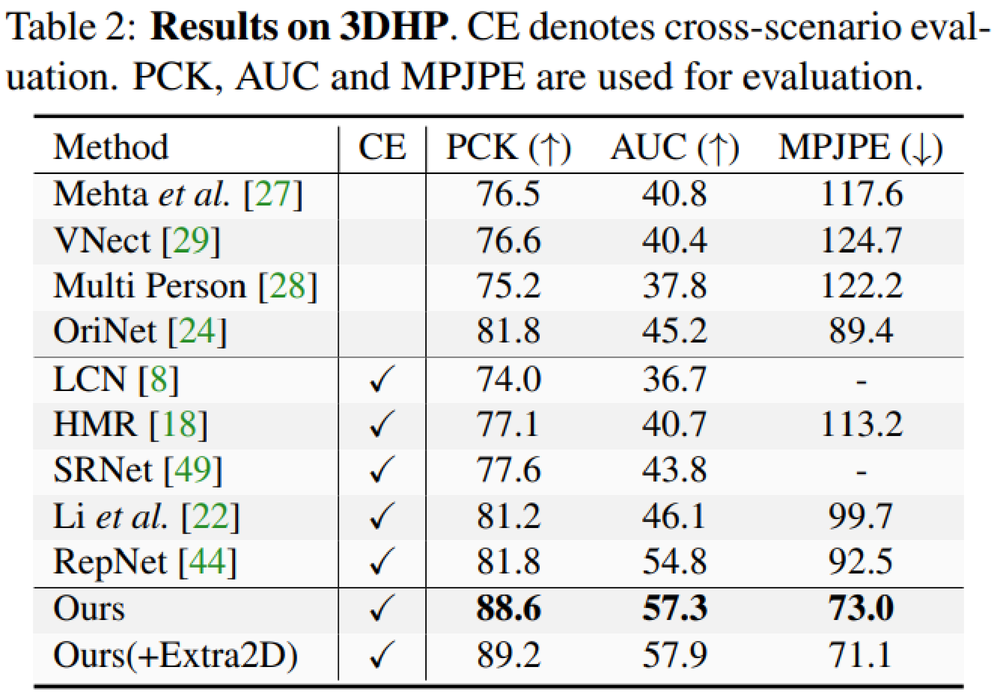


Table 2: **Results on 3DHP**. CE denotes cross-scenario evaluation. PCK, AUC and MPJPE are used for evaluation.

Method	CE	PCK (↑)	AUC (↑)	MPJPE (↓)
Mehta *et al.* [27]		76.5	40.8	117.6
VNect [29]		76.6	40.4	124.7
Multi Person [28]		75.2	37.8	122.2
OriNet [24]		81.8	45.2	89.4
LCN [8]	✓	74.0	36.7	-
HMR [18]	✓	77.1	40.7	113.2
SRNet [49]	✓	77.6	43.8	-
Li *et al.* [22]	✓	81.2	46.1	99.7
RepNet [44]	✓	81.8	54.8	92.5
Ours	✓	**88.6**	**57.3**	**73.0**
Ours(+Extra2D)	✓	89.2	57.9	71.1

Table 2: Results on 3DHP. CE denotes cross-scenario evaluation.
表 2：3DHP 的結果。 CE 表示跨場景評估。

PCK, AUC and MPJPE are used for evaluation.
PCK、AUC 和 MPJPE 用於評估。

Results on 3DHP (cross-scenario) 3DHP 上的結果（跨場景）

We then evaluate how PoseAug facilitates model generalization to cross-scenario datasets.
然後我們評估 PoseAug 如何促進模型泛化到跨場景數據集。

We compare PoseAug against various state-of-theart methods, including the latest one using offline data augmentation [22], the ones exploiting complex network architecture [8, 49] and weakly-supervised learning [18, 44] and the ones trained on the training set of 3DHP [27, 29, 28, 24].
我們將 PoseAug 與各種最先進的方法進行比較，包括最新的使用離線數據增強的方法 [22]、利用複雜網絡架構的方法 [8, 49] 和弱監督學習的方法 [18, 44] 以及經過訓練的方法 3DHP 的訓練集 [27, 29, 28, 24]。

From Table 2, we can observe our method achieves the best performance w.r.t. all the metrics, outperforming previous approaches by a large margin.
從表 2 中，我們可以觀察到我們的方法實現了最佳性能 w.r.t. 所有指標都大大優於以前的方法。

This verifies the effectiveness of PoseAug in improving model generalization to unseen scenarios.
這驗證了 PoseAug 在改進模型對未知場景的泛化方面的有效性。

Moreover, PoseAug can further improve the performance (from 73.0 to 71.1 in MPJPE) by using additional in-

the-wild 2D poses (MPII) to train the 2D discriminator.
此外，PoseAug 可以通過使用額外的野外 2D 姿勢 (MPII) 來訓練 2D 鑑別器，進一步提高性能（從 MPJPE 中的 73.0 到 71.1）。

This demonstrates its extensibility in leveraging extra 2D poses to further enrich the diversity of augmented data.
這證明了它在利用額外的 2D 姿勢以進一步豐富增強數據的多樣性方面的可擴展性。

Results on 3DPW (cross-scenario) 3DPW 的結果（跨場景）

We train four 3D pose estimators [26, 52, 3, 33] without and with PoseAug on H36M and compare their generalization performance on 3DPW.
我們在 H36M 上訓練了四個 3D 姿態估計器 [26, 52, 3, 33]，在沒有和有 PoseAug 的情況下，比較它們在 3DPW 上的泛化性能。

As shown in Table 4, on average, PoseAug brings 12:6% improvements for all the models.
如表 4 所示，平均而言，PoseAug 為所有模型帶來了 12:6% 的改進。

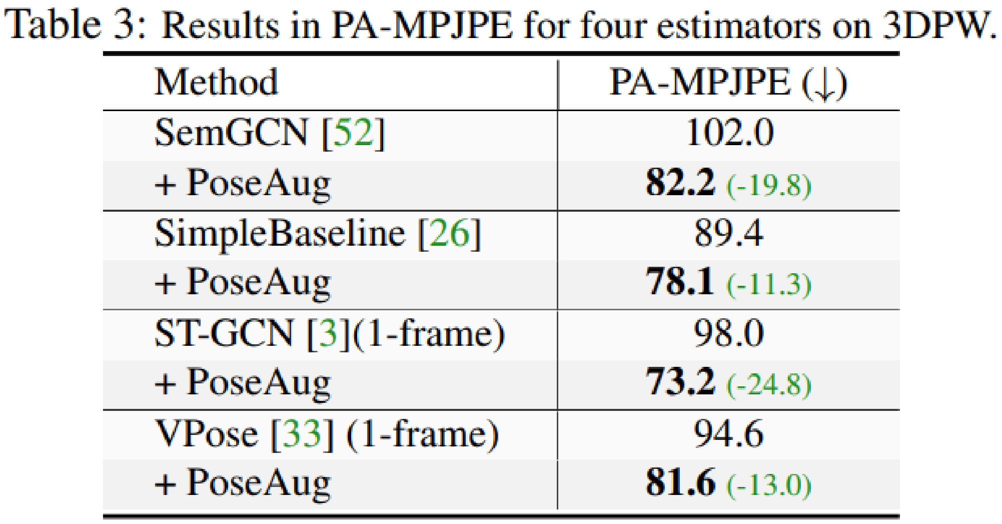


Table 3: Results in PA-MPJPE for four estimators on 3DPW.

Method	PA-MPJPE (↓)
SemGCN [52]	102.0
+ PoseAug	**82.2** (-19.8)
SimpleBaseline [26]	89.4
+ PoseAug	**78.1** (-11.3)
ST-GCN [3](1-frame)	98.0
+ PoseAug	**73.2** (-24.8)
VPose [33] (1-frame)	94.6
+ PoseAug	**81.6** (-13.0)

Table 3: Results in PA-MPJPE for four estimators on 3DPW.
表 3：PA-MPJPE 中 3DPW 上四個估計器的結果。

Qualitative results 定性結果

For subjective evaluation, we choose four challenging datasets, i.e., LSP, MPII, 3DHP and 3DPW, with large varieties of postures, body sizes, and view points between their data and the data from H36M.
對於主觀評估，我們選擇了四個具有挑戰性的數據集，即 LSP、MPII、3DHP 和 3DPW，它們的數據與來自 H36M 的數據之間的姿勢、體型和視角各不相同。

Results are shown in Fig. 5. We can see our method performs fairly well, even for those unseen difficult poses.

結果如圖 5 所示。我們可以看到我們的方法表現相當好，即使是那些看不見的困難姿勢。

4.3. Analysis on PoseAug - PoseAug 分析

Applicability to different estimators 適用於不同的估算器

Our PoseAug framework is generic and applicable to different 3D pose estimators.
我們的 PoseAug 框架是通用的，適用於不同的 3D 姿勢估計器。

To demonstrate this, we employ four representative 3D pose estimators as backbones:
為了證明這一點，我們使用了四個有代表性的 3D 姿態估計器作為主幹：

1) SemGCN [52], a graphbased 3D pose estimation network;
1) SemGCN [52]，一個基於圖的 3D 姿態估計網絡；

2) SimpleBaseline [26], an effective MLP-based network;
2) SimpleBaseline [26]，一個有效的基於 MLP 的網絡；

3) ST-GCN [3] (1-frame), a pioneer network that uses GCN-based architecture to encode global and local joint relations; and
3）ST-GCN [3]（1-frame），一種使用基於 GCN 架構的先驅網絡，對全局和局部聯合關係進行編碼； 和

4) VPose [33] (1-frame), a fully-convolutional network with SOTA performance.
4) VPose [33] (1-frame)，一個具有 SOTA 性能的全卷積網絡。

We train these models on the H36M dataset using 2D poses from four different 2D pose detectors, including CPN [7], DET [13], HR-Net [40] and groundtruth (GT).
我們使用來自四個不同 2D 姿勢檢測器的 2D 姿勢在 H36M 數據集上訓練這些模型，包括 CPN [7]、DET [13]、HR-Net [40] 和 groundtruth (GT)。

We evaluate these models on the test set of H36M and 3DHP w.r.t. MPJPE metric.
我們在 H36M 和 3DHP w.r.t. 的測試集上評估這些模型。 MPJPE 指標。

On H36M, we use the corresponding 2D poses for evaluation; while on 3DHP, we evaluate these models with GT 2D poses to filter out the influence of 2D pose detectors.
在 H36M 上，我們使用相應的 2D 姿勢進行評估； 而在 3DHP 上，我們使用 GT 2D 姿勢評估這些模型，以濾除 2D 姿勢檢測器的影響。

The results are shown in Table 4.
結果如表 4 所示。

We can see PoseAug brings clear improvements to all models on both H36M and more challenging 3DHP datasets.
我們可以看到 PoseAug 為 H36M 和更具挑戰性的 3DHP 數據集上的所有模型帶來了明顯的改進。

Notably, they obtain more than 13.1% average improvement on 3DHP when trained with PoseAug.
值得注意的是，當使用 PoseAug 進行訓練時，他們在 3DHP 上獲得了超過 13.1% 的平均改進。

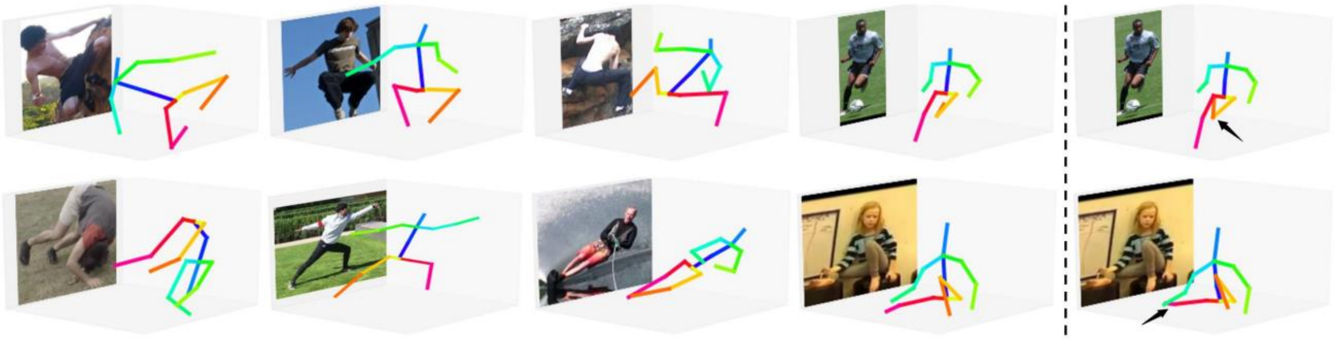


Figure 5: Example 3D pose estimations from LSP, MPII, 3DHP and 3DPW. Our results are shown in the left four columns. The rightmost column shows results of *Baseline*—VPose [33] trained w/o PoseAug. Errors are highlighted by black arrows.

Figure 5: Example 3D pose estimations from LSP, MPII, 3DHP and 3DPW.
圖 5：來自 LSP、MPII、3DHP 和 3DPW 的 3D 姿態估計示例。

Our results are shown in the left four columns.
我們的結果顯示在左邊的四列中。

The rightmost column shows results of Baseline—VPose [33] trained w/o PoseAug.
最右邊的列顯示了 Baseline-VPose [33] 在沒有 PoseAug 的情況下訓練的結果。

Errors are highlighted by black arrows.
錯誤由黑色箭頭突出顯示。

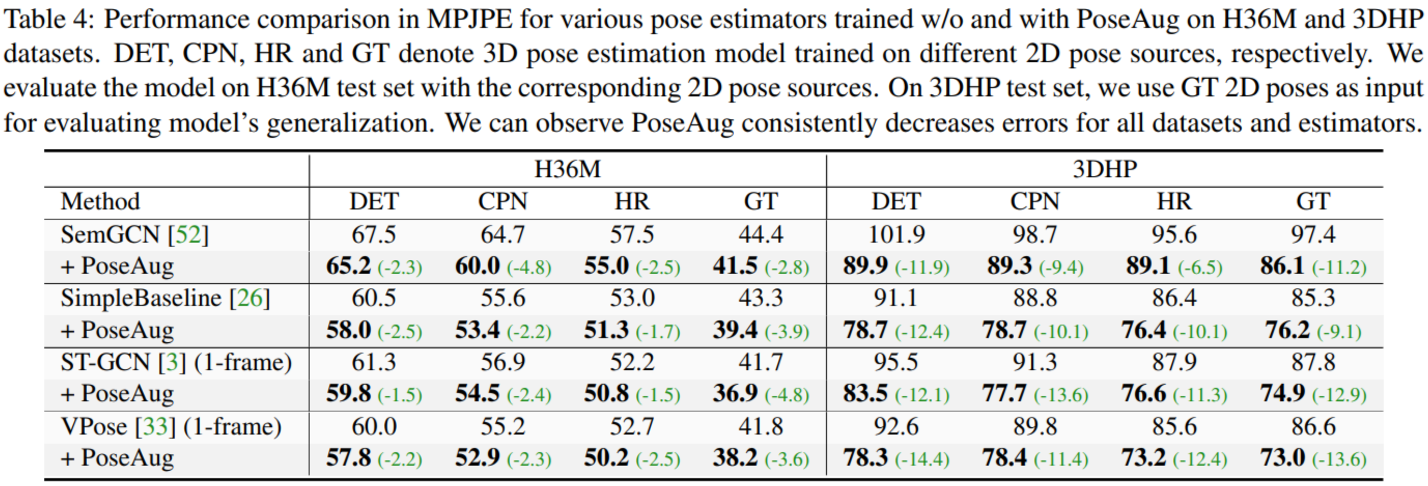


Table 4: Performance comparison in MPJPE for various pose estimators trained w/o and with PoseAug on H36M and 3DHP datasets. DET, CPN, HR and GT denote 3D pose estimation model trained on different 2D pose sources, respectively. We evaluate the model on H36M test set with the corresponding 2D pose sources. On 3DHP test set, we use GT 2D poses as input for evaluating model's generalization. We can observe PoseAug consistently decreases errors for all datasets and estimators.

	H36M				3DHP			
Method	DET	CPN	HR	GT	DET	CPN	HR	GT
SemGCN [52]	67.5	64.7	57.5	44.4	101.9	98.7	95.6	97.4
+ PoseAug	**65.2** (-2.3)	**60.0** (-4.8)	**55.0** (-2.5)	**41.5** (-2.8)	**89.9** (-11.9)	**89.3** (-9.4)	**89.1** (-6.5)	**86.1** (-11.2)
SimpleBaseline [26]	60.5	55.6	53.0	43.3	91.1	88.8	86.4	85.3
+ PoseAug	**58.0** (-2.5)	**53.4** (-2.2)	**51.3** (-1.7)	**39.4** (-3.9)	**78.7** (-12.4)	**78.7** (-10.1)	**76.4** (-10.1)	**76.2** (-9.1)
ST-GCN [3] (1-frame)	61.3	56.9	52.2	41.7	95.5	91.3	87.9	87.8
+ PoseAug	**59.8** (-1.5)	**54.5** (-2.4)	**50.8** (-1.5)	**36.9** (-4.8)	**83.5** (-12.1)	**77.7** (-13.6)	**76.6** (-11.3)	**74.9** (-12.9)
VPose [33] (1-frame)	60.0	55.2	52.7	41.8	92.6	89.8	85.6	86.6
+ PoseAug	**57.8** (-2.2)	**52.9** (-2.3)	**50.2** (-2.5)	**38.2** (-3.6)	**78.3** (-14.4)	**78.4** (-11.4)	**73.2** (-12.4)	**73.0** (-13.6)

Table 4: Performance comparison in MPJPE for various pose estimators trained w/o and with PoseAug on H36M and 3DHP datasets.

表 4：MPJPE 中在 H36M 和 3DHP 數據集上使用 PoseAug 訓練的各種姿勢估計器的性能比較。

DET, CPN, HR and GT denote 3D pose estimation model trained on different 2D pose sources, respectively.
DET、CPN、HR 和 GT 分別表示在不同 2D 位姿源上訓練的 3D 位姿估計模型。

We evaluate the model on H36M test set with the corresponding 2D pose sources.
我們使用相應的 2D 位姿源在 H36M 測試集上評估模型。

On 3DHP test set, we use GT 2D poses as input for evaluating model’s generalization.
在 3DHP 測試集上，我們使用 GT 2D 姿勢作為評估模型泛化的輸入。

We can observe PoseAug consistently decreases errors for all datasets and estimators.
我們可以觀察到 PoseAug 持續減少所有數據集和估計器的錯誤。

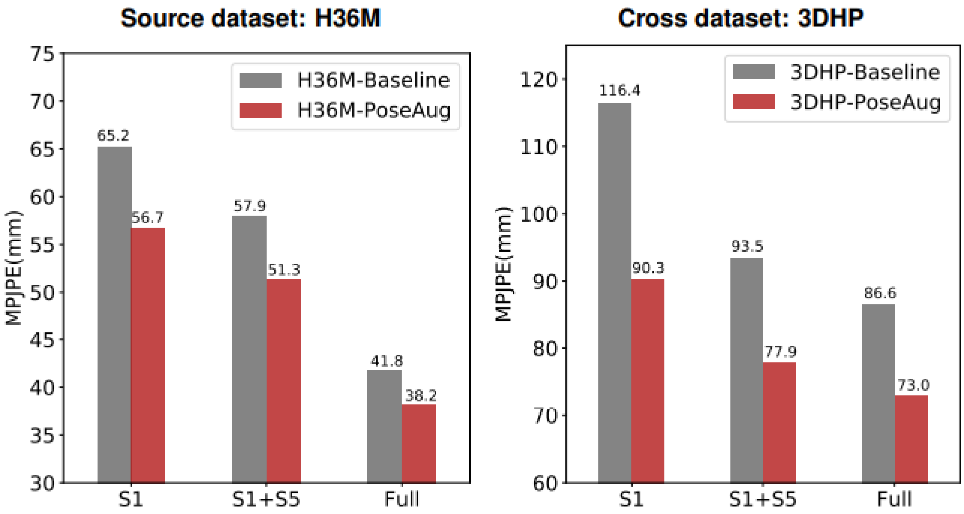


Figure 6: Ablation study on limited data setup. We report MPJPE for evaluation. Best viewed in color.

Figure 6: Ablation study on limited data setup.
圖 6：對有限數據設置的消融研究。

We report MPJPE for evaluation.
我們報告 MPJPE 進行評估。

Best viewed in color.
最好用彩色觀看。

Effectiveness for limited training data cases 有限訓練數據案例的有效性

3D pose annotations are expensive to collect, making limited training data a common challenge.
3D 姿勢註釋的收集成本很高，這使得有限的訓練數據成為一個普遍的挑戰。

To demonstrate the effectiveness of our method on addressing such cases, we use pose data from H36M S1 and S1+S5 for model training which only contain 16% and 41% training samples, respectively.
為了證明我們的方法在解決此類情況方面的有效性，我們使用來自 H36M S1 和 S1+S5 的姿態數據進行模型訓練，它們分別只包含 16% 和 41% 的訓練樣本。

The results in Fig. 6 show PoseAug consistently improves model performance with varying amounts of training data, on both H36M and 3DHP.
圖 6 中的結果顯示，在 H36M 和 3DHP 上，PoseAug 使用不同數量的訓練數據持續提高模型性能。

Meanwhile, the improvements brought by our method are more significant for cases with less training data (e.g., MPJPE in 3DHP, S1: 116.4!90.3, Full: 86.6 ! 73.0).
同時，對於訓練數據較少的情況（例如，3DHP 中的 MPJPE，S1：116.4！90.3，Full：86.6！73.0），我們的方法帶來的改進更為顯著。

Moreover, in cross-scenario generalization, our method trained with only S1 achieves the comparable result (MPJPE: 90.3) to baseline trained using full dataset (MPJPE: 86.6), and our method trained with S1+S5 can outperform baseline trained using full dataset by a large margin (77.9 vs 86.6 in MPJPE).
此外，在跨場景泛化中，我們僅使用 S1 訓練的方法獲得了與使用完整數據集訓練的基線（MPJPE：86.6）相當的結果（MPJPE：86.6），我們使用 S1+S5 訓練的方法可以優於使用完整數據集訓練的基線大幅提升（MPJPE 為 77.9 對 86.6）。

Analysis on the augmentor 增強器分析

We then check the effectiveness of each module in augmentor.
然後我們檢查增強器中每個模塊的有效性。

Table 5 summarizes the results.
表 5 總結了結果。

By gradually adding the BA, RT and BL operations, the pose estimation error can be monotonically decreased from 41.8/86.6 to 38.8/73.5 (on H36M/3DHP).
通過逐漸加入 BA、RT 和 BL 操作，位姿估計誤差可以從 41.8/86.6 單調降低到 38.8/73.5（在 H36M/3DHP 上）。

Moreover, incorporating the error feedback guidance can further improve performance to 38.2 for H36M and 73.0 for 3DHP.
此外，結合錯誤反饋指導可以進一步將 H36M 的性能提高到 38.2，將 3DHP 的性能提高到 73.0。

These verify the effectiveness of each module of the augmentor in producing more effective augmented samples.

這些驗證了增強器的每個模塊在生成更有效的增強樣本方面的有效性。

Among these modules, RT contributes the most to cross-scenario performance, which implies it benefits data diversity most effectively.
在這些模塊中，RT 對跨場景性能的貢獻最大，這意味著它最有效地有利於數據多樣性。

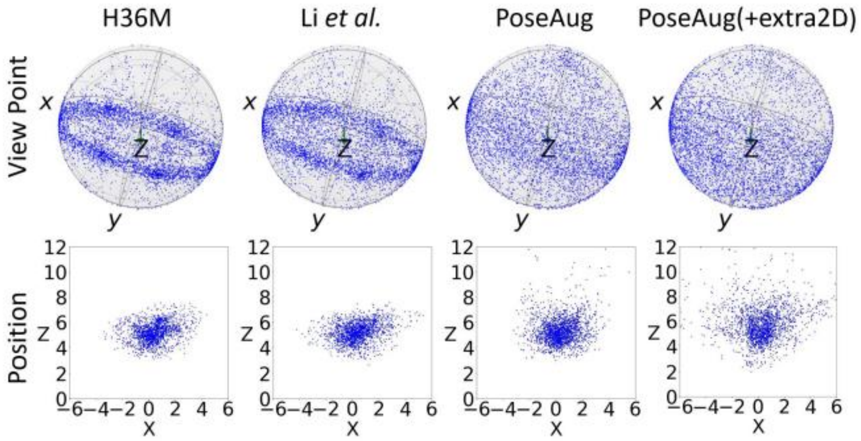


Figure 7: Distribution on view point (top row) and position (bottom row) for original data H36M, and augmented data from Li *et al.* [22], PoseAug (3rd column) and PoseAug with extra 2D poses. This distribution shows PoseAug significantly improves diversity of view point and position.

Figure 7: Distribution on view point (top row) and position (bottom row) for original data H36M, and augmented data from Li et al. [22], PoseAug (3rd column) and PoseAug with extra 2D poses.
圖 7：原始數據 H36M 的視點（頂行）和位置（底行）分佈，以及來自 Li et al. [22]、PoseAug（第三列）和 PoseAug 的具有額外 2D 姿勢的增強數據。

This distribution shows PoseAug significantly improves diversity of view point and position.
這種分佈表明 PoseAug 顯著提高了視點和位置的多樣性。

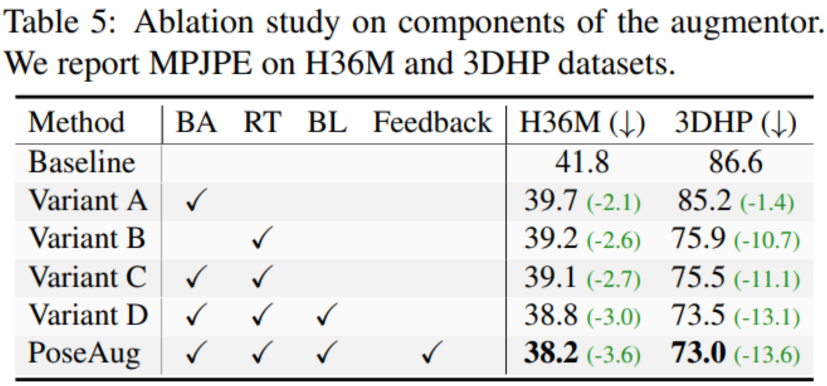


Table 5: Ablation study on components of the augmentor. We report MPJPE on H36M and 3DHP datasets.

Method	BA	RT	BL	Feedback	H36M (↓)	3DHP (↓)
Baseline					41.8	86.6
Variant A	✓				39.7 (-2.1)	85.2 (-1.4)
Variant B		✓			39.2 (-2.6)	75.9 (-10.7)
Variant C	✓	✓			39.1 (-2.7)	75.5 (-11.1)
Variant D	✓	✓	✓		38.8 (-3.0)	73.5 (-13.1)
PoseAug	✓	✓	✓	✓	**38.2** (-3.6)	**73.0** (-13.6)

Table 5: Ablation study on components of the augmentor.
表 5：增強器組件的消融研究。

We report MPJPE on H36M and 3DHP datasets.
我們在 H36M 和 3DHP 數據集上報告 MPJPE。

Analysis on diversity improvement 多樣性提升分析

To demonstrate effectiveness of PoseAug in enhancing data diversity, considering RT operation which augments the view point and position contributes the most to cross-scenario performance, as shown in Table 5, we make diversity analysis on view point and position distribution.
為了證明 PoseAug 在增強數據多樣性方面的有效性，考慮到增強視點和位置的 RT 操作對跨場景性能的貢獻最大，如表 5 所示，我們對視點和位置分佈進行了多樣性分析。

Fig. 7 demonstrates the distributions of view point and position of H36M and the augmented data generated by Li et al. [22] and our method.
圖 7 展示了 H36M 的視點和位置分佈以及 Li 等人生成的增強數據。Li et al. [22] 和我們的方法。

For H36M data, one can observe their view points concentrate near to the xz-plane with a limited diversity along the y-axis; and their positions form a small and concentrated cluster, also showing a limited diversity.
對於 H36M 數據，可以觀察到它們的視點集中在 xz 平面附近，沿 y 軸的多樣性有限； 它們的位置形成一個小而集中的集群，也表現出有限的多樣性。

This explains why the model trained on H36M hardly generalizes to in-the-wild scenarios.
這解釋了為什麼在 H36M 上訓練的模型很難推廣到野外場景。

Similarly, we observe small divergence for the view point and position distribution of augmented data from Li et al. [22].
類似地，我們觀察到來自 Li et al. [22] 的增強數據的視點和位置分佈的小差異。

This implies the diversity improvement from the handcrafted rule is limited.
這意味著手工規則的多樣性改進是有限的。

Comparably, our PoseAug can offer more plausible view points and positions using the learnable augmentor, with a much greater diversity.
相比之下，我們的 PoseAug 可以使用可學習的增強器提供更合理的觀點和位置，並且具有更大的多樣性。

In addition, the diversity on human positions can be further improved with extra 2D poses, which also explains its resulted improved generalization ability in Table 2.
此外，通過額外的 2D 姿勢可以進一步提高人體位置的多樣性，這也解釋了其在表 2 中提高泛化能力的原因。

Analysis on the discriminator 判別器分析

We here demonstrate the effectiveness of plausibility guidance from the 2D and 3D discriminators.
對鑑別器的分析我們在這裡展示了來自 2D 和 3D 鑑別器的合理性指導的有效性。

Table 6 summarizes the results.
表 6 總結了結果。

By adding one of the 2D or 3D discriminators, the performance of baseline can be boosted by 2.2/5.8 and 2.2/7.0 on H36M/3DHP, respectively.
通過添加 2D 或 3D 鑑別器之一，基線在 H36M/3DHP 上的性能可以分別提高 2.2/5.8 和 2.2/7.0。

Including both discriminators into PoseAug training can further boost the performance by 3.6/13.6 on H36M/3DHP, which clearly verify the effectiveness of both discriminators and also the importance of plausibility (in augmented poses) for estimator performance.
在 PoseAug 訓練中包含兩個判別器可以進一步將 H36M/3DHP 的性能提高 3.6/13.6，這清楚地驗證了兩個判別器的有效性以及合理性（在增強姿勢中）對估計器性能的重要性。

Analysis on part-aware KCS (PA-KCS) 部分感知 KCS (PA-KCS) 分析

To verify its effectiveness, we replace it in PoseAug with KCS [44].
為了驗證其有效性，我們在 PoseAug 中將其替換為 KCS [44]。

Table 7 summarizes the results. PA-KCS clearly outperforms KCS on both 3DHP and 3DPW.
表 7 總結了結果。 PA-KCS 在 3DHP 和 3DPW 上明顯優於 KCS。

This verifies our PA-KCS provides better guidance than KCS during training.
這驗證了我們的 PA-KCS 在培訓期間提供了比 KCS 更好的指導。

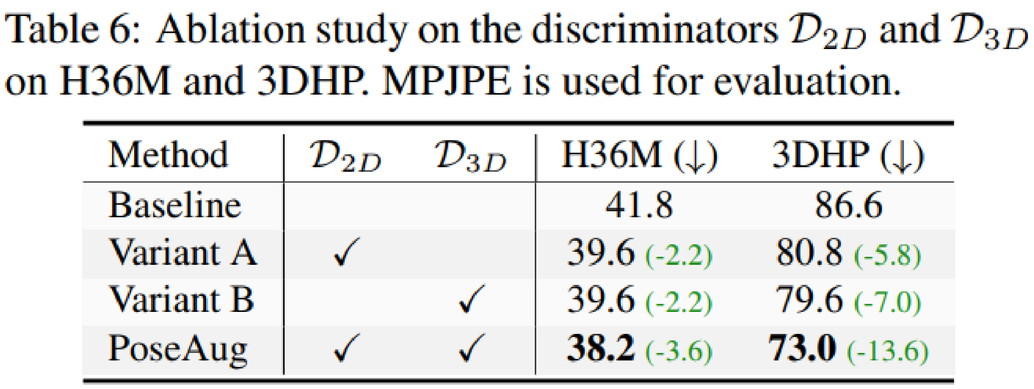


Table 6: Ablation study on the discriminators $\mathcal{D}_{2D}$ and $\mathcal{D}_{3D}$ on H36M and 3DHP. MPJPE is used for evaluation.

Method	$\mathcal{D}_{2D}$	$\mathcal{D}_{3D}$	H36M (↓)	3DHP (↓)
Baseline			41.8	86.6
Variant A	✓		39.6 (-2.2)	80.8 (-5.8)
Variant B		✓	39.6 (-2.2)	79.6 (-7.0)
PoseAug	✓	✓	**38.2** (-3.6)	**73.0** (-13.6)

Table 6: Ablation study on the discriminators D2D and D3D on H36M and 3DHP.
表 6：對 H36M 和 3DHP 上的鑑別器 D2D 和 D3D 的消融研究。

MPJPE is used for evaluation.

MPJPE 用於評估。

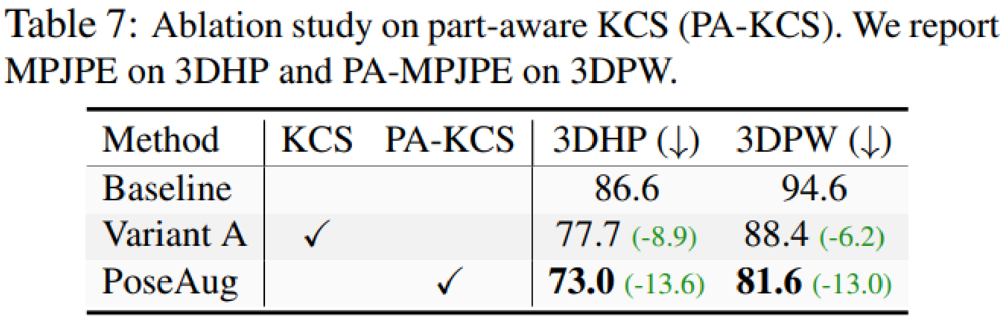


Table 7: Ablation study on part-aware KCS (PA-KCS). We report MPJPE on 3DHP and PA-MPJPE on 3DPW.

Method	KCS	PA-KCS	3DHP (↓)	3DPW (↓)
Baseline			86.6	94.6
Variant A	✓		77.7 (-8.9)	88.4 (-6.2)
PoseAug		✓	**73.0** (-13.6)	**81.6** (-13.0)

Table 7: Ablation study on part-aware KCS (PA-KCS).
表 7：部分感知 KCS (PA-KCS) 的消融研究。

We report MPJPE on 3DHP and PA-MPJPE on 3DPW.
我們在 3DHP 上報告 MPJPE，在 3DPW 上報告 PA-MPJPE。

5. Conclusion 結論

In this paper, we develop an auto-augmentation framework, PoseAug, that learns to enrich the diversity of training data and improves performance of the trained pose estimation models.
在本文中，我們開發了一個自動增強框架 PoseAug，它學習豐富訓練數據的多樣性並提高訓練姿勢估計模型的性能。

The PoseAug effectively integrates three components including the augmentor, estimator and discriminator and makes them fully interacted with each other.
PoseAug 有效地集成了增強器、估計器和鑑別器三個組件，並使它們彼此充分交互。

Specifically, the augmentor is designed to be differentiable and thus can learn to change major geometry factors of the 2D-3D pose pair to suit the estimator better by taking its training error as feedback.
具體來說，增強器被設計為可微分的，因此可以通過將其訓練誤差作為反饋來學習改變 2D-3D 姿勢對的主要幾何因素以更好地適應估計器。

The discriminator can ensure the plausibility of augmented data based on a novel part-aware KCS representation.
鑑別器可以確保基於新的部分感知 KCS 表示的增強數據的合理性。

Extensive experiments justify PoseAug can augment diverse and informative data to boost estimation performance for various 3D pose estimators.
大量實驗證明 PoseAug 可以增加多樣化和信息豐富的數據，以提高各種 3D 姿態估計器的估計性能。

Acknowledgement 致謝

This research was partially supported by AISG-100E-2019-035, MOE2017-T2-2-151, NUS ECRA FY17 P08 and CRP20-2017-0006.
這項研究得到了AISG-100E-2019-035、MOE2017-T2-2-151、NUS ECRA FY17 P08和CRP20-2017-0006的部分支持。

JZ would like to acknowledge the support of NVIDIA AI Tech Center (NVAITC) to this research project.
JZ感謝NVIDIA AI技術中心(NVAITC)對本研究項目的支持。