Human Motion Forecasting in Dynamic Domain Shifts: A Homeostatic Continual Test-time Adaptation Framework

Qiongjie Cui[1], Huaijiang Sun[1⋆]
Weiqing Li[1], Jianfeng Lu[1], and Bin Li[2]

[1] Nanjing University of Science and Technology, Nanjing, China
[2] Tianjin AiForward Science and Technology Co., Ltd., China
cuiqiongjie@126.com sunhuaijiang@njust.edu.cn

Abstract. Existing motion forecasting models, while making progress, struggle to bridge the gap between the source and target domains. Recent solutions often rely on an unrealistic assumption that the target domain remains stationary. Due to the ever-changing environment, however, the real-world test distribution may experience ongoing/continual shifts over time, leading to catastrophic forgetting and error accumulation when adapting to evolving domains. To solve these challenges, this work introduces HoCoTTA, a framework for homeostatic continual test-time adaptation. It aligns with the knowledge distillation and parameter isolation paradigm, enabling the identification of domain-invariant and domain-specific knowledge, where the former is shared (to be retained) in continual TTA across domains, while the latter needs to be updated. Specifically, we propose a multi-domain homeostasis assessment to estimate the uncertainty of the current model parameter when faced with novel-domain samples. Then, the Fisher information matrix is computed to measure the parameter sensitivity, with larger indicating the domain-sensitive parameter, and vice versa. Moreover, we propose an isolated parameter optimization strategy to update those domain-specific parameters to adapt to the new-domain, while preserving the invariant ones. In our experimental result, HoCoTTA outperforms the state-of-the-art approaches on several benchmarks, especially excelling in addressing continuous domain drifts, achieving a large improvement.

Keywords: Deterministic Human Motion Forecasting · Domain Generalization · Test-time Domain Adaptation · Parameter Isolation Update

1 Introduction

Given a series of historical poses, human motion forecasting aims to forecast the future pose as close as possible to the actual one, which has great potential in autonomous driving and human-robot cooperation [14, 35, 43, 50, 54, 60].

⋆ Corresponding author

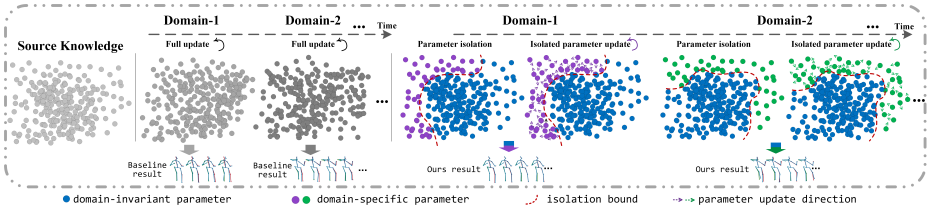

Fig. 1: In contrast to the standard TTA, which updates the full model at test-time, our HoCoTTA is able to identify the domain-invariant and domain-sensitive parameter, and retain the former while adapting the latter to the new domain. It therefore alleviates the catastrophic forgetting and adapt to the continually-changing target distributions, achieving a closer prediction result (green-red skeleton) to the ground truth (blue skeleton) against the state-of-the-art baseline [12].

Recently, this compelling topic has gained increased attention, emerging as a promising research direction [2,4,10,23,30,47,71]. Deep end-to-end networks have been sought-after to tackle this issue, which typically default to the training and test data are under the same distribution [25, 57, 64]. However, this assumption is often violated in real-world setups due to a large distribution gap between the source and target domains, such as the presence of novel motion patterns during testing.

Researchers attempt to use test-time adaptation (TTA) to address this issue [9, 11, 12, 26, 59, 66, 74], which considers the target domain to be stationary, and is expected to update the source-trained model for adaptation at test-time. Despite encouraging results, they usually face a dynamically-evolving environment over time, where the distribution of target motion sequences is not static but continuously changing. Stated in a different way, for sequentially arriving motion samples in the real deployment scenario, the continuous distribution shifts are inevitable. We notice that the existing approaches [11, 12] are sub-optimal in adapting to the continually-changing target distributions, leading to a large prediction error, which restricts their practical applications.

To address this issue, this work proposes a novel homeostatic continual test-time adaptation (HoCoTTA) framework, which is able to adapt to the continually-changing target distributions, and alleviate the catastrophic forgetting and error accumulation. Following the knowledge distillation paradigm [12,33,40,69], our HoCoTTA involves a teacher $\boldsymbol{\theta}_T$ and a student network $\boldsymbol{\theta}_S$ with identical architectures, selectively derived from existing or newly-designed motion forecasting networks. In contrast to the student, the teacher is preceded by a multi-domain augmenter $\mathcal{A}_\phi$, which is trained to generate novel-domain augmentations for each sample $\boldsymbol{X}^{(0)}$. Then, both augmented and original samples are fed into the teacher to produce the corresponding intermediate predictions in parallel. Comparing these predictions with the original one, the uncertainty matrix is computed to gauge the model's confidence level in its predictions under distribution shifts. Furthermore, the Fisher information matrix [39, 62] is computed to measure the parameter sensitivity, with larger indicating the domain-sensitive

parameter, and vice versa. We therefore leverage the τ-quantile to isolate the domain-sensitive and domain-invariant parameters. Then, we propose an isolated parameter optimization strategy to update those domain-specific parameters to adapt to the new-domain, while preserving the invariant ones, as shown in Fig.1. This strategy helps alleviate excessive forgetting of previously-learned information, and avoid the errors accumulation, when adapting to continuously changing target distributions. Therefore, the better prediction results are achieved.

Our main contributions are summarized as follows: 1) We propose the homeostatic continual test-time adaptation (HoCoTTA) framework to address the realism of the non-stationary target distribution in test motion sequences. 2) We propose to access the model's uncertainty, allowing the isolation of the domain-sensitive and domain-invariant parameters, which is able to alleviate the catastrophic forgetting. 3) Experiments on several benchmarks show that our HoCoTTA outperforms the state-of-the-art approaches, especially excelling in addressing continuous domain drifts, achieving a significant improvement.

2 Related Work

Human Motion Forecasting. Recent progresses have revealed the huge potential of deep learning-based approaches for deterministic motion prediction, establishing them as the prevailing technique [4, 14, 45, 56, 76]. Earlier works typically rely on the variants of RNNs to formulate this task as a sequence regression problem, aiming to mapping the past observed sequence to the future ones [17,21,50]. Despite capturing temporal correlation, RNNs suffer from the static predicted pose, and significant discontinuity. Forward networks are therefore introduced to address this issue, especially graph convolution networks with a higher interpretability, which are able to capture the semantic connectivity of 3D human skeletons, and gradually becoming the current dominant [13, 14, 34, 35, 48, 71].

Despite great progress most approaches assume the source and target domains are identical, which is a harsh condition in practical applications, where the target domain often differs from the source one. To address it, recent TTA works [11, 12] are proposed, and adapt the pre-trained model to unseen target domains at test-time. However, these methods still rely on an unrealistic assumption that the target domain remains stationary. In contrast, this work embraces a more realistic scenario, in which the target distributions is not only different from the source one, but also undergoes continuous changes over time.

Unsupervised Domain Adaptation (UDA) is a widely-recognized method in computer vision aimed at training models capable of generalizing to unseen domains [15, 19, 29, 73]. In the UDA context, diverse augmentation strategies are employed on both source and target domain data to mitigate the distribution gap, encompassing color/light shifts, rotations, and cropping [16, 36, 67], all designed to transfer the domain-invariant knowledge gained from these augmentations to aid the better adaptation to the target domain [51, 52]. In contrast to conventional UDA, our approach assesses the prediction uncertainty resulting from multiple augmentations in the new domains, allowing us to identify and

update domain-sensitive parameters. Moreover, we also consider data privacy, ensuring that source domain data is not required during test-time adaptation.

Test-time Adaptation (TTA) is an alternative, situated within the source-free domain adaptation paradigm [27, 31, 41, 42, 72, 74, 75]. It permits the fine-tuning of pre-trained models during inference, adapting them to specific test samples and facilitating more customized final decisions. TENT [68] commences with a source pre-trained model and exclusively updates the BN parameters, which is accomplished by minimizing the entropy in test predictions. AdaContrast [8] introduces weak and strong augmentation to enable the contrastive learning that refines the pseudo labeling of the target domain. [65] proposes an adversarial augmentation module to further enhance the knowledge distillation. We note that, whereas, the standard TTA needs to access to the full test data, which is often unrealistic for online applications of human motion forecasting.

Continual TTA. Typical TTAs often overlook the changing target distributions over time, leading to the proposal of continual test-time adaptation (CTTA) to address this issue [6, 18, 18, 20, 58, 61, 69]. In particular, [69] introduces the CoTTA model, a teacher-student framework, and incorporates a random restoration strategy to mitigate the catastrophic forgetting. Adhering to the teacher-student paradigm, [6] proposes a probabilistic version of CTTA (called PETAL), which regularizes the model update at inference time to prevent model drift. [58] tracks the progress of continual learning [5,44], and proposes a pruning-based approach to investigate the domain-specific capacity. Our method follows the above progress and breaks it further, and proposes to accurately identify the domain-sensitive and domain-invariant parameters, and update them separately, which is able to alleviate the catastrophic forgetting of past domains and error accumulation in continual adaptation.

3 Proposed Approach

3.1 Problem Formulation

Suppose $\boldsymbol{X}_{1:T} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T] \in \mathcal{X}$ is the past observed poses of a person, and human motion forecasting aims to generate the future sequence $\boldsymbol{Y}_{1:\Delta T} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{\Delta T}] \in \mathcal{Y}$, with each frame being the 3D coordinates of N joints. Most existing models [14, 45, 50, 76] are trained on the source domain $\mathcal{S} = \{(\boldsymbol{X}, \boldsymbol{Y})^{(i)}\}_{i=1}^{|\mathcal{S}|}$, w.r.t. training data size $|\mathcal{S}|$, and apply it to the target domain $\mathcal{T}$, under the assumption of $\mathcal{S} = \mathcal{T}$. Recent advances [11, 12] in motion prediction challenge this assumption, suggesting that the target domain differs from the source one: $\mathcal{S} \neq \mathcal{T}$, though still assuming an idealized setting: the target domains are static, *i.e.*, $\mathcal{T}_1 = \mathcal{T}_2 = ... = \mathcal{T}_n$.

Considering environmental changes and individual behavioral patterns, our work introduces a novel and more realistic scenario, in which target domains and source one is not same: $\mathcal{S} \neq \mathcal{T}_{1:n}$, and target domains are subject to ongoing change: $\mathcal{T}_1 \neq \mathcal{T}_2 \neq ... \neq \mathcal{T}_n$, $n > 1$. Moreover, we propose a homeostatic continual test-time adaptation (HoCoTTA) to solve it.

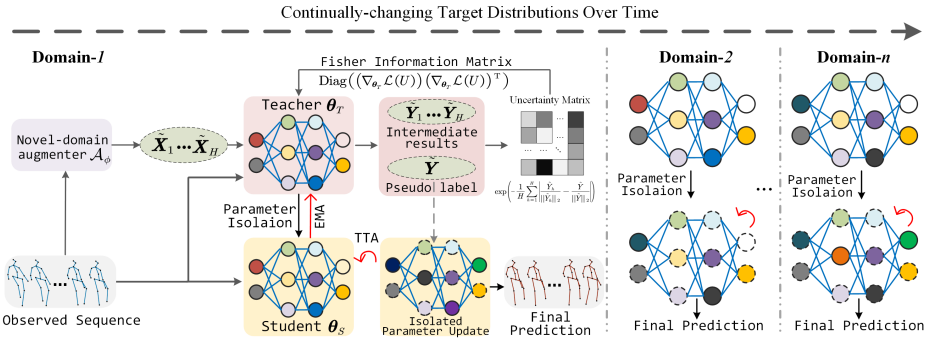

Fig. 2: Homeostatic continual test-time adaptation (HoCoTTA). It is first trained on the source data to achieve a base model, and then adapted to n target domains. The network is composed of a teacher $\boldsymbol{\theta}_T$ and a student $\boldsymbol{\theta}_S$, with identical architectures. The teacher is preceded by a trained novel-domain augmenter $\mathcal{A}_\phi$, which is able to generate diverse augmentations and estimate the model's uncertainty for new-domains. Then, Fisher information matrix is computed to measure the parameter sensitivity, where the larger indicates the domain-sensitive parameter, and vice versa. We leverage the τ-quantile to separately isolate and update the domain-sensitive and domain-invariant parameters. It alleviate the catastrophic forgetting and error accumulation in adapting to continually-changing target samples, and therefore brings the better prediction.

Our HoCoTTA aligns with the well-known knowledge distillation paradigm, incorporating a teacher $\boldsymbol{\theta}_T$ and a student network $\boldsymbol{\theta}_S$, with identical networks. In contrast to the existing methods [11,12,45,48,70], it offers several advantages: 1) The teacher network is equipped with a novel-domain augmenter, providing the capability to identify the model's sensitivity and homeostasis concerning new domains. 2) Then, the Fisher information matrix is estimated to isolate domain-sensitive and domain-invariant parameters in the student network. 3) In the subsequent test-time adaptation, the domain-sensitive parameters are updated, while the domain-invariant ones are preserved, to alleviate the catastrophic forgetting and error accumulation, as shown in Fig.2.

3.2 Multi-Domain Homeostasis Assessment

The teacher $\boldsymbol{\theta}_T$ is front-loaded with a multi-domain augmenter $\mathcal{A}_\phi$, and multiple novel-domain samples are fed into it to generate the corresponding intermediate predictions. Then, using uncertainty estimation, the model's steady-state/homeostasis for the new domain data is assessed.

Multi-domain augmenter. Recent advances show that the augmentation is an effective strategy to improve the cross-domain generalization ability [12, 52, 69]. Motivated by them, our multi-domain augmenter $\mathcal{A}_\phi$ is proposed, which falls into the adversarial learning paradigm, w.r.t. the trainable parameter ϕ.

Specifically, $\mathcal{A}_\phi$ is trained to augment H new-domain samples: $\tilde{\boldsymbol{X}}_h = \mathcal{A}_\phi(\boldsymbol{X})$, aiming for as much diversity as possible. Here, $\tilde{\boldsymbol{X}}_h$ is an augmented sample from

the set $\{\tilde{\boldsymbol{X}}_h\}_{h=1}^{H}$, and $\boldsymbol{X}$ is the original one. We denote cos_sim as the cosine similarity. The training objective has two terms, where the variability loss $\mathcal{L}_{var}$ encourages the diversity among the H augmentations, defined as:

$$\max_{\mathcal{A}_\phi} \mathcal{L}_{var}(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_j) = \frac{1}{H(H-1)} \sum_{i=1}^{H} \sum_{j \neq i} cos\_sim\left(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_j\right). \tag{1}$$

Then, the content consistency loss $\mathcal{L}_{con}$ ensures the similarity of motion contexts between the original and augmented samples:

$$\min_{\mathcal{A}_\phi} \mathcal{L}_{con} = \frac{1}{H} \sum\nolimits_{h=1}^{H} \left\| \tilde{\boldsymbol{X}}_h - \boldsymbol{X} \right\|_2. \tag{2}$$

These objectives collectively guide the training of $\mathcal{A}_\phi$ to produce diverse augmentations while maintaining content consistency with the original samples.

Intuitively, $\mathcal{A}_\phi$ can be conceptualized as a diverse motion generation model. For simplicity, it is implemented as the similar architecture in [47], expect for min-max adversarial training and a more streamlined architecture (9 GCN layers). Once $\mathcal{A}_\phi$ is trained, it is used to generate H augmented samples $\{\tilde{\boldsymbol{X}}_h\}_{h=1}^{H}$, and then fed into the teacher $\boldsymbol{\theta}_T$ to generate the corresponding intermediate predictions $\{\tilde{\boldsymbol{Y}}_h\}_{h=1}^{H}$. In this work, the hyperparameter H is set as 24.

Homeostasis assessment, also called uncertainty estimation [38, 53, 55], is widely-used to quantify the model's confidence level in its predictions under distribution shifts. While the confidence score is a common measure to assess prediction reliability, it tends to fluctuate irregularly and becomes unreliable in continual changing environment. To overcome this issue, we draw inspiration from the homeostasis mechanism in biological systems [38], which can determine the sensitivity of the model and then assess the model's stability in handling new domains.

Concretely, given the intermediate predictions $\{\tilde{\boldsymbol{Y}}_h\}_{h=1}^{H}$ and the original one $\{\tilde{\boldsymbol{Y}}\}$ obtained from $\mathcal{A}_\phi$, we compute the difference between $\{\tilde{\boldsymbol{Y}}_h\}_{h=1}^{H}$ and $\tilde{\boldsymbol{Y}}$ to form H sets of probability distribution matrix, and take the average of them as the uncertainty matrix $\boldsymbol{U}$, defined as:

$$\boldsymbol{U} = \exp\left(-\frac{1}{H} \sum_{h=1}^{H} \left| \frac{\tilde{\boldsymbol{Y}}_h}{\|\tilde{\boldsymbol{Y}}_h\|_2} - \frac{\tilde{\boldsymbol{Y}}}{\|\tilde{\boldsymbol{Y}}\|_2} \right| \right). \tag{3}$$

We note that $\boldsymbol{U}$ is an $N \times \Delta T$ matrix, where each element indicates the model's uncertainty score of the corresponding joint. The smaller the value, the lower homeostasis the model (more sensitive to the novel domain), and vice versa, which aids in the isolation of domain-sensitive parameters.

3.3 Domain Parameter Isolation

In dynamic environments, where target domain data continually changes and exhibits different distribution shifts over times, effective domain transfer becomes

crucial. To reduce error accumulation and catastrophic forgetting, it is necessary to isolate the different knowledge, and manage or utilize them separately.

For this purpose, the data-driven domain parameter isolation is proposed. Fisher information matrix (FIM) has proven to be an effective tool for measuring the importance of the model parameter. Building on recent advancements [6,39], FIM is therefore used to identify which parameters in the teacher model are sensitive to the new domains, and which ones are more homeostatic. Then, the results of parameter isolation is transferred to the student network, and allows for a separated optimization strategy in continual TTA phase.

For the sake the simplicity, we use the same symbol $\boldsymbol{\theta}_T \in \mathbb{R}^P$ to represent the flattened parameter of the teacher model, as well as $\boldsymbol{\theta}_S \in \mathbb{R}^P$ for the student, where P is the parameter dimension. The FIM $\boldsymbol{F}(\boldsymbol{\theta})$ is typically defined as the expectation of the second derivative of the log-likelihood function, given by:

$$\boldsymbol{F}(\boldsymbol{\theta}) = \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_T} \log p(\boldsymbol{U}|\boldsymbol{\theta}_T) \nabla_{\boldsymbol{\theta}_T} \log p(\boldsymbol{U}|\boldsymbol{\theta}_T)^\intercal\right], \tag{4}$$

where $\boldsymbol{F}(\boldsymbol{\theta})$ is a $P \times P$ matrix, and $p(\boldsymbol{U}|\boldsymbol{\theta}_T)$ is the probability density function of $\boldsymbol{U}$. We note that, $\boldsymbol{F}(\boldsymbol{\theta})$ is a positive semi-definite matrix, and its diagonal elements represents the importance of each parameter, while the off-diagonal elements denote the correlation between parameters.

For the lower calculation and higher interpretability, we assume the parameters are independent, and utilize the diagonal elements of $\boldsymbol{F}(\boldsymbol{\theta})$ to form an approximation $\boldsymbol{I}(\boldsymbol{\theta})$:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{Diag}\left(\left(\nabla_{\boldsymbol{\theta}_T} \mathcal{L}(\boldsymbol{U})\right)\left(\nabla_{\boldsymbol{\theta}_T} \mathcal{L}(\boldsymbol{U})\right)^\intercal\right). \tag{5}$$

We note that $\boldsymbol{I}(\boldsymbol{\theta}) \in \mathbb{R}^P$ has the identical dimension with $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_S$. $\mathcal{L}$ is simply implemented as the L2 norm.

Because $\boldsymbol{U}$ estimates the uncertainty against the H novel domains, the elements of $\boldsymbol{I}(\boldsymbol{\theta})$ indicate the sensitivity of the model parameter, where some are sensitive to the new domains, and the rest is more homeostatic. To achieve it, we use a binary mask $\boldsymbol{m} \in \{0,1\}^P$ to isolate the parameters:

$$\boldsymbol{m} \sim \mathrm{Bernoulli}(\alpha), \tag{6}$$

where α is the probability of 1. Here, based on $\boldsymbol{I}(\boldsymbol{\theta})$ of FIM, each element $\boldsymbol{m}_p$ of $\boldsymbol{m}$ is set using the following rule:

$$\boldsymbol{m}_p = \begin{cases} 1, \text{ if } \boldsymbol{I}(\boldsymbol{\theta})_p > \tau\text{-quantile} \\ 0, \text{ otherwise} \end{cases}, \quad p = 1.,,,.P, \tag{7}$$

where τ-quantile is a threshold, and we set $\tau = 0.2$ in all experiments. We note that, the above rule means that the top τ-quantile parameters, w.r.t. $\boldsymbol{m} \otimes \boldsymbol{\theta}$, with the largest values in $\boldsymbol{I}(\boldsymbol{\theta})$ are selected as the domain-sensitive ones (need to be update for current domain), and the rest, $(1 - \boldsymbol{m}) \otimes \boldsymbol{\theta}$, as the domain-invariant ones (need to be retained).

3.4 Homeostatic Test-time Adaptation

In continual TTA, we propose to isolate the full model into two parts: domain-sensitive and domain-invariant parameters. To manage and update them separately, we propose a novel homeostatic test-time parameter optimization strategy, which is implemented as the following two steps:

Isolated parameter optimization. Since $\boldsymbol{m}$ indicates those domain-sensitive weights in teacher model $\boldsymbol{\theta}_T$ as well as those stable ones, we further transfer it to the student network $\boldsymbol{\theta}_S$, and isolate the parameters as well. For a sample $\boldsymbol{X}^{(0)}$ from unseen domains, it is fed into both non-optimized student and teacher networks, to attain $\tilde{\boldsymbol{Y}}_S^{(0)} = \boldsymbol{\theta}_S(\boldsymbol{X}^{(0)})$ and $\tilde{\boldsymbol{Y}}_T^{(0)} = \boldsymbol{\theta}_T(\boldsymbol{X}^{(0)})$, where $\tilde{\boldsymbol{Y}}_T^{(0)}$ is regard as the pseudo ground truth. Then, we update the domain-sensitive parameters $\boldsymbol{\theta}_S \otimes \boldsymbol{m}$ to attain the adapted parameter of the student network, with a single gradient decent step:

$$\mathring{\boldsymbol{\theta}}_S^{(0)} \leftarrow \boldsymbol{\theta}_S^{(0)} - \eta \nabla_{\boldsymbol{\theta}_S^{(0)}} \mathcal{L}_{pred}(\tilde{\boldsymbol{Y}}_S^{(0)}, \tilde{\boldsymbol{Y}}_T^{(0)}) \otimes \boldsymbol{m}, \tag{8}$$

where $\eta = 0.001$ is the learning rate. Here, motivated by [48], the loss function $\mathcal{L}_{pred}$ is defined as the weighted sum of L_2 distance and bone length loss:

$$\lambda_1 \|\tilde{\boldsymbol{Y}}_S^{(0)} - \tilde{\boldsymbol{Y}}_T^{(0)}\|_2 + \lambda_2 \mathcal{L}_{bone}(\tilde{\boldsymbol{Y}}_S^{(0)}, \tilde{\boldsymbol{Y}}_T^{(0)}), \tag{9}$$

where $\mathcal{L}_{bone}$ is to compute the difference of bone length between two sequences, and $\lambda_1 = 0.9$, $\lambda_2 = 0.1$. Then, a forward pass is performed to predict the future motion $\mathring{\boldsymbol{Y}}_S^{(0)} = \mathring{\boldsymbol{\theta}}_S^{(0)}(\boldsymbol{X}^{(0)})$ using the adapted student model. After the isolated parameter optimization, the domain-invariant parameters $\boldsymbol{\theta}_S \otimes (1 - \boldsymbol{m})$ are preserved, and the domain-sensitive ones $\boldsymbol{\theta}_S \otimes \boldsymbol{m}$ are updated.

Exponential moving average. Next, the exponential moving average (EMA) is proposed to update the teacher network $\boldsymbol{\theta}_T$, with a momentum factor $\alpha = 0.99$:

$$\mathring{\boldsymbol{\theta}}_T^{(0)} \leftarrow \alpha \boldsymbol{\theta}_T^{(0)} + (1 - \alpha) \mathring{\boldsymbol{\theta}}_S^{(0)}, \tag{10}$$

where $\mathring{\boldsymbol{\theta}}_S^{(0)}$ is the adapted student network. Moreover, for the next sample $\boldsymbol{X}^{(1)}$, the model adaptation begins with $\mathring{\boldsymbol{\theta}}_S^{(0)} \rightarrow \boldsymbol{\theta}_S^{(1)}$ and $\mathring{\boldsymbol{\theta}}_T^{(0)} \rightarrow \boldsymbol{\theta}_T^{(1)}$ as the initial parameters, and the above process is repeated. The overall procedure of our HoCoTTA is summarized in Algorithm 1.

4 Experiments

4.1 Benchmark Datasets

(1) Human3.6M [28] is a well-known dataset, containing $\approx$ 3.6M frames of 15 action categories performed by 7 human subjects. **(2) CMU Mocap** [1]. Following [13, 32, 48], 8 daily action categories from CMU Mocap are selected. **(3) GRAB** [63] is a newly-introduced benchmark, including $\approx$1.6M poses of 29

Algorithm 1 Homeostatic Continual Test-time Adaptation

Require: multi-domain augmenter $\mathcal{A}_\phi$; teacher network $\boldsymbol{\theta}_T^{(i)}$, student network $\boldsymbol{\theta}_S^{(i)}$; learning rate η, binary mask $\boldsymbol{m}$, momentum factor α, τ-quantile;
Input: n samples from unseen domains $\{\boldsymbol{X}^{(i)}\}_{i=1}^n$;
Output: final predictions $\{\mathring{\boldsymbol{Y}}_S^{(i)}\}_{i=1}^n$;
1: for *each* i **do**
▷ *multi-domain homeostasis assessment*
2: augment H novel-domains $\{\tilde{\boldsymbol{X}}_h\}_{h=1}^H = \mathcal{A}_\phi(\boldsymbol{X}^{(i)})$;
3: $\{\tilde{\boldsymbol{Y}}_h\}_{h=1}^H = \boldsymbol{\theta}_T^{(i)}(\{\tilde{\boldsymbol{X}}_h\}_{h=1}^H)$, $\tilde{\boldsymbol{Y}} = \boldsymbol{\theta}_T^{(i)}(\boldsymbol{X}^{(i)})$;
4: compute uncertainty matrix $\boldsymbol{U}$ using Eq. (3);
5: compute fisher information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ using Eq. (5);
▷ *domain parameter isolation*
6: $\boldsymbol{m}_p \leftarrow \begin{cases} 1, \text{ if } \boldsymbol{I}(\boldsymbol{\theta})_p > \tau\text{-quantile} \\ 0, \text{ otherwise} \end{cases}$;
▷ *homeostatic test-time adaptation*
7: update the student: $\mathring{\boldsymbol{\theta}}_{\boldsymbol{S}}^{(\boldsymbol{i})} \leftarrow \boldsymbol{\theta}_S^{(i)} \otimes \boldsymbol{m}$ using Eq. (9);
8: update the teacher model: $\mathring{\boldsymbol{\theta}}_{\boldsymbol{T}}^{(\boldsymbol{i})} \leftarrow \boldsymbol{\theta}_T^{(i)}$ using Eq. (10);
9: $\mathring{\boldsymbol{Y}}_S^{(i)} \leftarrow \mathring{\boldsymbol{\theta}}_S^{(i)}(\boldsymbol{X}^{(i)})$; ▷ *make the final prediction*
10: $\boldsymbol{\theta}_S^{(i+1)} = \mathring{\boldsymbol{\theta}}_S^{(i)}$, $\boldsymbol{\theta}_T^{(i+1)} = \mathring{\boldsymbol{\theta}}_T^{(i)}$;
11: end for

actions from 10 human subjects. Compared with Human3.6M, GRAB are more diverse and involve interaction with the physical world, which is, therefore, more challenging. Each pose of all 3 datasets is specified by 3D coordinates of 17 joints, and normalized to $[-1, 1]$. All methods are implemented to predict the next 1 second frames, with the observed length of 1 second.

4.2 Baselines, Experimental Setups and Evaluation Metrics

Our HoCoTTA is compared with 7 recent approaches, categorized in 5 groups:

Baselines. 1) RNN-based: Resi. sup. [49] transforms the motion prediction into a sequence-to-sequence generation task; **2) GCN-based**: LTD [48], MSRGCN [14], PGBIG [46], and SPGSN [34] are the representative GCN-based baseline approaches, emerged in recent years; **3) MLP-based**: siMLPe [22] propose a variant of multilayer perceptron, achieving encouraging results; **4) TTA-based**: H/P-TTP [12] emerges in last year and use TTA to resolve the domain gap, which achieves the state-of-the-art performance.

Experimental setups. Recent progresses have proven that the performance of siMLPe [22] is superior than others under the typical experimental setting of human motion prediction, and it is open-source, which is therefore chosen as the backbone of both teacher and student networks in our HoCoTTA. Therefore, we design the following 3 experimental setups:

1) Setup-1 (generative predictive ability:) follows the standard data splitting that is consistent with the standard motion prediction task [13, 45, 48].

Metric	Methods	Human3.6M [28]					CMU Mocap [1]					GRAB [63]			
Time (milliseconds)		80	160	320	400	1000	80	160	320	400	1000	200	400	600	1000
MPJPE [mm]	Resi. sup. [49]	34.7	62.0	101.1	115.5	165.0	24.7	44.2	76.3	88.7	139.3	56.1	90.2	163.8	289.4
	LTD [48]	12.7	26.1	52.3	63.5	114.3	9.9	18.0	33.6	41.0	81.9	38.3	68.7	101.6	197.3
	MSRGCN [14]	12.1	25.6	51.6	62.9	114.2	8.7	15.8	30.6	38.1	79.0	32.2	60.2	96.3	178.6
	PGBIG [46]	10.3	22.7	47.4	58.5	110.3	8.2	15.4	30.1	37.3	76.7	30.1	53.9	92.2	157.2
	SPGSN [34]	10.4	22.3	47.1	58.3	109.6	8.3	14.8	28.6	37.0	77.8	27.4	50.6	91.3	144.5
	siMLPe† [22]	9.6	21.7	46.3	57.3	109.4	8.3	14.6	27.8	37.2	76.6	27.1	51.5	88.4	137.5
	H/P-TTP† [12]	9.8	21.1	47.2	55.6	103.7	8.0	13.1	28.5	35.3	74.4	26.5	47.4	85.5	138.0
	HoCoTTA‡	**9.2**	**20.5**	**46.0**	**52.8**	**98.4**	**7.8**	**12.7**	**24.2**	**35.0**	**71.1**	**24.1**	**45.2**	**81.0**	**131.8**
P-MPJPE [mm]	Resi. sup. [49]	23.4	45.1	78.5	96.6	111.0	14.1	26.2	39.7	56.6	86.2	27.6	43.2	110.3	144.7
	LTD [48]	8.9	17.2	36.9	53.1	101.5	6.5	12.1	20.0	31.2	66.0	21.8	37.1	83.2	128.8
	MSRGCN [14]	8.7	17.9	32.6	55.7	101.3	6.1	12.0	19.3	32.5	65.2	20.0	35.4	82.3	131.3
	PGBIG [46]	7.2	15.3	28.3	51.2	97.2	5.7	11.3	18.6	31.4	63.8	18.3	34.2	79.8	127.5
	SPGSN [34]	7.1	14.3	24.2	51.8	96.7	5.9	11.1	18.7	31.1	62.2	18.0	33.3	77.2	129.0
	siMLPe† [22]	6.7	14.0	23.7	49.3	94.2	5.4	10.0	17.4	29.1	60.5	18.1	33.4	70.2	124.4
	H/P-TTP† [12]	6.7	12.8	23.0	49.4	90.2	**5.2**	**9.3**	17.1	28.9	60.1	16.6	32.0	64.9	117.3
	HoCoTTA‡	**6.4**	**11.7**	**21.5**	**46.6**	**84.3**	**5.2**	9.4	**15.8**	**25.3**	**54.8**	**16.5**	**31.3**	**62.5**	**112.2**
PCK@150mm [%]	Resi. sup. [49]	64.8	62.3	60.0	57.5	50.3	76.4.	73.4	71.3	69.9	67.3	70.0	67.4	52.3	50.9
	LTD [48]	79.9	77.3	76.4	70.4	66.0	84.2	81.5	80.3	77.2	75.2	81.8	77.3	71.3	62.9
	MSRGCN [14]	85.4	83.0	82.1	75.5	70.1	86.7	82.5	81.1	78.3	76.2	84.7	79.7	75.3	65.6
	PGBIG [46]	88.5	84.2	83.0	77.3	69.6	88.8	83.2	81.5	78.0	77.0	84.3	82.2	75.8	66.4
	SPGSN [34]	87.8	84.7	85.2	80.1	71.2	88.4	85.1	82.0	77.9	76.4	87.1	80.4	77.0	67.8
	siMLPe† [22]	88.4	86.6	85.0	83.4	72.7	90.0	88.2	85.8	83.7	77.5	86.9	82.6	82.1	69.1
	H/P-TTP† [12]	91.2	**89.4**	86.8	85.1	74.6	**91.3**	**89.4**	**87.6**	**84.7**	**79.4**	**88.0**	82.3	81.1	70.4
	HoCoTTA‡	**92.2**	**89.4**	**87.1**	**86.0**	**76.1**	**93.7**	**90.1**	87.4	**85.1**	**81.3**	87.7	**84.2**	**83.0**	**72.3**

Table 1: General predictive ability comparison under the experimental setup-1. It follows the common data spitting. We highlight the best results in **bold**, and the second best in underlined. ‡ indicates our results, † is that are from the original papers, and others are from [70]. For the baselines that do not report P-MPJPE and PCK@150mm, we leverage the same transformation as [12] to re-statistic.

2) Setup-2 (predictive ability for unseen subjects and catoegories): is designed to evaluate the performance for new/unknown human subjects and categories, similar to the existing TTA-based predictive approaches [11, 12]. **3) Setup-3 (predictive ability for novel datasets):** is newly-introduced in our work, where the source data is from Human3.6M [28] and the pre-trained model is expected to adapt to the new dataset of GRAB [63]. We note that the setup-3 is more challenging than setup-1 and setup-2, because the target distributions, and data acquisition conditions, are completely different from the source one, and the distribution shift is more pronounced.

Evaluation Metrics. 1) MPJPE [7, 28, 46]: serves as the main metric to measure the average Euclidean distance between the prediction and ground truth. **2) P-MPJPE**: Procrustes aligned MPJPE (P-MPJPE) [37] aligns the predicted pose to the ground truth pose by a rigid transformation known as Procrustes Analysis (PA), removing errors independent of poses. **3) PCK@150mm:** Percentage of Correct 3D Keypoint (PCK) [3, 24] quantifies the proportion of predicted joints with MPJPE smaller than a predefined threshold of 150mm.

4.3 Generative Predictive Ability Evaluation

Our HoCoTTA mainly considers improving the prediction performance under the out-of-distribution setting in the complicated deployment scenarios; how-

		Predictive Ability for Unseen Categories									Predictive Ability for Unseen Subjects					
Time (ms)		160	400	1000	160	400	1000	200	400	1000	160	400	1000	200	400	1000
Actions		Human3.6M			CMU Mocap			GRAB			CMU Mocap			GRAB		
MPJPE	PGBIG [46]	27.8	62.1	114.2	15.2	45.7	86.4	27.5	59.3	155.3	27.4	58.0	108.3	30.4	55.3	158.6
	SPGSN [34]	26.3	63.6	115.6	14.8	43.6	88.5	34.7	56.4	150.8	27.2	58.3	107.0	31.2	56.8	159.3
	siMLPe [22]	25.3	60.5	112.8	15.8	44.9	84.7	36.4	57.8	151.4	25.6	55.3	102.5	30.1	57.2	155.9
	H/P-TTP [12]	24.7	53.6	102.5	**13.4**	41.0	77.9	31.1	51.2	140.2	24.7	56.4	102.8	28.6	52.3	135.5
	HoCoTTA‡	**22.1**	**52.7**	**99.6**	13.5	**39.6**	74.6	**30.2**	**49.8**	**136.2**	**22.9**	**53.1**	**98.4**	**27.2**	**48.0**	**130.1**
P-MPJPE	PGBIG [46]	16.4	59.7	99.7	17.3	29.4	69.7	20.2	38.7	130.3	13.3	52.8	89.8	21.8	40.6	137.8
	SPGSN [34]	14.2	60.7	100.4	14.7	26.9	67.8	18.9	35.8	126.4	13.8	54.2	90.5	22.9	42.7	133.9
	siMLPe [22]	14.0	58.3	98.8	15.2	27.8	65.4	18.5	36.3	123.8	12.3	51.4	88.6	21.3	40.4	127.9
	H/P-TTP [12]	13.8	50.4	90.3	13.1	25.8	60.0	17.1	**32.5**	120.1	12.6	50.1	85.5	20.2	39.1	124.1
	HoCoTTA‡	**13.2**	**49.4**	**85.1**	**11.0**	**24.9**	**58.3**	**17.0**	33.7	**116.7**	**12.2**	**48.7**	**81.2**	**18.0**	**36.8**	**118.6**
PCK @150mm	PGBIG [46]	77.3	72.2	67.4	72.5	71.2	68.9	77.8	73.2	66.4	75.7	71.3	65.7	77.5	71.2	63.4
	SPGSN [34]	76.6	74.0	67.8	74.8	72.1	71.3	81.0	75.9	65.4	76.4	70.8	66.4	81.8	73.3	66.5
	siMLPe [22]	75.3	73.1	68.9	77.7	74.2	71.0	80.2	77.5	67.8	80.1	75.2	68.5	84.0	80.0	67.7
	H/P-TTP [12]	79.0	75.1	73.3	83.6	81.4	75.3	84.4	81.3	70.4	80.5	74.3	70.9	83.5	79.4	69.6
	HoCoTTA‡	**83.8**	**79.2**	**75.1**	**86.4**	**82.6**	**77.8**	**86.5**	**82.4**	**72.5**	**86.0**	**82.1**	**76.6**	**86.4**	**82.7**	**72.1**

Table 2: Predictive ability for new categories or subjects. We see that our HoCoTTA brings the major improvement over the SoTA H/P-TTP method, indicating that the domain shift across action categories and human subjects can be calibrated.

ever, for the sake of fairness, it still needs to be evaluated under the common data splitting. Due to the diversity and stochasticity of human motion, even within the same dataset/domain, the distribution of the test samples remains a certain degree of difference from the training ones, representing a form of distribution shift. Therefore, we follows the widely-used data splitting of the used 3 benchmarks to evaluate our HoCoTTA and baselines, which calls the generative predictive ability evaluation, referring to as the experimental setup-1. Table 1 reports the average results of MPJPE, P-MPJPE and PCK@150mm of all 8 methods over the samples of each time step. From the results, we observe that our HoCoTTA achieves the overall best performance on all 3 metrics, which underscores the effectiveness of our proposed HoCoTTA. It also evidences that the common data splitting indeed remains a distribution gap between source training and target testing, which is not considered by the most standard motion prediction approaches. By contrast, our HoCoTTA effectively resolve the out-of-distribution problem in a sequence of test samples, and thus achieves the superior general predictive performance.

4.4 Predictive Ability for New Subjects/Categories

Consistent with the current TTA-based approaches [11,12], we proceed to evaluate the predictive ability for new subjects and categories, which is referred to as the experimental setup-2. In the real-world deployment scenarios, for the human motion prediction task, the target data often involves new human subjects and categories, diverging from the source training data. Compared to the setup-1, the setup-2 is more significant. Therefore, we construct the following 2 experiments,

		Predictive Ability for Novel Datasets: Human3.6M → GRAB											
	Time (ms)	200	400	800	1000	200	400	800	1000	200	400	800	1000
	Actions	A1 passing				A2 eating				A3 drinking			
MPJPE	PGBIG [46]	40.8	76.3	111.5	143.9	34.2	73.7	117.4	168.7	37.6	44.6	91.8	150.7
	SPGSN [34]	43.7	75.6	110.0	146.7	35.6	75.3	122.8	171.4	33.4	45.7	92.2	153.5
	siMLPe [22]	40.2	69.7	109.2	140.5	34.5	70.4	118.5	172.1	34.7	46.8	90.4	147.9
	H/P-TTP [12]	30.1	45.4	89.8	121.4	29.7	53.1	98.7	152.4	27.7	39.6	76.3	133.6
	HoCoTTA‡	**26.0**	**37.7**	**70.4**	**116.2**	**27.5**	**47.3**	**90.2**	**147.3**	**24.8**	**36.9**	**72.6**	**128.5**
P-MPJPE	PGBIG [46]	27.1	35.8	68.4	89.4	25.0	41.1	82.7	107.8	22.7	37.0	73.4	104.2
	SPGSN [34]	25.5	36.5	68.4	86.4	24.3	40.2	83.1	110.4	21.8	36.8	71.3	100.2
	siMLPe [22]	25.6	37.8	67.9	87.5	24.0	38.4	80.3	106.4	22.2	37.1	70.2	96.7
	H/P-TTP [12]	18.4	33.2	61.4	82.2	20.2	35.5	73.2	98.6	18.3	33.6	**62.4**	89.7
	HoCoTTA‡	**17.7**	**30.3**	**59.6**	**78.5**	**18.9**	**34.1**	**70.4**	**95.4**	**19.1**	**32.3**	**62.4**	**85.8**
PCK @150mm	PGBIG [46]	62.2	57.3	54.6	51.1	60.4	52.7	52.1	50.0	59.0	53.4	50.3	47.8
	SPGSN [34]	63.2	56.8	55.0	51.4	58.8	53.4	51.0	49.6	57.8	56.0	50.7	49.7
	siMLPe [22]	62.3	57.7	54.3	50.3	61.2	55.3	53.3	51.6	58.3	55.7	52.4	50.4
	H/P-TTP [12]	67.2	63.5	62.3	57.8	65.2	57.4	56.6	55.2	63.3	60.7	57.7	55.8
	HoCoTTA‡	**69.9**	**65.7**	**63.4**	**60.5**	**67.3**	**59.1**	**58.4**	**56.8**	**65.7**	**61.3**	**59.0**	**58.1**
	Actions	A4 lifting				A5 on				A6 squeeze			
MPJPE	PGBIG [46]	35.6	64.6	109.3	167.2	28.3	38.0	72.7	130.4	22.7	32.5	57.1	104.2
	SPGSN [34]	37.1	60.3	115.7	158.4	28.1	38.6	71.6	126.5	23.5	30.7	55.4	101.3
	siMLPe [22]	37.8	67.3	119.4	162.2	27.5	37.5	73.4	127.5	22.1	33.2	57.3	103.6
	H/P-TTP [12]	31.0	44.7	91.2	138.5	24.5	34.7	60.4	102.3	18.0	29.5	50.2	96.3
	HoCoTTA‡	**25.1**	**40.7**	**84.2**	**132.2**	**21.0**	**32.3**	**58.4**	**98.5**	**16.9**	**26.4**	**48.6**	**92.3**
P-MPJPE	PGBIG [46]	27.5	47.2	80.6	115.3	22.9	35.7	64.3	85.2	25.7	40.0	76.4	84.6
	SPGSN [34]	28.6	46.0	79.5	113.2	23.5	38.2	67.4	87.5	23.4	38.7	72.8	80.0
	siMLPe [22]	27.4	45.8	77.7	110.3	24.6	40.3	70.3	92.9	22.2	40.0	73.4	82.2
	H/P-TTP [12]	21.4	37.1	68.9	100.1	18.0	33.3	64.6	80.1	17.1	32.2	63.5	73.4
	HoCoTTA‡	**20.2**	**35.7**	**66.3**	**96.4**	**16.2**	**30.4**	**62.1**	**77.4**	**16.0**	**31.7**	**59.9**	**70.3**
PCK @150mm	PGBIG [46]	57.2	50.8	46.7	45.5	67.2	62.3	61.2	59.6	73.1	68.6	65.1	63.0
	SPGSN [34]	57.3	51.4	47.0	45.6	66.7	62.3	59.7	58.4	72.2	67.8	64.6	62.7
	siMLPe [22]	56.3	50.5	46.9	44.3	67.8	65.6	62.5	60.7	73.5	68.9	65.2	63.5
	H/P-TTP [12]	60.0	55.8	49.7	47.7	72.4	**70.3**	67.2	65.1	**80.2**	75.3	71.4	67.7
	HoCoTTA‡	**62.0**	**59.5**	**52.4**	**51.3**	**73.3**	70.2	**67.9**	**66.8**	80.0	**77.7**	**73.4**	**70.5**

Table 3: Comparison of the predictive ability for novel datasets, where the source model is trained on the Human3.6M [28], and the target is from the GRAB [63] dataset. The average results are reported for each action category.

where the model is expected to adapt to a new subject or action category during test-time, having been trained on the other subjects or categories. We compare 4 approaches of siMLPe [22] PGBIG [46], SPGSN [34], and H/P-TTP [12]—as they achieve the better general prediction results than the others 3 approaches under the experimental setup-1. The comparison results are provided in Table 2, which evaluates the average results of multiple multi-faceted adaptations for all samples of each unseen subject or category. From the results, we observe that the H/P-TTP and HoCoTTA achieve the better prediction, indicating that the domain shift can be calibrated by the test-time adaptation. In addition, the proposed HoCoTTA brings significant improvement over H/P-TTP, attributed to the preservation of domain-invariant parameters during the continual test-time adaptation phase, thereby avoiding error accumulation.

4.5 Predictive Ability for New Datasets

In this experiment, we tackle the most challenging scenario, where the target data comes from a new dataset. Due to the distinct data acquisition conditions

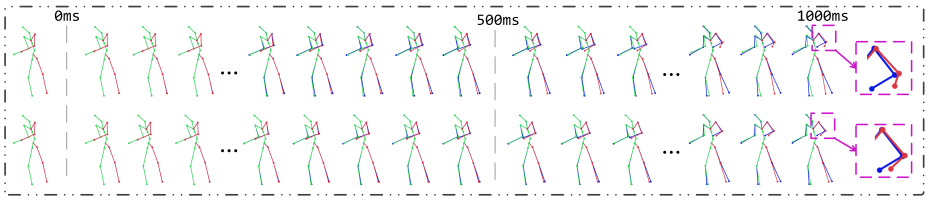

Fig. 3: Qualitative comparison: H/P-TTP [12] (top) *v.s.* our HoCoTTA (bottom). The green-red skeleton denotes the ground truth, and the blue is the prediction. Important details are highlighted in the purple box. It is clear that our HoCoTTA achieves the closer result to the ground truth.

and environments, the distribution shift between the source and target data is more significant than in the previous 2 experimental setups. Specifically, we train the model on the Human3.6M [28], and then adapt it to the GRAB dataset [63], referred to as experimental setup-3. Unlike the setup-1 and setup-2, the experimental setup-3 is newly-introduced in our work, and the existing approaches are not explicitly designed for this scenario. The GRAB dataset contains 26 action sub-types, which are more diverse and complex than the Human3.6M dataset, and we categorize them into 6 cases, according to the similarity of the action types, including A1 passing, A2 eating, A3 drinking, A4 lifting, A5 on, and A6 squeeze. As shown in Table 3, the results of all 5 approaches are reported, where the average performance of each time step across all target samples is evaluated. Under the evidence of the results, we observe that the proposed approach achieves the better performance than the other competitors. This superiority mainly stems from that our HoCoTTA is able to isolate the domain-specific knowledge and retain the cross-domain shared knowledge, which facilitates the continual model adaptation to the various distribution of the new dataset.

Fig. 3 also illustrates a qualitative comparison of the *wineglass-drink* activity under the experimental setup-3.

4.6 Ablation Studies

Here, we conduct the ablation studies to validate the effectiveness of the important aspects in our HoCoTTA. All experiments are designed under the setup-3, to consider the more challenging scene of adapting to a new dataset–GRAB [63]. Moreover, for simplicity, only the MPJPE metric is used.

We first investigate **(1) the proposed isolated parameter update** strategy, which is the key to our HoCoTTA. From the Table 4(left), we observe that the isolated parameter update strategy brings a major improvement in the prediction performance, compared to the previous full parameter update strategy.

In continual TTA phase, the domain-sensitive parameters of student are updated using the isolated parameter optimization strategy, with a learning rate η. Then, with a momentum factor τ, the EMA strategy is used to update the teacher network. To investigate **(2) the impact of the learning rate** η and **(3) momentum factor** τ, we conduct the 2 ablation experiments, keeping the

parameter update strategy	200	400	800	1000
full parameter update	25.7	39.4	75.3	127.0
isolated parameter update	**23.6**	**36.9**	**70.7**	**119.2**

η	α	200	400	800	1000
0.0005	**0.99**	24.1	37.1	73.8	124.5
	0.985	24.2	41.1	75.0	121.2
0.001	**0.99**	**23.6**	**36.9**	**70.7**	**119.2**
	0.995	23.7	37.3	**73.0**	122.7
0.0015	**0.99**	25.5	40.6	73.2	124.5

Table 4: Effect of our isolated parameter update strategy (left), learning rate η, and momentum factor α in the proposed HoCoTTA (right).

H	τ-quantile	200	400	800	1000
16	**0.2**	24.7	41.1	73.4	123.4
	0.15	24.0	38.2	73.1	122.1
24	**0.2**	**23.6**	**36.9**	**70.7**	**119.2**
	0.2	25.6	37.0	73.6	123.2
32	**0.2**	23.2	37.4	72.0	120.3

λ_1	λ_2	200	400	800	1000
0.85	**0.1**	25.3	37.9	73.0	122.7
	0.08	24.2	37.1	71.5	120.6
0.9	**0.1**	**23.6**	**36.9**	**70.7**	**119.2**
	0.12	24.7	38.0	72.3	121.1
0.95	**0.1**	24.4	40.3	73.7	121.6

Table 5: Effect of our isolated parameter update strategy (left), learning rate η, and momentum factor α in the proposed HoCoTTA (right).

other hyperparameters fixed. As reported in Table 4(right), the SoTA prediction performance is obtained when $\eta = 0.001$ and $\tau = 0.99$.

(4) Number of the novel-domain augmentations H has the significance for accessing the model's homeostasis. From Table 5, we observe the when $H = 24$, the best result is achieved, and a lower value leads to the performance degradation, and more augmentations bring no improvement.

(5) Top τ-quantile positions in FIM are treated as the domain-sensitive parameter that needs to be isolated and updated in our continual TTA, and the other parameters are preserved. To decide its value, we run an ablation analysis, and the optimal value is found to be $\tau = 0.2$, as in Table 5 (left).

In Table 5 (right), we analyze **(6) the impact of the loss function weights** λ_1 and λ_2 in Eq. (9), finding a balanced performance when $\lambda_1 = 0.9$ and $\lambda_2 = 0.1$.

5 Conclusion

To sum up, this work introduces a novel framework, homeostasis continual test-time adaptation (HoCoTTA), which tackles the challenging out-of-distribution issue in dynamic deployment scenarios for 3D human pose forecasting. Before making decisions, our HoCoTTA leverages the homeostasis assessment to evaluate the model's uncertainty with respect to the novel domains, and is able to estimate the domain-specific knowledge and isolate it from the cross-domain shared knowledge. It facilitates to preserve the cross-domain shared knowledge to eliminate the catastrophic forgetting, and avoid the error accumulation in continual test-time adaptation. Experiments across various benchmarks show the superior performance of our approach compared to state-of-the-art methods, particularly in the demanding scenario of adapting to a new dataset. These findings underscore the practical significance and robustness of the proposed HoCoTTA approach in the realistic scene.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62176125, 62306141), and in part by the Natural Science Foundation of Jiangsu Province (BK20220939).

References

1. CMU Graphics Lab: Carnegie-Mellon Motion Capture (Mocap) Database (2003), http://mocap.cs.cmu.edu
2. Aliakbarian, S., Saleh, F., Petersson, L., Gould, S., Salzmann, M.: Contextually Plausible and Diverse 3D Human Motion Prediction. In: ICCV. pp. 11333–11342 (2021)
3. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D Human Pose Estimation: New Benchmark and State of The Art Analysis. In: CVPR. pp. 3686–3693 (2014)
4. Barsoum, E., Kender, J., Liu, Z.: HP-GAN: Probabilistic 3D Human Motion Prediction via GAN. In: CVPR. pp. 1418–1427 (2018)
5. Bayasi, N., Hamarneh, G., Garbi, R.: Culprit-prune-net: Efficient continual sequential multi-domain learning with application to skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 165–175. Springer (2021)
6. Brahma, D., Rai, P.: A probabilistic framework for lifelong test-time adaptation. In: CVPR. pp. 3582–3591 (2023)
7. Cai, Y., Huang, L., Wang, Y., Cham, T.J., Cai, J., Yuan, J., Liu, J., Yang, X., Zhu, Y., Shen, X., et al.: Learning Progressive Joint Propagation for Human Motion Prediction. In: ECCV. pp. 226–242. Springer (2020)
8. Chen, D., Wang, D., Darrell, T., Ebrahimi, S.: Contrastive test-time adaptation. In: CVPR. pp. 295–305 (2022)
9. Chen, L., Zhang, Y., Song, Y., Shan, Y., Liu, L.: Improved test-time adaptation for domain generalization. In: CVPR. pp. 24172–24182 (2023)
10. Corona, E., Pumarola, A., Alenya, G., Moreno-Noguer, F.: Context-aware Human Motion Prediction. In: CVPR. pp. 6992–7001 (2020)
11. Cui, Q., Sun, H., Lu, J., Li, B., Li, W.: Meta-auxiliary learning for adaptive human pose prediction. In: AAAI (2023)
12. Cui, Q., Sun, H., Lu, J., Li, W., Li, B., Yi, H., Wang, H.: Test-time personalizable forecasting of 3d human poses. In: ICCV. pp. 274–283 (2023)
13. Cui, Q., Sun, H., Yang, F.: Learning Dynamic Relationships for 3D Human Motion Prediction. In: CVPR. pp. 6519–6527 (2020)
14. Dang, L., Nie, Y., Long, C., Zhang, Q., Li, G.: MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction. In: ICCV. pp. 11467–11476 (2021)
15. Fang, Y., Yap, P.T., Lin, W., Zhu, H., Liu, M.: Source-free unsupervised domain adaptation: A survey. arXiv preprint arXiv:2301.00265 (2022)
16. Fleuret, F., et al.: Test time adaptation through perturbation robustness. In: NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications (2021)
17. Fragkiadaki, K., Levine, S., Felsen, P., Malik, J.: Recurrent Network Models for Human Dynamics. In: ICCV. pp. 4346–4354 (2015)

18. Gan, Y., Bai, Y., Lou, Y., Ma, X., Zhang, R., Shi, N., Luo, L.: Decorate the newcomers: Visual domain prompt for continual test time adaptation. In: AAAI. vol. 37, pp. 7595–7603 (2023)
19. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: ICML. pp. 1180–1189. PMLR (2015)
20. Gong, T., Jeong, J., Kim, T., Kim, Y., Shin, J., Lee, S.J.: Note: Robust continual test-time adaptation against temporal correlation. NeurIPS **35**, 27253–27266 (2022)
21. Gui, L.Y., Wang, Y.X., Liang, X., Moura, J.M.F.: Adversarial Geometry-Aware Human Motion Prediction. In: ECCV. pp. 786–803 (2018)
22. Guo, W., Du, Y., Shen, X., Lepetit, V., Alameda-Pineda, X., Moreno-Noguer, F.: Back to mlp: A simple baseline for human motion prediction. In: WACV. pp. 4809–4819 (2023)
23. Guo, X., Choi, J.: Human Motion Prediction via Learning Local Structure Representations and Temporal Dependencies. In: AAAI. pp. 2580–2587 (2019)
24. Habibie, I., Xu, W., Mehta, D., Pons-Moll, G., Theobalt, C.: In The Wild Human Pose Estimation using Explicit 2D Features and Intermediate 3D Representations. In: CVPR. pp. 10905–10914 (2019)
25. Hassan, M., Ceylan, D., Villegas, R., Saito, J., Yang, J., Zhou, Y., Black, M.J.: Stochastic scene-aware motion prediction. In: ICCV. pp. 11374–11384 (2021)
26. He, Y., Carass, A., Zuo, L., Dewey, B.E., Prince, J.L.: Autoencoder Based Self-supervised Test-Time Adaptation for Medical Image Analysis. Medical Image Analysis p. 102136 (2021)
27. Hu, M., Song, T., Gu, Y., Luo, X., Chen, J., Chen, Y., Zhang, Y., Zhang, S.: Fully Test-Time Adaptation for Image Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 251–260. Springer (2021)
28. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **36**, 1325–1339 (2014)
29. Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G.: Contrastive adaptation network for unsupervised domain adaptation. In: CVPR. pp. 4893–4902 (2019)
30. Kundu, J.N., Gor, M., Babu, R.V.: BiHMP-GAN: Bidirectional 3D Human Motion Prediction GAN. In: AAAI. vol. 33, pp. 8553–8560 (2019)
31. Kundu, J.N., Venkat, N., Babu, R.V., et al.: Universal source-free domain adaptation. In: CVPR. pp. 4544–4553 (2020)
32. Li, C., Zhang, Z., Sun Lee, W., Hee Lee, G.: Convolutional Sequence to Sequence Model for Human Dynamics. In: CVPR. pp. 5226–5234 (2018)
33. Li, J., Seltzer, M.L., Wang, X., Zhao, R., Gong, Y.: Large-scale domain adaptation via teacher-student learning. arXiv preprint arXiv:1708.05466 (2017)
34. Li, M., Chen, S., Zhang, Z., Xie, L., Tian, Q., Zhang, Y.: Skeleton-Parted Graph Scattering Networks for 3D Human Motion Prediction. In: ECCV. pp. 18–36. Springer (2022)
35. Li, M., Chen, S., Zhao, Y., Zhang, Y., Wang, Y., Tian, Q.: Dynamic Multiscale Graph Neural Networks for 3D Skeleton Based Human Motion Prediction. In: CVPR. pp. 214–223 (2020)
36. Li, S., Xie, M., Gong, K., Liu, C.H., Wang, Y., Li, W.: Transferable semantic augmentation for domain adaptation. In: CVPR. pp. 11516–11525 (2021)

37. Li, Y., Li, K., Jiang, S., Zhang, Z., Huang, C., Da Xu, R.Y.: Geometry-driven self-supervised method for 3d human pose estimation. In: AAAI. vol. 34, pp. 11442–11449 (2020)
38. Liu, J., Yang, S., Jia, P., Lu, M., Guo, Y., Xue, W., Zhang, S.: Vida: Homeostatic visual domain adapter for continual test time adaptation. arXiv preprint arXiv:2306.04344 (2023)
39. Liu, J., Yuan, H., Lu, X.M., Wang, X.: Quantum fisher information matrix and multiparameter estimation. Journal of Physics A: Mathematical and Theoretical **53**(2), 023001 (2020)
40. Liu, Q., Lin, L., Shen, Z., Yang, Z.: Periodically exchange teacher-student for source-free object detection. In: ICCV. pp. 6414–6424 (2023)
41. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: CVPR. pp. 1215–1224 (2021)
42. Liu, Y., Kothari, P., Van Delft, B., Bellot-Gurlet, B., Mordan, T., Alahi, A.: Ttt++: When does self-supervised test-time training fail or thrive? NeurIPS **34**, 21808–21820 (2021)
43. Liu, Z., Lyu, K., Wu, S., Chen, H., Hao, Y., Ji, S.: Aggregated Multi-GANs for Controlled 3D Human Motion Prediction. In: AAAI. pp. 2225–2232 (2021)
44. Lodagala, V.S., Ghosh, S., Umesh, S.: Pada: Pruning assisted domain adaptation for self-supervised speech representations. In: 2022 IEEE Spoken Language Technology Workshop (SLT). pp. 136–143. IEEE (2023)
45. Ma, T., Nie, Y., Long, C., Zhang, Q., Li, G.: Progressively Generating Better Initial Guesses Towards Next Stages for High-Quality Human Motion Prediction. In: CVPR. pp. 6437–6446 (2022)
46. Ma, T., Nie, Y., Long, C., Zhang, Q., Li, G.: Progressively generating better initial guesses towards next stages for high-quality human motion prediction. In: CVPR. pp. 6437–6446 (2022)
47. Mao, W., Liu, M., Salzmann, M.: Generating smooth pose sequences for diverse human motion prediction. In: ICCV. pp. 13309–13318 (2021)
48. Mao, W., Liu, M., Salzmann, M., Li, H.: Learning Trajectory Dependencies for Human Motion Prediction. In: ICCV. pp. 9489–9497 (2019)
49. Martinez, J., Black, M.J., Romero, J.: On Human Motion Prediction using Recurrent Neural Networks. In: CVPR. pp. 2891–2900 (2017)
50. Martínez-González, A., Villamizar, M., Odobez, J.M.: Pose Transformers (POTR): Human Motion Prediction with Non-Autoregressive Transformers. In: ICCV. pp. 2276–2284 (2021)
51. Melas-Kyriazi, L., Manrai, A.K.: Pixmatch: Unsupervised domain adaptation via pixelwise consistency training. In: CVPR. pp. 12435–12445 (2021)
52. Orbes-Arteaga, M., Varsavsky, T., Sorensen, L., Nielsen, M., Pai, A., Ourselin, S., Modat, M., Cardoso, M.J.: Augmentation based unsupervised domain adaptation. arXiv preprint arXiv:2202.11486 (2022)
53. Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J.V., Lakshminarayanan, B., Snoek, J.: Can you trust your model's uncertainty. evaluating predictive uncertainty under dataset shift (2019)
54. Piergiovanni, A., Angelova, A., Toshev, A., Ryoo, M.S.: Adversarial Generative Grammars for Human Activity Prediction. In: ECCV. pp. 507–523. Springer (2020)
55. Roy, S., Trapp, M., Pilzer, A., Kannala, J., Sebe, N., Ricci, E., Solin, A.: Uncertainty-guided source-free domain adaptation. In: ECCV. pp. 537–555. Springer (2022)
56. Ruiz, A.H., Gall, J., Moreno-Noguer, F.: Human Motion Prediction via Spatio-Temporal Inpainting. In: CVPR. pp. 7134–7143 (2018)

57. Saadatnejad, S., Rasekh, A., Mofayezi, M., Medghalchi, Y., Rajabzadeh, S., Mordan, T., Alahi, A.: A generic diffusion-based approach for 3d human pose prediction in the wild. In: ICRA. pp. 8246–8253. IEEE (2023)
58. Sanyal, S., Babu, R.V., et al.: Continual domain adaptation through pruning-aided domain-specific weight modulation. In: CVPR. pp. 2456–2462 (2023)
59. Shin, C., Kim, T., Lee, S., Leey, S.: Test-Time Adaptation for Out-Of-Distributed Image Inpainting. In: ICIP (2021)
60. Sofianos, T., Sampieri, A., Franco, L., Galasso, F.: Space-Time-Separable Graph Convolutional Network for Pose Forecasting. In: ICCV. pp. 11209–11218 (2021)
61. Sójka, D., Cygert, S., Twardowski, B., Trzciński, T.: Ar-tta: A simple method for real-world continual test-time adaptation. In: ICCV. pp. 3491–3495 (2023)
62. Spall, J.C.: Monte carlo computation of the fisher information matrix in nonstandard settings. Journal of Computational and Graphical Statistics **14**(4), 889–909 (2005)
63. Taheri, O., Ghorbani, N., Black, M.J., Tzionas, D.: GRAB: A Dataset of Whole-Body Human Grasping of Objects. In: ECCV (2020)
64. Tanke, J., Zhang, L., Zhao, A., Tang, C., Cai, Y., Wang, L., Wu, P.C., Gall, J., Keskin, C.: Social diffusion: Long-term multiple human motion anticipation. In: ICCV. pp. 9601–9611 (2023)
65. Tomar, D., Vray, G., Bozorgtabar, B., Thiran, J.P.: Tesla: Test-time self-learning with automatic adversarial augmentation. In: CVPR. pp. 20341–20350 (2023)
66. Varsavsky, T., Orbes-Arteaga, M., Sudre, C.H., Graham, M.S., Nachev, P., Cardoso, M.J.: Test-time Unsupervised Domain Adaptation. In: MICCAI. pp. 428–436. Springer (2020)
67. Volpi, R., Morerio, P., Savarese, S., Murino, V.: Adversarial feature augmentation for unsupervised domain adaptation. In: CVPR. pp. 5495–5504 (2018)
68. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020)
69. Wang, Q., Fink, O., Van Gool, L., Dai, D.: Continual test-time domain adaptation. In: CVPR. pp. 7201–7211 (2022)
70. Xu, C., Tan, R.T., Tan, Y., Chen, S., Wang, X., Wang, Y.: Auxiliary tasks benefit 3d skeleton-based human motion prediction. In: ICCV. pp. 9509–9520 (2023)
71. Xu, S., Wang, Y.X., Gui, L.Y.: Diverse human motion prediction guided by multi-level spatial-temporal anchors. In: ECCV. pp. 251–269. Springer (2022)
72. Yang, S., Wang, Y., Van De Weijer, J., Herranz, L., Jui, S.: Generalized source-free domain adaptation. In: ICCV. pp. 8978–8987 (2021)
73. Ye, J., Fu, C., Zheng, G., Paudel, D.P., Chen, G.: Unsupervised domain adaptation for nighttime aerial tracking. In: CVPR. pp. 8896–8905 (2022)
74. Yuan, L., Xie, B., Li, S.: Robust test-time adaptation in dynamic scenarios. In: CVPR. pp. 15922–15932 (2023)
75. Zhang, M., Levine, S., Finn, C.: Memo: Test time robustness via adaptation and augmentation. NeurIPS **35**, 38629–38642 (2022)
76. Zhang, Y., Black, M.J., Tang, S.: We are more than our joints: Predicting how 3d bodies move. In: CVPR. pp. 3372–3382 (June 2021)