

Source-free Domain Adaptive Human Pose Estimation

摘要（现领域问题+本方法）

我们进一步提出了一个新的框架，该框架由三个模型组成：源模型、中间模型和目标模型，从源保护和目标相关的角度探讨了任务。源保护模块在抗噪声的同时更有效地保留了源信息，目标相关模块通过构建新的空间概率空间降低了空间表征的稀疏性，并在此空间基础上提出了针对特定姿态的对比学习和信息最大化。

介绍（细说问题+细说本方法和他方法解决问题+贡献（任务+创新+性能））

合成数据集可以生成很多标记数据，但是差异大，不好用，域适应降低差距。

具体来说，在传统的数据处理任务中，源数据可以参与自适应过程，从而专注于减少源和目标之间的域转移。然而，在无源数据分析中，在自适应过程中只能使用预训练的源模型，从而导致源域的灾难性遗忘。

直接应用源预训练模型也会由于域移位而引入噪声。因此，平衡源模型知识的吸收与源模型噪声的抵抗是至关重要的。

每个关键点与其他关键点之间的距离都是基于骨骼长度的，关键点K的数量(例如，人体姿势14个，手部姿势21个)明显小于图像中的像素数量(例如 $512 \times 512 \approx 0.26M$)，这使得它成为一项具有挑战性的任务。

源模型保留源信息，目标模型从目标数据中吸收知识，中间模型与源模型和目标模型相互作用，隐式地减小源和目标之间的域间隙，从而用于最终的推理。在源保护模块中，我们在源模型和中间模型之间进行知识传递，防止源相关信息的灾难性遗忘。我们将知识从源模型转移到中间模型，同时保持源模型的特征提取器不变。针对噪声，我们提出了一种新的残差损失，并对源模型的回归量进行了微调。在目标相关模块中，我们在中间模型和目标模型之间执行，以减轻源域和目标域之间的域转移。我们为HPE构建了一个新的空间概率空间，以缓解稀疏性问题。

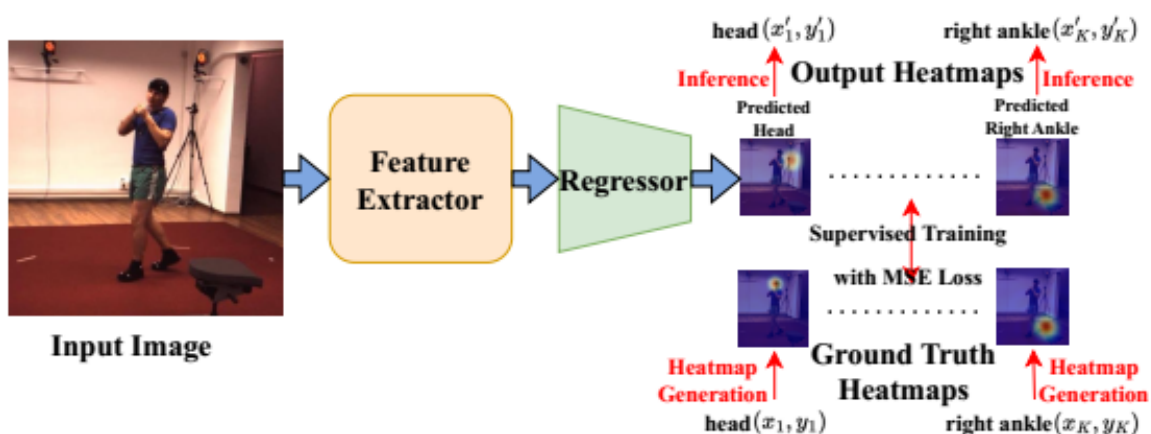
结论（贡献的总结版）

pass

本方法

2D HPE的准备

基于热图的二维HPE过程。该模型包括两个部分：特征提取器G和回归器F。将输入图像输入到模型中，得到K个关键点对应的K张输出热图。



Ground truth heatmaps 由一个以 ground truth joint location,为中心的二维高斯图生成的，每个关键点有一个热图。在推理阶段，训练良好的监督模型输出热图，预测每个像素处出现关节的概率。对于每个热图，概率最高的像素被认为是关节的预测位置。

问题情况

源域和目标域共享相同的标签空间，但位于不同的分布中(例如合成与真实)。**训练过程分为两个阶段:预训练阶段和适应阶段。**在预训练阶段，使用标记的源域数据集 S 来训练源模型 f^{sr} 。然而，无源域自适应HPE与一般域自适应HPE的区别在于自适应阶段，**即只有源模型 f^{sr} 和目标域数据集 T 可用于训练，源域数据集不能再使用。**我们的目标是开发一个在目标数据集上表现良好的模型。

源模型预训练

该网络由特征提取器 G (如ResNet主干)和姿态回归器 F 组成，因此源模型可以表示为 $f^{sr} = F^{sr}(G^{sr}(\cdot))$ 。

预训练的总体目标

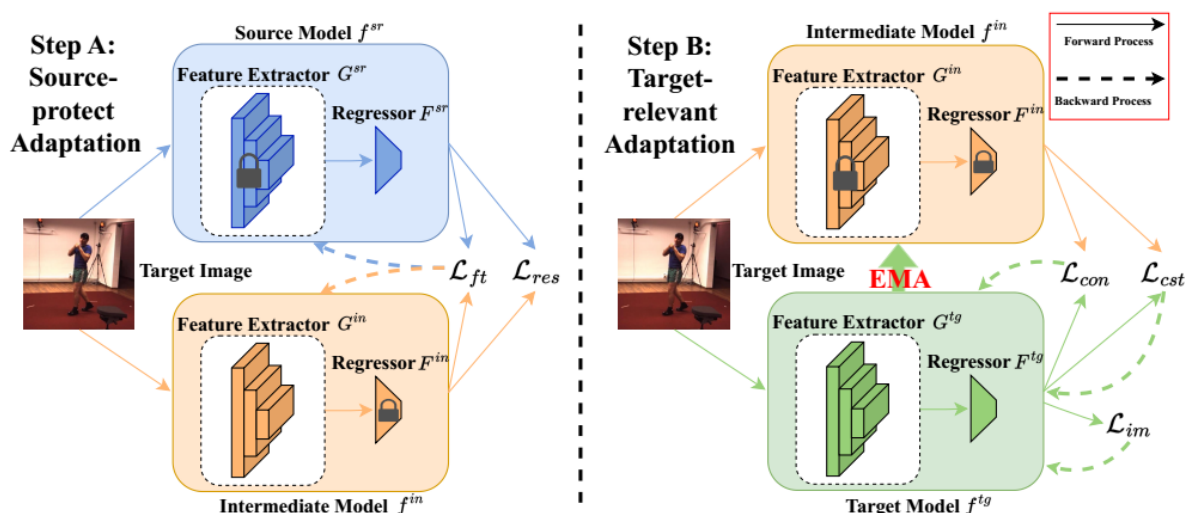
$$\mathcal{L}_{pretrain} = \mathbb{E}_{(x_i^s, y_i^s) \in S} \mathcal{L}_{mse}(F^{sr}(G^{sr}(x_i^s)), T(y_i^s)), \quad (1)$$

其中 \mathcal{L}_{mse} 是预测热图表示与真实情况之间的均方误差(MSE)损失。

适应框架

它包括三个模型，每个模型都有一个特征提取器和一个回归器。适应包括步骤A和步骤B两个步骤。步骤A适用于源保护适应，步骤B适用于目标相关适应。**实线箭头表示关联损失的正向更新过程，虚线箭头表示关联损失的反向更新过程。**EMA是指从目标模型到中间模型更新中间模型权重的指数移动平均过程。

在开始适应时， f^{in} 和 f^{tg} 从源模型 f^{sr} 初始化。对于单个适配迭代，执行两个步骤如图2所示。

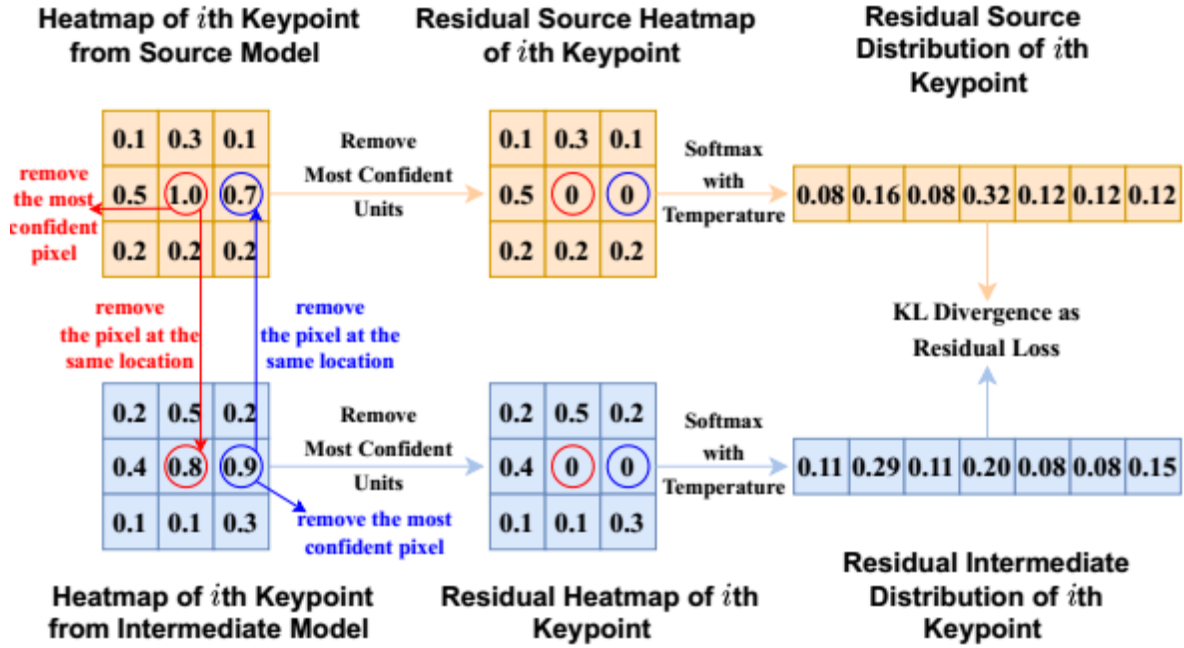


在步骤A中，我们的目标是克服源域上的灾难性遗忘，同时消除由于域漂移而从源模型中产生的噪声。因此，源模型的特征提取器 G^{sr} 是固定的，以保留源知识，而其回归量 F^{sr} 是不断更新的，以提高回归精度，从而降低输出中的噪声。相反，中间模型的回归量 F^{in} 是固定的，但其特征提取器 G^{in} 不断更新。**这是因为我们希望特征提取器在表示级别上学习更多的源知识，而源模型中由于域移位而产生的噪声可以被抑制。**

在步骤B中，我们的目标是在中间模型和目标模型之间以目标相关的方式进行自适应。在此过程中，使用了几种特定姿势的技术来最小化这两个模型之间的差异。**虽然 f^{in} 和 f^{tg} 都在这一步更新，但 f^{tg} 是通过反向传播更新的，而 f^{in} 是通过 f^{tg} 的EMA更新的，因为在训练步骤上平均模型权重往往比直接使用最终权值产生更准确的模型。**

源保护模块（基于向量的）

这里我们选择源模型和中间模型在同一位置生成的两张热图进行说明。为简单起见，这里将热图的比例尺设置为 3×3 ，而实际尺寸为 64×64 。残余损失如下：



首先，我们引入在源模型上校准的微调损失：

$$\mathcal{L}_{ft} = \mathbb{E}_{x_i^t \in \mathcal{T}} \mathcal{L}_{mse}(F^{sr}(G^{sr}(x_i^t)), F^{in}(G^{in}(x_i^t))). \quad (2)$$

回归器正在更新：

$$\theta_{F^{sr}}^{t+1} = \theta_{F^{sr}}^t - \lambda^{sr} \frac{\partial \mathcal{L}_{ft}}{\partial \theta_{F^{sr}}^t}, \quad (3)$$

式中， $\theta_{F^{sr}}^s$ 为源模型回归量 F^s 的权值， t 为训练步长， λ^s 为源模型的学习率。

方程2的应用促进了源模型和中间模型之间的相互知识转移。**除了热图中置信度最高的像素外，其他像素还包含有利于知识转移的有价值信息，这就是我们在图3中提出残差损失的原因。**

从图中的示例中，移除源热图中最自信的像素(置信度= 1.0)，同时移除红色标记的中间热图中相同位置置信度为0.8的像素。用蓝色标记的配对遵循相同的规则，但基于中间热图。

然后建立自适应残差热图，分别记为 $\hat{h}_{i,j}^{sr}$ 和 $\hat{h}_{i,j}^{in}$ 。此外，我们利用KL散度保证了 $\hat{h}_{i,j}^{in}$ 接近 $\hat{h}_{i,j}^{sr}$ ，从而保留了源信息。基于这些，我们可以推导出剩余损失为：

$$\mathcal{L}_{res} = \mathbb{E}_{x_i^t \in \mathcal{T}} \mathbb{E}_{j \in \mathcal{P}} D_{KL}(\sigma(\hat{h}_{i,j}^{sr}/\tau) || \sigma(\hat{h}_{i,j}^{in}/\tau)), \quad (4)$$

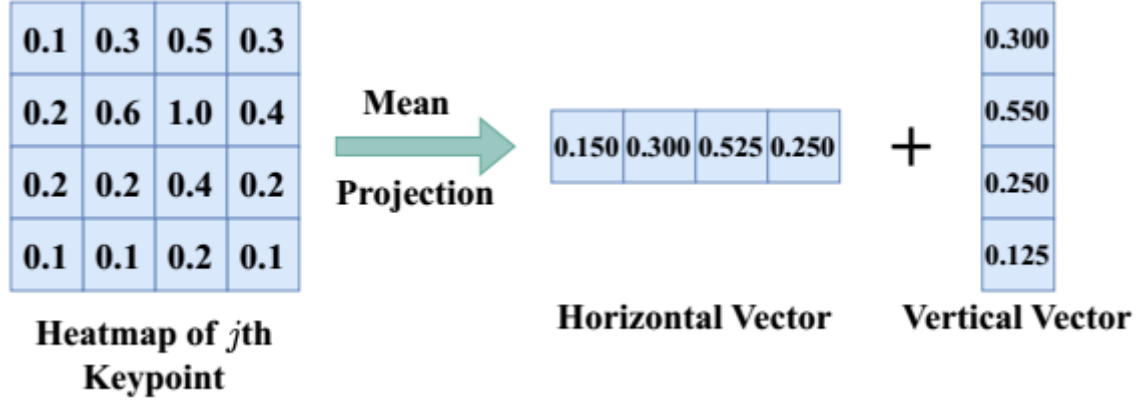
其中 D_{KL} 为KL散度， $\sigma(\cdot)$ 为softmax函数。此外， τ 是用于缩放残差热图的温度，经验设置为0.3。结合微调损失和残差损失，得到更新中间模型的优化过程和目标函数：

$$\mathcal{L}_{in} = \mathcal{L}_{ft} + \alpha \mathcal{L}_{res}, \quad \theta_{G^{in}}^{t+1} = \theta_{G^{in}}^t - \lambda^{in} \frac{\partial \mathcal{L}_{in}}{\partial \theta_{G^{in}}^t}, \quad (5)$$

其中 t 是训练步长, λ^{in} 是 G^{in} 的学习率。此外, α 是这两种损失之间的权衡超参数。

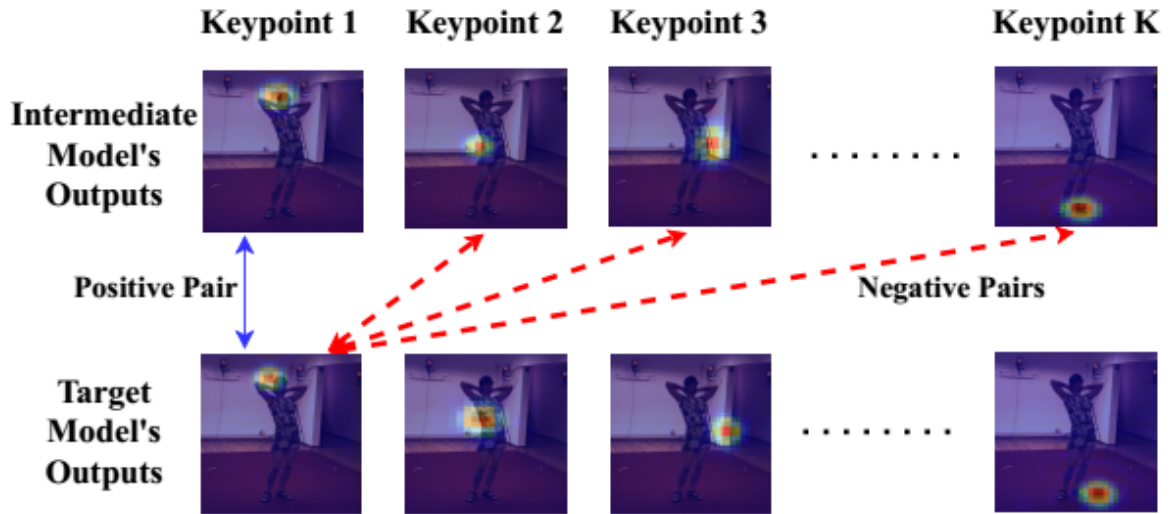
Target-relevant Modules

热图投影的图示:为了降低热图的稀疏性, 我们将热图投影成一个水平向量和一个垂直向量。为简单起见, 这里将热图的比例尺设置为 4×4 , 将矢量的比例尺设置为 4×1 , 而实际热图的大小为 64×64 , 矢量的大小为 64×1 。



我们从水平和垂直两个方向引入热图的投影。对于中间模型从第 i 个样本的第 j 个关键点生成的热图 $h_{i,j}^{in}$, 其投影可以表示为 $\text{proj}(h_{i,j}^{in}) = (v_{i,jx}^{in}, v_{i,jy}^{in})$, 其中 $v_{i,jx}^{in}$ 为水平向量, $v_{i,jy}^{in}$ 为垂直向量。这个矢量对我们提出的以下模块和损失的基础。

首先, 我们提出了中间模型和目标模型之间的特定姿势对比损失。根据[12]的观察, 错误的预测通常位于其他关键点的位置。在这种情况下, 改进回归可以通过最小化同一位置的热图差异和最大化不同位置的热图差异来实现。这就是我们在图5中定义正负对的方法。



将中间模型和目标模型生成的两张热图的相似度分别表示为:

$$\text{sim}(h_{i,j}^{in}, h_{i,k}^{tg}) = \frac{1}{2} \left(\frac{v_{i,jx}^{in} \cdot v_{i,kx}^{tg}}{\|v_{i,jx}^{in}\| \|v_{i,kx}^{tg}\|} + \frac{v_{i,jy}^{in} \cdot v_{i,ky}^{tg}}{\|v_{i,jy}^{in}\| \|v_{i,ky}^{tg}\|} \right). \quad (6)$$

在相似度的基础上, 我们将整体的特定姿势对比损失定义为:

$$\mathcal{L}_{cst} = -\mathbb{E}_{x_i^t \in \mathcal{T}} \mathbb{E}_{j \in \mathcal{P}} \log \frac{\exp(\text{sim}(h_{i,j}^{in}, h_{i,j}^{tg}))}{\sum_{k \in \mathcal{P}} \exp(\text{sim}(h_{i,j}^{in}, h_{i,k}^{tg}))}. \quad (7)$$

接下来，我们提出了一种姿态特定的自我监督技术，称为姿态特定信息最大化。信息最大化对无源领域自适应分类非常有帮助[21]，因为它使目标输出个体确定性和全局多样性，从而隐式地减轻了领域差距。然而，由于热图的稀疏性，改进是有限的。在投影向量的帮助下，我们将信息最大化损失细化为：

$$\mathcal{L}_{im} = \mathbb{E}_{x_i^t \in \mathcal{T}} (\mathcal{L}_{entx}(x_i^t) + \mathcal{L}_{enty}(x_i^t) - \mathcal{L}_{div}(x_i^t)), \quad (8)$$

$$\begin{cases} \mathcal{L}_{entx}(x_i^t) = -\mathbb{E}_{j \in \mathcal{P}} [\sigma(v_{i,jx}^{tg}) \log \sigma(v_{i,jx}^{tg})], \\ \mathcal{L}_{enty}(x_i^t) = -\mathbb{E}_{j \in \mathcal{P}} [\sigma(v_{i,jy}^{tg}) \log \sigma(v_{i,jy}^{tg})], \\ \mathcal{L}_{div}(x_i^t) = -[\mathbb{E}_{j \in \mathcal{P}} \sigma(h_{i,j}^{tg})] \log [\mathbb{E}_{j \in \mathcal{P}} \sigma(h_{i,j}^{tg})]. \end{cases} \quad (9)$$

这里 \mathcal{L}_{entx} 是水平向量的熵， \mathcal{L}_{enty} 是垂直向量的熵。公式8中的最后一项是促进多样性的目标。此外，我们引入了两个模型输出之间的一致性损失：

$$\mathcal{L}_{con} = \mathbb{E}_{(x_i^s, y_i^s) \in \mathcal{S}} \mathcal{L}_{mse}(F^{in}(G^{in}(x_i^s)), F^{tg}(G^{tg}(x_i^s))). \quad (10)$$

综上所述，我们可以得出目标模型的总体目标损失函数为：

$$\mathcal{L}_{tgt} = \mathcal{L}_{con} + \beta \mathcal{L}_{cst} + \gamma \mathcal{L}_{im}, \quad (11)$$

其中 β 和 γ 是权衡参数。

此外，在这一步中，中间模型通过指数移动平均策略(EMA)进行更新：

$$\theta_{fin}^{t+1} = \eta \theta_{fin}^t + (1 - \eta) \theta_{ftg}^t, \quad (12)$$

其中 t 表示训练步长， η 表示平滑系数默认设置为0.999。

注释：**热图**中颜色渐变的根据通常是像素点的数值大小来确定的。在计算机视觉中，热图通常用于表示一些连续的数值分布，例如像素点的梯度、深度、置信度等。通常情况下，数值越大的像素点，对应的颜色就越深，数值越小的像素点，对应的颜色就越浅，从而形成一种颜色渐变的效果。

正向更新：对于每个真实框（ground truth），计算其与所有锚框（anchor）的相关系数，得到一个相关系数矩阵。将相关系数矩阵作为模型的输出，与真实标签进行比较，计算损失值。根据损失值更新模型参数。

反向更新：计算损失函数对模型参数的梯度。根据梯度下降算法，对模型参数进行反向更新。

实验

数据集。我们使用三个人体姿势数据集和三个手部姿势数据集来验证我们的方法。SURREAL[36]作为**源人体姿势数据集**，包含600万张合成人体姿势图像。Human3.6M[11]是**目标人体姿态数据集**之一，由360万张图像组成，分布在7个折叠中。根据前人的研究[12,16]，将S1、S5、S6、S7、S8指定为训练集，将S9、S11指定为测试集。另一个目标数据集是Leeds Sports Pose [15] (LSP)，这是一个包含2000张图像的真实数据集，我们使用所有这些图像进行自适应。我们专注于两个领域适应任务: SURREAL \rightarrow Human3.6M and SURREAL \rightarrow LSP for human pose tasks.

Rendered Hand Pose Dataset 是具有43986张合成手图像的源手数据集，其中41258张用于训练，其余2728张用于验证。

Hand-3D-Studio (H3D)是两个目标手数据集之一，包括22,000个真实世界的帧。对于训练，我们使用了18,800帧，并保留其余的用于测试，遵循先前研究论文[12,16]中概述的方案。

另一个目标手数据集FreiHand[43]提供了130k张真实世界的图像，我们将它们全部用于我们的自适应任务。

实现细节。我们采用Simple Baseline[39]和ResNet-101[9]作为HPE骨干网，遵循之前的工作[12,16]。在预训练过程中，我们使用Adam[17]优化器进行了40个epoch，每个epoch有500次迭代，初始学习率为 $1e-4$ 。学习率在25次时降至 $1e-5$ 。对于适应过程，我们执行了40个epoch和30,000次迭代。在步骤A中，我们选择 F^{sr} 的初始学习率为 $1e-4$ ， G^{in} 的初始学习率为 $1e5$ 。步骤B设置 G^{tg} 的初始学习率为 $1e-4$ ， F^{tg} 的初始学习率为 $1e-3$ 。我们对学习率调度器使用了与[12]中相同的退火策略。对于超参数，我们选择 $\alpha = 0.7$ ， $\beta = 0.5$ ， $\gamma = 0.85$ 。并利用中间模型进行最终推理。

结果

Table 1. PCK@0.05 on RHD \rightarrow H3D Task

Method	SF	MCP	PIP	DIP	Fin	All
Source-only	-	67.4	64.2	63.3	54.8	61.8
CC-SSL [25] (CVPR'20)	×	81.5	79.9	74.4	64.0	75.1
MDAM [19] (CVPR'21)	×	82.3	79.6	72.3	61.5	74.1
RegDA [12] (CVPR'21)	×	79.6	74.4	71.2	62.9	72.5
UniFrame [16] (ECCV'22)	×	86.7	84.6	78.9	68.1	79.6
RegDA-SF [12] (CVPR'21)	✓	71.2	66.8	66.2	58.5	66.9
UniFrame-SF [16] (ECCV'22)	✓	84.8	84.2	77.0	68.0	77.9
SHOT [21] (ICML'20)	✓	77.5	69.8	70.6	64.2	73.7
MMT [6] (ICLR'20)	✓	78.7	74.1	72.5	65.3	75.3
SHOT++ [22] (TPAMI'22)	✓	85.3	85.1	78.2	67.6	78.9
Ours	✓	88.4	89.2	80.9	71.4	82.2

Table 2. PCK@0.05 on RHD \rightarrow FreiHand Task

Method	SF	MCP	PIP	DIP	Fin	All
Source-only	-	35.2	50.1	54.8	50.7	46.8
CC-SSL [25] (CVPR'20)	\times	37.4	48.2	50.1	46.5	43.8
MDAM [19] (CVPR'21)	\times	32.3	48.1	51.7	47.3	45.1
RegDA [12] (CVPR'21)	\times	40.9	55.0	58.2	53.1	51.1
UniFrame [16] (ECCV'22)	\times	43.5	64.0	67.4	62.4	58.5
RegDA-SF [12] (CVPR'21)	\checkmark	38.6	52.9	57.6	54.3	49.5
UniFrame-SF [16] (ECCV'22)	\checkmark	40.6	62.5	61.0	60.2	55.7
SHOT [21] (ICML'20)	\checkmark	40.4	61.3	60.5	58.1	53.3
MMT [6] (ICLR'20)	\checkmark	39.6	60.4	60.0	57.8	52.6
SHOT++ [22] (TPAMI'22)	\checkmark	41.0	62.8	62.7	59.3	55.8
Ours	\checkmark	43.7	65.9	66.6	63.1	58.8

Table 3. PCK@0.05 on SURREAL \rightarrow Human3.6M Task

Method	SF	Sld	Elb	Wrist	Hip	Knee	Ankle	All
Source-only	-	69.4	75.4	66.4	37.9	77.3	77.7	67.3
CC-SSL [25] (CVPR'20)	\times	44.3	68.5	55.2	22.2	62.3	57.8	51.7
MDAM [19] (CVPR'21)	\times	51.7	83.1	68.9	17.7	79.4	76.6	62.9
RegDA [12] (CVPR'21)	\times	73.3	86.4	72.8	54.8	82.0	84.4	75.6
UniFrame [16] (ECCV'22)	\times	78.1	89.6	81.1	52.6	85.3	87.1	79.0
RegDA-SF [12] (CVPR'21)	\checkmark	70.6	82.0	69.8	43.3	79.1	79.4	71.5
UniFrame-SF [16] (ECCV'22)	\checkmark	74.3	84.8	72.5	45.9	81.5	85.2	75.1
SHOT [21] (ICML'20)	\checkmark	73.0	83.2	71.9	46.7	79.8	82.4	73.3
MMT [6] (ICLR'20)	\checkmark	73.2	83.5	72.4	45.1	80.8	83.9	73.9
SHOT++ [22] (TPAMI'22)	\checkmark	75.0	85.2	76.7	45.3	82.8	85.0	75.3
Ours	\checkmark	77.9	88.8	80.4	52.3	84.2	86.9	78.7

Table 4. PCK@0.05 on SURREAL \rightarrow LSP Task

Method	SF	Sld	Elb	Wrist	Hip	Knee	Ankle	All
Source-only	-	51.5	65.0	62.9	68.0	68.7	67.4	63.9
CC-SSL [25] (CVPR'20)	\times	36.8	66.3	63.9	59.6	67.3	70.4	60.7
MDAM [19] (CVPR'21)	\times	61.4	77.7	75.5	65.8	76.7	78.3	69.2
RegDA [12] (CVPR'21)	\times	62.7	76.7	71.1	81.0	80.3	75.3	74.6
UniFrame [16] (ECCV'22)	\times	69.2	84.9	83.3	85.5	84.7	84.3	82.0
RegDA-SF [12] (CVPR'21)	\checkmark	54.8	70.5	67.6	65.4	73.2	70.0	66.5
UniFrame-SF [16] (ECCV'22)	\checkmark	68.4	80.5	79.1	82.7	80.8	81.0	78.8
SHOT [21] (ICML'20)	\checkmark	63.5	72.7	66.5	78.4	79.7	73.2	72.4
MMT [6] (ICLR'20)	\checkmark	60.9	70.9	70.3	81.1	79.3	72.8	71.5
SHOT++ [22] (TPAMI'22)	\checkmark	69.5	81.7	80.9	84.0	82.3	79.7	79.9
Ours	\checkmark	70.7	85.4	83.8	86.6	85.2	85.0	83.2

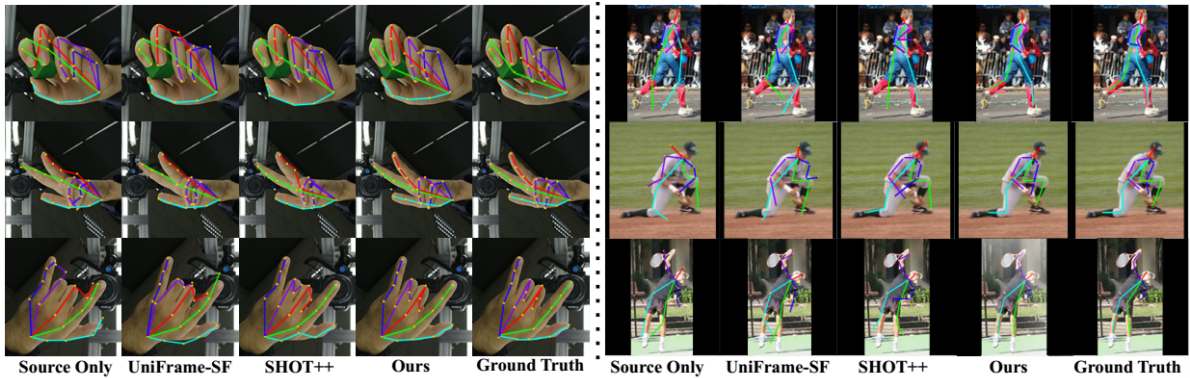


Figure 6. Qualitative results on H3D dataset (left) and LSP dataset (right).

表一和表二是手部任务。表三和表四是人体任务。

为了评估所提出方法的准确性，我们采用了先前工作中使用的正确关键点百分比(PCK)度量。我们将正确预测的比率设置为5%，并在表1-4中报告PCK@0.05。

除了整体关键点精度外，我们还将关节划分为几个部分段，并使用特定指标衡量其性能。对于21个关键点的手部骨骼，我们使用掌指关节(MCP)、近端指间关节(PIP)、远端指间关节(DIP)和指尖(Fin)指标。对于18个关键点的人体骨骼，我们选择肩膀(Sld)，肘部(Elb)，手腕，臀部，膝盖和脚踝指标。这些特定于细分市场的指标可以对模型的性能进行更详细的评估。

消融实验

结构

源保护模块(SP)和目标相关模块(TR)，这里我们重点讨论它们的功能。此外，我们使用MMT[6]作为基线

Table 5. Ablation of Frameworks on RHD → H3D

Method	MCP	PIP	DIP	Fin	All
MMT [6]	78.7	74.1	72.5	65.3	75.3
MMT [6] +TR	81.3	77.5	76.0	67.7	79.1
SP+TR (Ours)	88.4	89.2	80.9	71.4	82.2

Table 6. Ablation of Frameworks on SURREAL → Human3.6M

Method	Sld	Elb	Wrist	Hip	Knee	Ankle	All
MMT [6]	73.2	83.5	72.4	45.1	80.8	83.9	73.9
MMT [6] +TR	74.4	86.5	75.7	48.9	81.3	85.1	76.2
SP+TR (Ours)	77.9	88.8	80.4	52.3	84.2	86.9	78.7

损失

我们在基线中保留了源保护模块中的Lf t和目标相关模块中的Lcon。

Table 7. Ablation of Losses on RHD \rightarrow H3D

Method	MCP	PIP	DIP	Fin	All
Baseline	85.3	86.8	78.5	68.1	79.4
\mathcal{L}_{res}	85.5	87.9	79.3	70.7	80.1
\mathcal{L}_{cst}	85.5	89.0	78.4	70.1	80.9
\mathcal{L}_{im}	85.8	87.2	78.7	69.8	80.3
$\mathcal{L}_{cst} \& \mathcal{L}_{im}$	86.9	88.8	80.1	70.3	81.4
$\mathcal{L}_{res} \& \mathcal{L}_{cst} \& \mathcal{L}_{im}$	88.4	89.2	80.9	71.4	82.2

Table 8. Ablation of Losses on SURREAL \rightarrow Human3.6M

Method	Sld	Elb	Wrist	Hip	Knee	Ankle	All
Baseline	75.4	86.1	76.8	46.5	83.0	85.0	75.8
\mathcal{L}_{res}	76.0	86.9	77.5	47.8	83.3	85.2	76.6
\mathcal{L}_{cst}	76.6	87.4	78.8	50.6	83.5	86.0	77.2
\mathcal{L}_{im}	76.2	87.1	78.3	49.7	83.2	85.7	76.9
$\mathcal{L}_{cst} \& \mathcal{L}_{im}$	77.3	87.7	79.2	51.2	83.6	86.4	78.0
$\mathcal{L}_{res} \& \mathcal{L}_{cst} \& \mathcal{L}_{im}$	77.9	88.8	80.4	52.3	84.2	86.9	78.7

我们比较了基于热图的方法和基于向量的方法的两种损失函数:对比学习(\mathcal{L}_{cst})和信息最大化(\mathcal{L}_{im})。我们将基于热图的对比学习和信息最大化分别表示为HBCL和HBIM, 将基于向量的对比学习和信息最大化分别表示为VBCL和VBIM。

Table 9. Ablation of Sparsity Reduction on RHD \rightarrow H3D

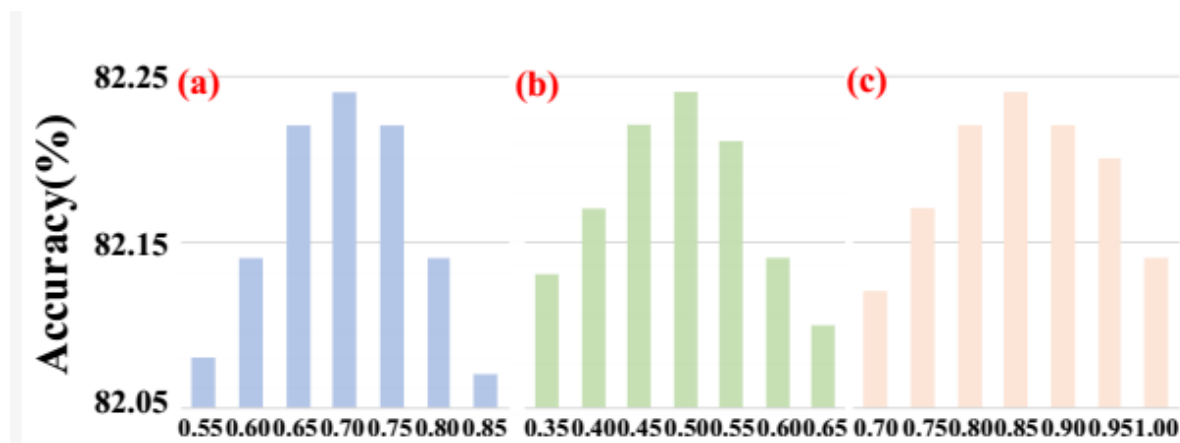
Method	MCP	PIP	DIP	Fin	All
MMT [6]	78.7	74.1	72.5	65.3	75.3
MMT [6] + HBCL	77.4	73.0	70.7	65.5	74.2
MMT [6] + VBCL	81.7	76.6	74.4	67.9	78.7
MMT [6] + HBIM	79.3	74.7	72.9	65.8	75.9
MMT [6] + VBIM	80.3	76.1	73.6	66.5	77.5

Table 10. Ablation of Sparsity Reduction on SURREAL \rightarrow Human3.6M

Method	Sld	Elb	Wrist	Hip	Knee	Ankle	All
MMT [6]	73.2	83.5	72.4	45.1	80.8	83.9	73.9
MMT [6] + HBCL	71.7	82.0	71.8	44.6	80.2	82.4	73.0
MMT [6] + VBCL	75.6	87.0	75.1	49.2	83.3	86.3	76.7
MMT [6] + HBIM	73.9	84.6	72.7	46.0	81.1	83.0	74.3
MMT [6] + VBIM	75.3	86.4	74.3	48.6	82.4	85.7	76.0

参数分析

RHD→H3D参数分析(彩色效果最佳)。a:对 α 的分析。b: β 分析。c: γ 分析



由此可知, $\alpha = 0.7$, $\beta = 0.5$, $\gamma = 0.85$ 是最佳选择。



如图8所示, 结果证明了模型对未知数据集的有效泛化能力, 这意味着自适应过程提高了模型的泛化能力, 超出了原始训练数据集。

补充: 退火策略 (Simulated Annealing) 是一种优化算法, 常用于在大规模搜索空间中找到全局最优解。其基本思想是在搜索过程中逐渐减小温度, 从而使算法在早期能够接受较差的解, 避免陷入局部最优解。

具体而言, 退火策略包括以下步骤:

- (1) 初始化温度 T 和初始解 x ;
- (2) 在每个温度下进行迭代, 每次迭代时随机生成一个新解 x' ;
- (3) 计算新解的目标函数值 $f(x')$ 和当前解的目标函数值 $f(x)$;
- (4) 如果 $f(x') < f(x)$, 则接受新解 x' ;
- (5) 如果 $f(x') > f(x)$, 则以一定概率接受新解, 概率由Metropolis准则决定;
- (6) 降低温度 T ;

重复步骤2-6, 直到满足终止条件。

其中, Metropolis准则指出, 如果新解比当前解更劣, 则以概率 $\exp(-(f(x')-f(x))/T)$ 接受新解。这个概率会随着温度的降低而逐渐减小, 从而使算法在后期更加趋向于接受更优的解。