SONY

NEURAL INFORMATION PROCESSING SYSTEMS

CHASE: Learning Convex Hull Adaptive Shift for Skeleton-based Multi-Entity Action Recognition

NeurIPS 2024

Yuhang Wen, Mengyuan Liu*, Songtao Wu, Beichen Ding*

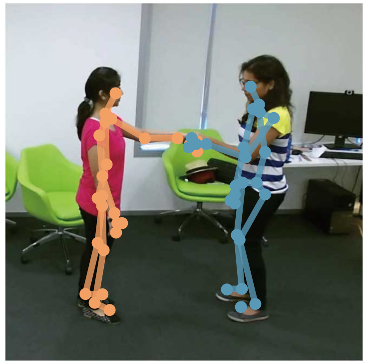
Sun Yat-sen University, Peking University, Sony R&D Center China

https://github.com/Necolizer/CHASE

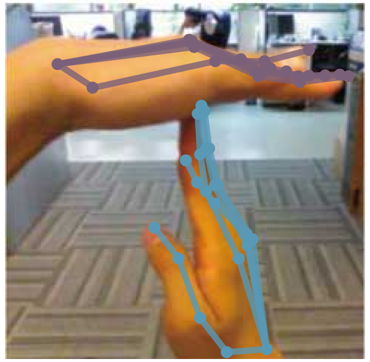
GitHub arXiv

1. Motivation

Multi-Entity Actions

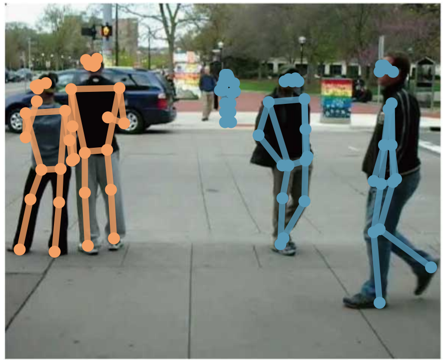


Person-Person Interactions

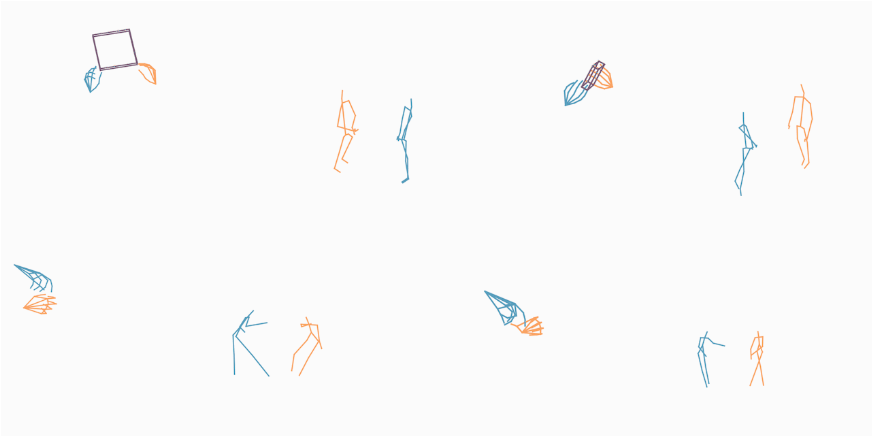


Hand-Hand Interactions

Hand-Object Interactions



Group Activities

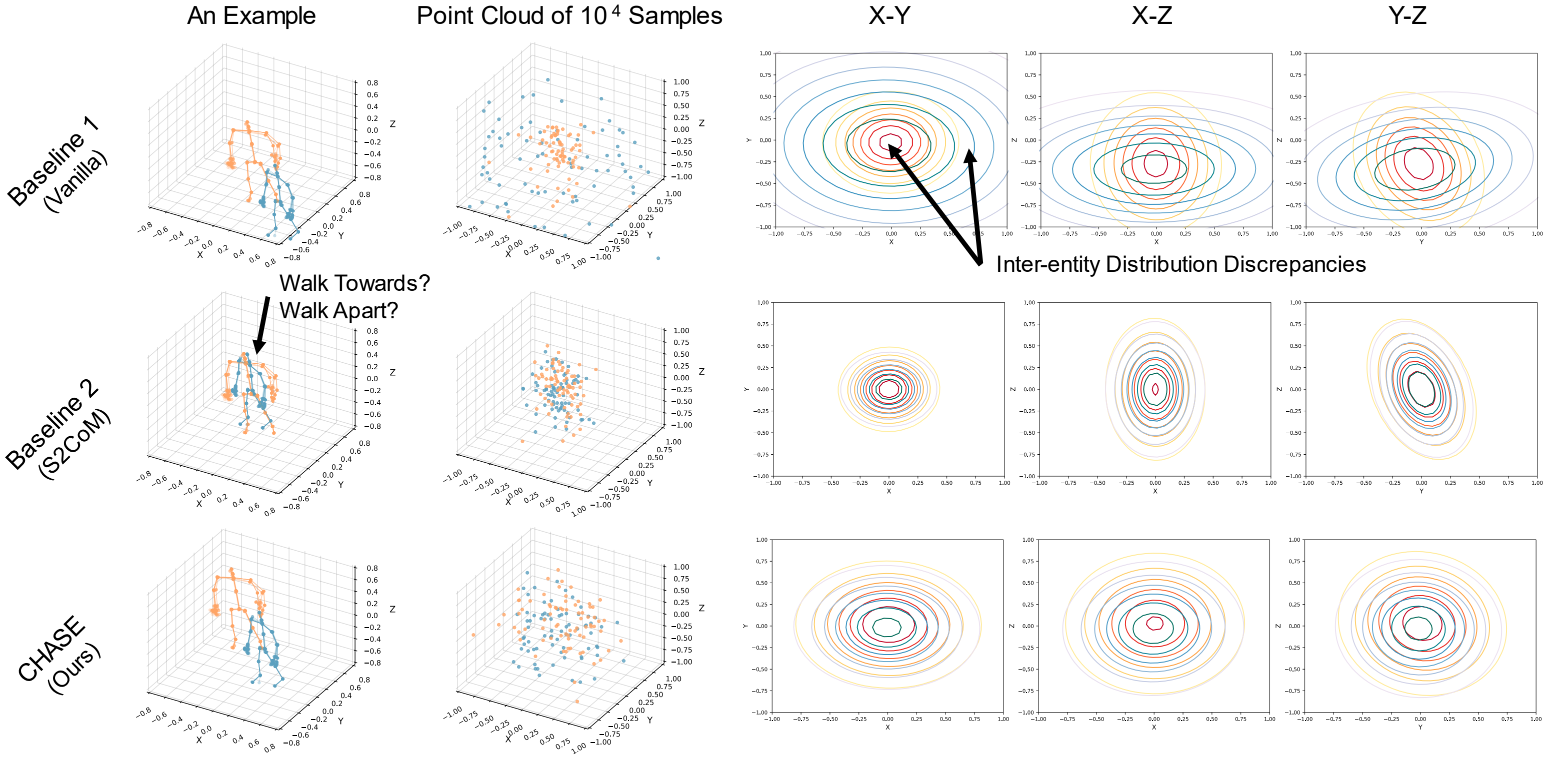


··· and more.

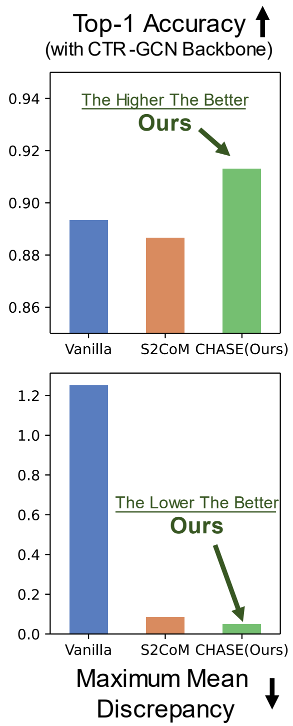
There are many existing benchmarks on interaction recognition.

- But why did almost all skeleton-based methods **limit themselves to one specific type of interactions**?
- Can we treat all these 3D interactive skeletal data **in a general view**?
- More importantly, is there a way we could **solve this general multi-entity problem in a unified manner**?

1. Motivation



(a) Estimated Distributions of Point Clouds from Skeleton Sequences



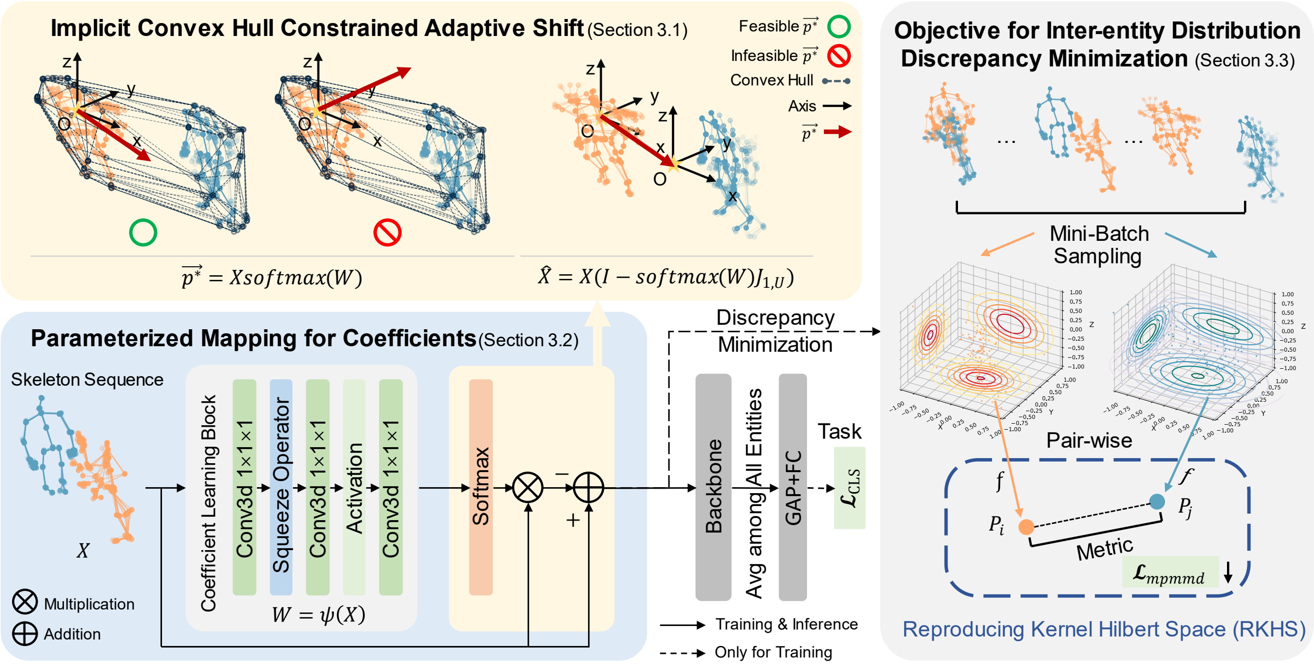
(b) Experimental Results

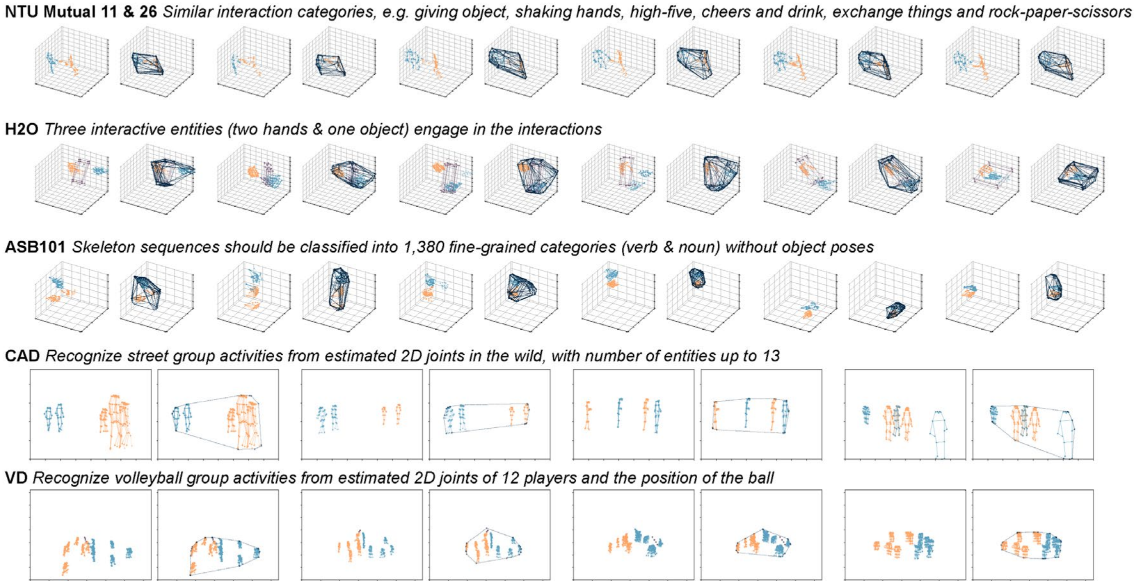
We aim to recognize multi-entity actions using single-entity classifiers with late fusion strategy, which is a unified way to solve this general (in-the-wild) interaction learning problem.

However, we discover **inter-entity distribution discrepancies (entity bias)** in multi-entity skeletons. This is the crux towards better understanding of multi-entity actions. **It explains:**

- **why multi-entity action modeling usually diverges from the single-entity one**
- **why models tailored for individuals get unsatisfactory performance in this scenario**

2. Method





CHASE consists of **a learnable parameterized network** and **an auxiliary objective**.

1) Implicit Convex Hull Constrained Adaptive Shift

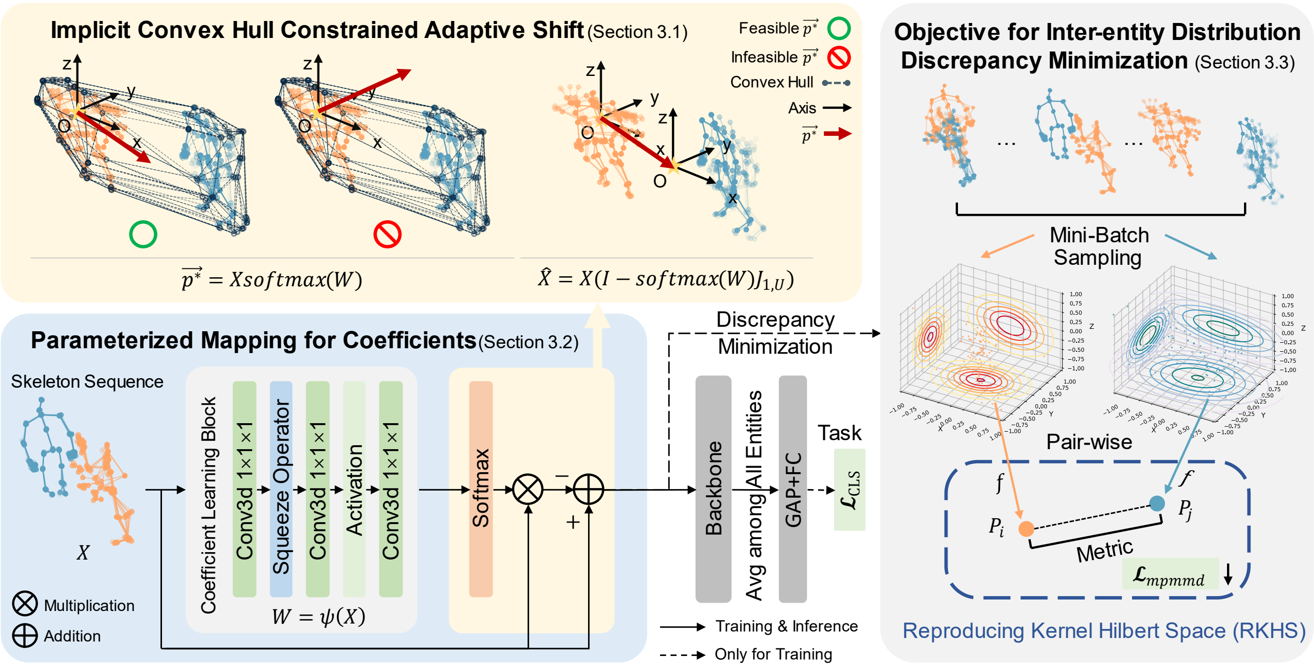
$$\vec{p^*} = X\text{softmax}(W)$$

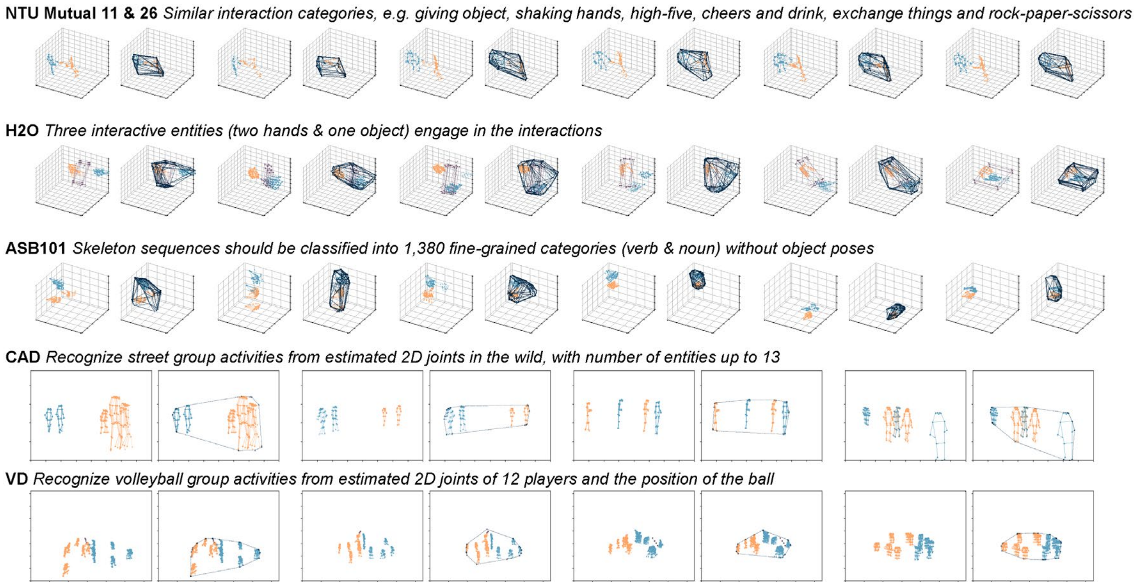
X: Spatiotemporal U keypoints of a multi-entity skeleton sequence. W: A learnable weight matrix. S: Convex Hull of X.

A proof to a simple yet crucial proposition: the new origin is a point that lies in the minimal convex set containing X. It ensures sample-adaptivity and plausibility.

$$\hat{X} = X(I - \text{softmax}(W)J_{1,U}) \qquad \tilde{S} = \left\{ \sum_{i=1}^{U} \tilde{\alpha}_i \vec{p}_i \,\middle|\, \vec{p}_i \in X, \sum_{i=1}^{U} \tilde{\alpha}_i = 1, \tilde{\alpha}_i \in (0,1) \right\} \subset S$$

2. Method





2) Parameterized Mapping for Coefficients

$$W = \psi(X) = W_3\delta(W_2\phi(W_1X + b))$$

A lightweight parameterization of the mapping from skeleton sequences to their specific coefficients in convex combinations, further improving sample-adaptivity.

3) Mini-batch Pair-wise Maximum Mean Discrepancy

$$\mathrm{MMD}(P, Q) = \sup_{\|f\|_{\mathcal{H}} \leq 1} (\mathbb{E}[f(x)] - \mathbb{E}[f(y)])$$

$$\mathbb{E}_{r(z)}[\mathrm{MMD}(z)] = \sum_{i=1}^{E-1} \sum_{j=i+1}^{E} \mathrm{MMD}(P^i, P^j)/\mathrm{C}(E, 2)$$

$$\mathbb{E}_{r(z)}[\mathrm{MMD}(z)] \approx \frac{1}{M} \sum_{m=1}^{M} \mathrm{MMD}(z_m)$$

An auxiliary objective aimed at minimizing the inter-entity distribution discrepancies.

3. Experiments

Method	Venue	NTU Mutual 26(%)		NTU Mutual 11(%)	
		X-Sub	X-Set	X-Sub	X-View
GDCN [11]	TPAMI'23	85.80	92.10	-	-
SkeleTR [76]	ICCV'23	87.80	88.30	94.80	97.70
ISTA-Net [35]	IROS'23	90.56 (±0.08)	91.72 (±0.30)	-	-
AHNet-Large [83]	PR'24	86.43	86.64	90.85	93.38
me-GCN [77]	arXiv'24	90.00	90.00	95.50	98.20
CTR-GCN [36]	ICCV'21	89.32 (±0.06)	90.19 (±0.17)	95.94 (±0.36)	98.32 (±0.29)
+ CHASE (Ours)	-	**91.30** ↑1.98 (±0.22)	**92.34** ↑2.15 (±0.10)	**96.45** ↑0.51 (±0.05)	**98.83** ↑0.51 (±0.13)
InfoGCN [37](k=1)	CVPR'22	90.22 (±0.13)	91.13 (±0.16)	95.51 (±0.10)	97.76 (±0.22)
+ CHASE (Ours)	-	**91.86** ↑1.64 (±0.05)	**92.41** ↑1.28 (±0.34)	**96.35** ↑0.84 (±0.18)	**98.25** ↑0.49 (±0.25)
STSA-Net [40]	Neuro.'23	88.41 (±0.01)	90.19 (±0.11)	95.96 (±0.09)	98.47 (±0.09)
+ CHASE (Ours)	-	**89.77** ↑1.36 (±0.18)	**91.54** ↑1.35 (±0.12)	**96.63** ↑0.68 (±0.10)	**98.73** ↑0.26 (±0.08)
HD-GCN [38](CoM=1)	ICCV'23	88.25 (±0.44)	90.08 (±0.12)	95.58 (±0.10)	97.93 (±0.07)
+ CHASE (Ours)	-	**90.81** ↑2.56 (±0.13)	**92.06** ↑1.97 (±0.21)	**96.22** ↑0.64 (±0.05)	**98.31** ↑0.38 (±0.07)

Method	Venue	H2O(%)	ASB101(%)	CAD(%)	VD(%)
AT [26]	CVPR'20	-	-	-	92.30
ISTA-Net [35]	IROS'23	89.09 (±1.21)	28.01 (±0.06)	87.16 (±2.55)	91.40 (±0.23)
H2OTR [80]	CVPR'23	90.90	-	-	-
EffHandEgoNet [81]	arXiv'24	91.32	-	-	-
AHNet-Large [83]	PR'24	-	-	89.32	84.31
CTR-GCN [36]	ICCV'21	81.68 (±0.85)	27.83 (±0.45)	80.45 (±2.29)	92.66 (±0.21)
+ CHASE (Ours)	-	**91.05** ↑9.37 (±1.98)	**28.03** ↑0.21 (±0.30)	**89.61** ↑9.16 (±0.20)	**92.89** ↑0.24 (±0.15)
InfoGCN [37](k=1)	CVPR'22	76.24 (±3.93)	27.18 (±0.10)	83.07 (±0.46)	91.77 (±0.15)
+ CHASE (Ours)	-	**83.47** ↑7.23 (±2.89)	**27.36** ↑0.18 (±0.12)	**84.18** ↑1.11 (±2.91)	**92.00** ↑0.23 (±0.15)
STSA-Net [40]	Neuro.'23	92.29 (±0.52)	27.70 (±0.19)	80.20 (±3.60)	92.52 (±0.52)
+ CHASE (Ours)	-	**94.77** ↑2.48 (±1.36)	**27.81** ↑0.11 (±0.13)	**85.93** ↑5.73 (±2.46)	**92.78** ↑0.26 (±0.41)
HD-GCN [38](CoM=1)	ICCV'23	72.73 (±0.41)	27.31 (±0.36)	76.93 (±4.38)	91.32 (±0.02)
+ CHASE (Ours)	-	**81.61** ↑8.88 (±1.03)	**27.50** ↑0.19 (±0.24)	**82.39** ↑5.46 (±1.61)	**92.00** ↑0.68 (±0.07)

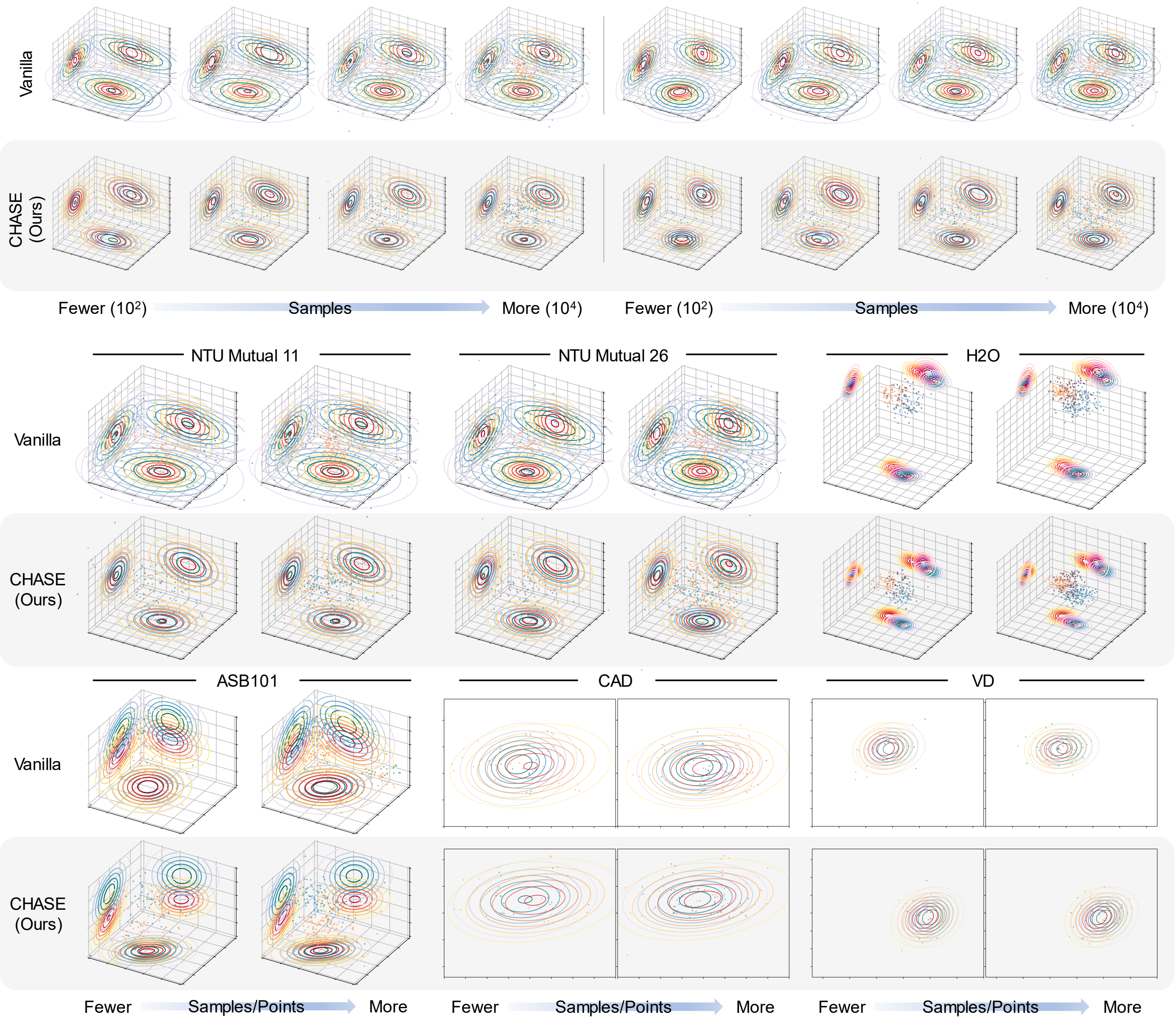
Set	Method	Avg KLD ↓	JSD ↓	BD ↓	HD ↓	MMD ↓
I	Vanilla	1.07 (±0.25)	0.19 (±0.04)	0.25 (±0.06)	0.46 (±0.06)	0.94 (±0.54)
	CHASE (Ours)	**0.39** (±0.09)	**0.08** (±0.02)	**0.10** (±0.02)	**0.30** (±0.03)	**0.05** (±0.02)
II	Vanilla	1.00 (±0.23)	0.18 (±0.04)	0.23 (±0.05)	0.45 (±0.05)	1.03 (±0.60)
	CHASE (Ours)	**0.45** (±0.08)	**0.10** (±0.02)	**0.11** (±0.02)	**0.32** (±0.03)	**0.07** (±0.02)
III	Vanilla	0.72 (±0.14)	0.14 (±0.02)	0.17 (±0.03)	0.39 (±0.04)	1.25 (±0.60)
	CHASE (Ours)	**0.41** (±0.08)	**0.08** (±0.02)	**0.10** (±0.02)	**0.30** (±0.03)	**0.05** (±0.04)
IV	Vanilla	0.75 (±0.14)	0.14 (±0.03)	0.17 (±0.03)	0.40 (±0.04)	1.15 (±0.56)
	CHASE (Ours)	**0.41** (±0.07)	**0.08** (±0.01)	**0.09** (±0.02)	**0.30** (±0.03)	**0.04** (±0.03)

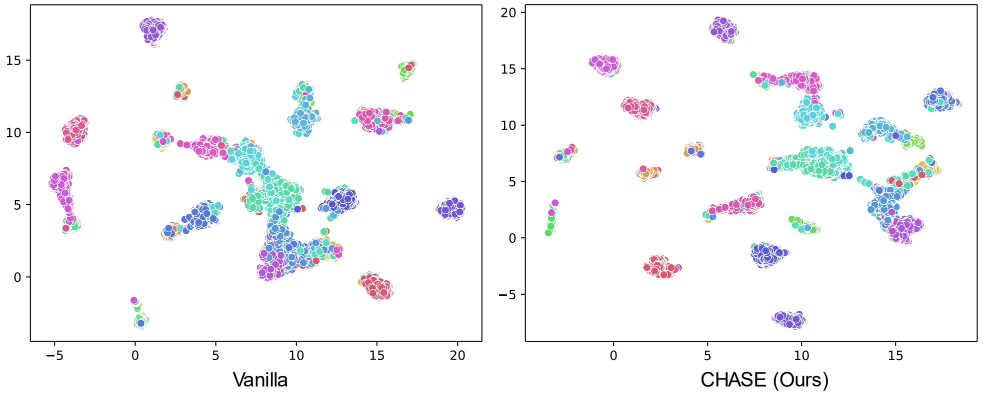
Method	Acc (%)	Δ (%)
Vanilla	89.32 (±0.06)	-
S2CoM	88.66 (±0.26)	−0.67
BatchNorm	89.06 (±0.16)	−0.27
ER [35]	89.34 (±0.15)	+0.02
Aug	89.72 (±0.04)	+0.40
S2CoM†/STD	90.29 (±0.06)	+0.97
S2CoM†	90.79 (±0.10)	+1.47
CHASE (Ours)	**91.30** (±0.22)	**+1.98**

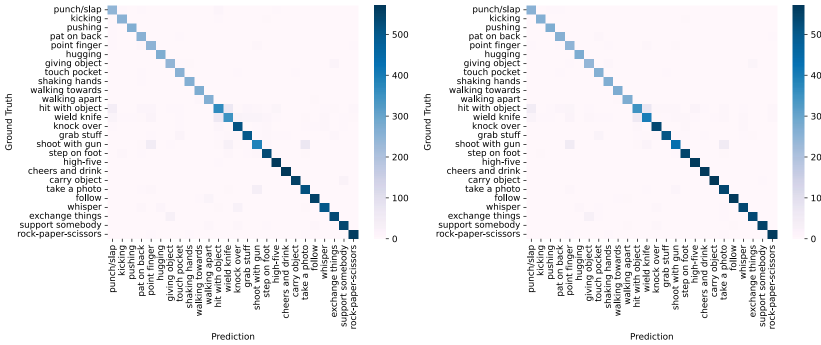
ICHAS		CLB	MPMMD	lr	Acc (%)	Δ (%)
AS	CHC					
✓	✓	✓	✓	0.1	**91.30** (±0.22)	-
✓		✓	✓	0.1	22.65 (±0.35)	−68.65
✓		✓	✓	0.01	86.99 (±0.16)	−4.32
✓	✓		✓	0.1	91.20 (±0.13)	−0.10
✓			✓	0.1	22.75 (±0.12)	−68.56
✓			✓	0.01	23.51 (±0.38)	−67.79
	✓	✓	✓	0.1	20.42 (±0.09)	−70.88
✓	✓	✓		0.1	91.17 (±0.18)	−0.13
				0.1	89.50 (±0.14)	−1.81

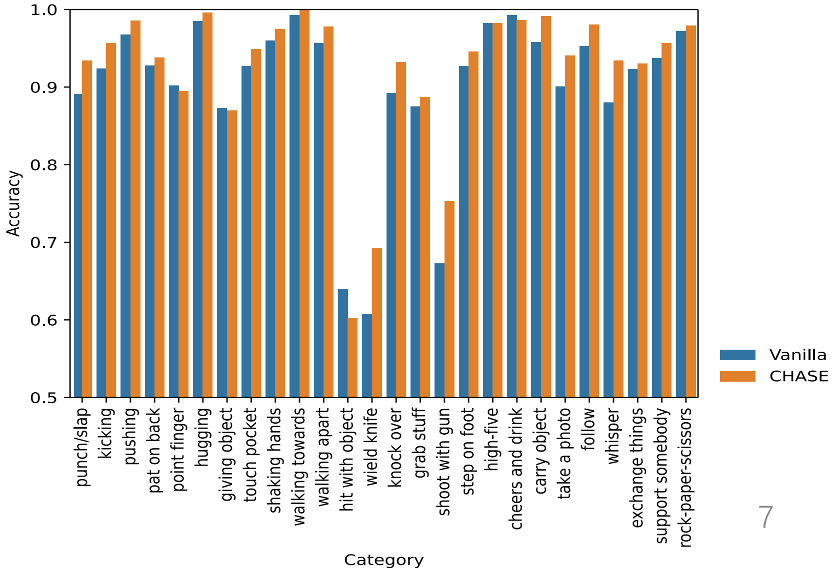
By adopting our proposed CHASE, we can boost the performance of the vanilla counterparts by a noticeable margin in most multi-entity scenarios. CHASE also significantly minimizes discrepancies across all evaluation metrics.

3. Experiments









4. Conclusions

- We discover an interesting observation in multi-entity skeletons: **Entity Bias**.
- Proposed a Convex Hull Adaptive Shift based multi-Entity action recognition method (CHASE), **serving as an additional normalization step for single-entity backbones**.
- Our main insight lies in **the adaptive repositioning of skeleton sequences to mitigate inter-entity distribution gaps**, thereby unbiasing the subsequent classifier and boosting its performance in multi-entity scenarios.

SONY

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you for listening

[NeurIPS 2024] CHASE: Learning Convex Hull Adaptive Shift for Skeleton-based Multi-Entity Action Recognition

Yuhang Wen, Mengyuan Liu*, Songtao Wu, Beichen Ding*

Sun Yat-sen University, Peking University, Sony R&D Center China

https://github.com/Necolizer/CHASE

GitHub

arXiv