One-shot skeleton-based action recognition on strength and conditioning exercises
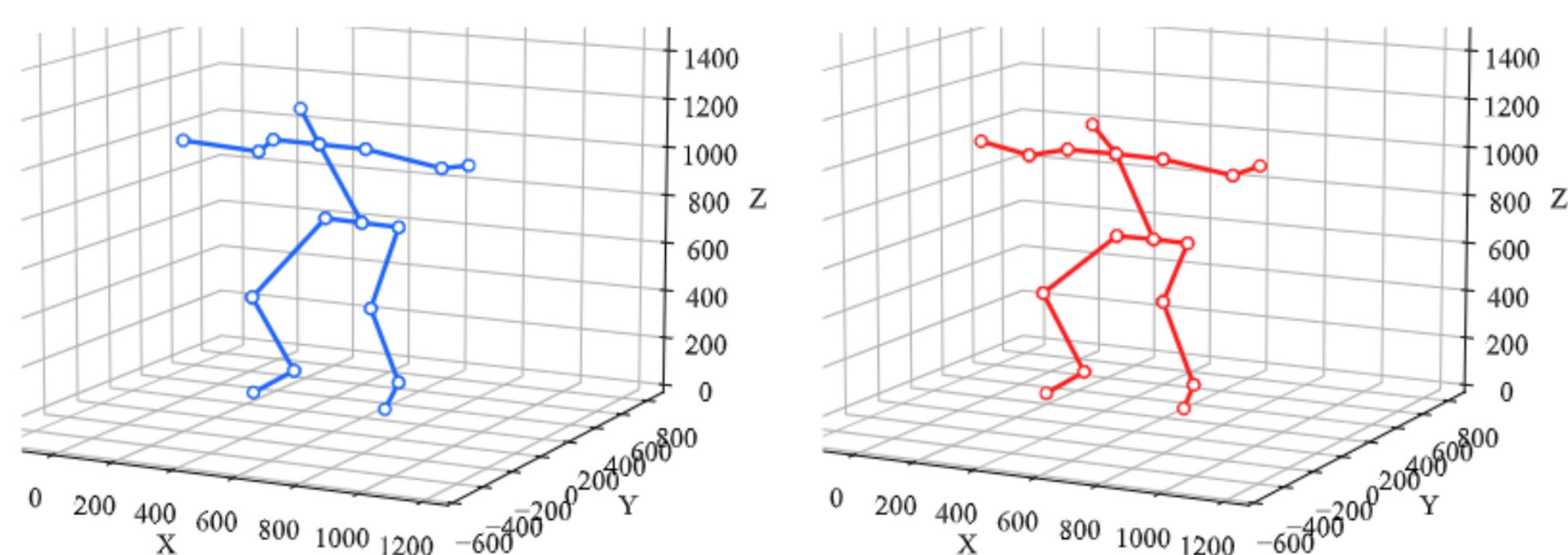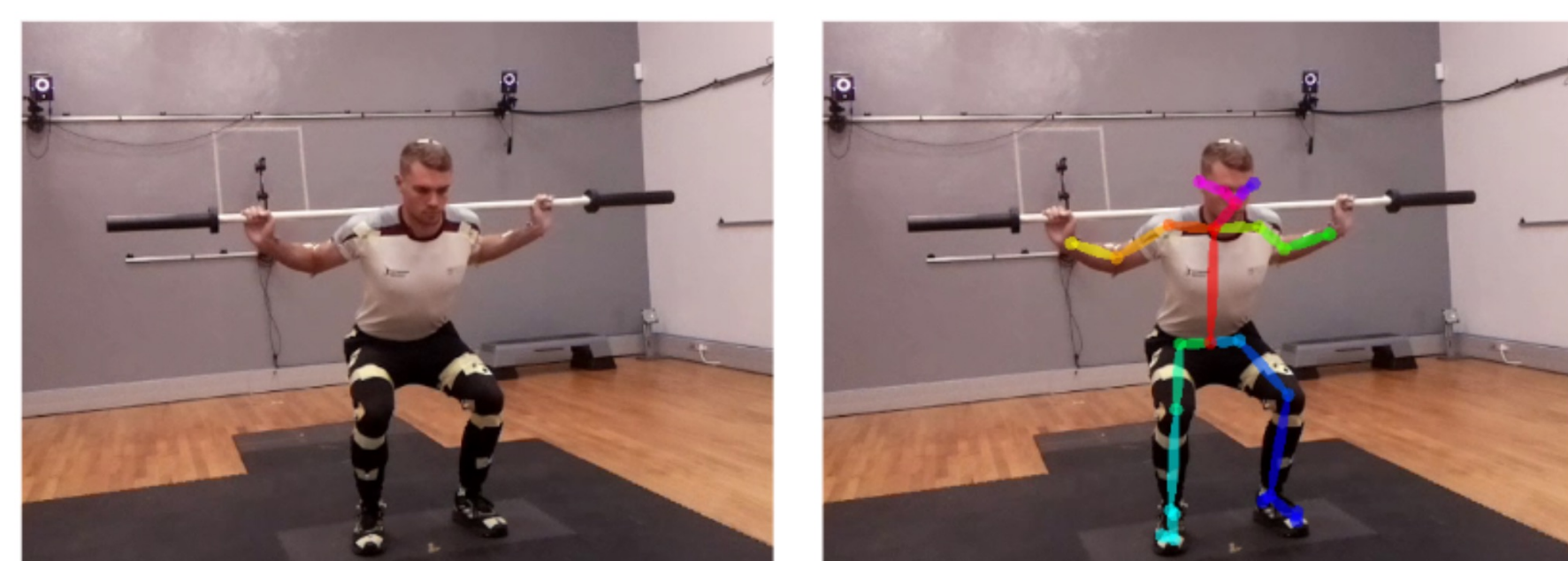
Michael Deyzel, Rensu P. Theart

Stellenbosch University, Stellenbosch, South Africa

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- There is a need for a **practical** system that can **identify** human physical activity to enable intelligent workout feedback and virtual coaching.
- Such a system should be able to classify an athlete's actions from only **limited examples** since it is not feasible to collect a large quantity of human data for every action of interest.
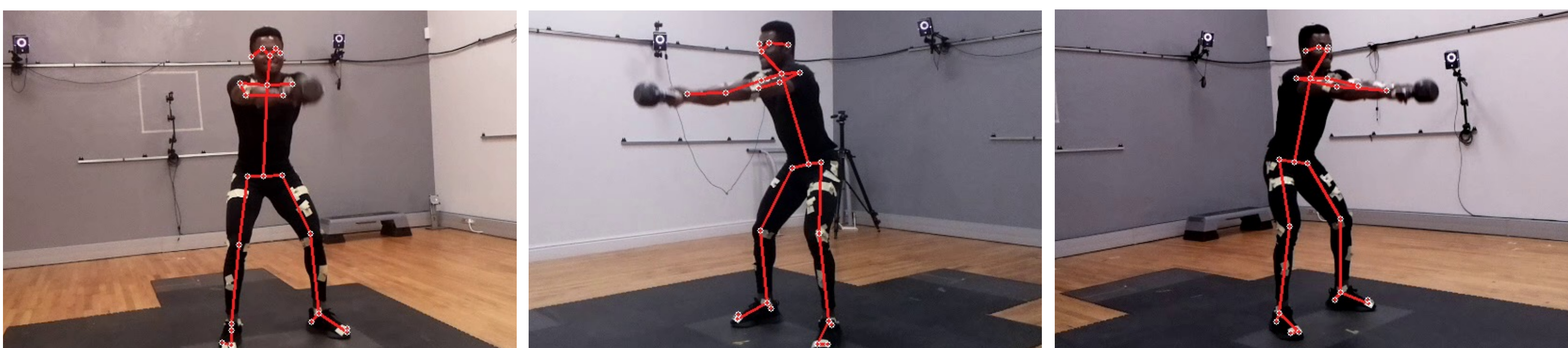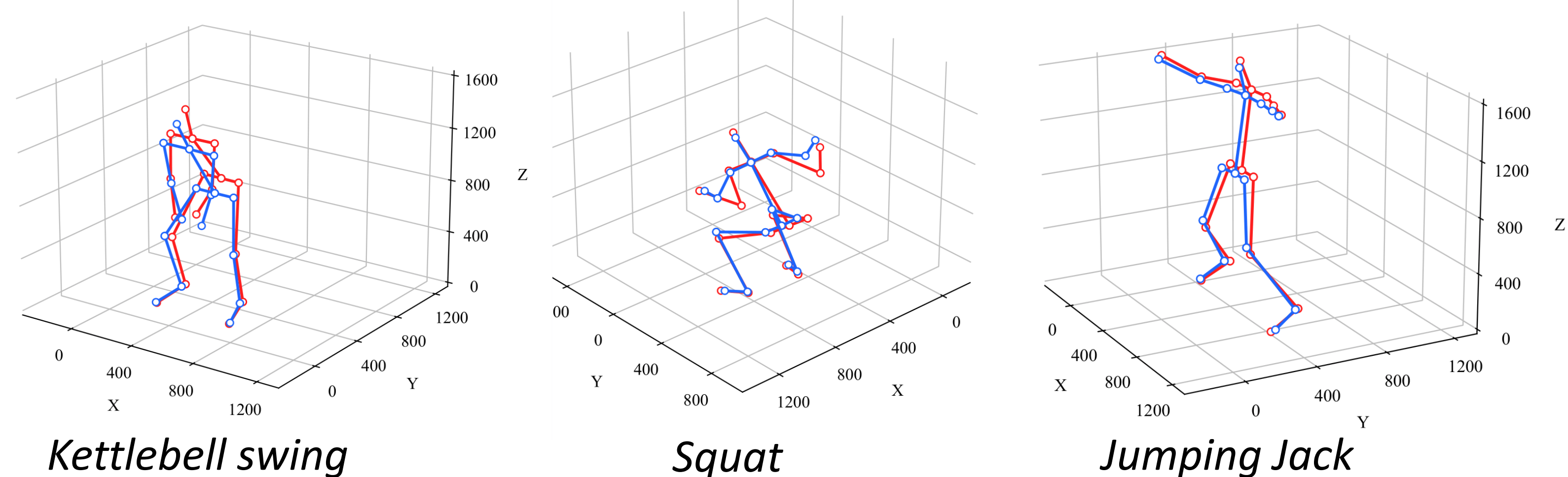
SU-EMD

Stellenbosch University Exercise Motion Dataset: a novel **dataset** of skeleton motion sequences of **seven** common **strength and conditioning** exercises as captured by both a markerless and marker-based motion capture system.
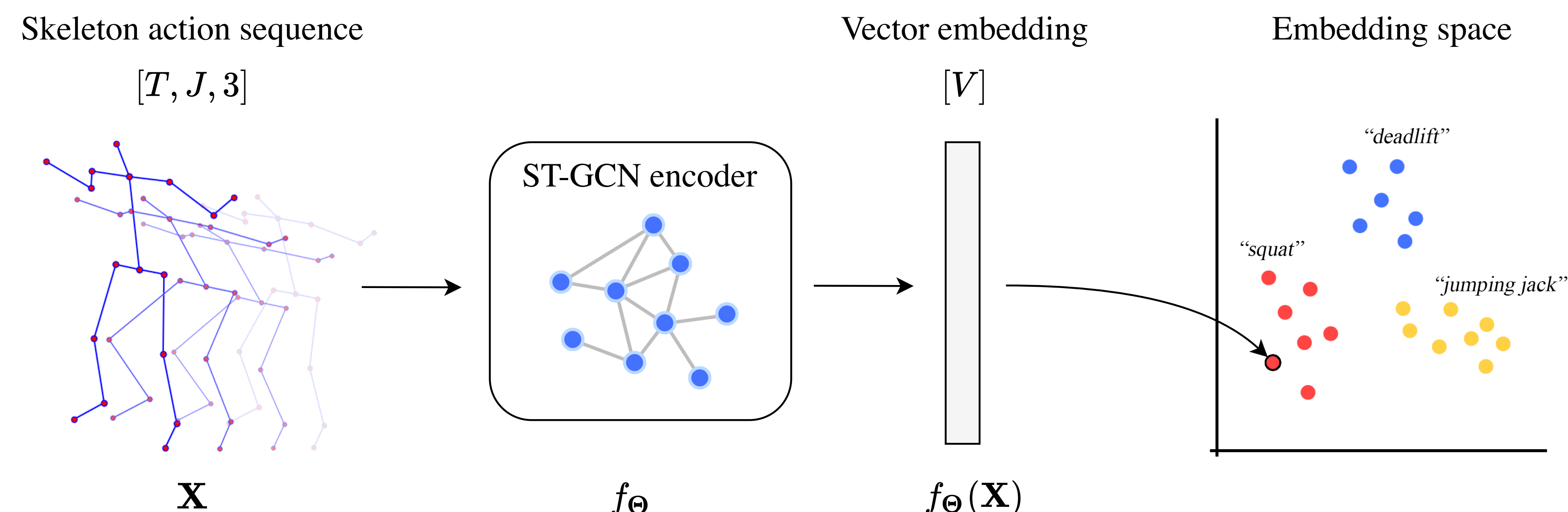
- 420 examples
- 4 subjects
- Skeleton compatible with **NTU RGB+D** and **OpenPose**.

Marker-based   Markerless

Kettlebell swing   *Squat*   *Jumping Jack*

Deep metric learning of exercise motion

Skeleton action sequence $[T, J, 3]$ → ST-GCN encoder → Vector embedding $[V]$ → Embedding space

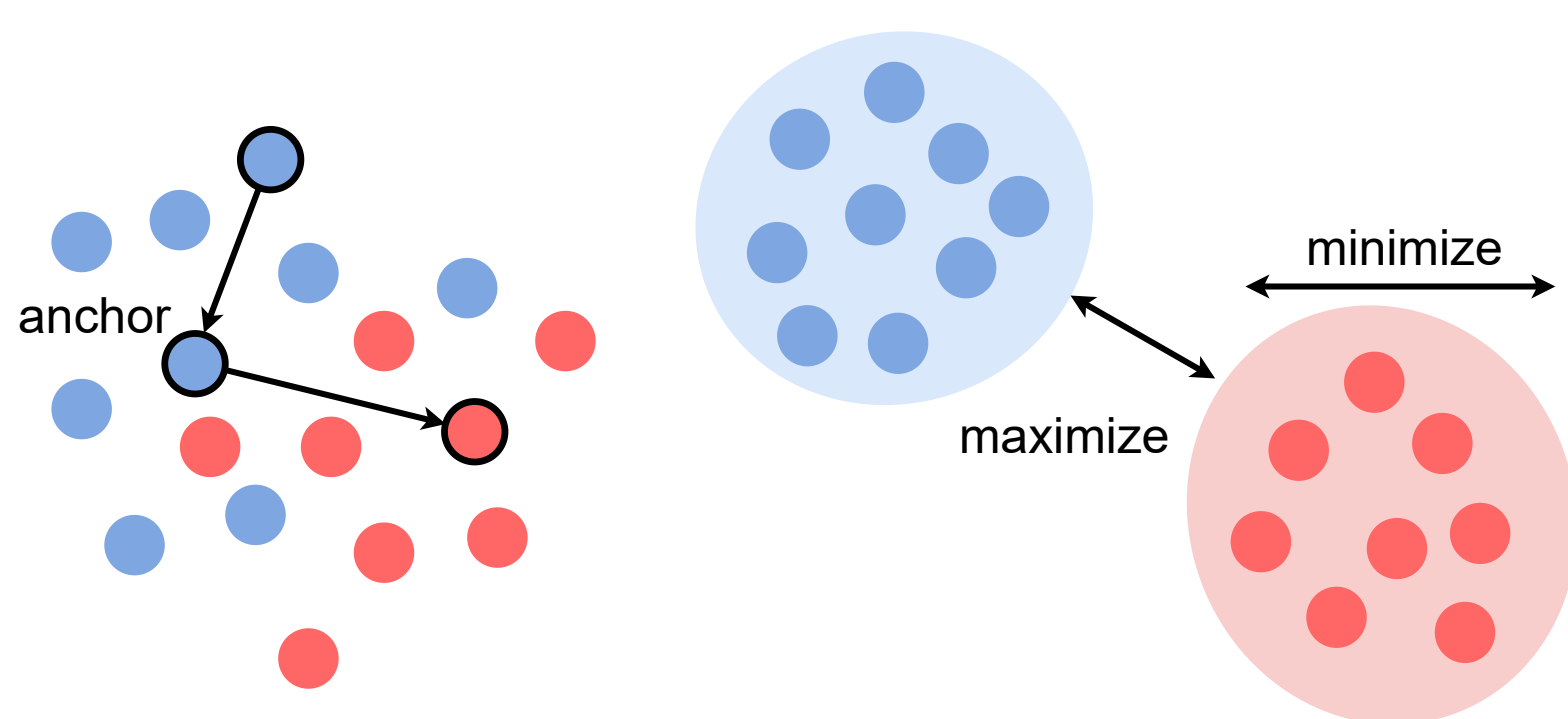$\mathbf{X}$ f_Θ $f_\Theta(\mathbf{X})$

Train a state-of-the-art spatio-temporal **graph convolutional network** (ST-GCN) with a one-shot metric learning paradigm to **learn skeletal motion features** to project dissimilar actions further away and similar actions closer together in the learned metric space.
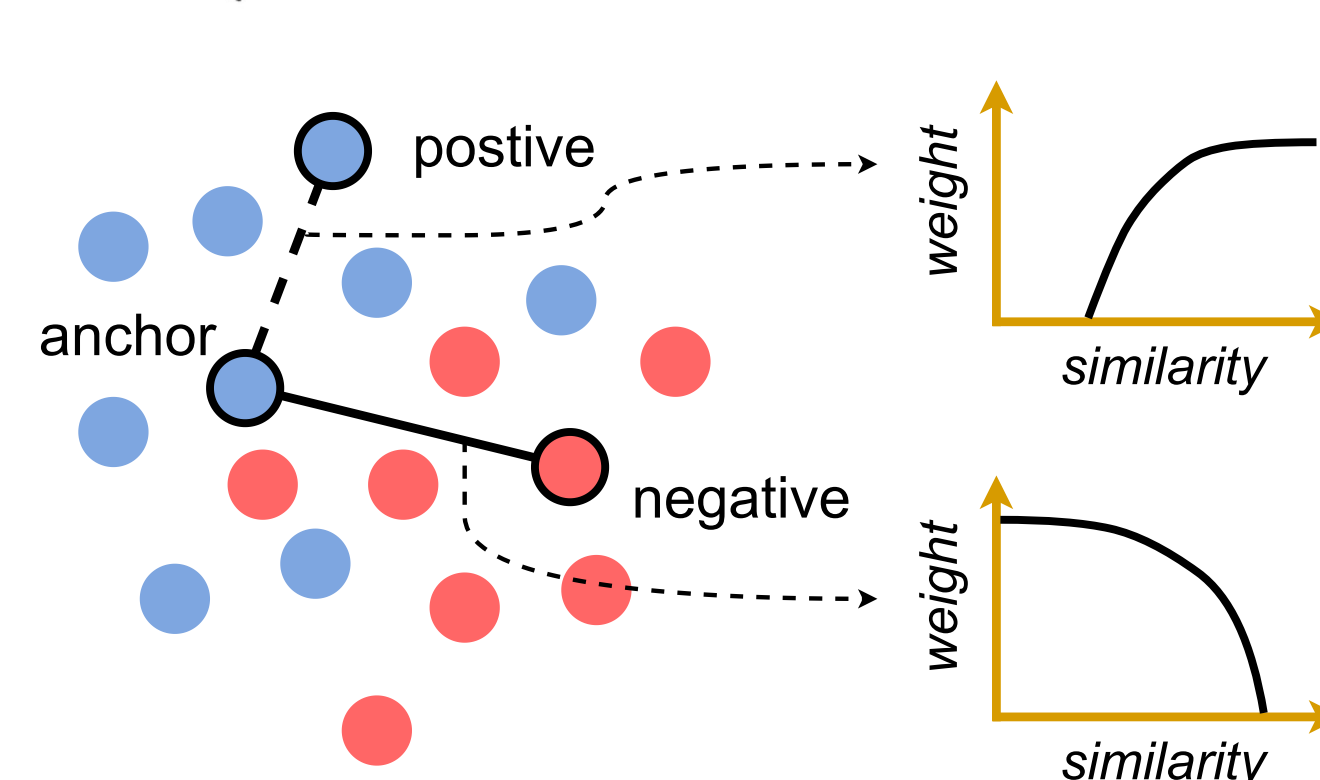
GCN feature extractor f_Θ

$\mathbf{S}$
Input skeleton sequence

$\mathbf{z}$
Feature vector
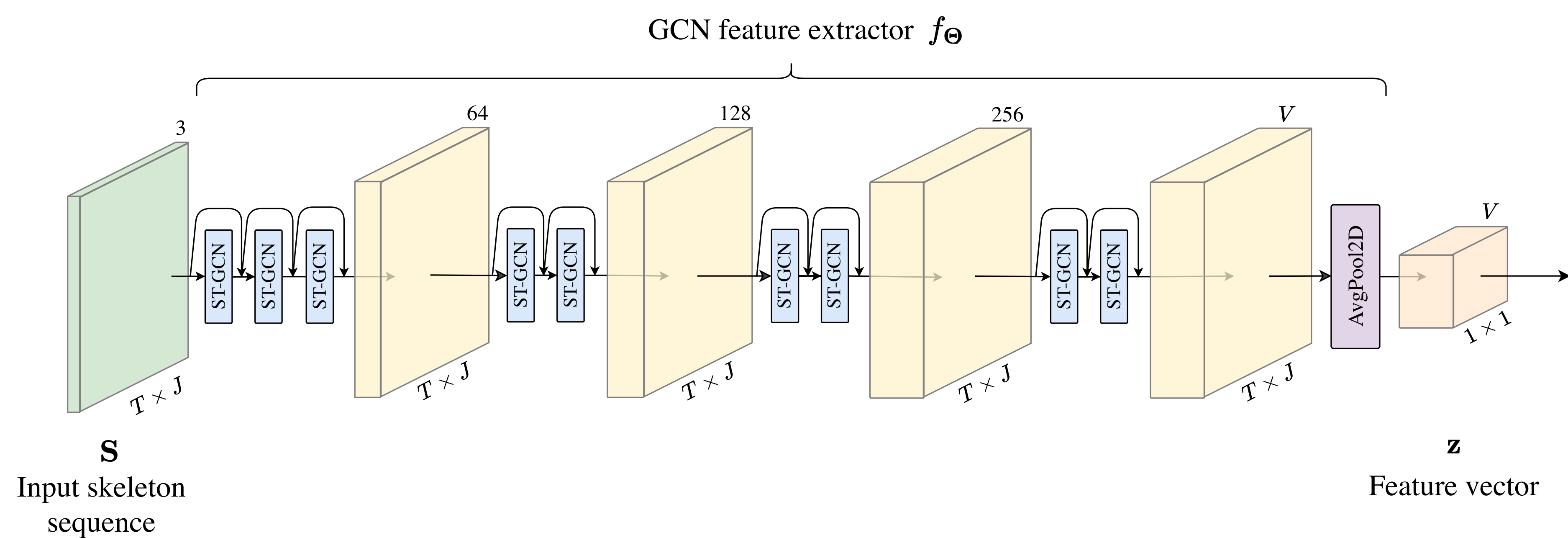
Triplet margin (TM) loss

$$\mathcal{L}_{\text{TM}} = \max\{d(\mathbf{z}_a, \mathbf{z}_p) - d(\mathbf{z}_a, \mathbf{z}_n) + \alpha, 0\}$$

anchor   minimize
maximize

Multi-similarity (MS) loss

$$\mathcal{L}_{\text{MS}} = \frac{1}{B} \sum_{i=1}^{B} \left\{ \frac{1}{\beta} \log[1 + \sum_{k \in \mathcal{P}_i} e^{-\beta(\mathbf{S}_{ik} - \lambda)}] + \frac{1}{\gamma} \log[1 + \sum_{k \in \mathcal{N}_i} e^{\gamma(\mathbf{S}_{ik} - \lambda)}] \right\}$$

positive
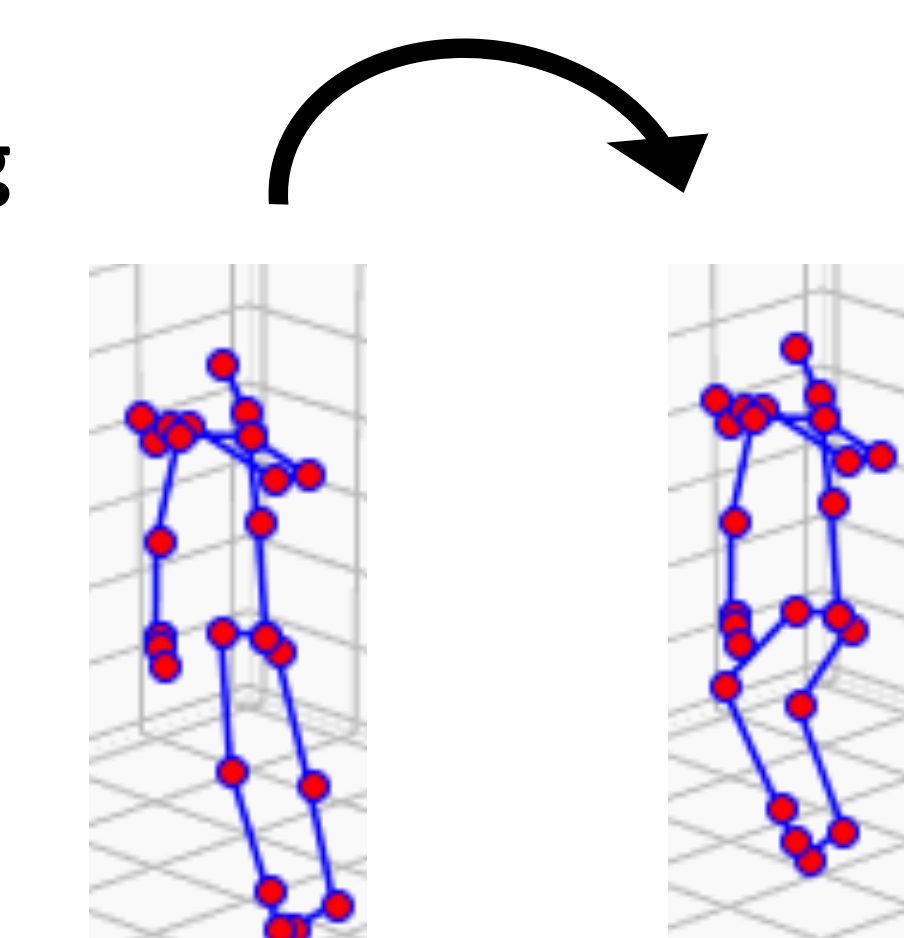anchor   weight
negative   similarity

Experiments

- Pre-train Metric GCN on NTU-RGB+D 120
- Validate on NTU-RGB+D 120 one-shot set
- **Augmentations**: random rotations, random moving, random frame dropping, **random pivoting**
- Vary embedding sizes (128, 512) and losses
- Easy positive sampling (EPS)

Random pivoting

Original   Pivot about knees and hips

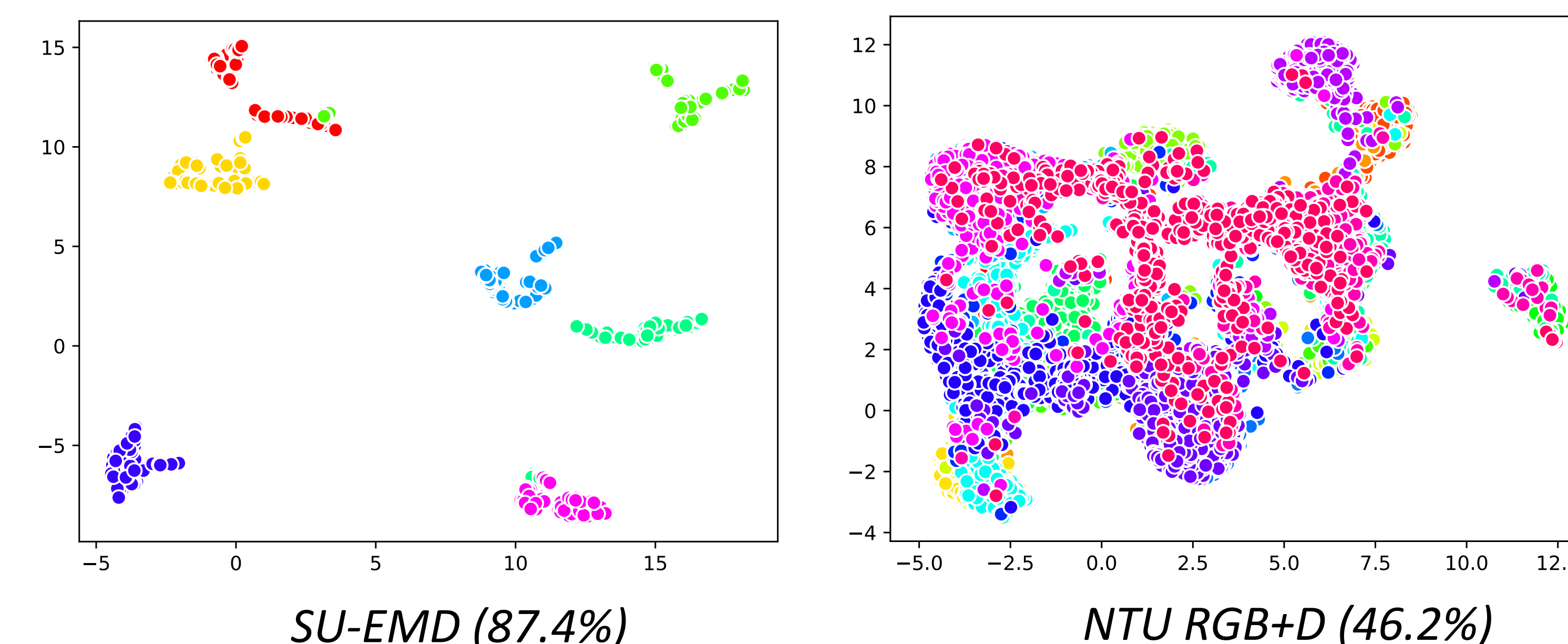| Loss	EPS	Augmentations	Accuracy@1 [%]	
			128	512
TM	No	Rot, Pivot, Drop, Move	36.6	36.7
TM	Yes	Rot, Pivot, Drop, Move	37.0	37.7
TM	Yes	None	37.9	37.5
MS	—	Rot, Pivot, Drop, Move	**46.2**	45.8
MS	—	None	44.1	43.8

Ablation study on NTU RGB+D one-shot test set

Results

- Metric GCN achieves classification accuracy of **87.4%** on all seven **never-before-seen** actions.

SU-EMD (87.4%)   *NTU RGB+D (46.2%)*

- One-shot metric learning can be used to classify sports actions of motion sequences in a virtual coaching system where users cannot provide many expert examples for the enrolment of new actions.