Subspace Kernel Learning on Tensor Sequences

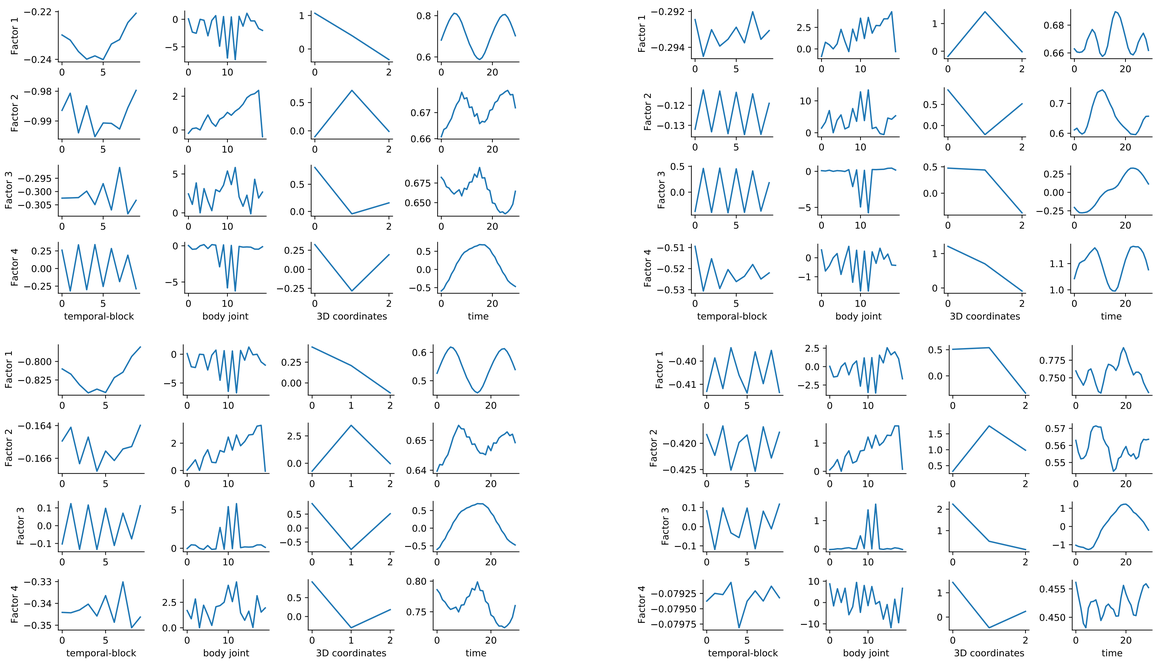
Lei Wang Xi Ding Yongsheng Gao Piotr Koniusz

Griffith UNIVERSITY Queensland, Australia · DATA 61 · CSIRO · UNSW SYDNEY

ICLR International Conference On Learning Representations

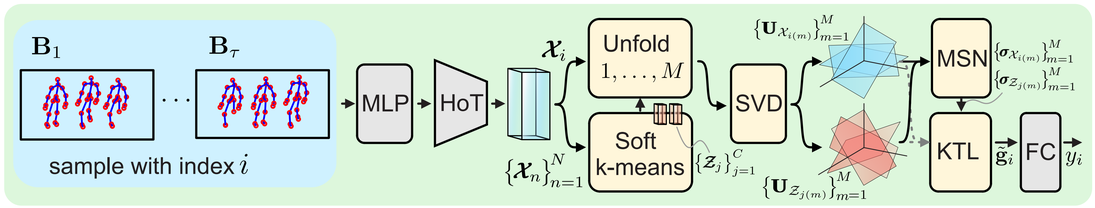
Motivation and key ideas

Motivation. Visualization of Tucker decomposition across tensor modes for the actions *draw x* (left column) and *draw tick* (right column). Each row corresponds to a different subject.



- Mode-wise factor matrices from the Tucker decomposition for the action called *"draw x"*.
- Each row shows one latent factor (from 4 leading factors), and each column corresponds to one tensor mode: *temporal block*, *body joints*, *3D coordinates*, and *time*.
- Structured patterns show interpretable, mode-specific information which motivates our approach.

Overview of the proposed Uncertainty-driven Kernel Tensor Learning (UKTL) pipeline for action recognition. For brevity, we use skeletons as an example.



- Each skeleton sequence is divided into temporal blocks $\mathbf{B}_1, \dots, \mathbf{B}_\tau$, embedded via an MLP, and processed by a Higher-order Transformer (HoT) to obtain feature tensor $\mathcal{X}_i$.
- These tensors undergo mode-m matricization $(1, \dots, M)$ and SVD to extract M subspaces per sample. Soft k-means clustering yields C Nyström pivots, each represented by also M subspaces.
- A Multi-mode SigmaNet (MSN) estimates uncertainty vectors over all subspaces, which are used to regularize kernel computations.
- The Nyström-approximated KTL maps inputs to compact, uncertainty-aware representations, $\tilde{\mathbf{g}}_i$ for final classification. The entire model is trained end-to-end.

Quantitative results

Single-modality evaluation. Results on NTU-60, NTU-120, and Kinetics-Skeleton. UKTL outperforms graph, hypergraph, and transformer models by using uncertainty-aware tensor kernels.

	Method	NTU-60		NTU-120		Kinetics-Skeleton	
		X-Sub(%)	X-View(%)	X-Sub(%)	X-Setup(%)	Top-1(%)	Top-5(%)
Graph-based	TCN	-	-	-	-	20.3	40.0
	ST-GCN	81.5	88.3	70.7	73.2	30.7	52.8
	AS-GCN	86.8	94.2	78.3	79.8	34.8	56.5
	2S-AGCN	88.5	95.1	82.5	84.2	36.1	58.7
	NAS-GCN	89.4	95.7	-	-	37.1	60.1
	Shift-GCN	90.7	96.5	85.9	87.6	-	-
	MS-G3D	91.5	96.2	86.9	88.4	38.0	60.9
	Sym-GNN	90.1	96.4	-	-	37.2	58.1
	SSL	92.8	96.5	84.8	85.7	-	-
	CTR-GCN	92.6	96.7	89.6	91.0	-	-
	FD-GCN	92.8	96.7	89.4	90.7	-	-
	DSDC-GCN	93.0	97.1	89.9	90.6	38.6	**63.4**
Hypergraph-based	Hyper-GNN	89.5	95.7	-	-	37.1	60.0
	DHGCN	90.7	96.0	86.0	87.9	37.7	60.6
	SD-HGCN	90.9	96.7	87.0	88.2	37.4	60.5
	Selective-HCN	90.8	96.6	-	-	38.0	61.1
	Hyper-GCN	91.4	95.5	87.0	88.7	-	-
Transformer-based	ST-TR	90.3	96.3	85.1	87.1	38.0	60.5
	STST	91.9	96.8	-	-	38.3	61.2
	MTT	90.8	96.7	86.1	87.6	37.9	61.3
	4s-GSTN	91.3	96.6	86.4	88.7	-	-
	MAMP	84.9	89.1	78.6	79.1	-	-
	STJD-MP	85.9	90.0	77.1	79.3	-	-
Tensor-based	*Backbone*	90.8	95.8	85.2	87.4	36.7	59.5
	+ KPCA (*baseline*)	92.0	96.8	88.6	90.1	37.1	59.8
	+ TPCA (*baseline*)	91.6	96.8	88.2	90.0	38.0	60.5
	+ KTL	92.5	97.1	88.8	90.3	38.9	61.9
	+ UKTL	**93.1**	**97.3**	**90.0**	**91.4**	**39.2**	62.3

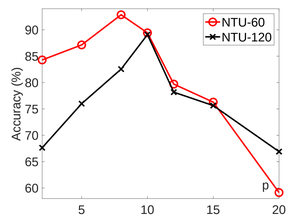
Multi-modal evaluation. Evaluation of single- and multi-modal performance on NTU-60 and NTU-120.

Modality	NTU-60		NTU-120	
	X-Sub (%)	X-View (%)	X-Sub (%)	X-Setup (%)
Skeleton	93.1	97.3	90.0	91.4
RGB	83.2	86.1	79.0	81.0
Depth	85.0	87.5	80.5	82.1
Skeleton + RGB	94.5	97.9	91.3	92.5
Skeleton + Depth	94.8	98.0	91.5	92.7
RGB + Depth	87.2	89.6	82.5	84.3
Skeleton + RGB + Depth	**95.5**	**98.5**	**92.8**	**94.0**

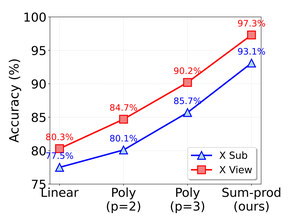
- Skeleton alone provides a strong, competitive baseline, while RGB and Depth demonstrate the framework's flexibility across modalities.
- Fusing multiple modalities consistently improves accuracy, showing that UKTL effectively captures complementary information and inter-modal correlations.

Qualitative results

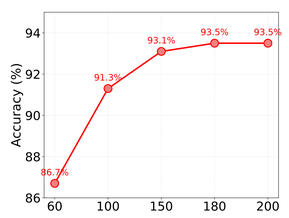
Ablation study evaluating the effects of subspace order (on NTU-60/120), and of kernel choice, Nyström pivots, and kernel composition (on NTU-60) within the UKTL framework.



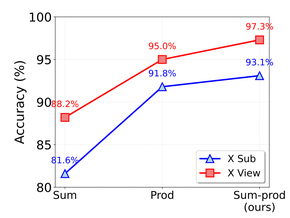
Subspace order p.



Kernel choice.



Number of pivots C.



Kernel composition.