Space-Time Domain Tensor Neural Networks An Application on Human Pose Classification

L-Università ta' Malta
Institute of Digital Games

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
UNIVERSITY OF WEST ATTICA

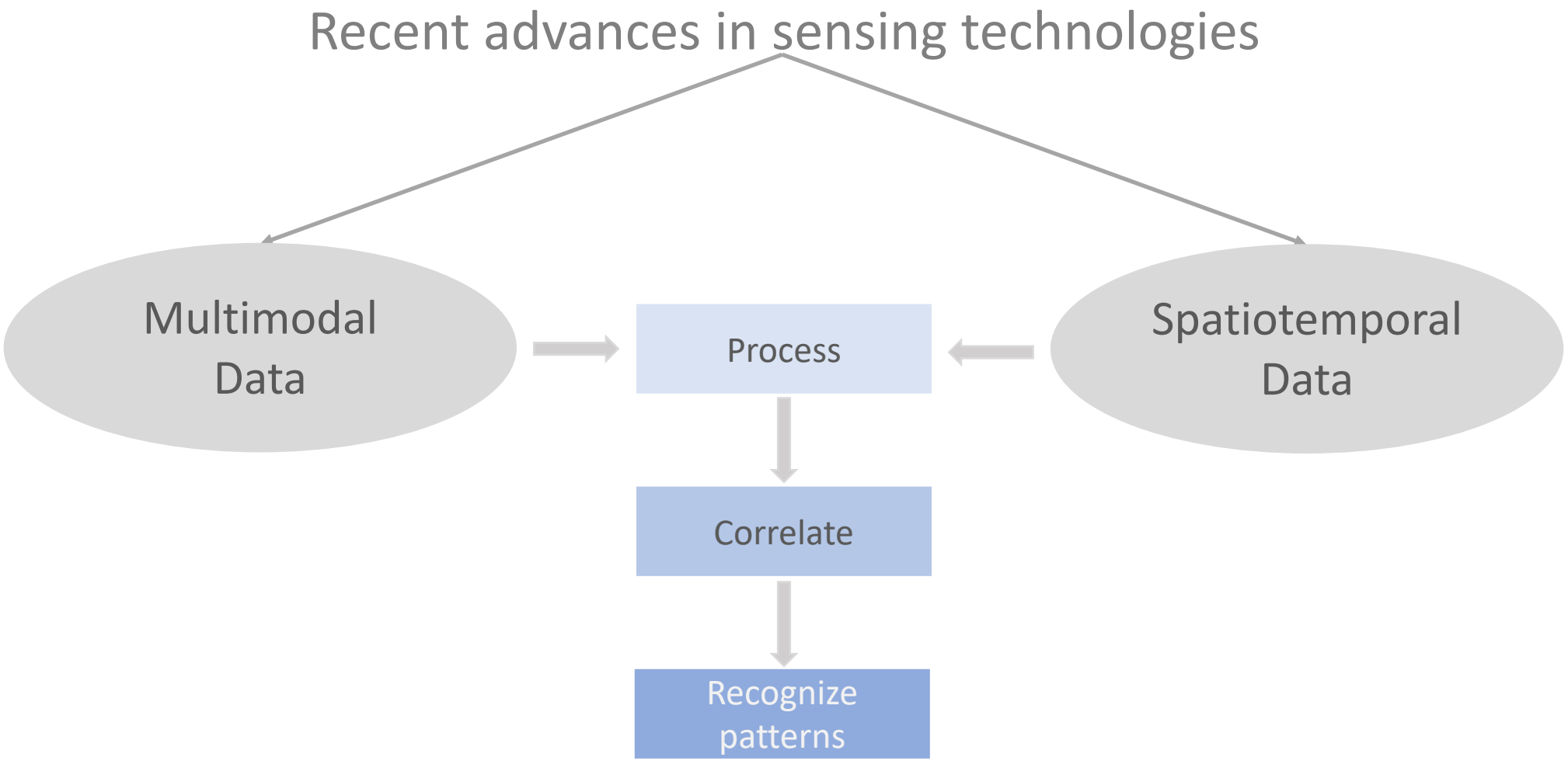
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

Konstantinos Makantasis[1], Athanasios Voulodimos[2], Anastasios Doulamis[3], Nikolaos Bakalos[3] and Nikolaos[2] Doulamis

[1] Institute of Digital Games, University of Malta, Malta
[2] Department of Informatics and Computer Engineering, University of West Attica, Greece
[3] School of Rural and Surveying Engineering, National Technical University of Athens, Greece

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

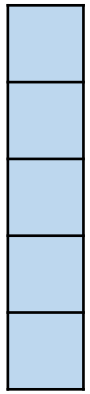
Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification



Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification

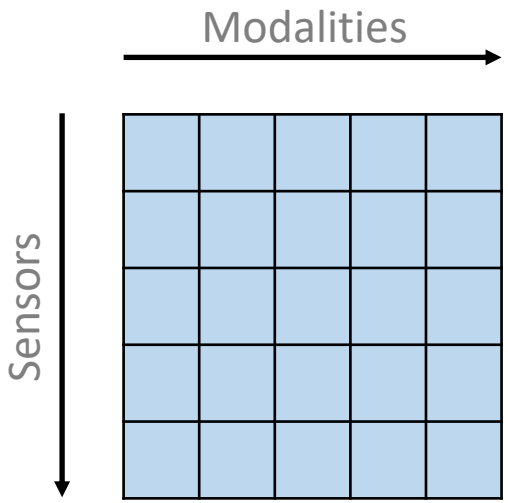
1 sensor, J modalities

Vector in $\mathbb{R}^J$



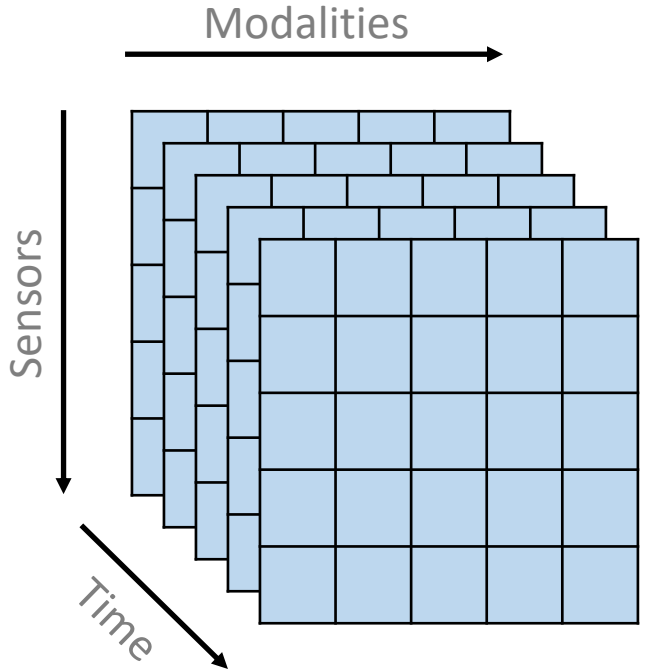
N sensors, J modalities

Matrix in $\mathbb{R}^{N \times J}$



N sensors, J modalities, T time instances

Tensor in $\mathbb{R}^{N \times J \times T}$



Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification

$\boldsymbol{S}_j \in \mathbb{R}^{N\times T}$: information for j-th modality

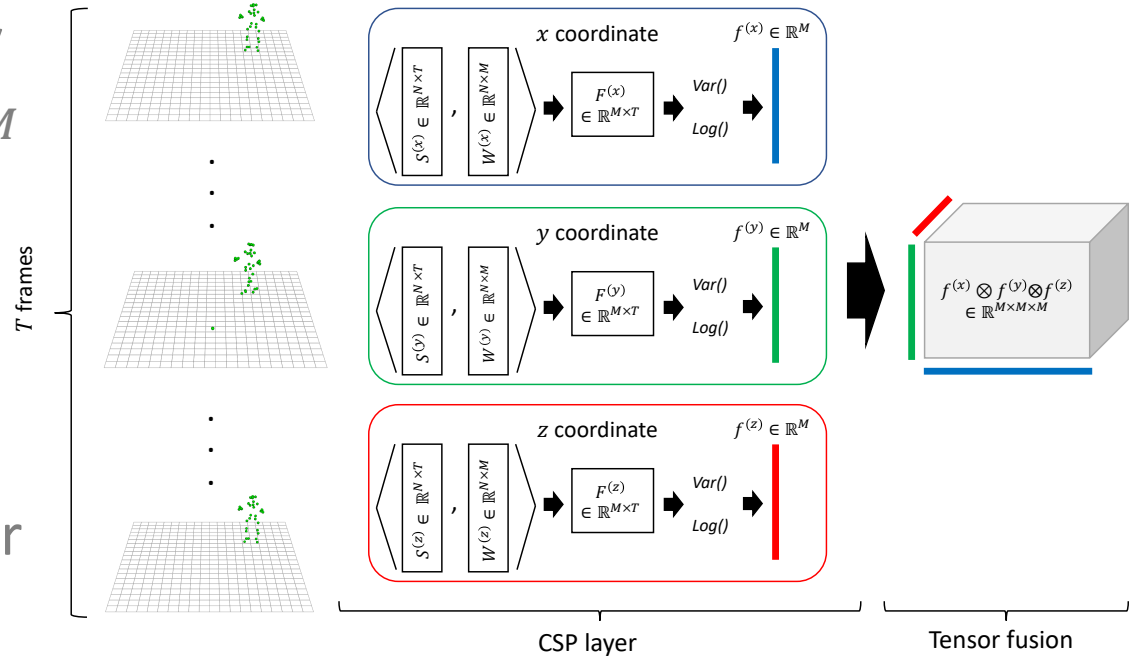
1. Project $\boldsymbol{S}_j$ in $\mathbb{R}^{M\times T}$ using $\boldsymbol{W}_j \in \mathbb{R}^{N\times M}$
2. Compute features

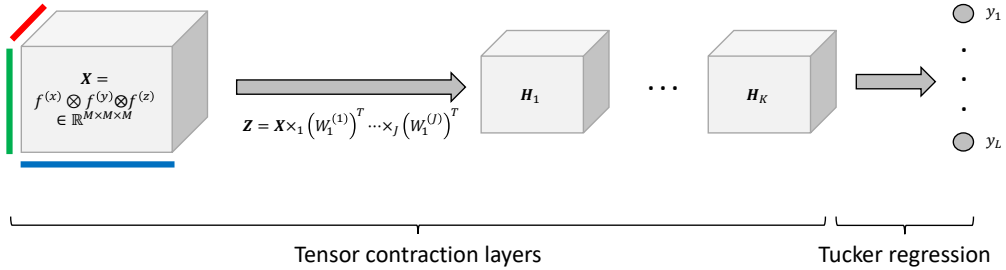
$$\boldsymbol{f}_j = \log\left[\frac{\mathrm{var}\left(\boldsymbol{W}^T\boldsymbol{S}_j^1\right)}{\sum_{i=1}^{M}\mathrm{var}\left(\boldsymbol{W}^T\boldsymbol{S}_j^i\right)}\cdots\frac{\mathrm{var}\left(\boldsymbol{W}^T\boldsymbol{S}_j^M\right)}{\sum_{i=1}^{M}\mathrm{var}\left(\boldsymbol{W}^T\boldsymbol{S}_j^i\right)}\right] \in \mathbb{R}^M$$

3. Fuse different modalities into a tensor

$$\mathcal{X} = \boldsymbol{f}_1 \otimes \cdots \otimes \boldsymbol{f}_J \in \mathbb{R}^{M\times M\times\cdots\times M}$$

4. Classify $\mathcal{X}$ using a tensor neural network

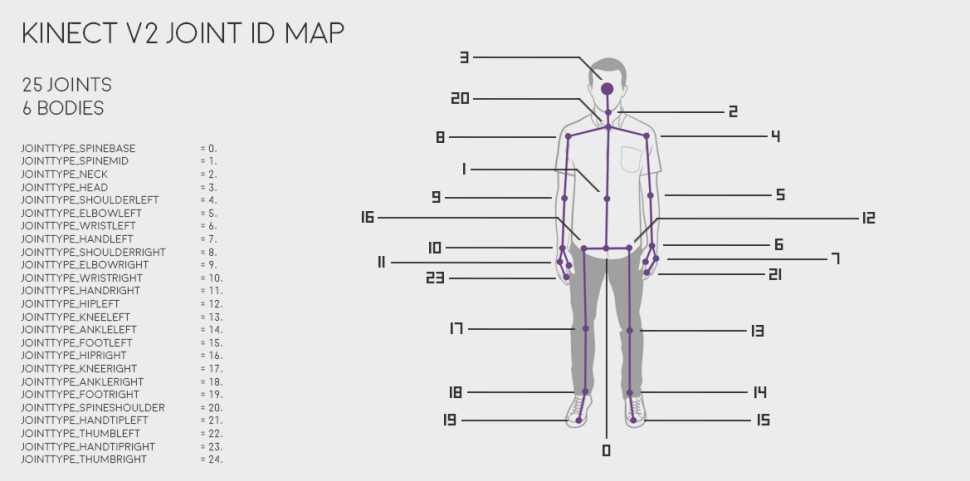




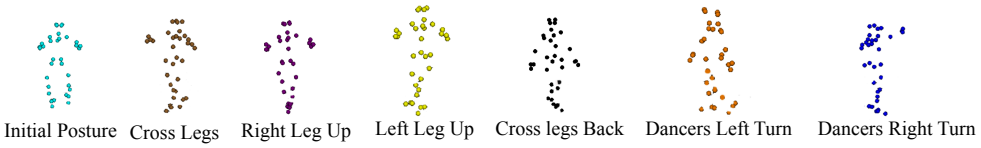
Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification

Dataset:

- Kinect II 3D skeleton data
- 10 sessions, 3 dancers, 4 dances
- Leave-one-session-out cross validation



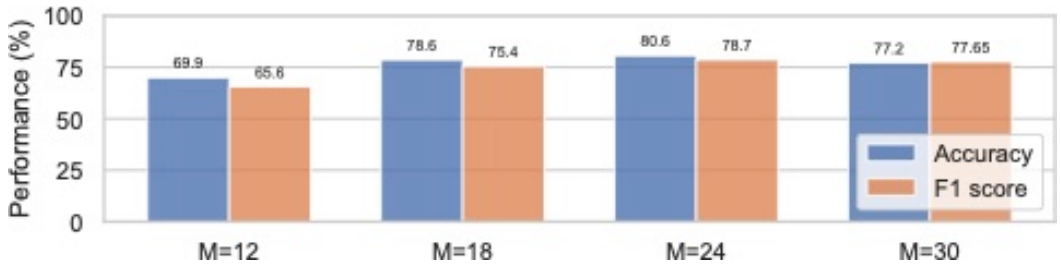
	Accuracy(%)	F1 Score (%)
LSTM	84.2%	82.0%
BOBi LSTM*	85.4%	80.7%
1D-CNN	91.1%	89.7%
Our approach	**91.6%**	**90.9%**



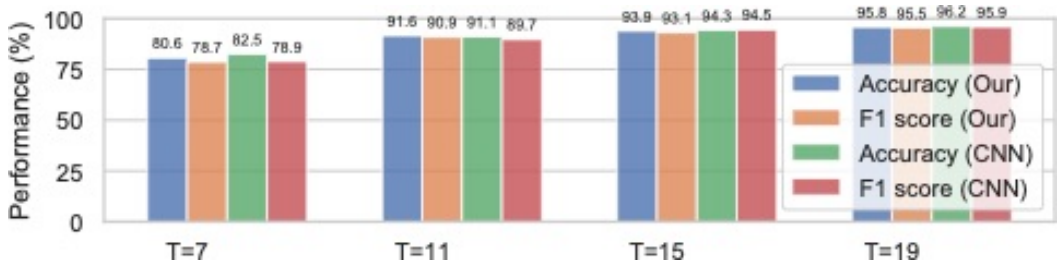
*I. Rallis et al. “Learning choreographic primitives through a bayesian optimized bi-directional lstm model,” in IEEE ICIP, 2019, pp. 1940–1944.

Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification

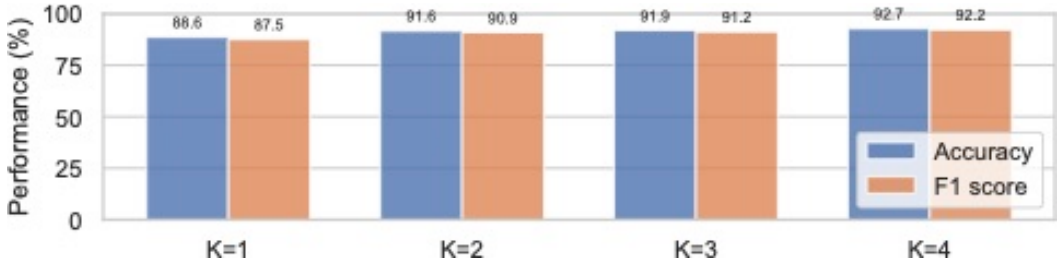
Effect of parameter M (T=7, 2TCL)



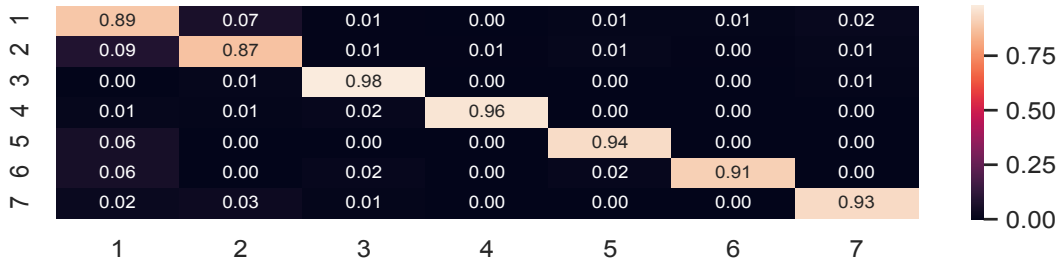
Effect of parameter T (M=24, 2TCL)



Effect of tensor network depth (T=11, M=24



Confusion matrix



Space-Time Domain Tensor Neural Networks
An Application on Human Pose Classification

Thank You

Take a picture to
download the full paper