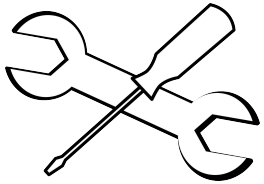
ICPR 2020 - Online

3D attention mechanism for fine-grained classification of table tennis strokes using a Twin Spatio-Temporal Convolutional Neural Networks

Pierre-Etienne Martin
Jenny Benois-Pineau
Renaud Péteri
Julien Morlier

Goal

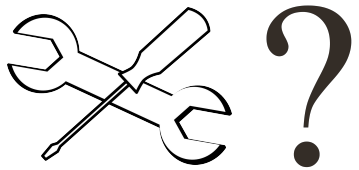
Improve athletes performances



through tools

for trainers and athletes

Goal

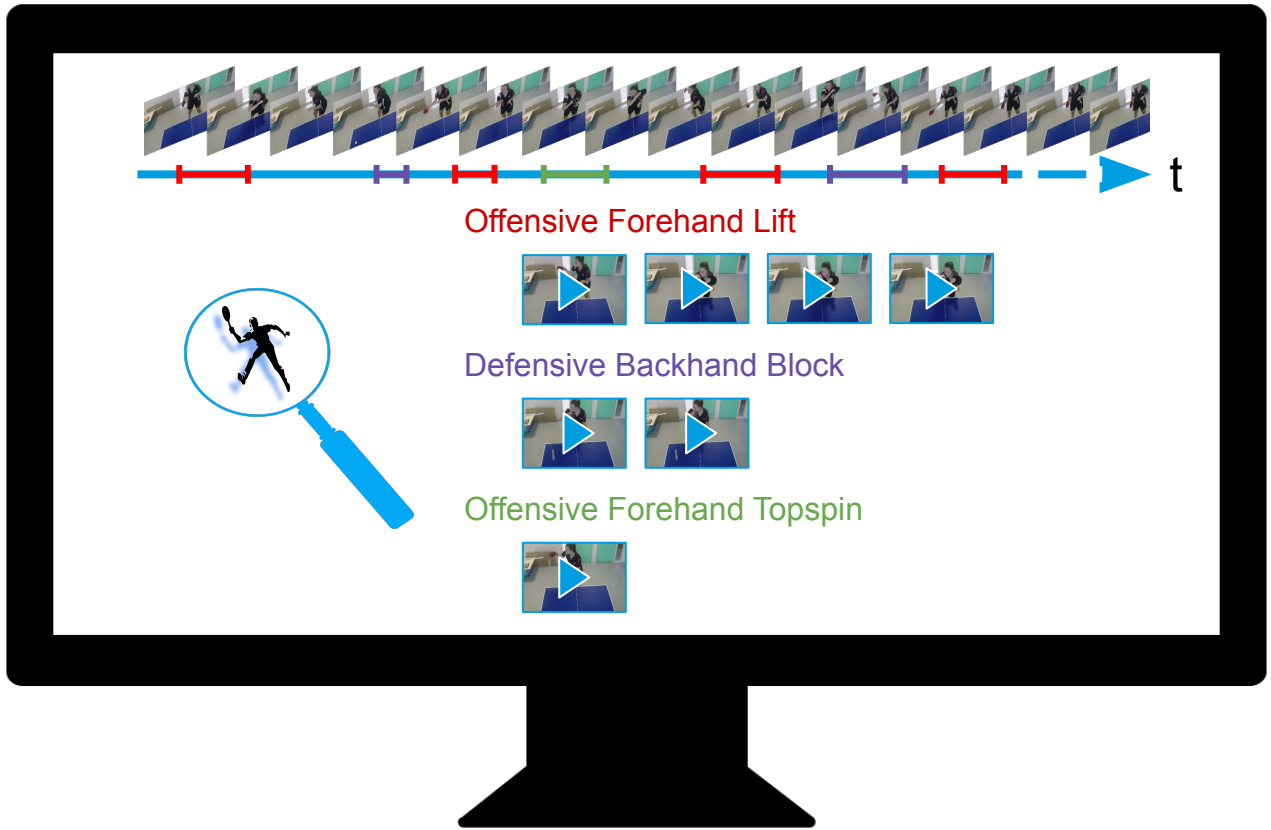


Input

Output

Offensive Backhand Flip

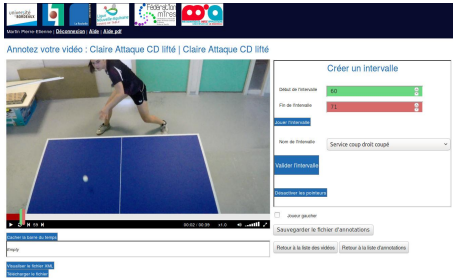
Goal


t
Offensive Forehand Lift
Defensive Backhand Block
Offensive Forehand Topspin

TTStroke-21 dataset [1]

- 241 videos at 25, 30 and 120 fps
- 4 209 annotations
- 3 428 strokes over 20 classes

Acquisition

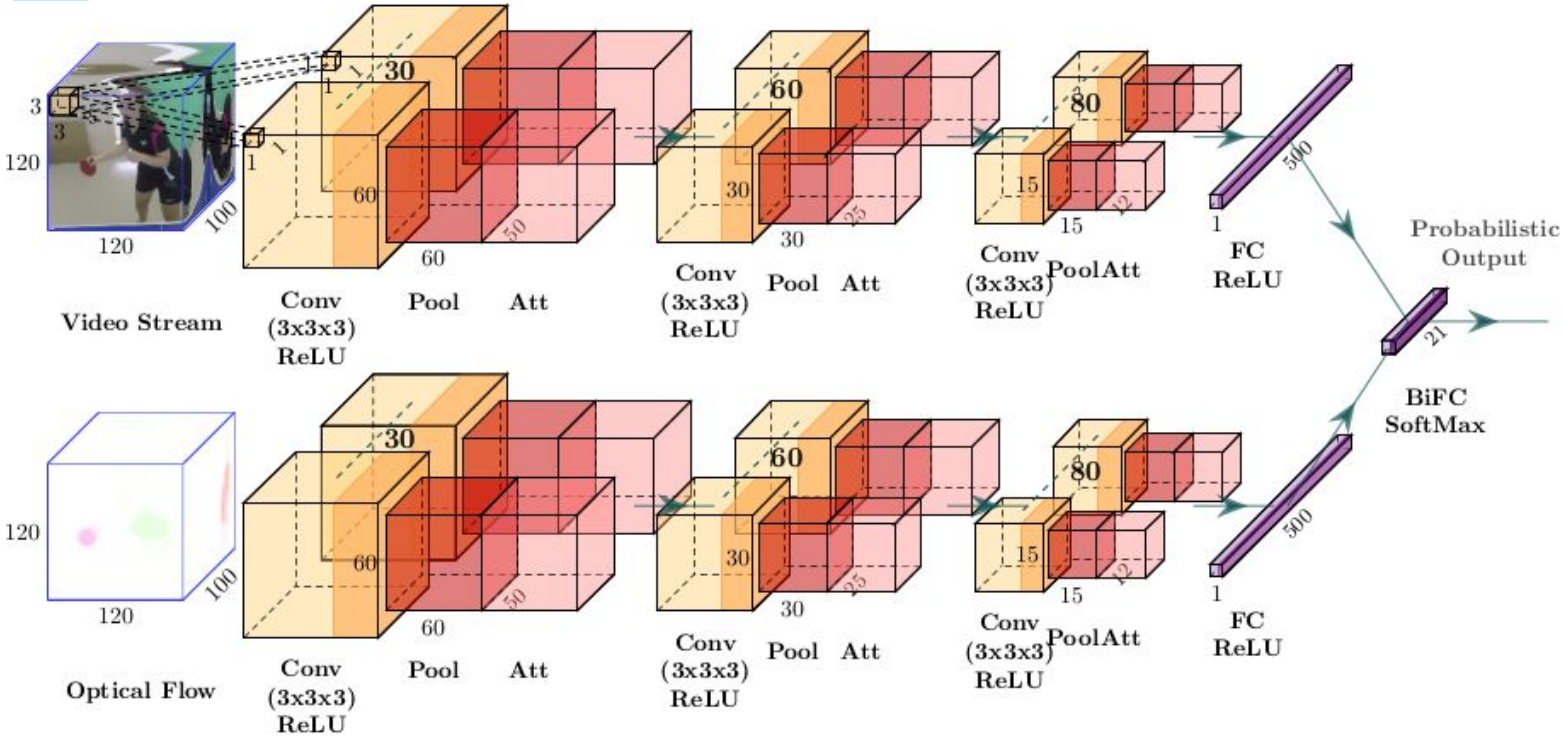


Annotation platform

Samples TTStroke-21

[1] P.-E. Martin et al., “Fine grained sport action recognition with Twin spatio-temporal convolutional neural networks”, in Multim. Tools Appl. 79, 27-28, 2020.

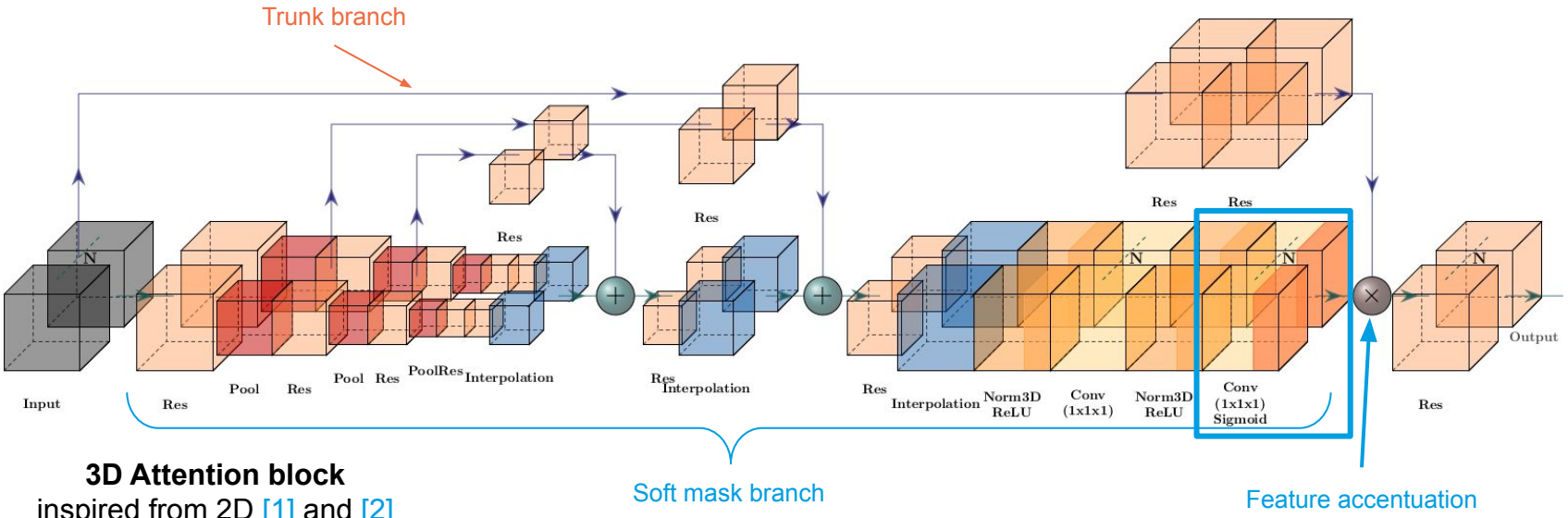
Improve with Attention Mechanism



T-STCNN with attention blocks

[1] P.-E. Martin et al., “Fine grained sport action recognition with Twin spatio-temporal convolutional neural networks”, in Multim. Tools Appl. 79, 27-28, 2020.

The 3D Attention Block



3D Attention block
inspired from 2D [1] and [2]

[1] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, “Residual attention network for image classification”, CVPR 2017, pp. 6450–6458.
[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, CoRR, vol. abs/1512.03385, 2015.

Results - Features accentuated

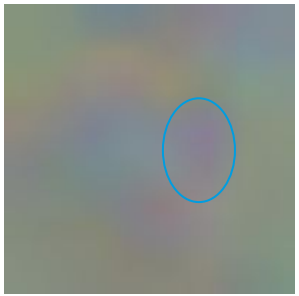
Soft mask branch output of:

table border

Attention block 1

player's body parts

Attention block 2



ball & racket

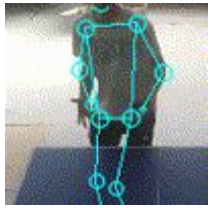
Attention block 3

Results - Classification accuracy (in %)

Models	Train	Validation	Test
RGB-I3D [1]	98	72.6	69.8
RGB-STCNN [2]	98.6	87	76.7
RGB-STCNN with Attention	96.9	88.3	85.6 93.2
Flow-I3D [1]	98.9	73.5	73.3
Flow-STCNN [2]	88.5	73.5	74.1 85.6
Flow-STCNN with Attention	96.4	83.5	79.7
Two Stream-I3D [1]	99.2	76.2	75.9
Twin-STCNN [2]	99	86.1	81.9
Twin-STCNN with Attention	**97.3**	**87.8**	**87.3** 95.8

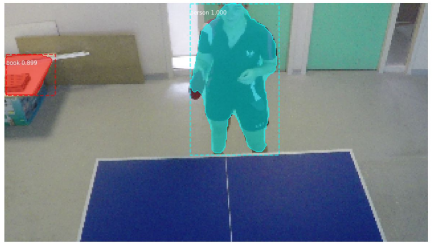
[1] J. Carreira and, A. Zisserman, “Quo vadis, action recognition? A new model and the kinetics dataset”, CoRR, vol. abs/1705.07750, 2017.
[2] P.-E. Martin et al., “Fine grained sport action recognition with Twin spatio-temporal convolutional neural networks”, in Multim. Tools Appl. 79, 27-28, 2020.

Future work and perspectives



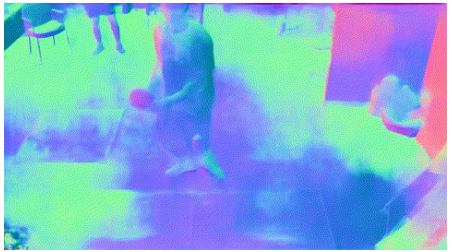
➔ **Ongoing work:**

- Attention mechanism observation during the training phase
- Pose [1] for classification
- UNet for segmentation



➔ **Later in 2021:**

- MediaEval 2021
- Pose + Depth [2] for stroke movement modeling with the aim to give a qualitative feedback to the players



[1] Newell Alejandro, Yang Kaiyu, Deng Jia: Stacked Hourglass Networks for Human Pose Estimation. ECCV 2016.
[2] Ramamonjisoa Michaël, Lepetit Vincent: SharpNet: Fast and Accurate Recovery of Occluding Contours in Monocular Depth Estimation. ICCV Workshops 2019.

LaBRI

université de BORDEAUX