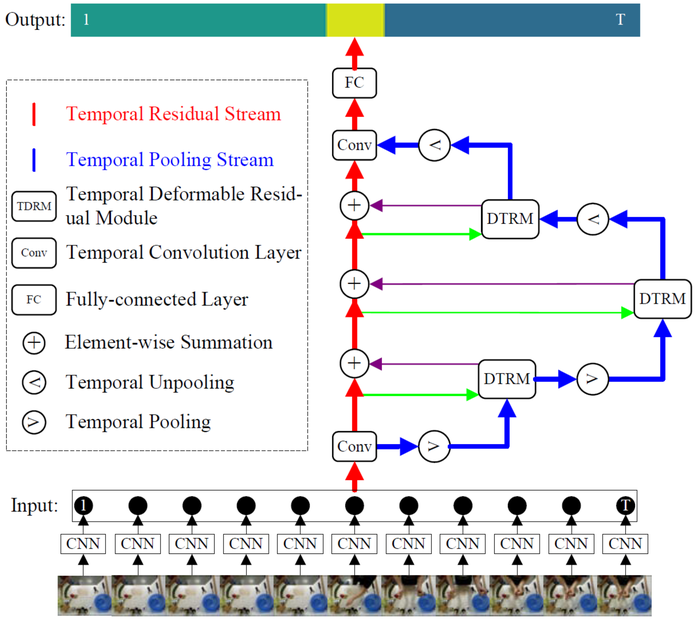
Temporal Deformable Residual Networks for Action Segmentation in Videos

Peng Lei and Sinisa Todorovic
Oregon State University

Problem: Temporal Action Segmentation in Videos

Key Ideas

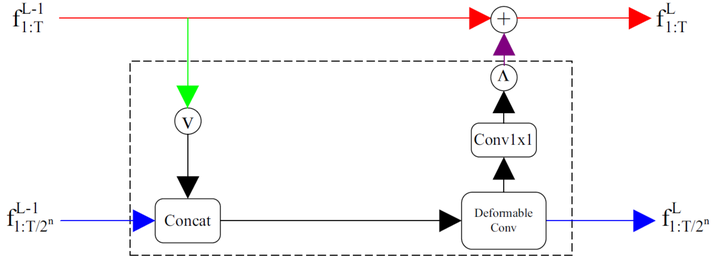
- Account for Local and Long-range Temporal Cues
- Two Temporal Processing Streams
  - Residual Stream
  - Pooling Stream
- Deformable Temporal Convolution



Contributions

1. Specify an easily trained fully-convolutional temporal residual network for robust action recognition and accurate action segmentation in videos
2. Show deformable temporal convolutions improve action segmentation over regular temporal convolutions
3. Outperform the state of the art in action segmentation on the 50Salads, GTEA and JIGSAWS datasets.

Deformable Temporal Residual Module



Our DTRM --- Two inputs two outputs
[Lea CVPR17][Ronneberger MICCAI15][He CVPR16] --- One input one output

Results

Table.1 Performance comparison with respect to the most related temporal convolution models.

Dataset	50Salads (mid)			GTEA			JIGSAWS		
Model	F1@10	Edit	Acc	F1@10	Edit	Acc	F1@10	Edit	Acc
ED-TCN [Lee CVPR17]	68.0	59.8	64.7	72.2	-	64.0	89.2	84.7	80.8
TUnet [Ronneberger MICCAI15]	59.3	50.6	60.6	67.1	60.3	59.9	85.9	79.8	80.2
TResNet [He CVPR16]	69.2	60.5	66.0	74.1	64.4	65.8	86.2	85.2	81.1
TDRN	**72.9**	**66.0**	**68.1**	**79.2**	**74.1**	**70.1**	**92.9**	**90.2**	**84.6**

Acknowledgement: This work was supported in part by DARPA XAI Award N66001-17-2-4029.

Table.2 Results on Results on 50 Salads (mid), GTEA and JIGSAWS.

Dataset	50Salads (mid)			GTEA			JIGSAWS		
Model	F1@10	Edit	Acc	F1@10	Edit	Acc	F1@10	Edit	Acc
Spatial CNN [Lea ECCV16]	32.3	24.8	54.9	41.8	-	54.1	-	37.7	74.0
Dilated TCN [Lea CVPR17]	52.2	43.1	59.3	58.8	-	65.8	-	-	-
ST-CNN [Lea ECCV16]	55.9	45.9	59.4	58.7	-	60.6	78.3	68.6	78.4
Bi-LSTM [Singh CVPR16]	62.6	55.6	55.7	66.5	-	55.5	77.8	66.8	77.4
ED-TCN [Lea CVPR17]	68.0	59.8	64.7	72.2	-	64.0	89.2	84.7	80.8
TRN	70.2	63.7	66.9	77.4	72.2	67.8	91.4	87.7	83.3
TDRN	**72.9**	**66.0**	**68.1**	**79.2**	**74.1**	**70.1**	**92.9**	**90.2**	**84.6**

Figure.1 Action segmentations for a sample test video from the 50Salads dataset. Colors indicate action classes.

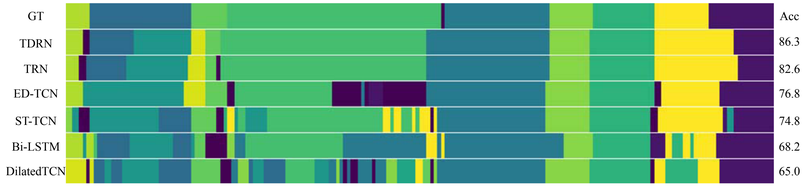


Figure.2 Action segmentations for a sample test video from the JIGSAWS dataset. Colors indicate action classes.

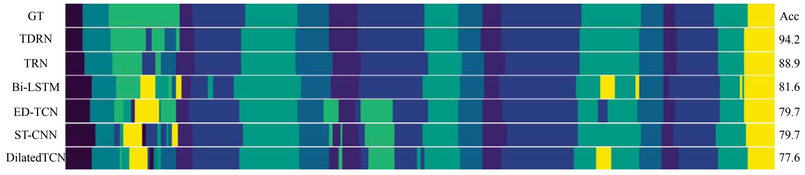


Figure.3 Action segmentations for a sample test video from the GTEA dataset. Colors indicate action classes.