

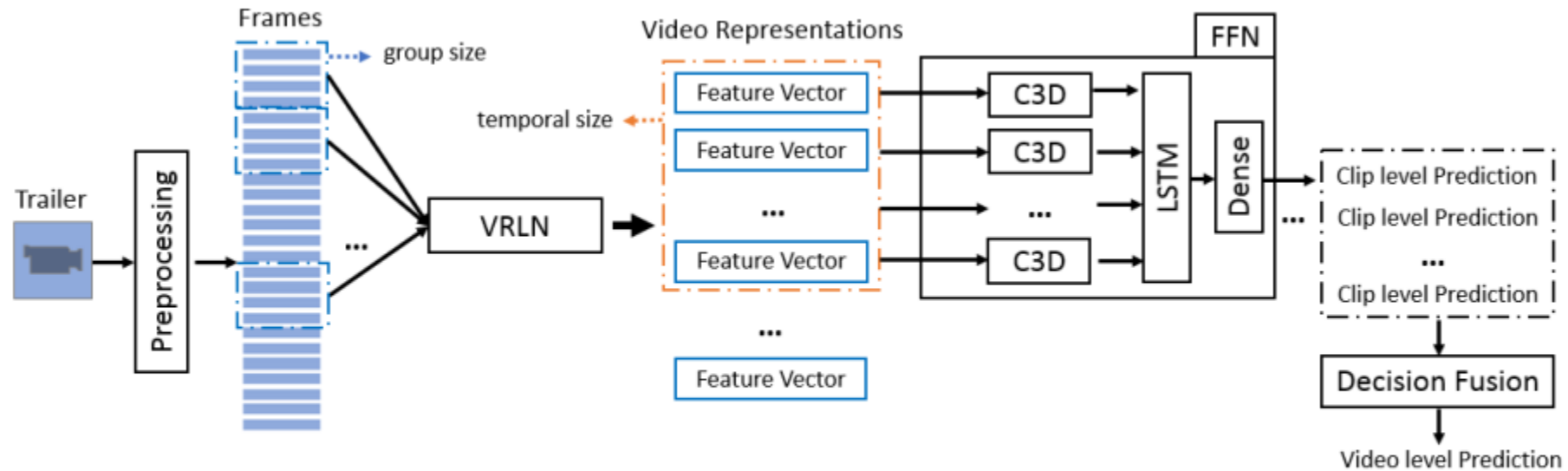
Video Representation Fusion Network For Multi-label Movie Genre Classification

Tianyu Bi, Dmitri Jarnikov, Johan Lukkien



Our Method

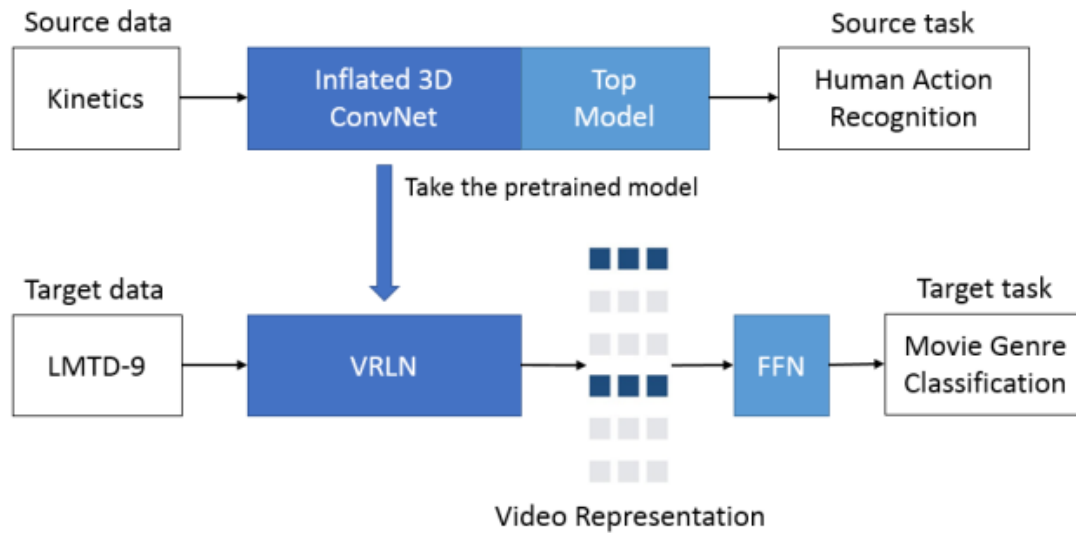
- Video Representation Fusion Network



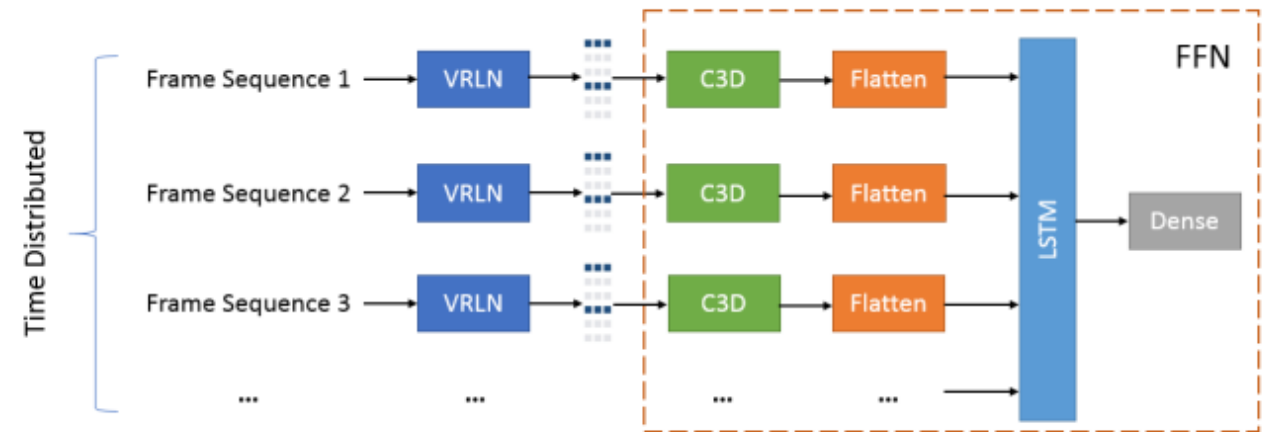
VRFN data flow. It consists of two parts: VRLN and FFN. VRLN generates video representations (group-level feature vectors) from a group of frames. FFN learns a higher-level video representation from the intermediate video representations. Finally, decision fusion is applied to obtain video-level prediction results.

Our Method

- Video Representation Fusion Network
 - Video Representation Learning Network
 - Feature Fusion Network



Creating VRLN by transfer learning.



FFN architecture.

Experimental Results

LMTD-9 DATASET [5].

Genre	Train	Validation	Test
Action	611	78	164
Adventure	432	51	108
Comedy	1109	148	301
Crime	477	59	121
Drama	1437	192	394
Horror	324	33	78
Romance	468	59	122
SciFi	229	26	57
Thriller	502	61	129
Total	2861	374	772

COMPARING RESULTS WITH STATE-OF-THE-ART METHODS [5].

Models	AUPRC (macro)	AUPRC (micro)	AUPRC (weighted)	Architecture
CTT-MMC-A[5]	0.618	0.712	0.683	CTT+FC
CTT-MMC-B[5]	0.599	0.704	0.661	CTT+2FC
CTT-MMC-C[5]	0.624	0.722	0.697	CTT+Maxout
VRFN2-64-10	0.645	0.747	0.721	VRLN+FFN

Thanks