

VID-Trans-ReID: Enhanced Video Transformers for Person Re-identification

Aishah Aleshaim and Toby P. Breckon, Durham University, UK

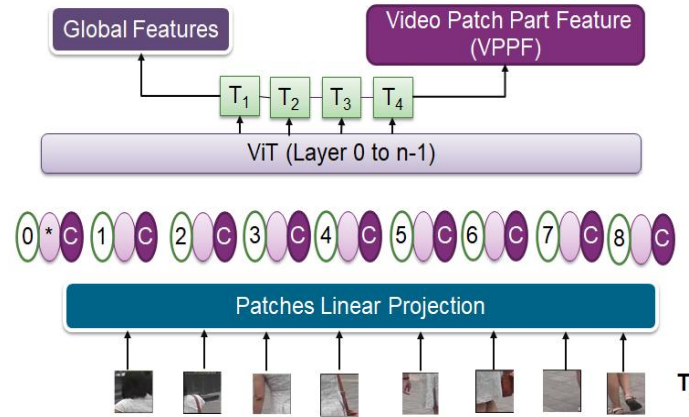
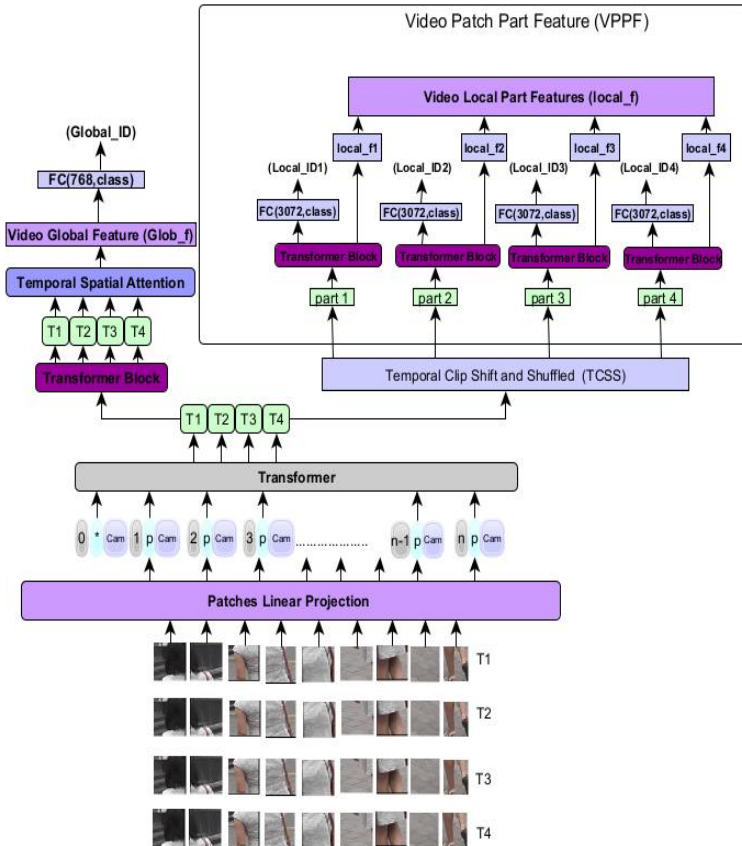
Motivation

Extracting a robust feature representation that is invariant to the challenges of pose and illumination variation across multiple camera viewpoints is a key challenge within video person Re-ID.

Method

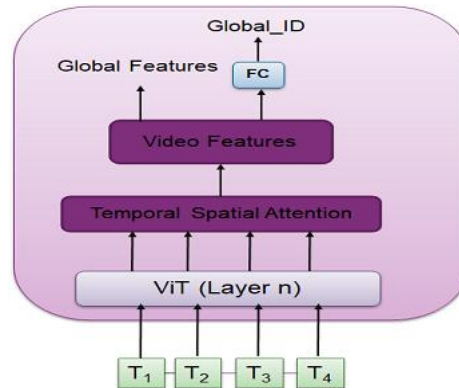
We use Vision Transformer (ViT) in place of 'traditional' CNN via:

- an enhanced ViT based architecture for video person Re-ID, incorporating both global video-level and local patch-based features.
- novel Temporal Clip Shift and Shuffle (TCSS) and Video Patch Part Feature (VPPF) modules that provide robust fine-grained feature extraction to boost overall video person Re-ID performance.



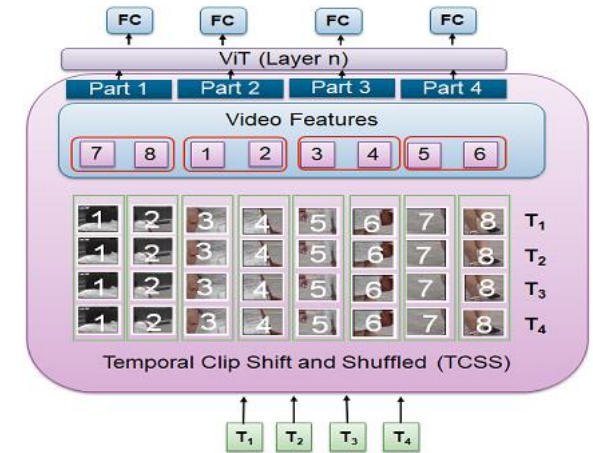
- Each video frame is split into n equal size overlapped patches.
- A learnable positional embedding is prepending to each patch to preserve spatial information.
- Given the multi-camera nature of video person Re-ID additional learnable camera embedding is added to represent the camera ID of a given view.
- Two types of video features extracted by the model:
 - Global video features
 - Local part video features

Global Video-level Features



In the global branch the model learns to produce video-level features over subsets of four adjacent frames. Frame-level features are then aggregated to video-level features using spatiotemporal attention.

Local Part Video Features



This local branch extracts fine-grained features using our novel Temporal Clip Shift and Shuffle (TCSS) and Video Patch Part Feature (VPPF) modules.

Results



Methods	Publication	MARS [66] Rank-1 (mAP)	iLIDS-VID [50]	PRID2011 [19]
PSTA [51]	ICCV 2021	91.5 (85.8)	91.5	95.6
SSN3D [24]	AAAI2021	90.1(86.2)	88.9	-
CTL [31]	CVPR2021	91.40(86.70)	89.70	-
Watching You [32]	CVPR 2021	91.0 (84.8)	90.4	96.2
BiCnet-TKS [22]	CVPR 2021	90.2 (86.0)	-	-
PiT [61]	IEEE TH 2022	90.22 (86.80)	92.07	-
SINet [5]	CVPR 2022	91.0 (86.2)	92.5	96.5
ViT [25] (baseline)	-	90.68(78.61)	38.67	74.16
VID-Trans-ReID (ours)	-	96.36 (90.25)	94.67	96.63

Online code access
<https://github.com/AishahAADU/VID-Trans-ReID/>

