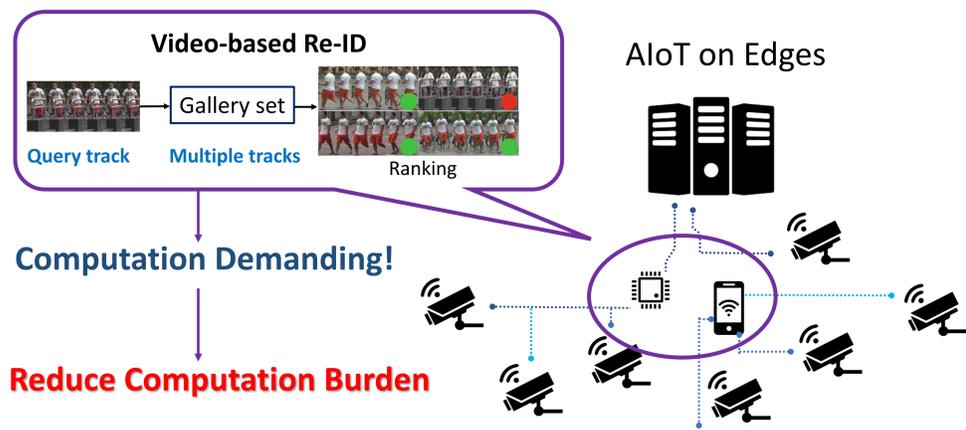




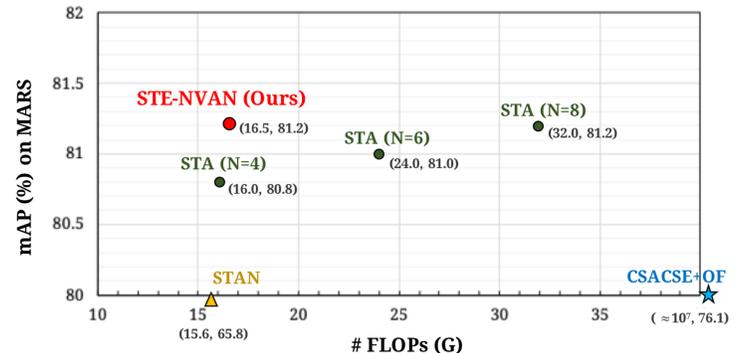
Codes are available

Goal



Contribution

- Introduce **non-local attention** operation into CNN to incorporate both **spatial and temporal** characteristics for re-ID.
- Significantly **reduce computation costs** and achieve satisfactory **trade-off** between performance and computation.

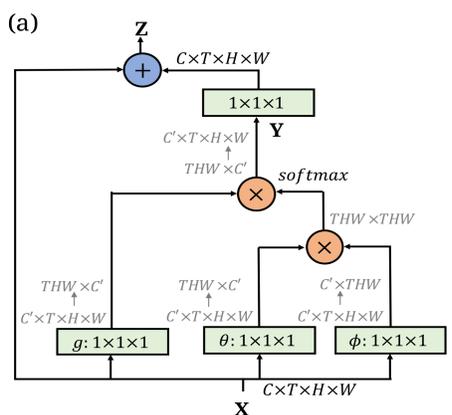


Proposed Method

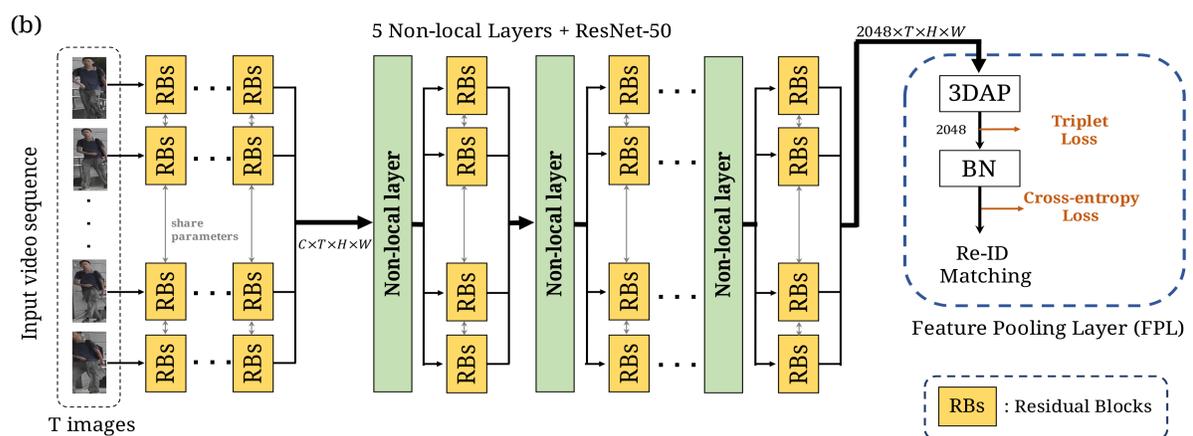
- Non-local Video Attention Network (NVAN)
 - Self-attention on spatial and temporal features. (R1: **90.0%**, mAP: **82.8%** on MARS benchmark)
 - **Computation demanding** in non-local layers.

- Input Sequence: RRS sampling (N=8) [1]
- Backbone: Non-local neural network [2]
- Output: **Feature pooling layer** with \mathcal{L}_{total}

$$\mathcal{L}_{total} = \mathcal{L}_{triplet} [3] + \mathcal{L}_{classification}$$

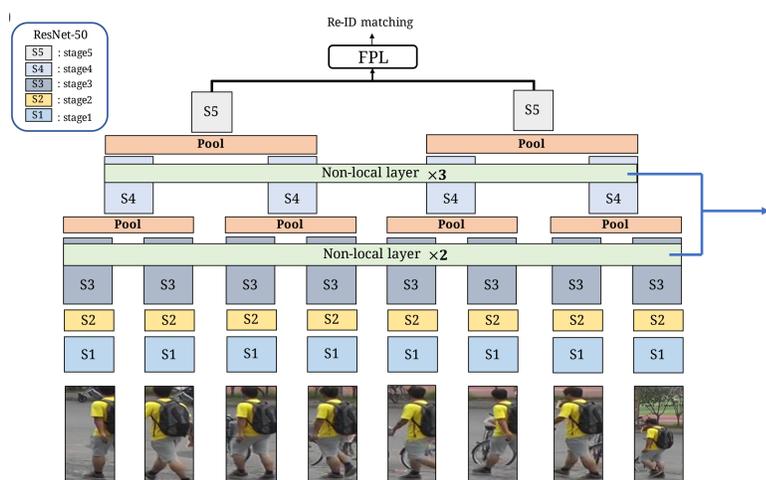


(a) Details of Non-local Attention Layer

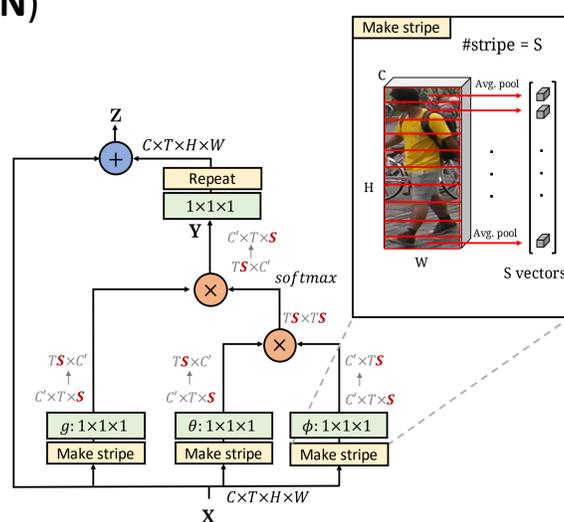


(b) Architecture of our Non-local Video Attention Network (NVAN)

- Spatially and Temporally Efficient NVAN (STE-NVAN)



1. Temporal Reduction → Hierarchical Structure



2. Spatial Reduction → Make Stripes of Features

Complexity:

$$O(C'T^2H^2W^2 + CC'THW)$$

$$O(C'T^2S^2 + CC'TS)$$

Performance Comparison

Methods	Source	MARS		DukeV	
		R1	mAP	R1	mAP
CNN+Kiss. [26]	ECCV16	65.0	45.6	-	-
SeeForest [45]	CVPR17	70.6	50.7	-	-
LatentParts [16]	CVPR17	70.6	50.7	-	-
TriNet [11]	arXiv17	79.8	67.7	-	-
ETAP-Net(supervised) [36]	CVPR18	80.8	67.4	83.6	78.3
STAN [17]	CVPR18	82.3	65.8	-	-
CSACSE+OF [1]	CVPR18	86.3	76.1	-	-
STA (N=8) [8]	AAAI19	86.2	81.2	96.0	95.0
NVAN (ours)	-	90.0	82.8	96.3	94.9
STE-NVAN (ours)	-	88.9	81.2	95.2	93.5

Ablation Study

Method	Feature Aggregation	MARS		DukeV		# FLOP
		R1	mAP	R1	mAP	
ResNet-50	FPL	87.3	79.1	95.0	92.7	30.4 G
ResNet-50	max-FPL	86.3	76.6	95.4	92.4	30.4 G
NVAN	FPL	90.0	82.8	96.3	94.9	60.0 G
NVAN+Spatial Reduc.	FPL	89.7	82.5	96.3	94.7	30.4 G
NVAN+Temporal Reduc.	FPL	89.2	81.2	95.6	93.7	40.4 G
STE-NVAN	FPL	88.9	81.2	95.2	93.5	16.5 G

[1] Shuang Li et al. Diversity regularized spatiotemporal attention for video-based person re-identification. CVPR, 2018.

[2] Xiaolong Wang et al. Non-local neural networks. CVPR, 2018.

[3] Alexander Hermans et al. In defense of the triplet loss for person re-identification. arXiv, 2017.