Multi-level Deep Learning Vehicle Re-identification using Ranked-based Loss Functions

Eleni Kamenou[1] | Jesus Martinez del Rincon[1] | Paul Miller[1] | Patricia Devlin-Hill[2]
[1]Centre for Secure Information Technologies (CSIT)
Queen's University Belfast, United Kingdom
[2]Thales UK, Belfast, Northern Ireland

Problem Definition

Vehicle Re-Identification (ReID)

Identifying a vehicle as it transits across different cameras with non-overlapping fields of view.

ReID as a retrieval task

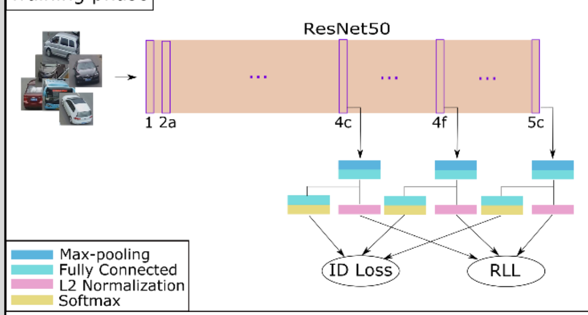
Given a query image of a vehicle, numerous gallery images are searched to find the same vehicle captured by other cameras.

Contribution

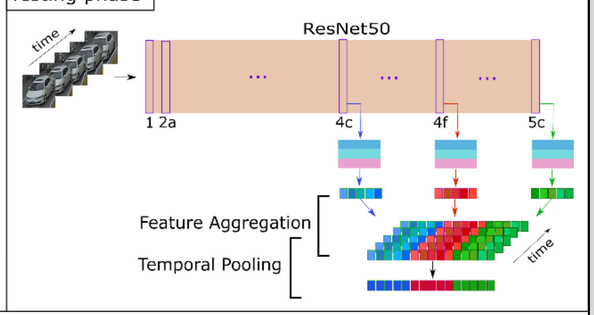
	Current Works	Our approach
Input	detect vehicle parts	raw vehicle images
Output	re-ranking	no post-processing
Data Format	partially video-based	fully video-based
Extra Annotation	vehicle type video timestamps	no extra annotation

Proposed Approach

Training phase



Testing phase



Ranked-List Loss Function

Aims to keep the Euclidean distance *(d)* between the query and a group of positive samples below a certain threshold *(α-m)*, while separating the query from the negative samples by a margin *(α)*.

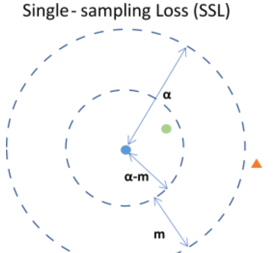
$$l_p(x_c^i, x_c^j) = \max\left(d_{ij} - (a-m),\ \ 0\right)$$

$$l_n(x_c^i, x_k^j) = \max\left(a - d_{ij},\ \ 0\right),\quad c \neq k$$

$$L_p(x_c^i) = \frac{1}{|P_{c,i}|}\sum_{x_c^j \in P_{c,i}} l_p(x_c^i, x_c^j)$$

$$L_n(x_c^i) = \frac{1}{|N_{c,i}|}\sum_{x_k^j \in N_{c,i}} l_n(x_c^i, x_k^j)$$

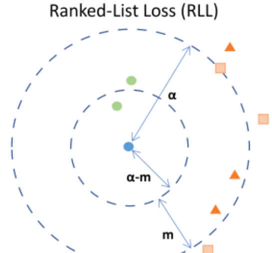
Single-sampling Loss (SSL)



$$L_{SSL} = l_p(x_c^i, x_c^j) + l_n(x_c^i, x_k^j)$$

$$L_{final} = w_1 * L_{SSL} + w_2 * L_{ID}$$

Ranked-List Loss (RLL)



$$L_{RLL}(x_c^i) = L_p(x_c^i) + L_n(x_c^i)$$

$$L_{final} = w_1 * L_{RLL} + w_2 * L_{ID}$$

P: the set of positive samples whose distance from the query is larger than *α-m*
N: the set of negative samples whose distance from the query is smaller than *α*

- positive sample of class #1
- anchor sample of class #1
- negative sample of class #2
- negative sample of class #3

c, k: vehicle identity classes

w_1,w_2: weighting factors

Multi-Level embeddings

Extracting features from 3 different layers of ResNet50.
At training phase: Applying the loss function to each feature vector separately
At testing phase: Feature aggregation

Temporal Pooling

Video feature representation: the mean over the feature vectors of all the individual frames in the video.
Similarity measuring: image-to-image, image-to-video and video-to-video

Experimental Results

Evaluation on 2 datasets

VeRi-776 (200 vehicles)
CityFlow-ReID (333 vehicles)

ReID Evaluation Metrics

mean Average Precision (mAP)
Rank-k scores (R-1, R-5)

	VeRi-776									CityFlow-ReID					
	Image to Image			Image to Tracklet			Tracklet to Tracklet			Image to Image			Image to Tracklet		
	mAP	R-1	R-5	mAP	R-1	R-5	mAP	R-1	R-5	mAP	R-1	R-5	mAP	R-1	R-5
Baseline(Triplet)	49.52	80.87	92.25	58.02	80.51	81.94	67.60	87.72	88.08	18.48	40.19	58.29	33.42	44.39	44.39
SSL	54.60	78.24	91.06	62.13	80.51	81.10	68.00	85.69	86.17	19.51	42.67	61.90	33.27	41.14	41.14
RLL	56.64	79.43	91.17	63.37	78.96	79.55	69.24	87.00	87.42	21.85	40.19	59.43	34.18	44.57	44.76
Multi-Level RLL	**62.48**	**88.64**	**94.75**	**68.25**	**87.90**	**88.31**	**73.34**	**91.06**	**91.41**	**24.59**	**47.07**	**62.36**	**36.44**	**46.48**	**46.48**

Single-level Embeddings

Multi-level Embeddings

Conclusions

A robust end-to-end vehicle ReID framework, which is able to effectively identify vehicles from both image and video data.

- Combination of features from different levels of the network allows stronger feature representations to be obtained.
- Temporal pooling provides robust video feature representations and extends our system to a fully video-based approach for vehicle ReID.

Multi-Level Embeddings	+ 6% mAP + 9% Rank-1
Temporal Pooling	+ 12% mAP

Contact: **Eleni Kamenou,** PhD candidate, CSIT | e-mail: ekamenou01@qub.ac.uk