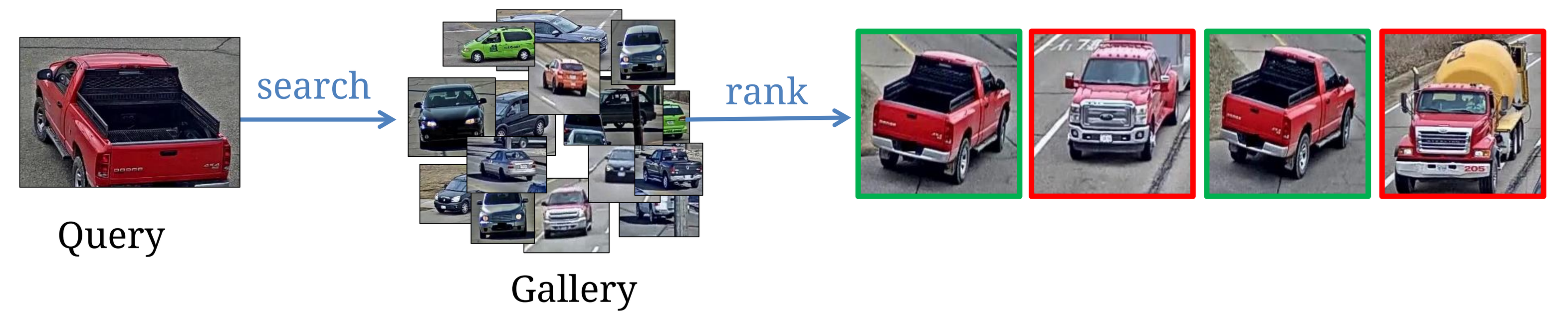
Supervised Joint Domain Learning for Vehicle Re-Identification

Chih-Ting Liu, Man-Yu Lee, Chih-Wei Wu,
Bo-Ying Chen, Tsai-Shien Chen, Yao-Ting Hsu, Shao-Yi Chien
NTU IoX Center, National Taiwan University

國立臺灣大學 National Taiwan University
NTU IoX Center
CVPR LONG BEACH CALIFORNIA June 16-20, 2019
NVIDIA AI CITY CHALLENGE

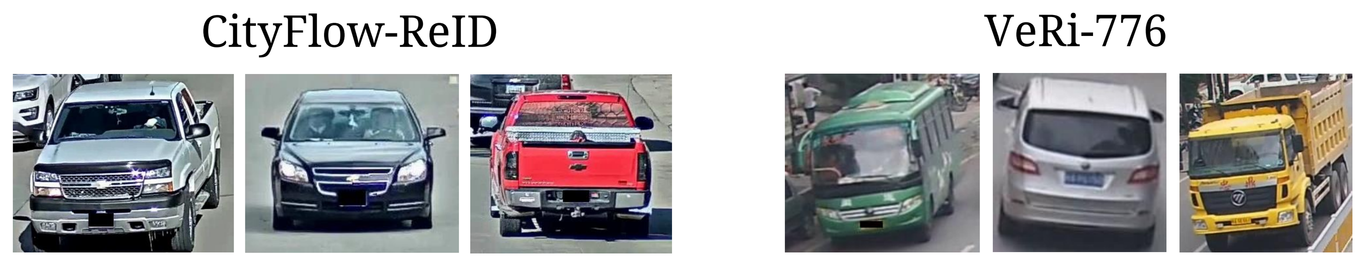
Track2 Problem Statement

- Vehicle Re-Identification
  - Tracking and identifying moving vehicles across videos captured at multiple locations

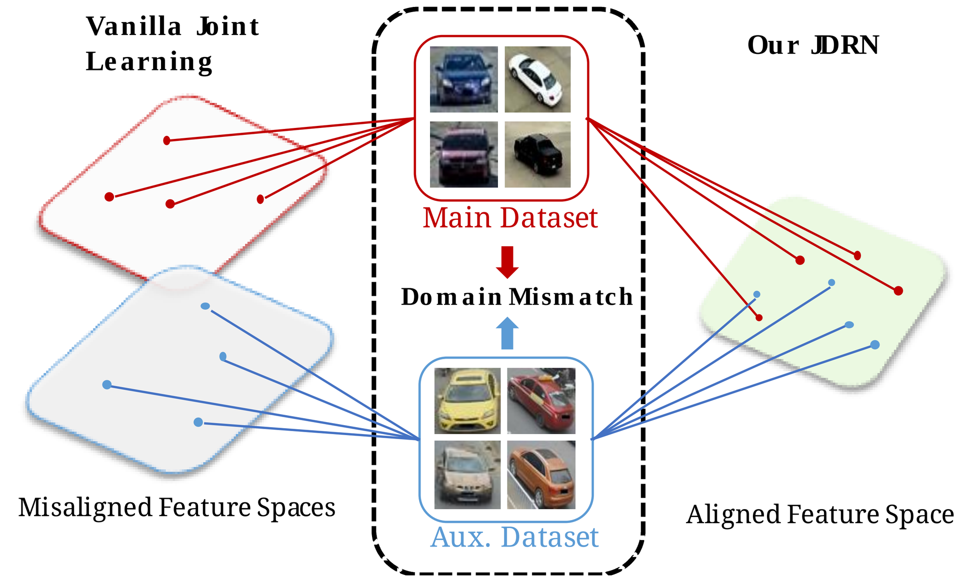


Introduction

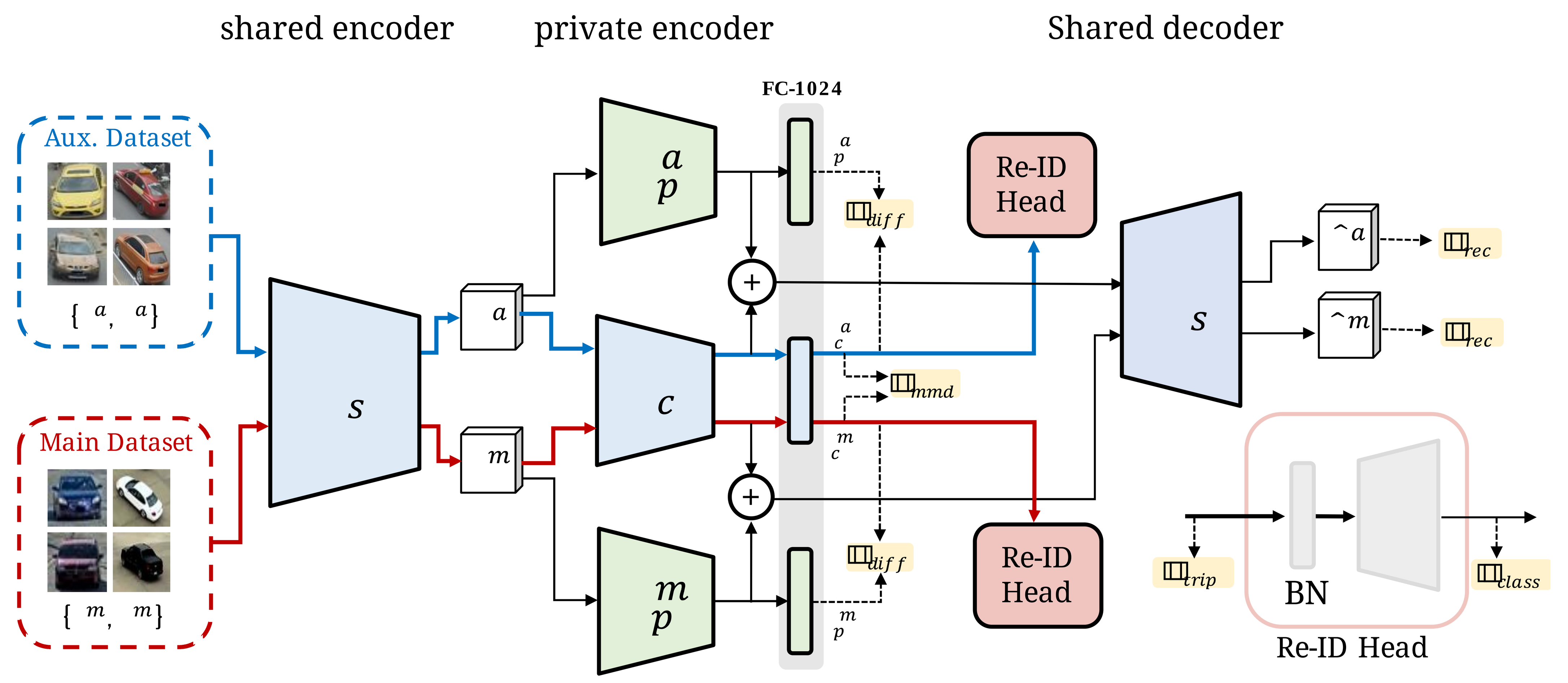
- Challenges
  - Limitation of existing datasets
    - datasets are either too small or with limited diversity
  - Data augmentation using other datasets → **Domain mismatch**



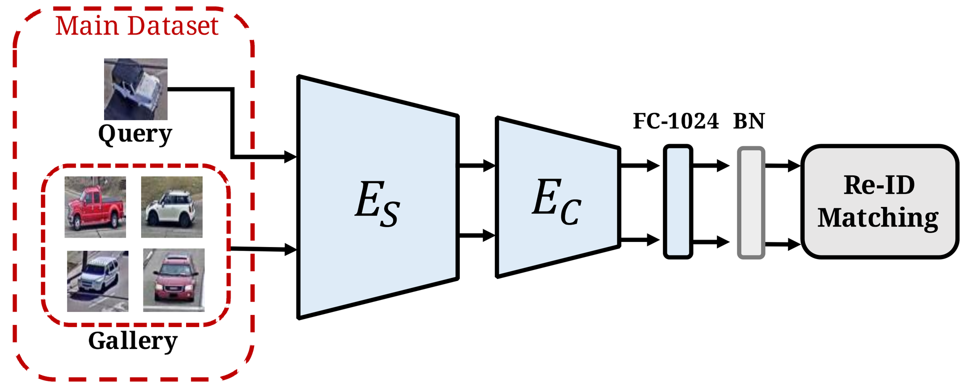
- Main Contribution
  - Jointly learn on multiple datasets to mitigate the limitation of single dataset
  - Address the domain mismatch problem using **Joint Domain Re-identification Network** (JDRN)
  - JDRN demonstrates promising performance on vehicle Re-ID datasets



Proposed method



- Re-ID Feature Learning
  - Cross-entropy loss $\mathcal{L}_{class} = -\sum_{i=1}^{n} y_i \cdot log\hat{y}_i$
  - Weighted triplet loss [2] $\mathcal{L}_{trip} = \sum_{a,p,n} F(w_p d(E(I_a), E(I_p)) - w_n d(E(I_a), E(I_n)))$, $w_p = \frac{e^{d(f_a, f_p)}}{\sum_{a,p} e^{d(f_a, f_p)}}, w_n = \frac{e^{-d(f_a, f_n)}}{\sum_{a,n} e^{-d(f_a, f_p)}}$
- Joint Domain Feature Learning
  - Difference loss $\mathcal{L}_{diff} = \| \mathbf{H}_c^{m\top} \mathbf{H}_p^m \|_F^2 + \| \mathbf{H}_c^{a\top} \mathbf{H}_p^a \|_F^2$
  - Reconstruction loss $\mathcal{L}_{rec} = \sum_{i=1}^{n_m} \| F_i^m - \hat{F}_i^m \|_2^2 + \sum_{i=1}^{n_a} \| F_i^a - \hat{F}_i^a \|_2^2$
  - MMD loss [3] $\mathcal{L}_{mmd} = \| \frac{1}{n_m} \sum_{i=1}^{n_m} \phi(f_{c,i}^m) - \frac{1}{n_a} \sum_{j=1}^{n_a} \phi(f_{c,j}^a) \|_{\mathcal{H}}^2$
- Re-ID inference stage
  - Each image in query and gallery set is fed into { Es, Ec } and the Re-ID head
  - Matching with cosine similarity



Experiment Results

(a) Comparison with state-of-the-arts on VeRi-776 dataset

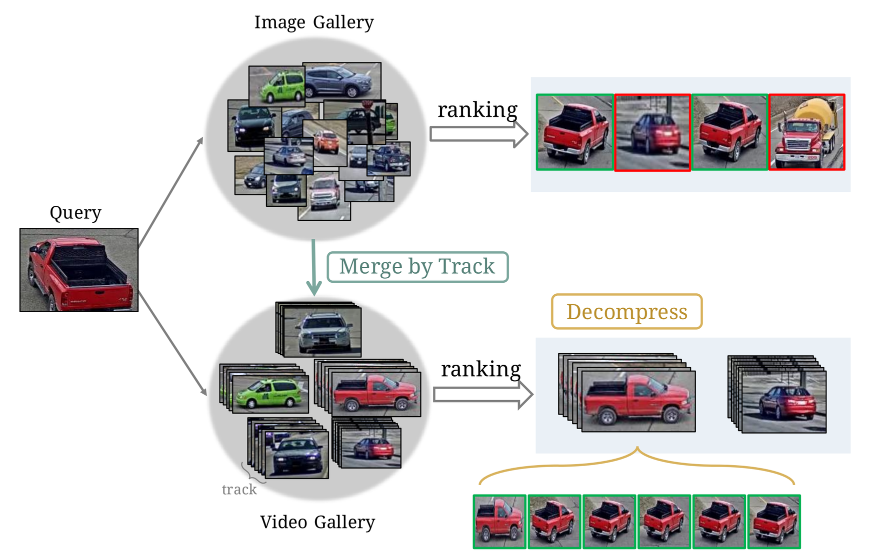
Method	Source	VeRi mAP
XVGAN [34]	BMVC17	24.65
FACT+Plate-SNN [14]	ECCV16	25.88
OIFE [24]	ICCV17	51.42
RNN-HA [26]	ACCV18	56.80
S-CNN+Path-LSTM [19]	ICCV17	58.27
VAMI+STR [35]	CVPR18	61.32
Our JDRN	-	69.08
Our JDRN + re-ranking	-	**73.10**

(b) Comparisons of different baseline with two training set configuration

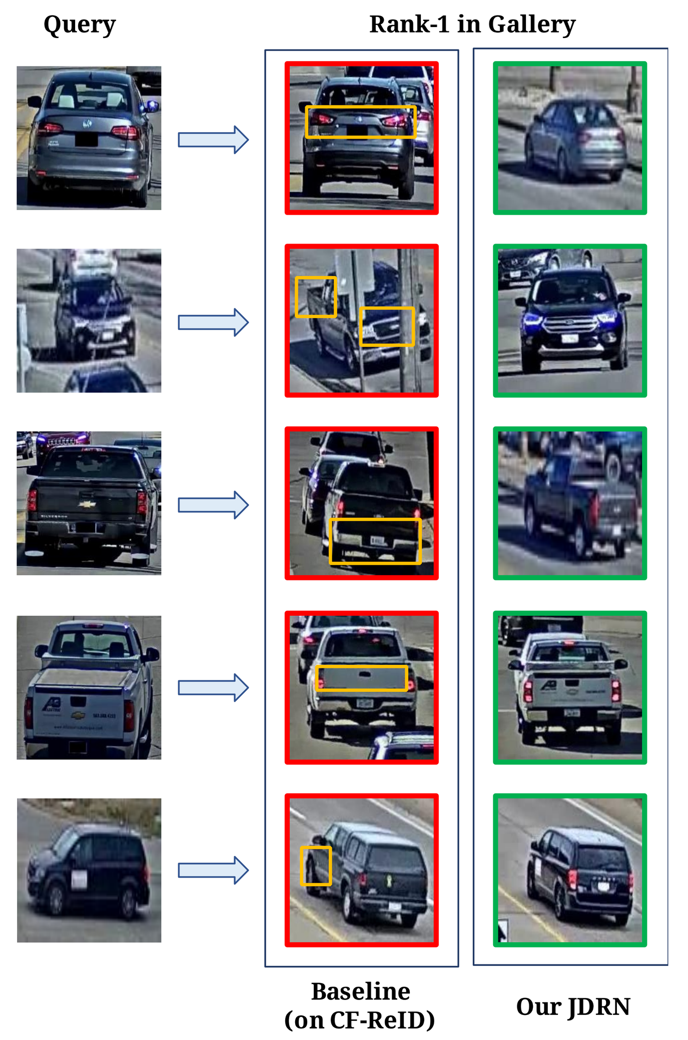
Method	Training set	CF-ReID mAP	VeRi mAP
Baseline (E)	Self	36.26	59.94
Baseline (E)	Joint	35.81	56.98
Baseline w/ $\mathcal{L}_{mmd}$	Joint	32.91	56.97
Our JDRN	Joint	**44.14**	**69.08**

(c) Submission on AIC2019 track2
- K-reciprocal re-ranking [4]
- Ensemble top 3 results
- Video-based inference scheme

Final score → **49.98% in mAP**



Visualization



Reference

[1] Yu-Jhe Li, *et al.* Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. CVPRW, 2018.

[2] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification.CVPR, 2018.

[3] Arthur Gretton *et al.* A fast, consistent kernel two sample test. NeurIPS, 2009.

[4] Zhun Zhong, *et al.* Re-ranking person re-identification with k-reciprocal encoding. CVPR, 2017.