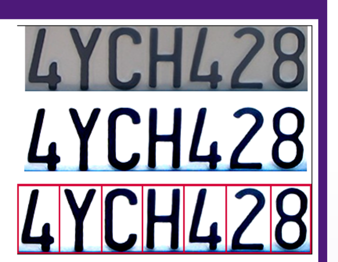
Domain-Invariant Learning in Vehicle Re-identification Task Powered by Deep Neural Networks

Author: Quang Truong   Advisor: Dr. Bo Mei
Department of Computer Science

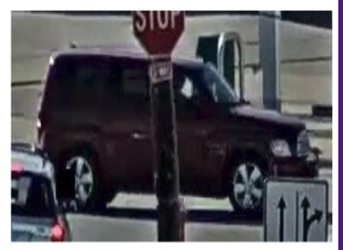
INTRODUCTION

Vehicle Re-identification (re-id) aims to retrieve matching vehicles across different cameras. Unlike person re-id where a person can be easily recognized by pose or face, vehicles are hard to be distinguished because they share similar attributes to each other. In our research project, we leverage the superior performance of deep neural networks (DNN) on high-level computer vision tasks to extract discriminative features of each vehicle. To further enhance the performance of DNNs, we develop a robust domain-invariant pipeline for multi-domain learning

MOTIVATION



- Vehicle re-id is achievable with License Plate Detection using Optical Character Recognition (OCR). However:
    + OCR requires high-resolution images.
    + License plate may be occluded or not presented in a particular view.



- Re-id datasets are relatively limited in general. The reason is that surveillance data is private. → Need to utilize the existing datasets.

➡ Domain- invariant Visual-based Vehicle re-id

CHALLENGES

Two main challenges:

- **Intra-class variability:** a vehicle may look different under various viewpoints due to environments.
- **Inter-class similarity:** two vehicles may look similar if they belong to the same production line.

Example of inter-class similarity

Example of intra-class variability

METHODOLOGY

- **Loss Function:**

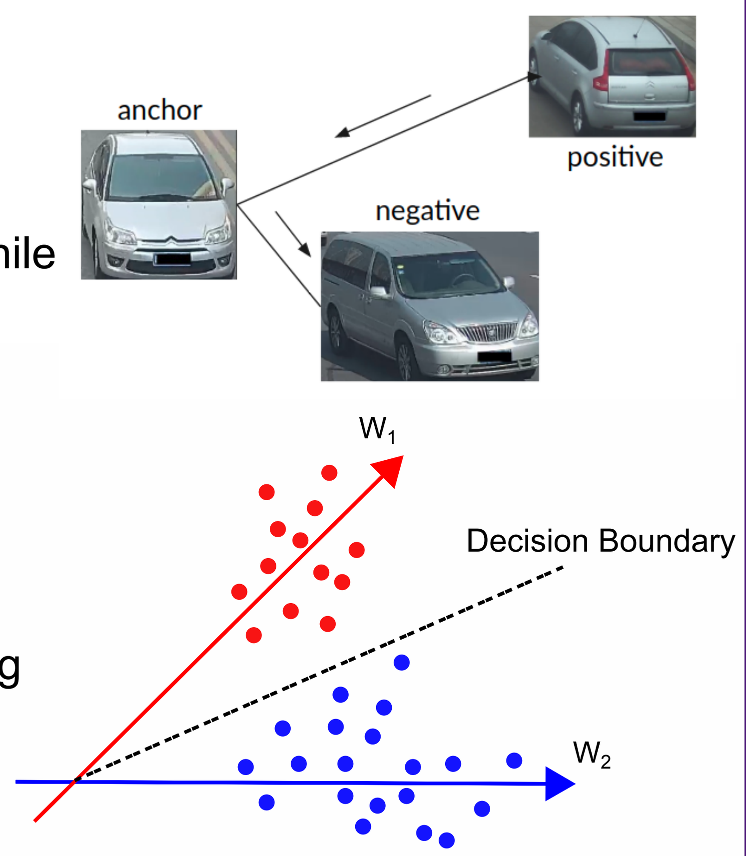
$$\mathcal{L}_{\text{TriSoft}} = \lambda_{\text{Triplet}} \cdot \mathcal{L}_{\text{Triplet}} + \lambda_{\text{Softmax}} \cdot \mathcal{L}_{\text{Softmax}}$$

  + Triplet Loss: addresses intra-class variability
  + Softmax Loss: addresses inter-class similarity

- **2.5-stage Progressive Learning:**
  + Domain Adaptation using CycleGAN (1 stage).
  + Filtering Policy (0.5 stage)
  + Domain-invariant Training with IBN-Net. (1 stage)

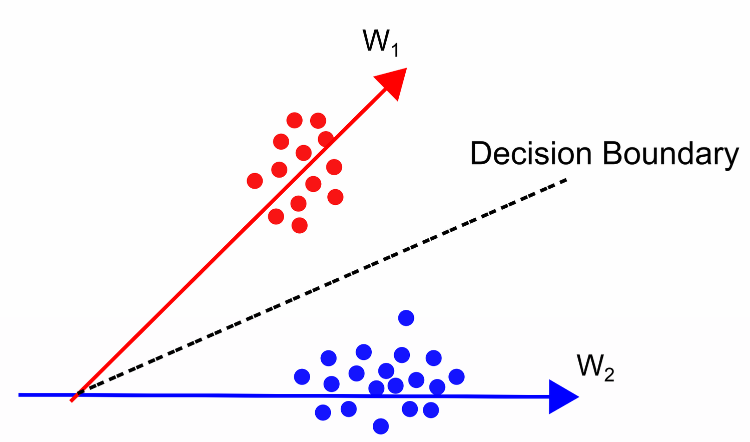
METRIC LEARNING

- **Triplet Loss:** minimizing the distance between anchor and its positive images while maximizing the distance between anchor and its negative images.



- **Softmax Loss (Cross Entropy Loss):** constructing hyperplanes to seperate classes (red and blue).

→ **Combination of Softmax Loss and Triplet Loss:** better discriminative capability due to larger margins between classes.

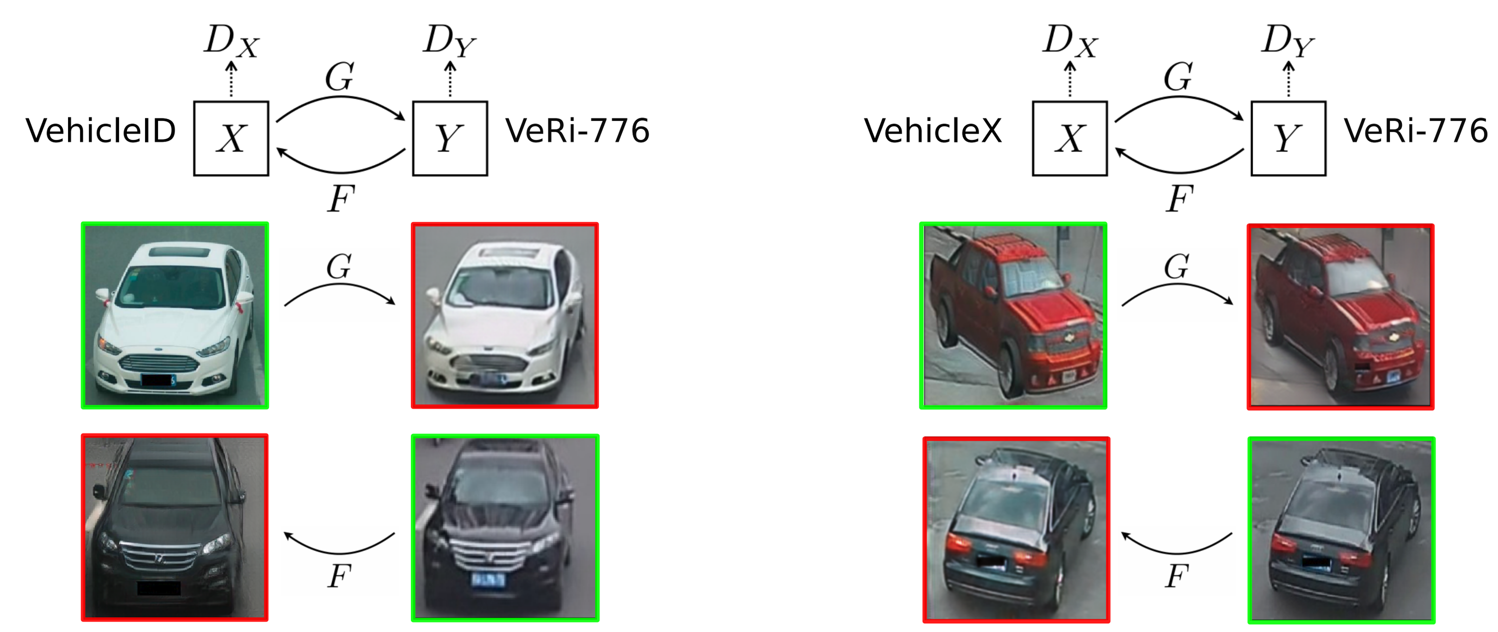


TECHNOLOGY

NVIDIA PyTorch python™

CYCLEGAN

Target: Learn mapping function from one domain to another using Cycle-Consistent Adversarial Networks to convert images.



Note: VehicleID and VeRi-776 are real datasets, VehicleX is synthetic dataset constructed by softwares. Green border denotes original images, while red border denotes images generated by CycleGAN.

FILTERING POLICY

- **Target:** CycleGAN may generate images that are unsuccessfully adapted. These images harm the learning progress. Filtering Policy is a novel filtering technique to filter out those images.

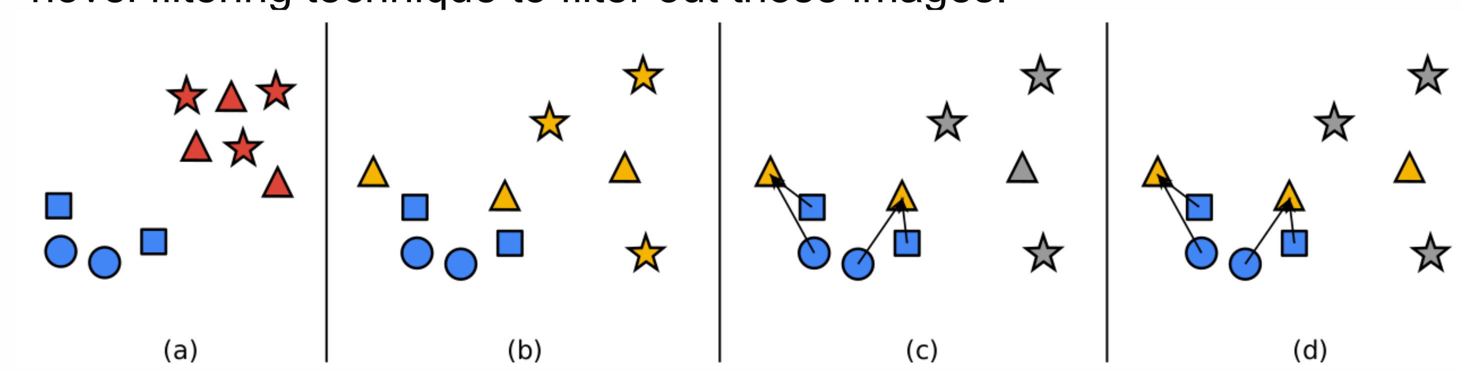
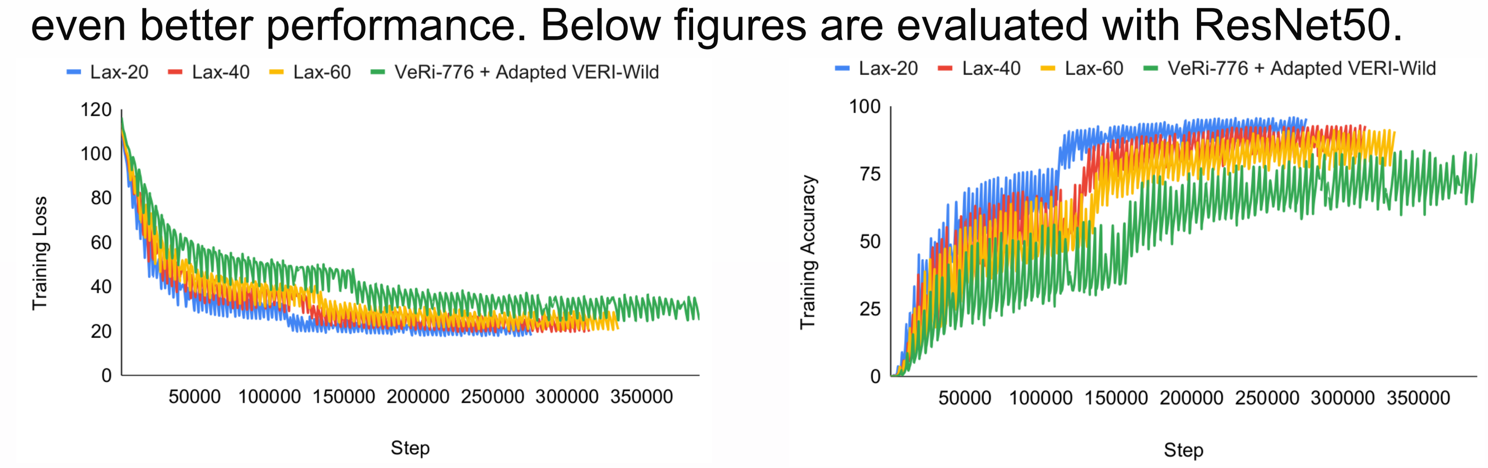


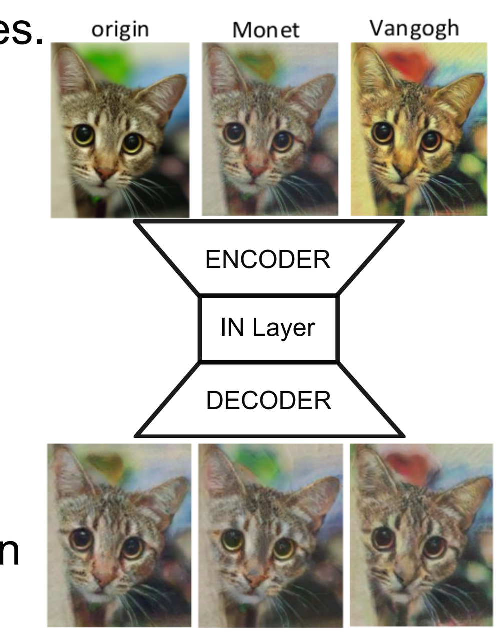
Illustration of dataset merging policies; (a) merging without any domain adaptation; (b) merging with domain adaptation, where triangles are succesfully adapted (close to blue) while stars are not (far from blue); (c) Rigid Filtering Policy where only easy samples are included; (d) Lax Filtering Policy where easy samples as well as hard samples having same identities with the easy ones are included.

- Filtering Policy leads to a quicker loss convergence but still keeps same or even better performance. Below figures are evaluated with ResNet50.



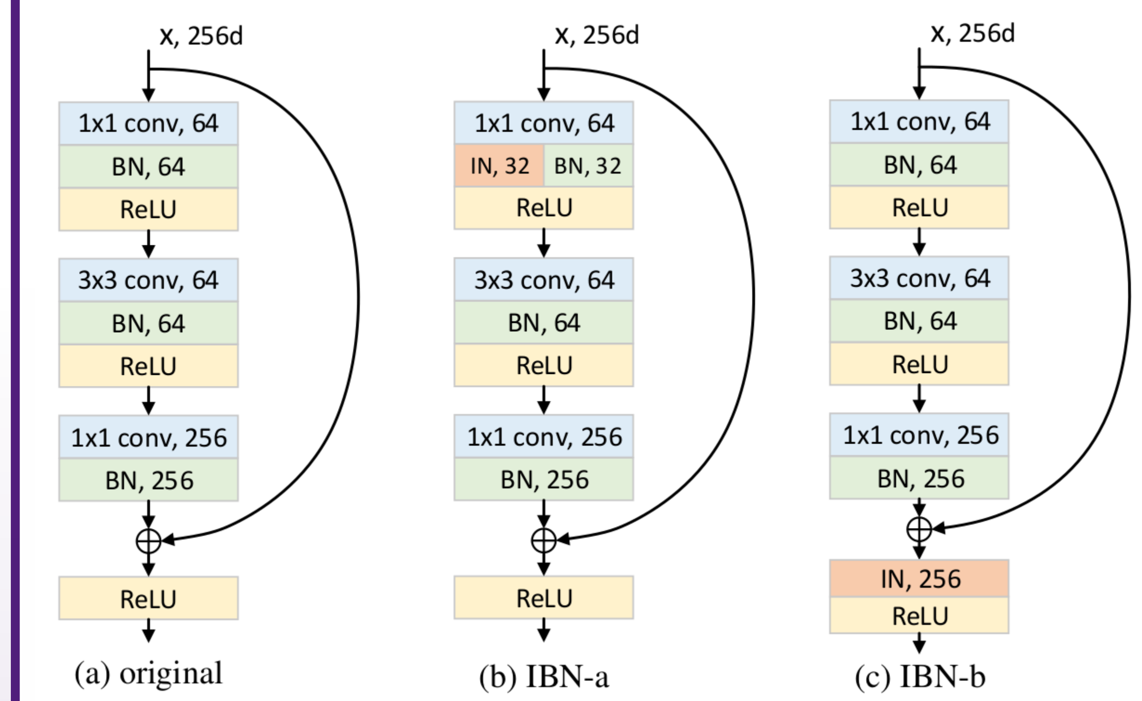
IBN-NET

- **Target:** eliminate apperance variance while maintaining discriminative features of images.



- Effect of Instance Norm (IN) Layer (image on the right): filtering out style and learning only contents of an image.
- However, IN Layer is found to have inferior performance on high-level vision tasks such as classification and re-id. Meanwhile, Batch Normalization (BN) Layer is leveraged in popular DNN architectures.

➡ Instance-Batch Normalization Block Architecture



RESULT & CONCLUSION

- Our proposed approach outperforms other state-of-the-art methods evaluated on VeRi-776 benchmark.

Method	#Images	#IDs	mAP (%)	Rank@1 (%)
OIFE + ST [3]	225,268	36,108	51.42	68.30
SAVER [4]	3,706,670	N/A	79.60	96.40
VehicleNet [5]	434,440	31,805	83.41	96.78
Ours	**355,566**	**31,947**	**85.78**	**97.14**

- In conclusion, we proposed a domain-invariant vehicle re-id pipeline that relies on state-of-the-art techniques in domain learning. We also prove that blind merging and domain adaptation alone are not sufficient to maximize the generalization capability of the model.

REFERENCES

[1] J. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networkss,” in ICCV 2017.
[2] P. Xingang, L. Ping, S. Jianping, and T. Xiaoou, “Two at once: Enhancing learning and generalization capacities via IBN-Net,” in ECCV 2018.
[3] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, J. Yan, S. Wang, H. Li, and X. Wang, “Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification,” in ICCV 2017.
[4] P. Khorramshahi, N. Peri, J. Chen, and R. Chellappa, “The devil is in the details: Self-supervised attention for vehicle re-identification,” in ECCV 2020.
[5] Z. Zheng, T. Ruan, Y. Wei, Y. Yang, and T. Mei, “VehicleNet: Learning robust visual representation for vehicle re-identification,” in IEEE Transactions on Multimedia.

ACKNOWLEDGEMENT

- I would like to thank my research advisor Dr. Mei who has always supported me since I came to TCU.
- This work is supported by SERC Grant.