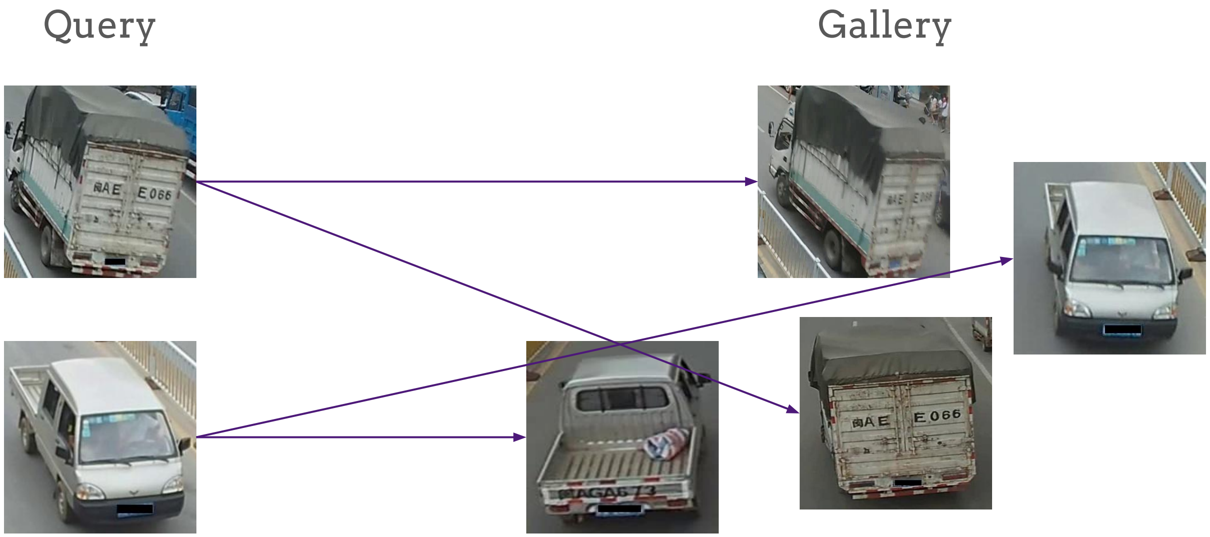
24th IEEE International Conference on Intelligent Transportation - ITSC 2021

Not All Data Matters: An Efficient Approach to Multi-Domain Learning in Vehicle Re-identification

Quang Truong, Bo Mei

Background

- Vehicle Re-identification is the task of retrieving matching vehicles across non-overlapping cameras. For example, we need to match each image in the query with each image in the gallery and rank the images based on similarity.



Source: VeRi-776 Dataset

Background

- Vehicle Re-identification is the task of retrieving matching vehicles across non-overlapping cameras. For example, we need to match each image in the query with each image in the gallery and rank the images based on similarity.

Source: VeRi-776 Dataset

Background

- One approach of vehicle re-identification is Optical Character Recognition (OCR). By detecting license plates and then extracting numbers and characters, we can match the images with corresponding information of the registered vehicles in database.

Source: https://upload.wikimedia.org/wikipedia/commons/9/9c/California_license_plate_ANPR.png

Background

- How can this approach work when the license plate is not visible?

Source: VeRi-776 Dataset

→ Second approach: Visual-based Vehicle Re-identification

- By leveraging a feature extractor, a visual-based vehicle re-identification system can rank the images based on appearance similarity.

Motivation

- Visual-based vehicle re-identification is a particularly challenging due to two reasons:
    a. Inter-class similarity: similar appearance of two vehicles because they belong to the same series of a particular brand.
    b. Intra-class variability: image variations across non-overlapping cameras even though they belong to the same car.

Example of inter-class similarity

Example of intra-class variability

Source: VeRi-776 Dataset

Motivation

- In order to overcome these challenges, extra features are usually needed. Common approaches are:
  a. Architecture Re-engineering
  b. Additional annotations or labels.

- Both methods require tremendous efforts. Therefore, we decide to tackle the problem by multi-domain learning.

- Specifically, to enhance the generalization capability of the model, we train our model on numerous vehicle datasets. Our model will learn generalizable and transferable visual representation across multiple domains.

Motivation

- However, the approach of multi-domain learning does not captivate much attention from researchers.
- Majority of multi-domain learning models only blindly merge datasets without being aware of significant gaps between domains.
- Our research explores how to construct an efficient multi-domain learning pipeline by investigating the followings:
  a. Effects of style transferring to reduce domain discrepancy.
  b. Image selection methods to filter out unnecessary images.
  c. Model options to learn style-invariant features.

Methods

- Our proposed approach can be summarized as a 2.5-stage progressive learning strategy:

    a. Domain Adaptation using CycleGAN (1 stage)

    b. Filtering Policy (0.5 stage)

    c. Domain-invariant Training with IBN-Net (1 stage)

Domain Adaptation using CycleGAN

- Domain distribution plays a vital role when learning visual representations. Therefore, we need to perform domain adaptation to translate images from the source domain to the target domain.

- CycleGAN is first introduced by Zhu *et. al.* as an unaligned image translation model for style transfer task [1]. We leverage CycleGAN as our domain adaptation model because of the following reasons:
  a. Performance superiority.
  b. The unavailability of paired vehicle datasets.

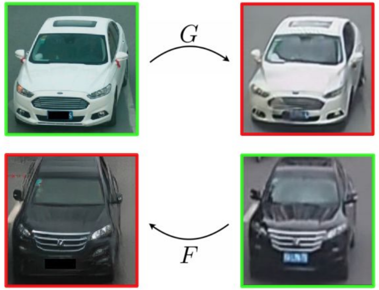
[1] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE ICCV.

Domain Adaptation using CycleGAN

Source: VeRi-776, VehicleID, and VehicleX Datasets

- In the above example, CycleGAN learns the mapping functions between source domains (VehicleID and VehicleX) and target domain (VeRi-776).
- Note: VehicleID and VeRi-776 are real datasets, while VehicleX is a synthetic dataset. Green border denotes original images, while red border denotes images generated by CycleGAN.

Domain Adaptation using CycleGAN



Source: VeRi-776, VehicleID, and VehicleX Datasets

- Observations:
  a. VehicleID to VeRi-776: image becomes blurry because VeRi-776 is a low-resolution dataset.
  b. VeRi-776 to VehicleID: image becomes sharper because VehicleID is a high-resolution dataset.
  c. VehicleX to VeRi-776: image looks more realistic with shades, the background is also adjusted to match the real background.
  d. VeRi-776 to VehicleX: image looks less realistic with no shade, the brightness is increased to match the synthetic background.

Filtering Policy

- Even though CycleGAN helps reduce domain discrepancy between the target domain and the source domain, illy-translated images are unavoidable. These images deteriorates the learning progress.

Examples of unsuccessful image translation (VehicleID → VeRi-776)

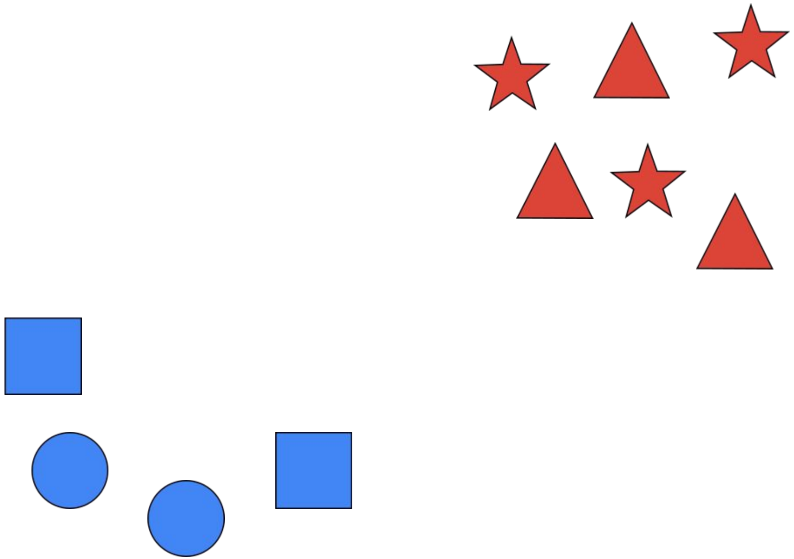
Filtering Policy

- Another problem of multi-domain learning is the increase in images and identities.
- Large amount of images and identities lead to a slow loss convergence:
  a. Cross-entropy Loss: relies on number of identities → Extreme Classification when there is a tremendous number of identities.
  b. Triplet Loss: relies on triplet samples (anchor, positive, negative) → More negative samples as a result of dataset expansion.

Filtering Policy

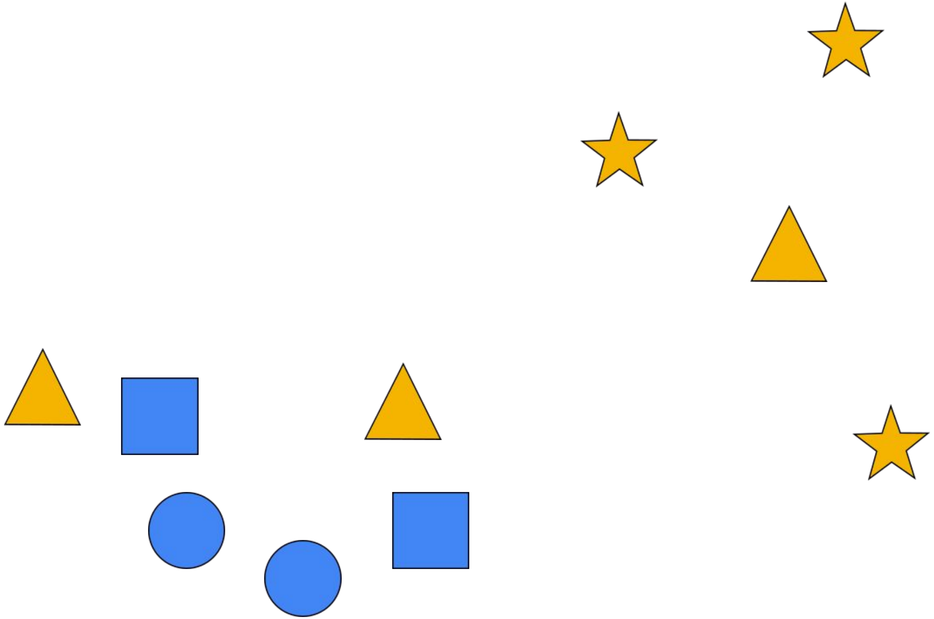
- We propose Filtering Policy to counter these problems.
- First, we compute the pairwise distance between each image in the target dataset and each image in the adapted source dataset.
- We proceed to either one of these filtering policies:
    a. Rigid Filter: only select the k-nearest adapted source images for each target image.
    b. Lax Filter: select the k-nearest adapted source images for each target image, and also include those images having same ground truth label with the k-nearest selected images.

Filtering Policy Visualization



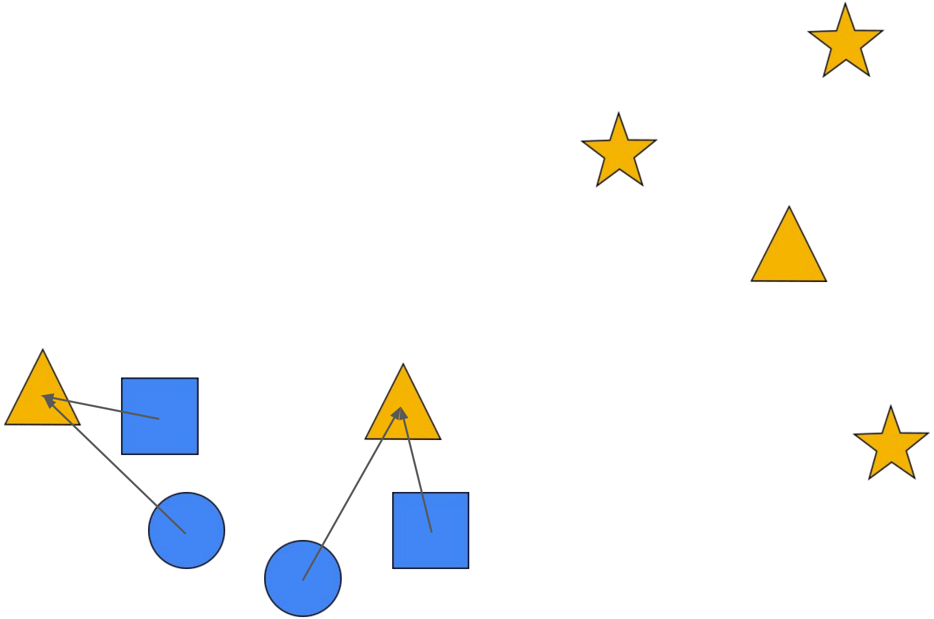
(a) Merging without any domain adaptation where blue and red are two domains

Filtering Policy Visualization



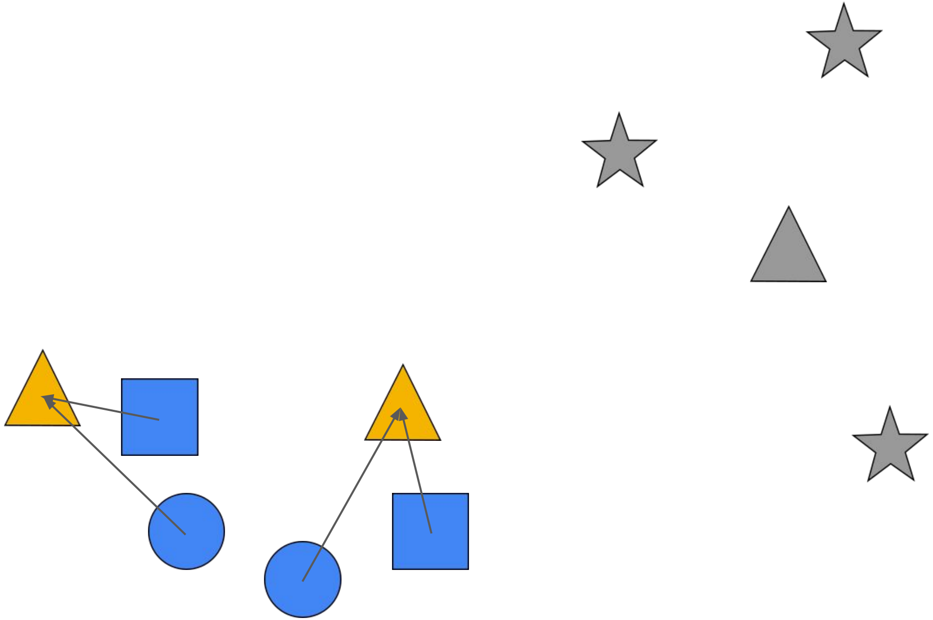
(b) Merging with domain adaptation,
where triangles are successfully adapted (close to blue) while stars are not (far from blue)

Filtering Policy Visualization



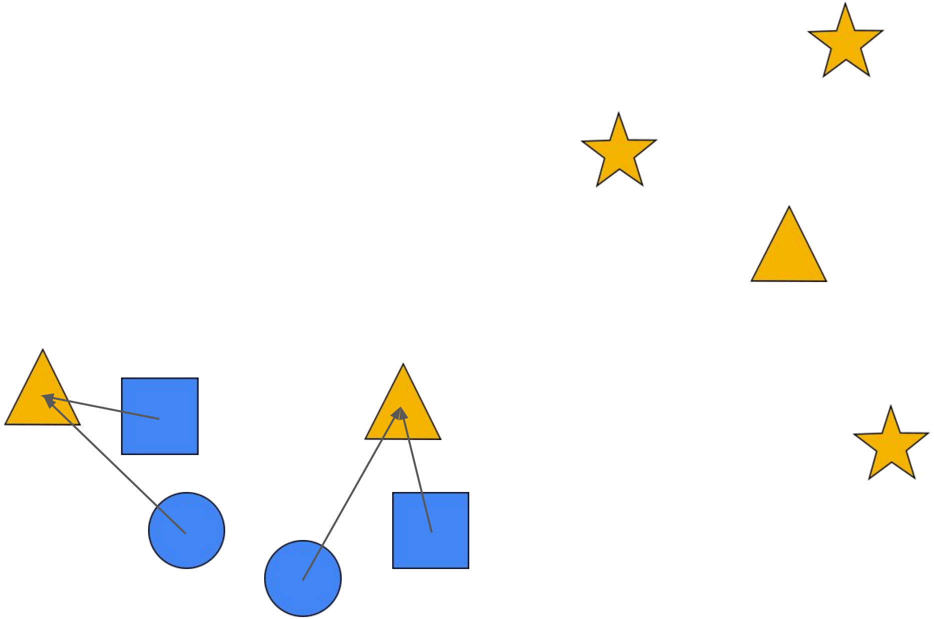
(c) Rigid-1 Filtering Policy where only easy samples
of the triangle (close to blue) are included

Filtering Policy Visualization



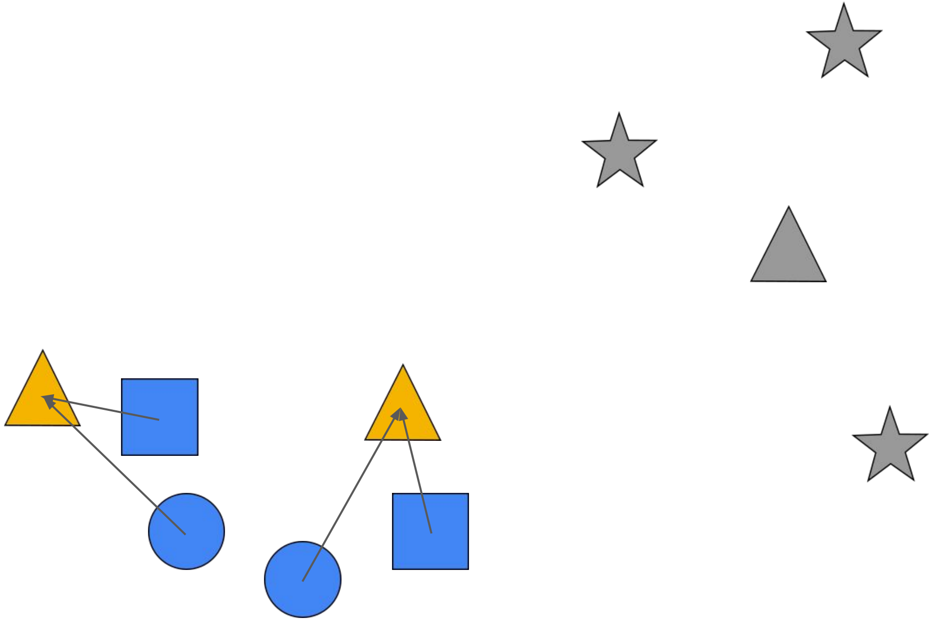
(c) Rigid-1 Filtering Policy where only easy samples of the triangle (close to blue) are included

Filtering Policy Visualization



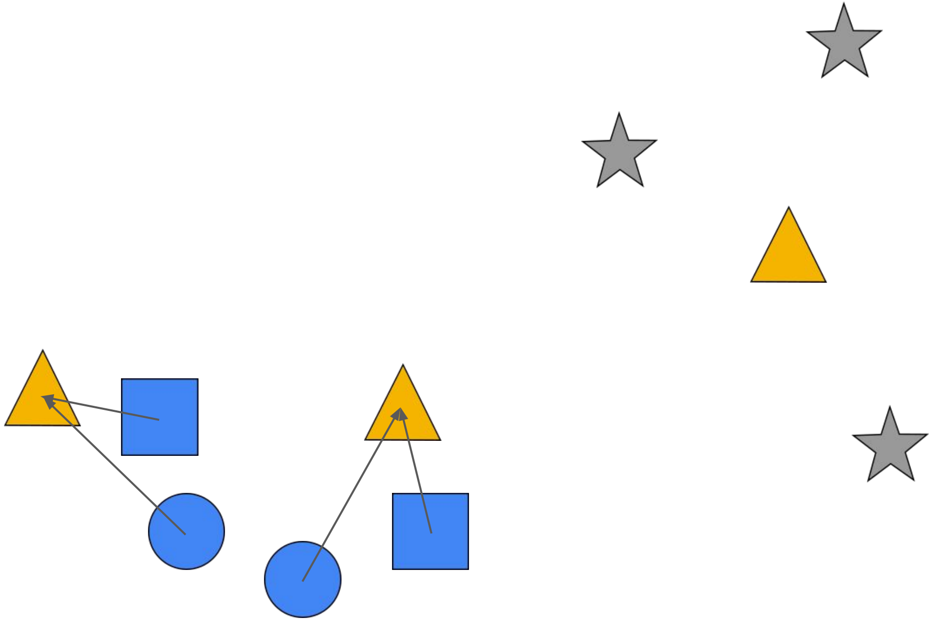
(d) Lax-1 Filtering Policy where all samples of the triangle are included

Filtering Policy Visualization



(d) Lax-1 Filtering Policy where all samples of the triangle are included

Filtering Policy Visualization



(d) Lax-1 Filtering Policy where all samples of the triangle are included

Domain-invariant Training with IBN-Net

- IBN-Net [2] is a domain-invariant convolution architecture that integrates Instance Normalization (IN) and Batch Normalization (BN) into one module for both learning appearance-invariant features and preserving discriminative features.
- By leveraging IBN-Net in our method, we would be able to mitigate the domain discrepancy between the source dataset and the target dataset.

[2] X. Pan, P. Luo, J. Shi, and X. Tang, "Two at once: Enhancing learning and generalization capacities via IBN-Net," in 2018 ECCV

Results

- We expand VeRi-776 by merging it with filtered adapted datasets such as VERI-Wild, VehicleX, and VehicleID. The newly formed dataset is called MegaVeRiVehicle.
- We then evaluate our baseline, ResNet50 [3], on this dataset. The result is illustrated below.

Policy	**Dataset**	**#Images**	**#IDs**	**mAP**	**CMC**	
					Rank@1	**Rank@5**
N/A	VeRi-776	37,778	576	77.95	95.59	97.91
Lax-60	+ Adapted VERI-Wild	231,870	23,852	84.04	96.36	98.15
Lax-2	+ Adapted VehicleX	73,732	1,329	79.07	95.53	98.57
Rigid-2	+ Adapted VehicleID	12,186	6,190	77.22	94.87	97.74
N/A	MegaVeRiVehicle	355,566	31,947	85.23	96.66	98.69

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE CVPR*.

Results

- We can further improve the performance by using ResNet50-IBN-a instead of ResNet50.

Model	**mAP**	**CMC**	
		Rank@1	**Rank@5**
ResNet50	85.23	96.66	98.69
ResNet50-IBN-a	85.78	97.14	98.27

- Our model uses less external data and has significantly better performance on VeRi-776.

Method	**#Images**	**#IDs**	**mAP**	**Rank@1**
OIFE+ST [4]	225,268	36,108	51.42	68.30
SAVER [5]	3,706,670	-	79.60	96.40
VehicleNet [6]	434,440	31,805	83.41	96.78
Ours	**355,566**	**31,947**	**85.78**	**97.14**

[4] 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 2017 IEEE ICCV.

[5] P . Khorramshahi, N. Peri, J. Chen, and R. Chellappa, “The devil is in the details: Self-supervised attention for vehicle re-identification,” in 2020 ECCV.

[6] 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.

Ablation Studies

- In this section, we demonstrate the effects of Domain Adaptation and Filtering Policy on the performance using only the baseline.
- By illustrating the performance difference when one of our core components is removed, we can clearly see the importance of domain distribution in learning visual representation.

Effects of CycleGAN

- We evaluate the baseline performance on unfiltered adapted datasets.
- Domain adaptation generally helps the performance.
- The only exception is VehicleID, which possesses significant differences in image settings compared to VeRi-776. Therefore, CycleGAN generates undesired images, leading to a poor performance.

Dataset	**mAP**	**CMC**	
		Rank@1	**Rank@5**
VeRi-776 only	77.95	95.59	97.91
VeRi-776 + VERI-Wild	83.85	96.31	98.39
VeRi-776 + Adapted VERI-Wild	84.04	96.48	98.03
VeRi-776 + VehicleX	78.63	96.31	98.45
VeRi-776 + Adapted VehicleX	78.82	96.25	98.57
VeRi-776 + VehicleID	77.24	94.46	97.38
VeRi-776 + Adapted VehicleID	72.02	90.64	96.42

Effects of Filtering Policy

- We control parameter *k* to illustrate how varying included images on different datasets may help the performance.
- By selecting *k* = 60 and choosing Lax Policy, we only use 231,870 out of 277,797 images (~83%) in the adapted VERI-Wild to have the best performance. Therefore, the total images for training is 269,648 (37,778 images in VeRi-776).

Policy	#Images	#IDs	mAP	CMC	
				Rank@1	Rank@5
Original	315,575	31,247	84.04	96.48	98.03
Lax-20	220,842	18,166	83.40	96.19	98.39
Lax-30	239,532	20,538	83.17	95.71	98.03
Lax-40	253,341	22,248	83.36	96.60	98.45
Lax-50	262,160	23,446	83.94	96.01	98.21
Lax-60	**269,648**	**24,428**	**84.04**	**96.36**	**98.15**
Lax-70	275,840	25,259	83.65	96.48	98.21

Effects of Filtering Policy on VeRi-776 + Adapted VERI-Wild

Effects of Filtering Policy

- By selecting $k = 2$ and choosing Lax Policy, we only include 73,732 out of 75,516 images in the adapted VehicleX (~98%).
- A significantly large portion of VehicleX is included because VehicleX is synthesized from VeRi-776.

Policy	#Images	#IDs	mAP	CMC	
				Rank@1	Rank@5
Original	113,294	1,938	78.82	96.25	98.57
Lax-1	108,237	1,846	78.42	96.07	98.21
Lax-2	**111,510**	**1,905**	**79.07**	**95.53**	**98.57**

Effects of Filtering Policy on VeRi-776 + Adapted VehicleX

Effects of Filtering Policy

- Unlike two previous experiments, because distribution gap between VehicleID and VeRi-776 is large, Rigid Policy is leveraged.
- Only 12,186 out of 113,346 images in the adapted VehicleID are included (~11%).

Policy	#Images	#IDs	mAP	CMC	
				Rank@1	Rank@5
Original	151,124	13,740	72.02	90.64	96.42
Lax-1	88,485	5,149	74.91	93.15	96.16
Lax-2	104,282	6,766	73.90	92.31	97.02
Rigid-1	45,385	5,149	77.12	94.70	97.74
Rigid-2	**49,964**	**6,766**	**77.22**	**94.87**	**97.74**
Rigid-3	53,787	7,819	75.90	93.80	97.44
Rigid-5	60,092	9,182	75.29	93.50	97.38

Conclusion

- Images should be translated to the target domain before learning.
- Filtering Policy helps reduce domain discrepancies across datasets.
- IBN-Net is particularly helpful in multi-domain learning.
- Not all data are conducive to the model performance because source images may not be well- aligned with the target dataset's distribution.

References

[1] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE ICCV.

[2] X. Pan, P. Luo, J. Shi, and X. Tang, "Two at once: Enhancing learning and generalization capacities via IBN-Net," in 2018 ECCV.

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE CVPR.

[4] 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 2017 IEEE ICCV.

[5] P . Khorramshahi, N. Peri, J. Chen, and R. Chellappa, "The devil is in the details: Self-supervised attention for vehicle re-identification," in 2020 ECCV.

[6] 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.

[7] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in 2016 IEEE ICME. (VeRi-776 dataset)

[8] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in 2016 IEEE CVPR. (VehicleID dataset)

[9] Y. Yao, L. Zheng, X. Yang, M. Naphade, and T. Gedeon, "Simulating content consistent vehicle datasets with attribute descent," in 2020 ECCV. (VehicleX dataset)

[10] Y. Lou, Y. Bai, J. Liu, S. Wang, and L.-Y. Duan, "VERI-Wild: A large dataset and a new method for vehicle re-identification in the wild," in 2019 IEEE CVPR. (VERI-Wild dataset)

Thank you for your attention!