

Enhancing Cross-Domain Vehicle Detection with Transfer Learning and Source-Similar Sample Integration

Chi-Han Chen¹, Shu-Fang Zhang¹, Hsin-Te Wu², Rung-Shiang Cheng^{3*}

¹ Department of Information Technology, Overseas Chinese University, Taiwan

² Department of Computer Science and Information Engineering, National Taitung University, Taiwan

³ Department of Artificial Intelligence and Computer Engineering, National Chin Yi University of Technology, Taiwan

ludwig1017@ocu.edu.tw, nia983516@gmail.com, wuhsinte@nttu.edu.tw, rscheng@ncut.edu.tw

Abstract

In practical street vehicle detection applications, models may require a large amount of data due to varying street conditions across different regions, influenced by factors such as shooting angles and weather changes. Even with a high-precision detection model, applying it to a new urban area requires incorporating data from the new domain into the training process. To leverage the knowledge of an existing model for a new task, transfer learning models are utilized to prevent over-reliance on previous knowledge during training, which might result in the inability to detect target samples in the new task. Common research and application methods include knowledge distillation and cross-domain adaptation. This paper introduces a training paradigm that involves incorporating a small amount of source-approximate samples from the old task into the new task, followed by fine-tuning to experiment with cross-domain learning applications. Experimental results demonstrate that our paradigm, when augmented with source-approximate data—samples with similar scene or weather characteristics to the source domain—exhibits higher adaptability for detecting vehicle objects in the target domain compared to models utilizing knowledge distillation.

Keywords: Street vehicle detection, Transfer learning, Cross-domain adaptation, Training paradigm, Fine-tuning training

1 Introduction

In deep learning, choosing the right dataset and model for different tasks is really important. In real-world applications, changes in scenes and object appearances often cause models to struggle with adapting to new domains. This issue is known as “domain shift.” Domain shift means that the model might not perform well in the new domain or could lose its ability to recognize features in the original domain.

For street vehicle detection, a large number of samples are needed for annotation and training so the model can learn how vehicles look in urban areas. However, in

practice, even if a vehicle detection model works well for one urban area (domain A), it might face problems when used in a different urban area (domain B). Training the model directly with data from domain B can lead to it being overly influenced by what it learned from domain A, making it hard for the model to accurately learn the vehicle features specific to domain B. This can result in decreased detection capability due to domain shift issues.

In transfer learning, the concept of “domain” involves the differences between the source domain and the target domain, which can include sample features or model scales. “Cross-domain learning” builds on this idea. For instance, in street detection tasks, different weather conditions or backgrounds represent domain changes. These changes present challenges for practical applications like vehicle detection, including variations in lighting, monitoring angles, or extreme weather conditions. BDD100k is selected as the target domain in this study due to its rich diversity in urban scenes, which differ significantly from the structured city layouts and foggy conditions found in the Cityscapes and Foggy Cityscapes datasets. This allows for a more realistic evaluation of cross-domain adaptability.

The goal of cross-domain detection models is to learn features from different domains while keeping the original detection capabilities intact, thus enhancing generalization. This approach helps models adapt better and handle various changes in new environments, improving their stability and accuracy across different scenes.

To tackle this problem, it’s necessary to update the model’s criteria for identifying vehicle features during training. Common methods include knowledge distillation and domain adaptation models. Knowledge distillation involves setting parameters during training to ensure the model isn’t affected by domain shift during the transfer process. Domain adaptation models update the model’s ability to recognize object features by learning consistent features across different domains. These methods all involve tweaking the model’s standards for object features.

This study introduces a new training paradigm inspired by SSDA-YOLO [4], aimed at improving object detection across different domains. Unlike SSDA-YOLO, which uses a teacher-student model with differing backbone sizes, the proposed method employs two YOLO models of the same size. The key innovation is transferring

*Corresponding Author: Rung-Shiang Cheng; Email: rscheng@ncut.edu.tw

DOI: <https://doi.org/10.70003/160792642025122607010>

knowledge (weights) from a model trained on a source domain to another model trained with samples from a new domain. This approach generates “source-similar samples” and “pseudo-target samples” based on scene and weather features. Unlike SSDA-YOLO, which requires equal samples from both domains, the new method needs fewer target domain samples. Experiments show superior detection in target domains, though detection in the source domain may suffer due to the lack of parameter adjustments. Future research may explore pre-processing the source domain model to reduce the sample size needed for training.

2 Related Works

2.1 Using Object Detection Model to Vehicle Detection

Vehicle detection models initially relied on hand-crafted features or traditional machine learning models for feature learning [5-7]. Influenced by the features of the trained samples and the inference capability of the model, these detection models faced limitations in widespread application across different street scenes, resulting in lower detection performance. Since 2010, with the introduction of deep learning models and Convolutional Neural Networks (CNNs), many researchers have proposed higher-performing detection models.

Object detection models can be categorized based on their architecture into: 1. Second-stage models based on Fast R-CNN [8-11], and 2. First-stage models based on YOLO (You Only Look Once) [12-15]. For the task of vehicle object detection, models need to possess real-time detection capabilities. As second-stage models require more time for inference, subsequent researchers have predominantly focused on using first-stage models to construct street vehicle detection models. Notably, YOLOv4 was proposed by researchers including Alexey Bochkovskiy and collaborators from the United States Institute of Research [15]. This model integrates the Cross Stage Partial Network (CSPNet) [16], providing efficient detection results suitable for tasks such as autonomous vehicle driving and traffic flow management in vehicle detection.

2.2 Domain Adaptation in Transfer Learning

Domain adaptation [17-18] is one of the methods in transfer learning [19-22]. Since 2014, many scholars have proposed various approaches, starting from initially incorporating loss functions into neural networks to update object feature parameters and enhance the model’s detection capabilities across different domains. Subsequently, researchers explored alternative methods, such as style transfer on source domain samples and extraction of object features from different domains, to adapt the features of objects in diverse domains [23-27]. Additionally, to reduce the training cost of the model in the target domain, some scholars utilized Teacher-Student Models [28-31] for domain adaptation. Among them, [4] introduced the SSDA-YOLO model, using YOLOv5 as the model framework and employing semi-supervised learning

methods for knowledge distillation training, thus enhancing the detection capabilities of the domain adaptation object detection model.

3 Problem Formulation and Problem Solution

The main objective of this paper is to propose a training method for a street vehicle detection model applicable to real-world cross-domain scenarios. To create an environment suitable for cross-domain applications, considerations include defining “different domains” in the street task and obtaining similar scenes. For experimental verification, three publicly available datasets, namely Foggy Cityscapes [1], Cityscapes [2], and BDD100k [3], were chosen as experimental samples for the following reasons.

Foggy Cityscapes are generated from synthetic foggy samples from the Cityscapes dataset and are commonly used in vehicle detection tasks as both source and target domain samples. Since this paper explores whether incorporating similar samples after model transfer can assist in fine-tuning for adaptation to different environments, both datasets are defined as source domain scenes. The target domain scenes are selected from BDD100k, offering diverse street views from different cities at various times, providing the model with a variety of target domain scenes rather than being restricted to a specific street. Additionally, the training set samples in BDD100k are augmented with artificially generated foggy images, aligning them with the source domain scenes. BDD100k offers more diverse street scenes than Cityscapes, including variations in lighting, weather, and urban environments, making it ideal for evaluating model adaptability in realistic cross-domain scenarios. The experimental samples thus include real sunny scenes and artificially generated foggy scenes from both source and target domains.

The Cityscapes dataset is a widely used public dataset for vehicle detection research, containing street view images from 50 cities, covering different times of day and weather conditions, with tens of thousands of images divided into 30 categories. These richly annotated images make it suitable for various computer vision tasks such as object detection, classification, and instance segmentation. Classic papers like Mask R-CNN have used this dataset to validate model capabilities in street scene detection.

Foggy Cityscapes extends the Cityscapes dataset by simulating foggy street scenes through a fogging model applied to Cityscapes samples. Many researchers use these two datasets to simulate cross-domain vehicle detection experiments, validating the adaptability of detection models in different street scenes.

In practical applications, acquiring street images with specific weather conditions can be challenging. Even if a successful object detection model is trained, re-training it for special weather conditions, compared to sunny street scenes, may lack a sufficient number of samples. Therefore, this paper investigates whether transferring the source model trained on Foggy Cityscapes and subsequently fine-

tuning it with sunny images from Cityscapes can train the target model without the need for an extensive collection

of special weather scene images.

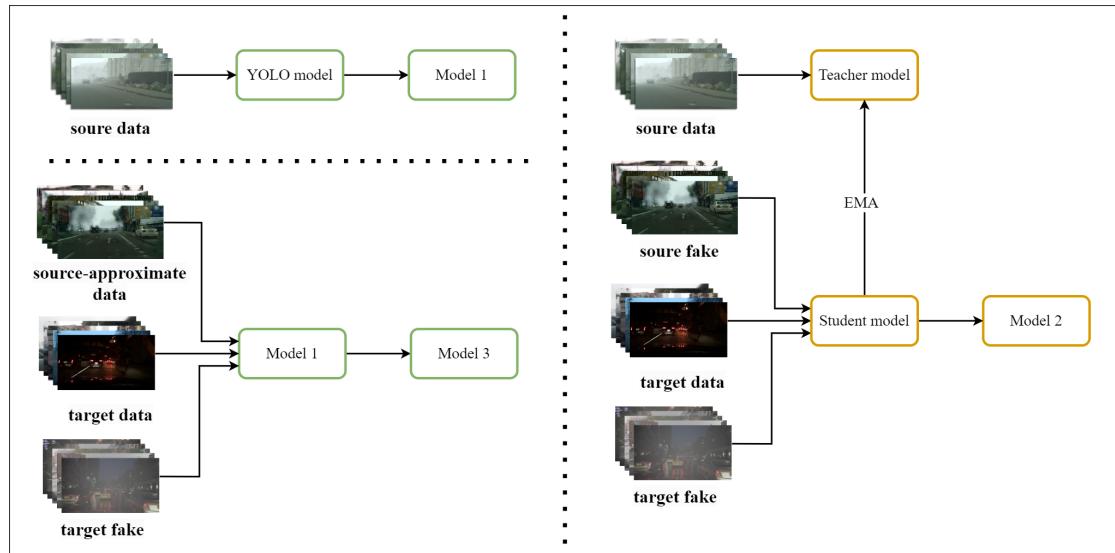


Figure 1. Training process diagram of our paradigm

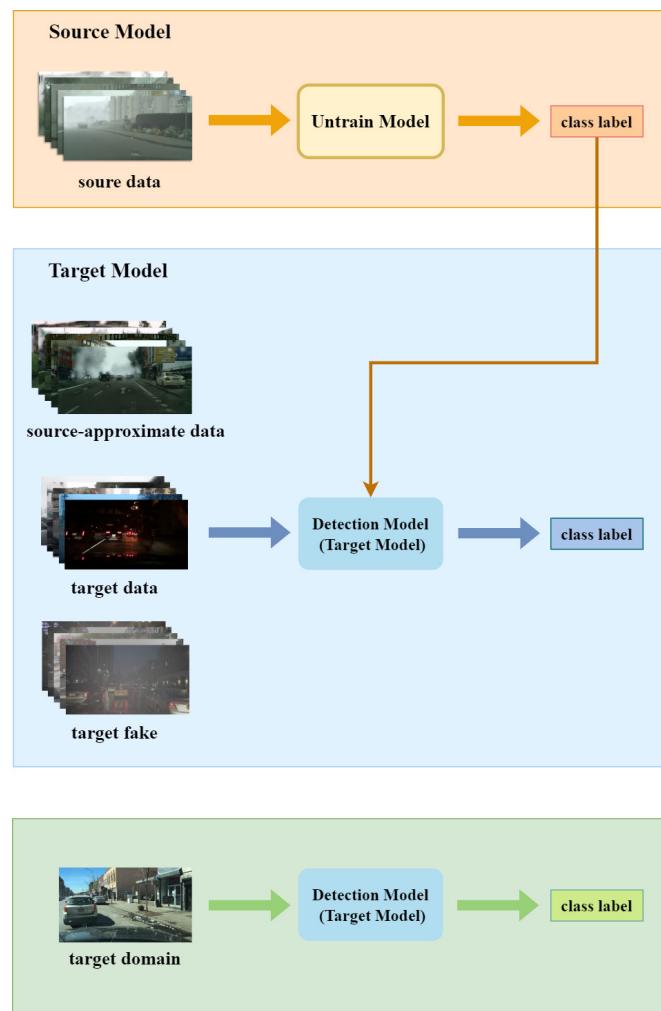


Figure 2. Our training paradigm diagram

Experiments were conducted with three models:

1. Model 1: Transferred the model and retrained without any additional samples.
2. Model 2: Transferred the model and utilized knowledge distillation to train on cross-domain data [4].
3. Model 3: Transferred the model and retrained by combining source-approximate with target domain street images.

The detection capabilities of the three methods for vehicle objects in the target domain were compared. The experimental results indicate that direct model transfer (Model 1) without fine-tuning is significantly impacted by domain shift, resulting in a substantial decrease in detection capabilities. However, fine-tuning the model after transfer with a small number of source-approximate data (Model 3) mitigates the effects of domain shift and exhibits higher adaptability compared to the knowledge distillation model (Model 2) applied to cross-domain tasks.

During training, samples are categorized into four types:

- Source data: Foggy street scenes from Foggy Cityscapes.
- Source-approximate data: Sunny street scenes from Cityscapes that share similar scene layouts with the foggy source domain.
- Target data: Real sunny scenes from BDD100k.
- Target fake data: Foggy images artificially generated from BDD100k samples.

“Target fake” refers to generated street scenes meant for target data, while “source-approximate” refers to street scenes similar to the source domain, helping to train the target model. As shown in Figure 1, the overall training process consists of source-model training and target-domain adaptation using the four types of samples described above. Here’s how the training process works:

Model 1: Trained with 1700 images of foggy street

scenes from the Foggy Cityscapes dataset. This helps the YOLOv5 model learn vehicle features in extreme weather conditions found in the source domain (Domain A). Consequently, the model becomes adept at identifying traffic objects in such challenging conditions.

Model 3: Uses the knowledge (weights) from Model 1 and trains with 800 images from Cityscapes as source-approximate data, 1700 sunny street scenes from BDD100k as target data, and 1700 foggy street scenes generated from BDD100k images as target fake data. This allows the model to learn vehicle features under different weather conditions in BDD100k street scenes.

During the testing phase (highlighted in green), validation set samples from the target domain (Domain B), including street scenes under various weather conditions, are used to evaluate the Target Model. This step checks the model’s ability to detect traffic objects in Domain B. As shown in Figure 2, the overall training paradigm consists of three stages: building the source model, adapting it to the target domain using mixed-domain samples, and evaluating the final detection performance on the target domain.

4 Experiment Result

To validate the proposed training method in this paper, we employed Foggy Cityscapes as the source domain scene and BDD100k as the target domain scene. The experiment compared the detection capabilities of three methods in cross-domain street vehicle detection tasks: (1) SSDA-YOLO model, (2) base object detection model and (3) our training paradigm. The primary differences among these models lie in how training samples are introduced and what type of samples are included (Figure 3). Each model was trained for 300 epochs, and the BDD100k validation set was used for testing. The actual training process is outlined below:

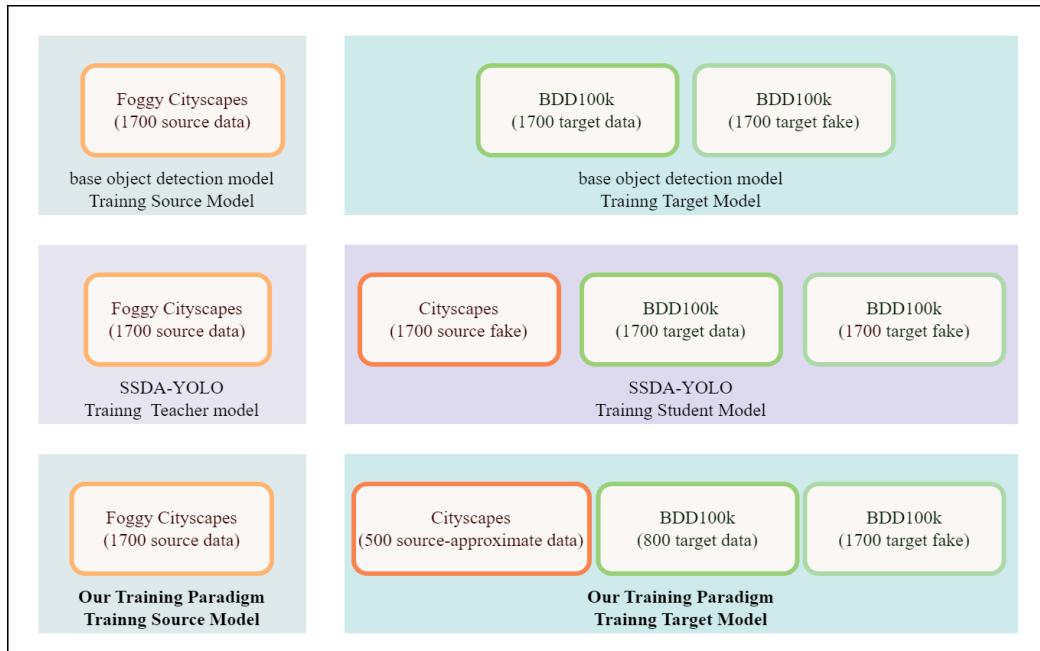


Figure 3. SSDA-YOLO with our training paradigm using the provided data

1. The SSDA-YOLO model consists of a teacher model and a student model, utilizing a total of 6800 street scenes for training.
2. The teacher model is trained with 1700 street scenes from Cityscapes (real sunny) and Foggy Cityscapes (generated foggy).
3. The student model is trained with 3400 street scenes from BDD100k (real sunny and generated foggy).
4. The base object detection model used is the YOLOv5 model, and it is divided into source data, target data and target fake for training model, incorporating a total of 5100 street scenes. The training configuration of YOLOv5 followed standard practices with default parameters; image resolution, batch size, and threshold values were consistent across all experiments to ensure fairness in comparison.
5. The source model is trained with 1700 street scenes from Foggy Cityscapes.
6. The target model is trained with 1700 street scenes from BDD100k, including real sunny scenes and generated foggy scenes.
7. The training paradigm we proposed utilizes the YOLOv5 model and is divided into source data, source-approximate data, target data and target fake for training model, incorporating a total of 4700 street scenes.
8. The source domain model is trained with 1700 source data (foggy street scenes from Foggy Cityscapes).
9. The target domain model is trained with 500 source-approximate data (sunny street scenes from Cityscapes), 800 target data (real sunny street scenes from BDD100k), and 1700 target fake (generated foggy street scenes from BDD100k).

Table 1 and Table 2 present the testing results of different models in the target domain. This study investigates whether a model trained on a single domain (Foggy Cityscapes) can adapt to another (BDD100k) by incorporating visually similar samples during fine-tuning.

Table 1. Compare the precision of different architectures for street detection

(Bold indicates the highest value; underline indicates the second highest value.)

Source domain: Foggy Cityscape -> Target domain: BDD100k						
Model	SSDA-YOLO		Base object detection model *		The proposed training paradigm	
category	mAP@.5	mAP@.5-.95	mAP@.5	mAP@.5-.95	mAP@.5	mAP@.5-.95
car	0.476	0.249	<u>0.650</u>	<u>0.370</u>	0.670	0.392
truck	<u>0.129</u>	0.063	0.294	<u>0.177</u>	0.318	0.207
person	<u>0.381</u>	<u>0.166</u>	0.333	0.124	0.419	0.178
bycycle	<u>0.209</u>	<u>0.093</u>	0.108	0.376	0.223	0.098
rider	0.228	0.106	0.132	0.048	0.228	<u>0.100</u>
motorcycle	<u>0.029</u>	0.117	0.022	0.007	0.110	<u>0.046</u>
bus	0.100	0.058	<u>0.260</u>	<u>0.167</u>	0.292	0.204
all	0.222	0.107	<u>0.257</u>	0.133	0.323	0.175

*After training the source model using Foggy Cityscapes, the model is directly transferred to the target domain for training without incorporating Cityscapes (source-approximate).

The aim is to enhance cross-domain detection performance without requiring large-scale retraining.

As shown in Table 1, when the model is directly applied to the target domain (BDD100k dataset) without any additional training, its detection performance in most categories is generally inferior to that of the model trained using the SSDA-YOLO teacher-student architecture and the proposed training paradigm. The base object detection model, which relies on original single-domain knowledge, exhibits relatively poor performance in most categories, showing unstable detection capabilities and often resulting in the lowest values.

Table 2 compares the detection results of SSDA-YOLO and the proposed training paradigm on the BDD100k dataset validation set. It can be observed that our training paradigm provides more accurate results in various street scenes compared to SSDA-YOLO. Under conditions such as daytime (with different lighting angles) and night-time (where vehicle objects are less visible), our method is more sensitive to the positions of vehicles and pedestrians, offering more precise detection results. This demonstrates the cross-domain adaptability advantage of the proposed method. By incorporating samples with similar scene features or weather conditions (one or the other) from the source domain into the training of the target domain samples, the detection model achieves higher adaptability to the visual features of vehicle objects during the training process.

This study demonstrates that by incorporating a small number of source-similar samples into the model and combining transfer learning with target domain fine-tuning, the trained model can achieve high mean Average Precision (mAP) across multiple categories, particularly in the vehicle and pedestrian categories (see Table 1). Additionally, this detection capability is unaffected by weather conditions (sunny, night-time, rainy), maintaining good detection performance even in low-light or low-visibility environments. The detection results validate the practicality of the proposed method for cross-domain applications.

Table 2. The detection results of SSDA-YOLO and our training paradigm in the BDD100k validation set

Image	SSDA-YOLO	The proposed training paradigm
1-1	1-2	1-3
2-1	2-2	2-3
3-1	3-2	3-3
4-1	4-2	4-3
5-1	5-2	5-3
6-1	6-2	6-3

5 Conclusion

This study uses three publicly available datasets—Foggy Cityscapes, Cityscapes, and BDD100k—to simulate cross-domain vehicle detection scenarios. The goal is to verify whether adding source-approximate samples during training helps the model adapt to the target domain.

Experiments evaluated three methods under various weather conditions and vehicle categories. Results show that our proposed training paradigm achieved the highest adaptability, followed by SSDA-YOLO, while the base detection model performed the worst. These findings confirm that integrating a small number of source-similar samples can enhance cross-domain detection performance without requiring complex knowledge distillation training.

6 Acknowledgements

The authors would like to thank the National Science and Technology Council, Taiwan (R. O. C.) for financially supporting this research under Contract No. NSTC 112-2221-E-240 -002 -MY3.

References

[1] C. Sakaridis, D. Dai, L. V. Gool, Semantic foggy scene understanding with synthetic data, *International Journal of Computer Vision*, Vol. 126, No. 9, pp. 973-992, September, 2018.
<https://doi.org/10.1007/s11263-018-1072-8>

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016, pp. 3213-3223.
<https://doi.org/10.1109/CVPR.2016.350>

[3] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, T. Darrell, Bdd100k: A diverse driving dataset for heterogeneous multitask learning, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, 2020, pp. 2633-2642.
<https://doi.org/10.1109/CVPR42600.2020.00271>

[4] H. Zhou, F. Jiang, H. Lu, SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection, *Computer Vision and Image Understanding*, Vol. 229, Article No. 103649, March, 2023.
<https://doi.org/10.1016/j.cviu.2023.103649>

[5] P. Viola, M. Jones, Robust real-time object detection, *Second International Workshop on Statistical and Computational Theories of Vision -- Modeling, Learning, Computing, and Sampling*, Vancouver, Canada, 2001, pp. 1-25.

[6] Z. Zheng, T. Ruan, Y. Wei, Y. Yang, VehicleNet: Learning Robust Feature Representation for Vehicle Re-identification, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Long Beach, CA, USA, 2019, pp. 1-4.

[7] X. Liu, W. Liu, H. Ma, H. Fu, Large-scale vehicle re-identification in urban surveillance videos, *2016 IEEE international conference on multimedia and expo (ICME)*, Seattle, WA, USA, 2016, pp. 1-6.
<https://doi.org/10.1109/ICME.2016.7553002>

[8] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, 2014, pp. 580-587.
<https://doi.org/10.1109/CVPR.2014.81>

[9] R. Girshick, Fast r-cnn, *Proceedings of the IEEE international conference on computer vision*, Santiago, Chile, 2015, pp. 1440-1448.
<https://doi.org/10.1109/ICCV.2015.169>

[10] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems*, Montreal, Canada, 2015, pp. 91-99.

[11] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, *Proceedings of the IEEE international conference on computer vision*, Venice, Italy, 2017, pp. 2961-2969.
<https://doi.org/10.1109/ICCV.2017.322>

[12] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016, pp. 779-788.
<https://doi.org/10.1109/CVPR.2016.91>

[13] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA, 2017, pp. 6517-6525.
<https://doi.org/10.1109/CVPR.2017.690>

[14] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, *arXiv preprint*, arXiv:1804.02767, April, 2018.
<https://arxiv.org/abs/1804.02767>

[15] A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint*, arXiv:2004.10934, April, 2020.
<https://arxiv.org/abs/2004.10934>

[16] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, I. H. Yeh, CSPNet: A new backbone that can enhance learning capability of CNN, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, WA, USA, 2020, pp. 1571-1580.
<https://doi.org/10.1109/CVPRW50498.2020.00203>

[17] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, *arXiv preprint*, arXiv:1412.3474, December, 2014.
<https://arxiv.org/abs/1412.3474>

[18] Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, *International conference on machine learning*, Lille, France, 2015, pp. 1180-1189.

[19] J. C. Jiang, B. Kantarci, S. Oktug, T. Soyata, Federated learning in smart city sensing: Challenges and opportunities, *Sensors*, Vol. 20, No. 21, Article No. 6230, November, 2020.
<https://doi.org/10.3390/s20216230>

[20] Y. Yao, G. Doretto, Boosting for transfer learning with multiple sources, *2010 IEEE computer society conference on computer vision and pattern recognition*, San Francisco, CA, USA, 2010, pp. 1855-1862.
<https://doi.org/10.1109/CVPR.2010.5539857>

[21] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, *Proceedings of Machine Learning Research, PMLR*, Vol. 37, pp. 97-105, 2015.

[22] M. A. Mohammed, A. Lakhani, K. H. Abdulkareem, D. A. Zebari, J. Nedoma, R. Martinek, S. Kadry, B. Garcia-Zapirain, Homomorphic federated learning schemes enabled pedestrian and vehicle detection system, *Internet of Things*, Vol. 23, Article No. 100903, October, 2023.
<https://doi.org/10.1016/j.iot.2023.100903>

[23] S. Lee, S. Cho, S. Im, Dranet: Disentangling representation and adaptation networks for unsupervised cross-domain adaptation, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Nashville, TN, USA, 2021, pp. 15247-15256.
<https://doi.org/10.1109/CVPR46437.2021.01500>

[24] X. Ma, T. Zhang, C. Xu, Gcan: Graph convolutional adversarial network for unsupervised domain adaptation, *Proceedings of the IEEE/CVF Conference on Computer*

Vision and Pattern Recognition, Long Beach, CA, USA, 2019, pp. 8258-8268.

<https://doi.org/10.1109/CVPR.2019.00846>

- [25] W. Tranheden, V. Olsson, J. Pinto, L. Svensson, Dacs: Domain adaptation via cross-domain mixed sampling, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 2021, pp. 1378-1388.
<https://doi.org/10.1109/WACV48630.2021.00142>
- [26] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, D. Erhan, Domain separation networks, *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, Spain, 2016, pp. 343-351.
- [27] K. Saito, D. Kim, S. Sclaroff, T. Darrell, K. Saenko, Semi-supervised domain adaptation via minimax entropy, *Proceedings of the IEEE/CVF international conference on computer vision*, Seoul, Korea (South), 2019, pp. 8049-8057.
<https://doi.org/10.1109/ICCV.2019.00814>
- [28] C. Buciluă, R. Caruana, A. Niculescu-Mizil, Model compression, *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Philadelphia, PA, USA, 2006, pp. 535-541.
<https://doi.org/10.1145/1150402.1150464>
- [29] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, *arXiv preprint*, arXiv:1503.02531, March, 2015.
<https://arxiv.org/abs/1503.02531>
- [30] G. Csurka, Domain adaptation for visual applications: A comprehensive survey, *arXiv preprint*, arXiv:1702.05374, March, 2017.
<https://arxiv.org/abs/1702.05374>
- [31] S. Razakarivony, F. Jurie, Vehicle detection in aerial imagery: A small target detection benchmark, *Journal of Visual Communication and Image Representation*, Vol. 34, pp. 187-203, January, 2016.
<https://doi.org/10.1016/j.jvcir.2015.11.002>

Biographies



Chi-Han Chen is a Ph.D. candidate in Computer Science and Engineering at National Yang Ming Chiao Tung University (NYCU), Hsinchu, Taiwan. He received the B.S. degree in Communications Engineering and the M.S. degree in Computer Science and Engineering from National Chiao Tung

University (NCTU), Hsinchu, Taiwan. He currently serves as an Intel Innovator. His research interests include deep learning with applications to computer vision, natural language processing, and embedded systems. He has over ten years of software development experience, from IoT to AIoT applications, and he founded an AI team that won the AIGO First Prize from the Industrial Development Bureau, Ministry of Economic Affairs. His current work focuses on intelligent computer vision systems.



Shu-Fang Zhang received the B.S. and M.S. degrees from Overseas Chinese University (OCU), Taichung, Taiwan. Her research interests include computer vision and industrial applications. In this study, she was responsible for data collection, the experimental design and implementation of vehicle detection, performance evaluation, and dataset construction/quality control, supporting ablation and error analyses to enhance reproducibility and validity.



Hsin-Te Wu is an Associate Professor of Department of Computer Science and Information Engineering from National Taitung University, Taiwan. He has served as Associate Editor for International Journal of Information Technologies and Systems Approach in Healthcare. He has served as Special Issue Guest Editor of Journal of Supercomputing and IET Networks. His research interests include computer networks, wireless network, Geospatial Data Analysis, Automatic Identification System, blockchain and Internet of things.



Rung-Shiang Cheng received his Ph.D. degree in Electrical Engineering from National Cheng Kung University, Taiwan, and is currently a Professor in the Department of Artificial Intelligence and Computer Engineering at National Chin-Yi University of Technology. His research interests include computer network simulation, wireless communications, AIoT applications, and edge intelligence. He has published more than thirty SCI-indexed journal papers and has received multiple Best Paper Awards at international conferences. He has led over forty government-funded research projects. His earlier work focused on network performance evaluation, communication protocol design, and wireless and mobile communications, while his recent research has shifted toward the integration of artificial intelligence, IoT applications, and intelligent systems in both academic and industrial contexts.