Domänenadaptation für feingranulare Fahrzeugklassifikation mittels Domain-Adversarial-Learning

Stefan Wolf [1,2]

[1] Lehrstuhl für Interaktive Echtzeitsysteme (IES)
Karlsruher Institut für Technologie KIT
stefan.wolf@kit.edu

Jürgen Beyerer [2,1]

Fraunhofer Institut für Optronik,
Systemtechnik und Bildauswertung IOSB
juergen.beyerer@iosb.fraunhofer.de

Schlüsselwörter: Feingranulare Klassifikation, Domänenadaption, Adversarial Learning

Im Sicherheitskontext ist das Auffinden von Tatfahrzeugen in Überwachungskameras ein wichtiger Baustein bei der Fahndung von Verdächtigen. Da Kennzeichen bei schweren Kriminaldelikten üblicherweise getauscht werden und Zeugenaussagen meist keine Kennzeichen enthalten, ist eine Suche nach dem Fahrzeugmodell eine Lösung, um den Suchraum an möglichen Fahrzeugen einzuschränken. Für eine automatische Suche, wird dabei ein Verfahren zur feingranularen Klassifikation von Fahrzeugmodellen benötigt. Mit einer der größten technischen Herausforderungen dafür ist die Unterstützung einer großen Breite an Fahrzeugmodellen, da dies einen fortlaufenden Datensammlungsprozess von Fahrzeugbildern in Überwachungskameras erfordern würde, der arbeitsintensiv und datenschutzkritisch wäre.

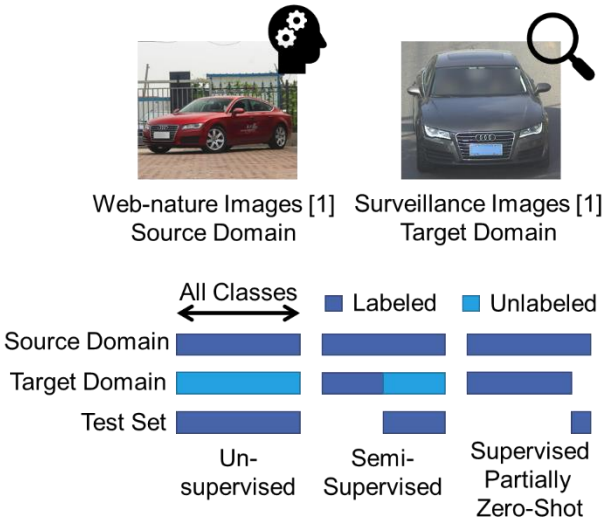

Abbildung 1: Für eine feingranulare Fahrzeugklassifikation sind Fahrzeugbilder aus dem Internet aufgrund der hohen Verfügbarkeit eine interessante Quelle für Lerndaten. Eine Herausforderung besteht darin, dass für den praktischen Einsatz im Sicherheitskontext andere Perspektiven relevant sind. Solche Domänenadaptionsszenarien werden meistens in einem unüberwachten oder semi-überwachten Szenario untersucht, das von einer breiten Verfügbarkeit an unannotierten Daten für die Zieldomäne ausgeht. Im Kontext dieses Vortrags wird der Fokus auf ein Szenario gelegt, in dem die Verfügbarkeit von Bildern in der Zieldomäne auf einen Teil der Klassen begrenzt ist.

Im Rahmen dieses Vortrags wird daher die Verwendung von im großen Umfang verfügbaren Web-Bildern als Lerndaten für diesen Einsatzzweck evaluiert. Die große Herausforderung dabei ist die Domänenlücke zwischen Web-Bildern und praxisrelevanten Szenarien wie bspw. Überwachungsbildern. Insbesondere moderne Deep-Learning-Ansätze haben bei Bildklassifikation sehr gute Ergebnisse gezeigt, neigen aber aufgrund ihrer starken Abhängigkeit von Lerndaten und der hohen Anzahl an Parametern auch dazu, domänenspezifische Merkmale zu extrahieren, welche für eine Inferenz in einer anderen Domäne kaum geeignet sind.

Daher wird der Einsatz von Techniken zur Domänenadaption untersucht, die die Genauigkeit bei einer Domänenlücke zwischen Quell- und Zieldomäne (bspw. Web- und Überwachungs-

bildern) deutlich steigern können. Hierbei hat sich insbesondere der Einsatz von Domain-Adversarial-Training [2] als vielversprechend herausgestellt. Eine Herausforderung dabei ist die Instabilität von Adversarial-Training, welche durch den Einsatz von Adversrial-Loss-Coupling [3] deutlich verbessert werden kann, wobei die Genauigkeit bei der Inferenz auf Überwachungsbildern durch einen stabileren Lernprozess noch weiter steigt [4]. Beim Loss-Coupling wird der Einfluss des Adversarial-Trainings auf die Gewichtanpassung dynamisch während des Trainings in Abhängigkeit von der Domäneninvarianz des Netzes angepasst.

Literatur

[1] Yang, L., Luo, P., Change Loy, C., & Tang, X. (2015). A large-scale car dataset for fine-grained categorization and verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3973-3981).

[2] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M. & Lempitsky, V. (2016). Domain-adversarial training of neural networks. Journal of machine learning research, 17(59), 1-35.

[3] Wiedemer, T., Wolf, S., Schumann, A., Ma, K., & Beyerer, J. (2022). Few-shot supervised prototype alignment for pedestrian detection on fisheye images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4142-4153).

[4] Wolf, S. (2023, July). A Baseline for Cross-Domain Fine-Grained Vehicle Classification in a Supervised Partially Zero-Shot Setting. In Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory (p. 87).