Teacher-Student Models for AI Vision at the Edge:
A Car Parking Case Study

Mbasa Joaquim Molo[1,2][a], Emanuele Carlini[2][b] Luca Ciampi[2][c],
Claudio Gennaro[2][d], and Lucia Vadicamo[2][e]

[1] *Department of Computer Science, University of Pisa, Pisa, Italy*

[2]*Institute of Information Science and Technologies (ISTI), Italian National Research Council (CNR), Pisa, Italy*
joaquim.molo@phd.unipi.it,{luca.ciampi, claudio.gennaro,emanuele.carlini,lucia.vadicamo}@isti.cnr.it

Abstract: The surge of the Internet of Things has sparked a multitude of deep learning-based computer vision applications that extract relevant information from the deluge of data coming from Edge devices, such as smart cameras. Nevertheless, this promising approach introduces new obstacles, including the constraints posed by the limited computational resources on these devices and the challenges associated with the generalization capabilities of the AI-based models against novel scenarios never seen during the supervised training, a situation frequently encountered in this context. This work proposes an efficient approach for detecting vehicles in parking lot scenarios monitored by multiple smart cameras that train their underlying AI-based models by exploiting knowledge distillation. Specifically, we consider an architectural scheme comprising a powerful and large detector used as a teacher and several shallow models acting as students, more appropriate for computational-bounded devices and designed to run onboard the smart cameras. The teacher is pre-trained over general-context data and behaves like an oracle, transferring its knowledge to the smaller nodes; on the other hand, the students learn to localize cars in new specific scenarios without using further labeled data, relying solely on the distilled loss coming from the oracle. Preliminary results show that student models trained only with distillation loss increase their performances, sometimes even outperforming the results achieved by the same models supervised with the ground truth.

1 Introduction

The emergence of AI-driven computer vision algorithms provides the opportunity to employ low-cost video cameras for visual sensing in Internet of Things (IoT) applications across various domains, ranging from face recognition (George et al., 2023) and crowd counting (Di Benedetto et al., 2022) to pedestrian detection (Cafarelli et al., 2022) and people/vehicle tracking (Foszner et al., 2023). Unlike cloud computing, which boasts nearly unlimited resources, edge computing in conjunction with IoT devices is characterized by the existence of compute nodes with limited computational capabilities and power allocation (Heckmann and Ravindran, 2023), promoting the

decentralization of data processing to the edge, where the data itself originates. However, despite offering advantages, such as latency reduction, lower costs, and reduced data traffic, this paradigm brings new challenges.

Specifically, AI vision models are mainly based on Deep Learning (DL) algorithms, sometimes requiring significant computational resources, especially for running in real-time requirements. Furthermore, state-of-the-art DL techniques rely on supervised learning, and they struggle when employed in new scenarios never seen during the training phase, a situation frequently encountered in the context of Edge AI. Naive solutions based on collecting new data are not only costly but sometimes even unfeasible. Data collection and curation necessitates manual labeling, often performed by human annotators with extensive domain expertise, exploiting time-consuming, expensive, and error-prone procedures. As a result, models are often trained by leveraging already existing big collections of labeled data and con-

---

[a] https://orcid.org/0000-0002-2096-1701
[b] https://orcid.org/0000-0003-3643-5404
[c] https://orcid.org/0000-0002-6985-0439
[d] https://orcid.org/0000-0002-3715-149X
[e] https://orcid.org/0000-0001-7182-7038

sequently suffer from shifts in data distributions when applied in new application scenarios. These emerging challenges create space for alternative solutions to efficiently train deep neural models for specific tasks when human-annotated data is limited (Ciampi et al., 2023) and with limited computational resources available (Ciampi et al., 2022).

In this work, we propose an efficient DL approach based on Knowledge Distillation (KD) (Hinton et al., 2015) for localizing vehicles in parking areas monitored by multiple smart cameras. KD has been introduced to obtain compressed models suitable for small devices, utilizing knowledge garnered from complex and large models. Specifically, in our scenario, we present a scheme comprising a powerful detector used as a teacher and several shallow models acting as students, tailored explicitly for devices with limited computational resources and intended to operate directly on smart cameras. The teacher is pre-trained on diverse generic datasets and behaves like an oracle, transferring its knowledge to the smaller nodes; on the other hand, the students rely only on the distilled loss coming from the oracle, and they learn to detect vehicles in new scenarios they are monitoring, without the need of additional labeled data. The preliminary results obtained in an experimental evaluation under different settings show that the student models trained only with the distillation loss coming from the teacher increase their performances, sometimes even outperforming the outcomes achieved by the same models supervised with the ground truth.

To summarize, the core contributions of this work are the following:

- We propose an approach based on knowledge distillation for detecting vehicles from smart cameras monitoring parking lots; our scheme includes a large detector acting as a teacher/oracle and several students, i.e., smaller models running on the edge devices. The latter learn to localize vehicles in their monitored areas by exploiting the distillation loss coming from the teacher without requiring additional labeled data.

- We perform an experimental evaluation considering, as the teacher, a large object detector pre-trained with data containing vehicles in general contexts and, as the students, a smaller version of the teacher. Results achieved by monitoring new scenarios demonstrate that the students increase their performances using the knowledge from the oracle, sometimes even outperforming the results obtained by the same models trained with the annotations.

The rest of the paper is organized as follows. Section 2 presents the related works. Section 3 illustrates the proposed approach. Section 4 discusses the setup of the experiments, including the models and datasets used, and discusses the results obtained. Section 5 draws the conclusions of this work.

2 Related works

Knowledge Distillation-based Computer Vision Applications. Several computer vision applications have recently exploited the Knowledge Distillation (KD) paradigm. For instance, in (Wang et al., 2019; Chen et al., 2021), knowledge distillation loss is described as a feature matching distance, necessitating both the teacher and the student models to share identical architectures. On the other hand, (Banitalebi-Dehkordi, 2021) proposed a technique to improve distillation effectiveness by utilizing unlabeled data, thus diminishing the need for labeled data. (Gan et al., 2019) presented a cross-modal auditory localization approach, leveraging a student-teacher training method to enable the transfer of object detection knowledge between vision and sound modalities. In the same angle, (Liu et al., 2021) introduced a framework that simultaneously utilizes domain adaptation and knowledge distillation to enhance efficiency in object detection, introducing a focal multi-domain discriminator to improve the performance of both teacher and student networks.

Finally, (Goh et al., 2023) investigated a self-supervised distillation framework to train efficient computer vision models focusing on self-supervised pre-training of teachers using a substantial dataset of unlabeled images.

Edge AI for Distributed Computer Vision Applications. In the last years, numerous works have been conducted in computer vision focusing on applications for distributed edge devices. For instance, a notable work is (Kang et al., 2020), which proposed a framework incorporating neural architecture search to address teacher-student capacity issues, optimizing structures for varying model sizes. The authors also introduced an oracle knowledge distillation loss, enabling the student to achieve high accuracy using oracle predictions. Another interesting study is (Bharadhwaj et al., 2022), which introduced a framework named Detect-Track-Count (DTC) designed to efficiently count vehicles on edge devices. The primary objective of this approach was to enhance the efficacy of compact vehicle detection models through the application of the ensemble knowledge distillation technique. Moreover, (Kruthiventi et al., 2017) employed knowledge distillation from a multi-modal pedestrian

detector. The network was designed to extract RGB and thermal-like features from RGB images, mitigating the necessity for expensive automotive-grade thermal cameras. On the other hand, (Alqaisi et al., 2023) focused on leveraging Docker containers to facilitate the deployment and management of computer vision applications on edge devices with limited resources, aiming at enabling the secure execution of multiple applications concurrently while ensuring flexibility and efficiency in deployment. The authors in (Seitbattalov et al., 2022) introduced a novel edge computing application utilizing Raspberry Pi and an OmniVision OV 5647 camera for efficient data pre-processing. The approach significantly reduces network traffic and computational burden on centralized servers, offering a cost-effective alternative to traditional solutions like FPGAs and personal computers.

In all of these works, the authors assumed that computer vision tasks were conducted using labeled data, and the models were trained after collecting data from various nodes, consolidating them at a central node to create a comprehensive representation of the entire data distribution. In contrast to previous works, our approach eliminates the need for extensive image annotation for edge nodes and their centralization, reducing the time and resources required for dataset preparation.

3 KD-driven Vehicle Detection on Edge Devices

This section describes our proposed KD-based approach for vehicle detection on edge devices monitoring parking lots. Usually, in a distributed camera network scenario, each camera collects data independently, resulting in diverse data distributions due to varying perspectives, illuminations, angles of view, backgrounds, and, in general, different contexts. This diversity poses a challenge or the deployment of pre-trained deep learning models on edge devices, leading to a significant drop in performance in scenarios not encountered during the learning phase. Directly fine-tuning these models on such varied data may be unfeasible as it requires manual image annotations, which is time-consuming, costly, and error-prone.

To mitigate these challenges, in this work, we propose a scheme that uses a centralized node serving as a teacher model (oracle) that supervises the training of the edge device models. The teacher is a powerful detector pre-trained over several general datasets for vehicle detection that may operate on a robust computing infrastructure, such as a central server or cloud computing platform; on the other hand, the stu-
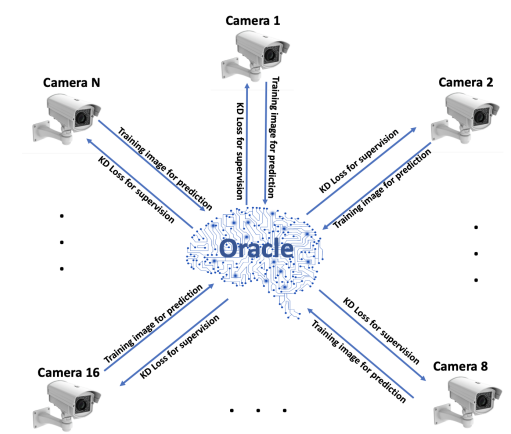

Figure 1: Overview of our proposed KD-based approach for vehicle detection in parking areas through smart cameras. The oracle is a powerful object detector trained with several general datasets for vehicle detection, acting as a teacher. Edge devices, i.e., smart cameras, rely on shallower models for analyzing images from their specific monitored areas. They are supervised only by the distilled knowledge from the oracle, without requiring extra labels.

dents are shallow models, more suitable for running on edge devices, and they learn to perform their job over their monitored region of interest by exploiting the distilled knowledge coming from the oracle, without requiring additional annotated data. The proposed scheme can be exploited in applications with a variable number of devices, as illustrated in Figure 1, showcasing its adaptability and scalability to different deployment scales. In this paper, we provide a practical case study of our paradigm by employing a network of nine smart cameras placed across various parking lots to monitor and detect vehicles in parking areas.

Our training approach is based on a KD paradigm. During the training of student models, the main loss function $\mathcal{L}$ comprises two components: the supervised *student loss* ($\mathcal{L}_{stu}$) and the *distillation loss* ($\mathcal{L}_{distil}$) (Hinton et al., 2015). Each loss is weighted by a scaling factor α (Zhou and Mosadegh, 2023) such that:

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{stu}(S_0, G_0) + (1 - \alpha) \cdot \mathcal{L}_{distil}(S_0, T_0), \quad (1)$$

where S_0, T_0, and G_0 represent the predictions of the student model, the predictions of the teacher model, and the targets used as ground truth by the student model, respectively. The hyper-parameter α balances the two components of the total loss.

In this work, we leverage a powerful pre-trained vehicle detector used as an oracle to train a specialized student model for each node using only the distillation loss, i.e., we set α to 0 in Equation (1) (see also Figure 2). This means that we are not using ground-truth labels during the training of the student model,
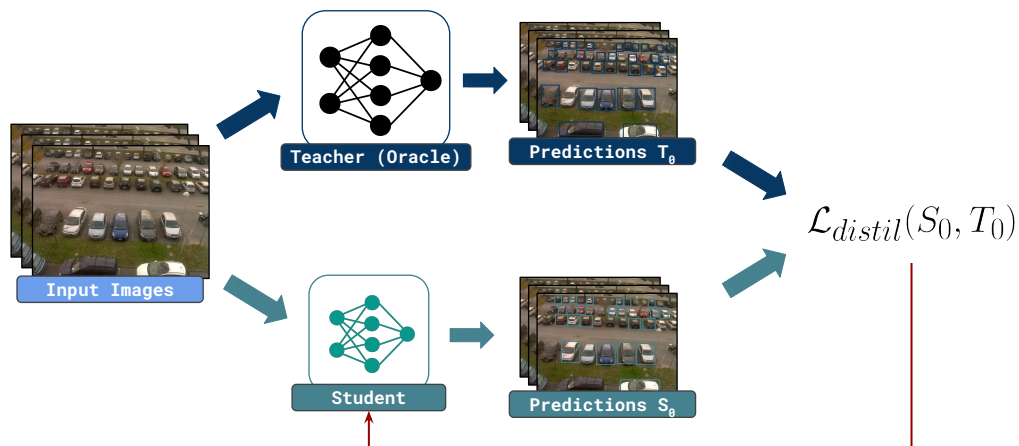
Figure 2: Knowledge distillation-based car detection approach, illustrating the training process of the student model supervised by a larger and more complex teacher model, typically known as the teacher-student paradigm. This figure visually represents how knowledge distillation transfers knowledge from the teacher to the student, resulting in a more compact yet efficient model for car detection tasks.

relying solely on the supervision coming from the distilled loss of the teacher. At inference time, the oracle is not needed anymore, and the student model operates independently.

4 Experimental Evaluation

This section details our experimental evaluation, covering the employed datasets and the experimental settings and presenting results with subsequent discussion.

4.1 The Datasets

For training and testing our proposed approach, we exploited several datasets. Standard data augmentation techniques (e.g., rotation, shear, brightness, noise, etc.) are always employed during the training.

To train the teacher model in some of our experiments, we combined images from several datasets for vehicle detection in general contexts already existing in the literature; specifically, we collected a total of 4,056 images from (i) the *Vehicles Image* dataset (Roboflow 100, 2023), (ii) the *PkLot* dataset (de Almeida et al., 2015), and (iii) the *Find a Car Park* dataset (Carr, 2019). We divided these data into training, validation, and test subsets, following common standards of 70%, 20%, and 10% of the total number of samples.

Additionally, we employed the *CNRPark-EXT* dataset (Amato et al., 2017), an extension of CNR-Park (Amato et al., 2016). This dataset serves a dual

purpose in the implementation of our approach. Initially, it is utilized for training and testing both the teacher and subsequent students, as detailed in the preliminary results in Section 4.3. Subsequently, in the second phase of our experimental setting, it is exclusively used for training and testing the students when the oracle is trained with the combined dataset described before. The CNRPark-EXT dataset encompasses parking lot images sourced from nine cameras capturing diverse weather conditions and offering various perspectives and angles of view. It is worth noting that this dataset was initially intended for parking lot classification. Therefore, the original bounding boxes do not cover the entire car but rather a portion of it since the authors wanted to prevent misclassifications caused by cars overlapping in adjacent spaces. This presented a challenge when evaluating our student models trained under the supervision of the oracle and tested on the CNRPark-EXT dataset. Indeed, given that the oracle was trained on datasets where the entire car was covered, in contrast to the CNRPark-EXT dataset, assessing the Intersection over Union (IoU) for student models on the latter led to poor performance and unsatisfactory results. To address this issue, we manually re-annotated the validation and testing sets within the CNRPark-EXT subsets[1], ensuring that the bounding boxes were adjusted to cover the entire car and, consequently, aligning the dataset with the oracle's training data configuration.

Finally, we utilized an additional dataset, named

---

[1] We make these new annotations publicly available at https://github.com/joaquimbasa/Teacher-Student-CarPark-Paper.git.

NDISPark (Ciampi et al., 2021), to evaluate our proposed approach. This database comprises 141 images collected from several distributed cameras in various parking lots, showcasing common challenges encountered in real-life scenarios. It is worth noting that we considered solely the training and validation sets, as bounding box annotations are unavailable for the test set since this dataset was initially created as a benchmark for the counting task.

4.2 Implementation details

Our implementation choice concerning the detector fell on the popular YOLOv5 architecture (Ultralytics, 2021) since it relies on the one-stage paradigm, thus ensuring a good trade-off between performance and efficiency. However, our proposed pipeline is model-agnostic, i.e., a different detector can be used without affecting the overall functioning. Specifically, we exploited the YOLOv5x object detector as the teacher model, where "x" denotes a deeper version of the standard YOLOv5 model; on the other hand, for the student models, we use a smaller and shallower version of the latter, denoted as YOLOv5n, more suitable for edge devices having limited computational requirement.

More in details, the YOLOv5x model we exploited as the teacher consists of a CNN with 322 layers, 86,173,414 parameters, and a computational load of 203.8 GFLOPs. We considered the COCO pre-trained version of this model, keeping its backbone frozen during the training stage and back-propagating the gradient only in the head layers. On the other hand, the YOLOv5n architecture that acts as the foundation for the student models is the most compact version of YOLOv5, consisting of 157 layers, 1,760,518 parameters, and 4.1 GFLOPs. Even in this case, during the training, its backbone is frozen, and only the head layers were fine-tuned by exploiting the distillation component of the KD loss.

YOLOv5 employs three distinct losses that constitute its final loss: a *Classification loss* ($\mathcal{L}_{cls}$), addressing classification errors, a *Confidence loss* ($\mathcal{L}_{obj}$), dealing with objectness errors, and a *Localization loss* ($\mathcal{L}_{bbox}$), focusing on localization errors (Terven and Cordova-Esparza, 2023). For both the classification and objectness losses, we used the Binary Cross Entropy with Logits Loss, while the Intersection Over Union loss is employed as the Localization loss.

4.3 Results and Discussion

We conducted two distinct sets of experiments. In the first set, we operated under the assumption of a
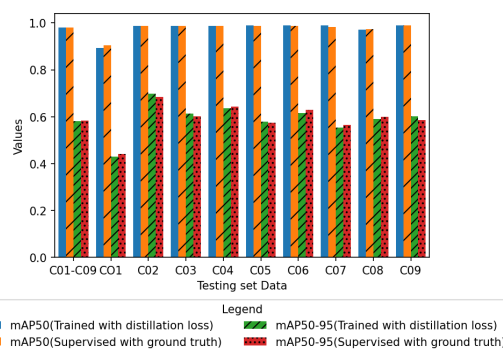

Figure 3: Comparison between the performance of YOLOv5n with knowledge distillation-based training and with supervised training on CNRPark-EXT. The student model was trained on the whole CNRPark-EXT training set and then tested on each camera.

teacher having some knowledge about the data distribution used at inference time by the student model. In other words, in this scenario, the teacher and the student were trained on the same distribution, specifically, the CNRPark-EXT dataset. In the second part of the experiment, our approach introduced the utilization of an oracle trained on a data distribution entirely different from the one used for training the students.

Preliminary results with the CNRPark-EXT dataset. In this initial set of experiments, as previously mentioned, our teacher model is assumed to be an almost perfect teacher over the considered scenario, showcasing outstanding performance with F1-Score, mAP50, and mAP50-95 scores of 0.99, 0.99, and 0.90, respectively. It is important to note that the performance evaluation in the whole of this study, including the one just given, is based on the MS COCO mean Average Precision (mAP), where the mAP50 is the Average Precision at a specific IoU threshold of 0.50, and mAP50-95 is the AP averaged over a range of IoU thresholds (0.50 to 0.95 with increments of 0.05) (Lin et al., 2014).

In the student training process, two cases are considered. On the one hand, we focused on training the student model utilizing exclusively the distillation loss ($\mathcal{L}_{distil}$), which constitutes one component of the loss function outlined in Equation (1). This aids in training the student model effectively, even when working with unlabeled data. On the other hand, we trained the student model by exploiting the supervised loss function ($\mathcal{L}_{stu}$), which makes use of the ground truth (i.e., α set to 1 in Equation (1)). This latter case can be regarded as an upper bound on the achievable performance for each dataset.

Figure 3 provides a comparative analysis of the

Table 1: Comparison between the mAP50 and mAP50-95 performance of specialized student models trained with distillation loss (Dist.) and supervised (Super.) approach on a single CNRPark-EXT camera.

Training Set	Testing Set	mAP50	mAP50-95	Mode
CNR-EXT C1	CNR-EXT C1	**0.95**	**0.50**	Super.
		0.91	0.45	Dist.
CNR-EXT C2	CNR-EXT C2	**0.99**	**0.77**	Super.
		0.99	0.73	Dist.
CNR-EXT C3	CNR-EXT C3	**0.99**	**0.66**	Super.
		0.99	0.60	Dist.
CNR-EXT C4	CNR-EXT C4	**0.99**	**0.66**	Super.
		0.99	0.61	Dist.
CNR-EXT C5	CNR-EXT C5	**0.99**	**0.60**	Super.
		0.98	0.55	Dist.
CNR-EXT C6	CNR-EXT C6	**0.99**	**0.64**	Super.
		0.99	**0.64**	Dist.
CNR-EXT C7	CNR-EXT C7	**0.99**	**0.60**	Super.
		0.98	0.53	Dist.
CNR-EXT C8	CNR-EXT C8	**0.98**	**0.63**	Super.
		0.97	0.57	Dist.
CNR-EXT C9	CNR-EXT C9	**0.99**	**0.66**	Super.
		0.99	0.60	Dist.

Table 2: Comparison of mAP50 and mAP50-95 for student models trained on a single CNRPark-EXT camera using distillation loss (Dist.) or supervised approach (Super.), then tested on the complete CNRPark-EXT test set.

Training Set	Test Set	mAP50	mAP50-95	Mode
CNR-EXT C1	CNR-EXT	0.70	0.23	Super.
		0.88	**0.37**	Dist.
CNR-EXT C2	CNR-EXT	0.60	0.20	Super.
		0.77	**0.31**	Dist.
CNR-EXT C3	CNR-EXT	0.79	0.32	Super.
		0.90	**0.43**	Dist.
CNR-EXT C4	CNR-EXT	0.78	0.35	Super.
		0.92	**0.46**	Dist.
CNR-EXT C5	CNR-EXT	0.79	0.36	Super.
		0.97	**0.47**	Dist.
CNR-EXT C6	CNR-EXT	0.78	0.36	Super.
		0.91	**0.47**	Dist.
CNR-EXT C7	CNR-EXT	0.80	0.37	Super.
		0.91	**0.46**	Dist.
CNR-EXT C8	CNR-EXT	0.83	0.40	Super.
		0.87	**0.45**	Dist.
CNR-EXT C9	CNR-EXT	0.82	0.36	Super.
		0.91	**0.45**	Dist.

results obtained against the testing sets of CNRPark-EXT in the two settings mentioned above. Specifically, we show the performance considering the whole CRNPark-EXT training set, as well as using the testing sets of individual subsets belonging to specific cameras.

The results indicate that the student model trained solely with the distillation loss exhibits a comparable level of performance to that of the same model supervised with labeled data. Moreover, it is worth noting that results for Camera 1 are lower due to a severe change in perspective compared to the other cameras.

Furthermore, we provide an experimental evaluation where the teacher supervises the training of the student model using only a specific subset of the data associated with a specific camera. Specifically, Table 1 presents a comprehensive view of the performance in terms of mAP exhibited by the student models. These models are trained using data from each camera subset and evaluated using the respective testing set data. Furthermore, their performance is assessed on the testing set encompassing the entire CNRPark-EXT dataset as shown in Table 2.

In the first scenario, the results in Table 1 show that a specialized student model tested on its specific data exhibits a slightly lower performance when using solely the distillation loss, in comparison to supervising it with the ground truth. However, in the second scenario, as illustrated in Table 2, the results show that testing the specialized model on diverse dis-

tributions encompassing various cameras showcases improved performance when employing only the distillation loss. This insight indicates that a highly accurate and knowledgeable teacher, with extensive experience from a wide-ranging data distribution, enables specialized models to maintain accurate predictions even when the configuration of the distribution of data changes. Thus, the teacher's comprehensive understanding of the data and ability to distill the knowledge to the specialized models enable effective training with unlabeled data, resulting in enhanced predictive capabilities.

Results using the oracle-teacher trained on a dataset agnostic to the test dataset. Further experiments were conducted using the combined dataset described in subsection 4.1 to pre-train the oracle and test the students against the NDISPark and the CNRPark-EXT datasets; thus, we assume to have a teacher agnostic to the data distribution used for the tests. We report the results in Table 3, considering only the training of the students with the distillation loss and comparing the performance against the ones obtained with the teacher. Concerning the NDISPark dataset, we combined all subsets due to its small size; on the other hand, for CNRParkEXT, as in the previous set of experiments, we considered the different subsets associated with specific cameras. As can be seen, we obtained moderate performance on the

Table 3: Results obtained through teacher-student distillation using the combined dataset and specific camera datasets, respectively.

Role	Test Dataset	Precision	Recall	mAP50	mAP50-95
Teach.	Combined Dataset	0.92	0.90	0.95	0.73
Teach.	NDISPark	**0.95**	**0.86**	**0.94**	**0.63**
Stud.		0.84	0.77	0.84	0.46
Teach.	CNR-EXT (all cameras)	0.82	**0.79**	0.76	0.23
Stud.		**0.86**	0.76	**0.83**	**0.26**
Teach.	CNR-EXT C1	0.75	0.61	0.60	0.31
Stud.		**0.76**	**0.75**	**0.88**	**0.37**
Teach.	CNR-EXT C2	0.78	0.64	0.71	0.42
Stud.		**0.84**	**0.79**	**0.88**	**0.59**
Teach.	CNR-EXT C3	0.76	0.76	0.80	0.25
Stud.		**0.87**	**0.85**	**0.91**	**0.39**
Teach.	CNR-EXT C4	0.82	0.76	0.81	0.24
Stud.		**0.87**	**0.79**	**0.86**	**0.36**
Teach.	CNR-EXT C5	0.83	0.72	0.77	0.20
Stud.		**0.89**	**0.82**	**0.84**	**0.29**
Teach.	CNR-EXT C6	0.82	0.77	0.79	0.23
Stud.		**0.85**	**0.82**	**0.86**	**0.31**
Teach.	CNR-EXT C7	0.79	0.70	0.73	0.21
Stud.		**0.86**	**0.76**	**0.82**	**0.27**
Teach.	CNR-EXT C8	0.84	0.72	0.75	0.22
Stud.		**0.92**	**0.82**	**0.87**	**0.28**
Teach.	CNR-EXT C9	0.82	0.74	**0.82**	0.28
Stud.		**0.85**	**0.84**	0.74	**0.42**

NDISPark dataset, but, on the other hand, concerning the CNRPark-EXT dataset, the student model sometimes even outperformed the oracle, demonstrating that the student can achieve reliable performance just exploiting the distillation loss from an oracle agnostic to the inference data distribution, without requiring any further annotations specific for the monitored scenario.

5 Conclusion

This work introduced a Knowledge Distillation-based approach tailored for computer vision applications at the edge. We focus on a scenario where several smart cameras, with limited computational capabilities, monitor parking areas by detecting the vehicles in their field of view. We proposed a teacher-student scheme where the teacher is a powerful and large detector acting as an oracle for the students, which in turn are shallow models more appropriate for computational-bounded devices. The teacher has extensive knowledge that transfers to the smaller nodes, which, on the other hand, learn to localize cars in new specific scenarios without using further labeled data, relying solely on the distilled loss from the oracle. This addresses challenges in Edge AI, where models on edge devices need to generalize knowledge to new scenarios and meet hardware constraints. We performed an experimental evaluation under different settings, considering a teacher pre-trained over different general-context datasets suitable for vehicle detection and, as the students, a smaller version of the teacher. The results demonstrate that students increase their performance only with knowledge from the oracle, sometimes even surpassing the results obtained by models trained with annotations.

In future work, we aim to extend our approach to various application scenarios and also explore comparisons or integrations with other distributed learning techniques, such as federated learning.

ACKNOWLEDGEMENTS

REFERENCES

Alqaisi, O. I., Şaman Tosun, A., and Korkmaz, T. (2023). Containerized computer vision applications on edge devices. In *2023 IEEE International Conference on Edge Computing and Communications (EDGE)*. IEEE.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., and Vairo, C. (2017). Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72:327–334.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., and Vairo, C. (2016). Car parking occupancy detection using smart camera networks and deep learning. In *2016 IEEE Symposium on Computers and Communication (ISCC)*. IEEE.

Banitalebi-Dehkordi, A. (2021). Knowledge distillation for low-power object detection: A simple technique and its extensions for training compact models using unlabeled data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE.

Bharadhwaj, M., Ramadurai, G., and Ravindran, B. (2022). Detecting vehicles on the edge: Knowledge distillation to improve performance in heterogeneous road traffic. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE.

Cafarelli, D., Ciampi, L., Vadicamo, L., Gennaro, C., Berton, A., Paterni, M., Benvenuti, C., Passera, M., and Falchi, F. (2022). *MOBDrone: A Drone Video Dataset for Man OverBoard Rescue*, page 633–644. Springer International Publishing.

Carr, D. (2019). Find a car park. https://www.kaggle.com/datasets/daggysheep/find-a-car-park. visited on 2023-10-17.

Chen, B.-C., Wu, Z., Davis, L. S., and Lim, S.-N. (2021). Efficient object embedding for spliced image retrieval. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Ciampi, L., Gennaro, C., Carrara, F., Falchi, F., Vairo, C., and Amato, G. (2022). Multi-camera vehicle counting using edge-ai. *Expert Systems with Applications*, 207:117929.

Ciampi, L., Santiago, C., Costeira, J., Falchi, F., Gennaro, C., and Amato, G. (2023). Unsupervised domain adaptation for video violence detection in the wild. In *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering*. SCITEPRESS - Science and Technology Publications.

Ciampi, L., Santiago, C., Costeira, J., Gennaro, C., and Amato, G. (2021). Domain adaptation for traffic density estimation. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications.

de Almeida, P. R., Oliveira, L. S., Britto, A. S., Silva, E. J., and Koerich, A. L. (2015). Pklot – a robust dataset for parking lot classification. *Expert Systems with Applications*, 42(11):4937–4949.

Di Benedetto, M., Carrara, F., Ciampi, L., Falchi, F., Gennaro, C., and Amato, G. (2022). An embedded toolset for human activity monitoring in critical environments. *Expert Systems with Applications*, 199:117125.

Foszner, P., Szczęsna, A., Ciampi, L., Messina, N., Cygan, A., Bizoń, B., Cogiel, M., Golba, D., Macioszek, E., and Staniszewski, M. (2023). Development of a realistic crowd simulation environment for fine-grained validation of people tracking methods. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications.

Gan, C., Zhao, H., Chen, P., Cox, D., and Torralba, A. (2019). Self-supervised moving vehicle tracking with stereo sound. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.

George, A., Ecabert, C., Shahreza, H. O., Kotwal, K., and Marcel, S. (2023). Edgeface: Efficient face recognition model for edge devices.

Goh, E., Ward, I. R., Vincent, G., Pak, K., Chen, J., and Wilson, B. (2023). Self-supervised distillation for computer vision onboard planetary robots. In *2023 IEEE Aerospace Conference*. IEEE.

Heckmann, O. and Ravindran, A. (2023). Evaluating kubernetes at the edge for fault tolerant multi-camera computer vision applications. In *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*. IEEE.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network.

Kang, M., Mun, J., and Han, B. (2020). Towards oracle knowledge distillation with neural architecture search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):4404–4411.

Kruthiventi, S. S. S., Sahay, P., and Biswal, R. (2017). Low-light pedestrian detection from rgb images using multi-modal knowledge distillation. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). Microsoft coco: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8693 LNCS:740–755.

Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., and Tang, J. (2021). Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, page 1–1.

Roboflow 100 (2023). Vehicles Dataset. https://universe.roboflow.com/roboflow-100/vehicles-q0x2v. visited on 2023-10-01.

Seitbattalov, Z. Y., Canbolat, H., Moldabayeva, Z. S., and Kyzyrkanov, A. E. (2022). An intelligent automatic number plate recognition system based on computer vision and edge computing. In *2022 International Conference on Smart Information Systems and Technologies (SIST)*. IEEE.

Terven, J. and Cordova-Esparza, D. (2023). A comprehensive review of yolo: From yolov1 and beyond.

Ultralytics (2021). YOLOv5: A state-of-the-art real-time object detection system. https://docs.ultralytics.com. Accessed: insert date here.

Wang, T., Yuan, L., Zhang, X., and Feng, J. (2019). Distilling object detectors with fine-grained feature imitation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Zhou, G. and Mosadegh, B. (2023). Distilling knowledge from an ensemble of vision transformers for improved classification of breast ultrasound. *Academic Radiology*.