Coaching a Teachable Student

Jimuyang Zhang   Zanming Huang   Eshed Ohn-Bar
Boston University
{zhangjim, huangtom, eohnbar}@bu.edu

Abstract

We propose a novel knowledge distillation framework for effectively teaching a sensorimotor student agent to drive from the supervision of a privileged teacher agent. Current distillation for sensorimotor agents methods tend to result in suboptimal learned driving behavior by the student, which we hypothesize is due to inherent differences between the input, modeling capacity, and optimization processes of the two agents. We develop a novel distillation scheme that can address these limitations and close the gap between the sensorimotor agent and its privileged teacher. Our key insight is to design a student which learns to align their input features with the teacher's privileged Bird's Eye View (BEV) space. The student then can benefit from direct supervision by the teacher over the internal representation learning. To scaffold the difficult sensorimotor learning task, the student model is optimized via a student-paced coaching mechanism with various auxiliary supervision. We further propose a high-capacity imitation learned privileged agent that surpasses prior privileged agents in CARLA and ensures the student learns safe driving behavior. Our proposed sensorimotor agent results in a robust image-based behavior cloning agent in CARLA, improving over current models by over 20.6% in driving score without requiring LiDAR, historical observations, ensemble of models, on-policy data aggregation or reinforcement learning.

1. Introduction

Learning internal representations for making intricate driving decisions from images involves a complex optimization task [4, 43, 54]. The inherent challenge for end-to-end training of driving agents lies in the immense complexity of learning to map high-dimensional visual observations into general and safe navigational decisions [13, 20, 70]. Even given millions of training examples [4], today's agents still fail to reliably learn an internal representation that can be used for robust processing of complex visual scenarios (e.g., dense urban settings with intricate layouts and dynamic obstacles) in a safe and task-driven manner [13, 20, 70].

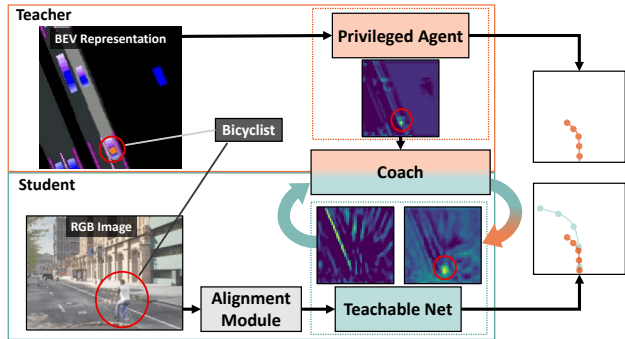To ease the challenging sensorimotor training task, re-

Figure 1. **Effective Knowledge Distillation for Sensorimotor Agents.** Our proposed CaT (Coaching a Teachable student) framework enables highly effective knowledge transfer between a privileged teacher and a sensorimotor (i.e., image-based) student. An alignment module learns to transform image-based features to the teacher's BEV feature space, where the student can then leverage extensive and direct supervision on its learned intermediate representations. The student model is optimized via a coaching mechanism with extensive auxiliary supervision in order to further scaffold the difficult sensorimotor learning task.

cent approaches decompose the task into stages, e.g., by first training a high-capacity privileged network with complete knowledge of the world and distilling its knowledge into a less capable vision-based student network [11, 13, 23, 47, 79]. However, due to the inherent differences between the inputs and architectures of the two agents, current methods rely on limited supervisory mechanisms from the teacher, i.e., exclusively through the teacher's output [11, 13] or knowledge distillation of a single final fully-connected layer [71, 79, 81]. Moreover, the privileged teacher's demonstration targets may be noisy or difficult for the student to imitate, given the limited perspective [27]. In this work, we sought to develop a more effective knowledge distillation paradigm for training a sensorimotor agent to drive. Our key insight is to enable more extensive supervision from the teacher by reducing the gap between internal modeling and learning capabilities between the two agents.

Our proposed approach for holistic knowledge distillation is informed by human instruction, which often involves structured supervision in addition to high-level demonstra-

tions, e.g., providing various hints to scaffold information in a way that the student can better understand [30]. When teaching others new and challenging skills, i.e., where a student may not be able to replicate the demonstration such as riding a bicycle or driving a vehicle, teachers may provide additional supervision regarding the underlying task structure and their own internal reasoning [64]. Analogously to our settings, the privileged teaching agent can potentially provide richer supervision when teaching a limited capacity sensorimotor student, i.e., through more careful and direct guidance of the underlying representation learning.

In our work, we introduce CaT, a novel method for teaching a sensorimotor student to drive using supervision from a privileged teacher. Our key insights are threefold: **1) Effective Teacher:** We propose to incorporate explicit safety-aware cues into the BEV space that facilitate a surprisingly effective teacher agent design. While prior privileged agents struggle to learn to drive in complex urban driving scenes, we demonstrate our learned agent to match expert-level decision-making. **2) Teachable Student via Alignment:** An IPM-based transformer alignment module can facilitate direct distillation of most of the teacher's features and better guide the student learning process. **3) Student-paced Coaching:** A coaching mechanism for managing difficult samples can scaffold knowledge and lead to improved model optimization by better considering the ability of the student. By holistically tackling the complex knowledge distillation task with extensive teacher and auxiliary supervision, we are able to train a state-of-the-art image-based agent in CARLA [21]. Through ablation studies, input design, and interpretable feature analysis, we also provide critical insights into current limitations in learning robust and generalized representations for driving agents.

2. Related Work

Knowledge Distillation: We study knowledge distillation methods originally developed for model compression and acceleration [12,17,29,31,35,36,52,55,74,75] in the context of training sensorimotor. While approaches for feature distillation have been minimally explored in this context, we are motivated by their success in other domains, including image classification [2,38,40,42,60], object detection [67], semantic segmentation [46,78], and natural language processing [3,68,73]. Yet, applying such techniques is not trivial given the drastically differing inputs between the privileged agent and student and the overwhelming sensorimotor task. Consequently, driving policy distillation methods only provide supervision either from the teacher's output [11,13] or the features of a single fully-connected layer [71,79,81], which (we hypothesize) does not provide sufficient hints to guide the student training. In this work, we provide novel mechanisms to close this gap and enable extensive supervision through deep feature distillation.

Imitation Learning to Drive: Recent approaches in imitation learning (IL) to drive can be traced to Pomerleau [54]. Recently, more elaborate IL-based approaches for driving have emerged [9,10,25,45,50,76,77,80,83]. Specifically, decomposing the imitation learning task into two stages, i.e., by first learning a privileged agent through behavior cloning [11,13] or reinforcement learning [8,62,71,79] and then training the sensorimotor agent to mimic the output of the privileged agent. Our study is motivated by such approaches, yet we explore the benefits of increased supervision by the privileged agent. Moreover, prior privileged agents produce noisy and sub-optimal demonstration supervision in complex urban scenarios (even when optimized with RL, as demonstrated by our analysis). While most current studies employ the RL-based teacher of [79], this can be problematic and inefficient in safety-critical conditions in the real-world. In our work, we introduce a novel privileged teacher which can greatly surpass prior agents while only relying on offline behavior cloning.

Intermediate Representation for Driving: Learning an effective 3D scene representation is crucial for safe autonomous driving. Researchers may obtain such representations by lifting the image to 3D using estimated depth and projecting the frustums into a BEV grid [32,53]. Alternatively, transformer-based architectures can also enable mapping a camera image to the BEV space [14,82], i.e., by attending to image-based information when populating a BEV-sampled grid. Related to our work is the study of Chen et al. [11], which learns a BEV representation from RGB image and LiDAR input. A motion planner is then used to generate future waypoints from the BEV. As this introduces significant challenges, we propose to leverage an IPM-based alignment module which can better structure to image-to-BEV projection task.

Curriculum and Self-paced Learning: Approaches for curriculum learning methods often structure learning from various difficulty samples [5,24,37,39,44,59,61,63,69,84]. To better consider the challenging sensorimotor task and limited capacity of the student, we adopt a curriculum (i.e., an imitation coach), with progressively more challenging samples. He et al. [27] also proposes training a coach across iterations of on-policy data collection [56]. In contrast, we implement and demonstrate the benefit of a coach over training iterations without a data aggregation part. This can facilitate more efficient model training. The approach is motivated by self-paced learning [41,72], where the model selects easy samples dynamically at each iteration based on a defined loss. In contrast to such studies, we only smooth the targets of the hard samples (based on the student), instead of discarding them. We empirically find this to result in improved coaching, potentially due to improved scaffolding of the difficult samples.

3. Method

Our goal is to decompose the challenging sensorimotor learning task and ease model optimization through effective supervision from a teacher agent. We first formulate the problem of privileged agent distillation in Sec. 3.1. Next, we address current limitations in privileged agent design to train a robust imitation learned teacher in Sec. 3.2. We then propose a transformer-based student architecture to resolve image-to-BEV feature alignment (Sec. 3.3). Finally, we train the sensorimotor student via extensive supervision and knowledge distillation (Sec. 3.4) via a progressive (i.e., coaching) mechanism (Sec. 3.5).

3.1. Formulation

Driving can be formulated as a sequential decision-making problem over a set of sensory observations $\mathcal{X}^s$, a set of actions (motor commands) $\mathcal{Y}$, and a policy function $f_{\boldsymbol{\theta}}^s \colon \mathcal{X}^s \to \mathcal{Y}$ for mapping observations to actions at each time step [27,50,56]. We consider the task of learning a sensorimotor student agent via knowledge distillation [13, 23] from a teacher agent $f_{\boldsymbol{\psi}}^t \colon \mathcal{X}^t \to \mathcal{Y}$ with privileged access to environmental information, e.g., ground-truth lanes, obstacles, and traffic light states. We parameterize the two agents as neural networks using weights $[\boldsymbol{\theta}, \boldsymbol{\psi}] \in \mathbb{R}^d$, and denote $\mathcal{F}^s(\cdot; \boldsymbol{\theta})$ and $\mathcal{F}^t(\cdot; \boldsymbol{\psi})$ to be the feature maps for the sensorimotor student network and teacher network, respectively. Given a dataset $\mathcal{D}$ comprising sensory and privileged observations and a loss function $\mathcal{L}$, the student can be optimized from the teacher using

$$\operatorname*{argmin}_{\boldsymbol{\theta}} \; \mathbb{E}_{(\mathbf{x}^s, \mathbf{x}^t) \sim \mathcal{D}}[\mathcal{L}(\mathcal{F}^s(\mathbf{x}^s; \boldsymbol{\theta}), \mathcal{F}^t(\mathbf{x}^t; \boldsymbol{\psi}))] \quad (1)$$

where the loss may be defined over a final layer, i.e., to match the output of the teacher [11,13], or over internal layers as well, i.e., feature distillation [71, 79, 81]. Given the challenging end-to-end sensorimotor learning task, the role of the teacher is to provide effective supervision to the student, i.e., informative targets in Eqn. 1. However, as the two agents operate from *drastically differing inputs*, the knowledge transfer from the teacher to the student can become ineffective. Specifically, the task of the student is significantly more challenging than that of the teacher, i.e., due to inherent noise and uncertainty. Indeed, there is currently a substantial gap between the performance of current sensorimotor learning agents and their teaching agents, even assuming access to expensive and high-quality 3D observations such as LiDAR [11, 13, 16] (not used in our work). The teaching process should also ideally consider any limited capacity of the student to facilitate effective optimization and learning [27]. Moreover, learning an optimal and robust teacher in itself can be a challenging task [13], even assuming the simplified learning task such as solved perception, as we demonstrate in Sec. 4 and addressed in Sec. 3.2.

Problem Setup and Agent Observations: We develop our CaT framework in the context of the CARLA simulator [21]. Our objective is to train a goal-conditional sensorimotor agent [16, 18] for mapping image and goal observations to vehicle throttle and steering control [13, 48]. We follow standard conditional imitation learning [11, 16] and assume access to three (non-overlapping) RGB camera views $\mathbf{I} = [\mathbf{I}_0, \mathbf{I}_1, \mathbf{I}_2] \in \mathbb{R}^{W \times H \times 3}$, a categorical navigational command $c \in \{1, \dots, 6\}$ (i.e., turn left, turn right, follow, forward, left lane changing and right lane changing) and an intermediate noisy goal $\mathbf{g} \in \mathbb{R}^2$ sampled from a GNSS (Global Navigation Satellite System) [11]. We note that both the command and GNSS observations can be easily obtained by today's vehicles, i.e., through generation of an A* plan [26] from a static map and coarse positioning system. The GNSS goals are sampled every 50-100 meters and reflect real-world measurement errors [11]. We compute a BEV $\mathbf{B} \in \{0, 1\}^{W_B \times H_B \times C_B}$ by rendering privileged (i.e., ground-truth) information from the underlying simulation, including 3D location and state of lanes, pedestrians, vehicles, and traffic lights [79]. While prior work has generally leveraged a standardized BEV representation [13, 32, 34, 57, 76], we find its design to be crucial to training an optimal teacher policy as will be discussed in Sec 3.2. We let $\mathbf{x}^s = (\mathbf{I}, \mathbf{g}, c) \in \mathcal{X}^s$ and $\mathbf{x}^t = (\mathbf{B}, \mathbf{g}, c) \in \mathcal{X}^t$ be the student and teacher observations, respectively. Given these observations at each time step, the agent learns to predict 10 future 2D waypoints in top-down vehicle coordinates for the next 2.5 seconds of driving. These are then given to a lateral and longitudinal PID controllers [13,48,66] to generate the final action.

Given this formulation, we next discuss our teacher agent design and training process. In particular, we aim to learn a privileged agent which can not only effectively solve the goal-oriented navigation task, but also *facilitate effective distillation and coaching* of a sensorimotor student policy, as will be discussed in Sec. 3.4.

3.2. Learning an Effective Teacher

In our formulation, learning a student agent begins with training an effective privileged agent $f_{\boldsymbol{\psi}}^t$. However, we find this itself to be a complex task. Even with complete knowledge of perfect BEV perception, current privileged agents are suboptimal, significantly under-performing CARLA's built-in autopilot [13, 79]. In this section, we uncover the underlying reason for this under-performance through the BEV design. Informed by this key finding, we train a highly effective imitation-learned teacher that surpasses prior privileged agents and matches expert-level performance in CARLA without requiring extensive data collection [56] or reinforcement learning [79]. Our proposed privileged agent can then enable effective supervision of the student in Sec. 3.3.
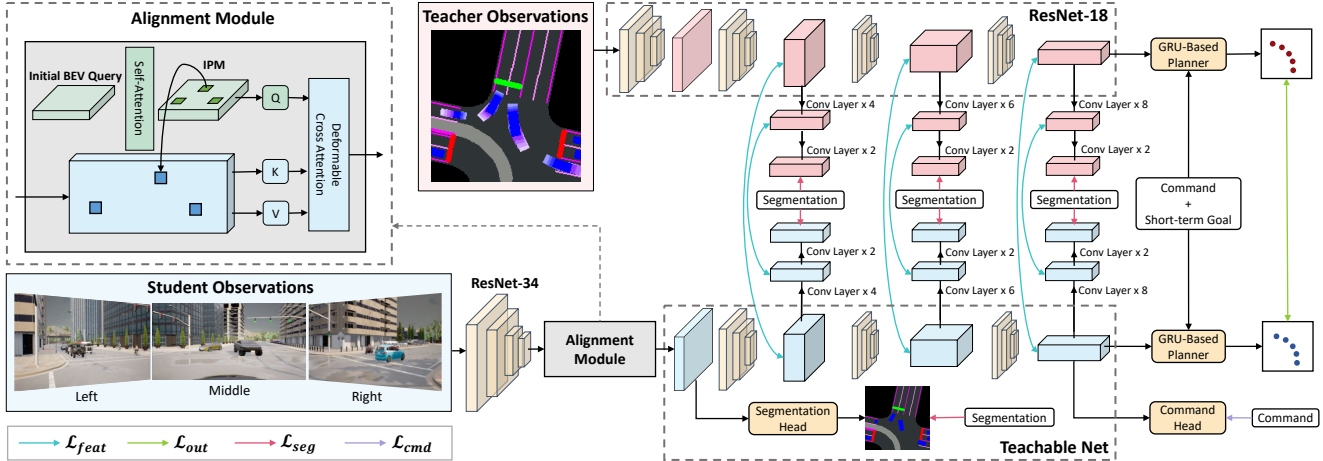
Figure 2. **Overview of Architecture and Distillation Process of CaT.** We first train a high-capacity privileged teacher agent (colored in red) which takes an augmented BEV representation as input, and produces a set of future waypoints conditioned on command and short-term goal. Our teachable student agent (colored in blue) maps the image features to the BEV space using an alignment module, which is trained by distillation of intermediate features (arrows colored in blue) and final output (arrow colored in green). Additionally, we leverage supervision from auxiliary segmentation (arrow colored in red) and command prediction tasks (arrow colored in purple) interlaced throughout the network to facilitate the learning of task-relevant features.

Teacher Training via Direct Expert Imitation: In imitation learning, an oracle (i.e., expert driver) $f^*: \mathcal{X}^t \rightarrow \mathcal{Y}$ is defined for demonstrating optimal actions $\mathbf{y}^*$. In CARLA [21], the expert is defined using longitudinal and lateral PID controllers which are carefully tuned and augmented by a collection of manually specified rules for handling diverse scenarios (e.g., sudden braking) over the underlying simulation state. The designed expert can then be used to generate trajectories as supervised data for learning the privileged policy function f^t_ψ with behavior cloning [18, 19, 48]. Our privileged agent is a ResNet-18 model [28], as shown in Fig. 2 (further increasing modeling capacity was not found to be beneficial). The ResNet model is followed by a GRU-based conditional waypoint predictor [11, 16]. Due to the complexity of our dense urban navigation task, e.g., with multi-lane roads, intersections, and merging, we incrementally refine the waypoints using two GRUs within each conditional branch. The first GRU regresses the preliminary set of waypoints via sequential waypoints directly from the embedded features. Subsequently, the second GRU takes the predicted waypoints together with the short-term goal and embedded features, to produce a refined set of waypoint targets.

Even assuming perfect perception, we find it difficult to train privileged agents that exhibit robust planning behavior in complex urban driving scenarios. Specifically, when using the most challenging CARLA benchmark (Longest6 [16], also detailed in Sec 4) using the standard BEV representation of Chen et al. [13] leads to 26% driving score compared to 72% by the rule-based expert. The rein-

forcement learning agent of Zhang et al. [79] achieves 60% on this task. We sought to explore the limits of behavior cloning for this task, as a noisy and sub-optimal teacher can hinder the training of the student model. In particular, we hypothesize that the reason for the poor performance lies in the increased complexity of the BEV state representation in dense scenarios. In such scenarios, learning to extract task-relevant details can become more challenging. To address this issue, we propose to introduce additional channels (akin to affordances [9, 22, 58]) that can more easily translate into safety-critical decision-making, simplify the learning task, and provide an expert-level teaching agent.

BEV with Safety Hints: Our BEV encodes state information into distinct channels, including drivable road regions, the desired route based on the command and layout, lane markings, and dynamic obstacles (see Fig. 3). Given the low performance of prior privileged agents, in particular around dynamic objects, the agent could benefit from more explicit safety-driven cues in the BEV. We propose to add two types of channels of (1) predicted agents' future and (2) entity attention. First, we utilize a kinematics bicycle model [16] in order to efficiently predict future trajectories of dynamic objects. This enables us to iteratively predict and represent short-term future position, orientation, and speed of agents (our supplementary contains additional information regarding this process). Secondly, we encode an explicit attention channel for highlighting *potential future infractions*, i.e., based on the kinematics bicycle model applied to the ego-vehicle. While this assumes that the current vehicle throttle and steering remain unchanged, the entity
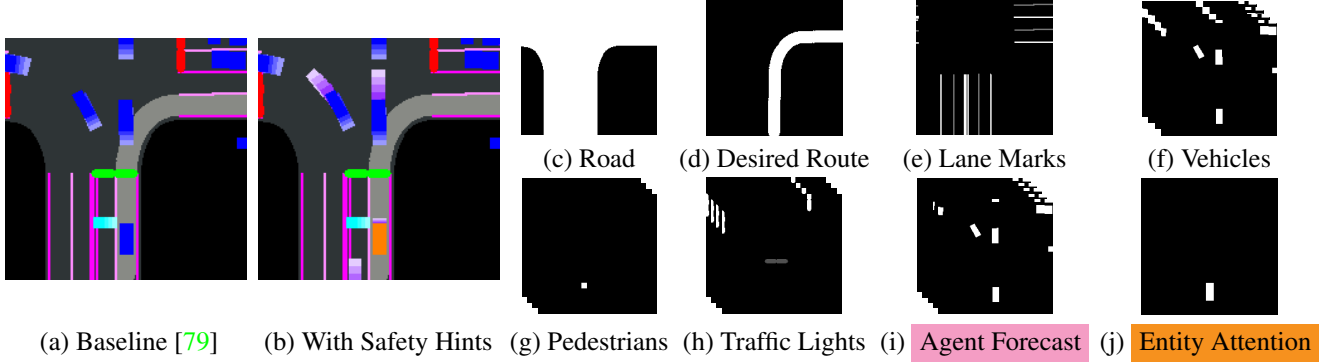
Figure 3. **Visualization of Our Proposed BEV with Safety Hints.** (a) and (b) visualize the baseline and our proposed BEV by encoding different channels into colors. Roads are shown in dark grey, Desired Route is shown in light grey, Lane Marks are shown in magenta, Vehicles are shown in blue with past trajectories shown in decreasing saturation with time into the past, Pedestrians are shown in light blue, Traffic Lights are shown in red or green according to its state, Agent Forecasts are shown in purple with decreasing saturation with time into the future, Entity Attention is shown in orange. (c) - (j) shows the different channels separately.

attention channel helps the agent better reason over safety while correlating states with actions. Surprisingly, the additions of the safety hints channels provide a strong teacher with offline behavior cloning alone, even improving over the rule-based expert (Table 1, 73% vs. 72%). We note that many of the remaining failures are due to simulation time-out in crowded intersections. Our boosted teacher agent will be critical for effective distillation in Sec. 3.4.

3.3. Learning a Teachable Student

Due to differing inputs and modeling capacities, it can be difficult to align the image-based student features and output with the BEV-based privileged teacher. We propose to alleviate misalignment by designing a teachable student network, i.e., a model that matches as much of the teacher architecture as possible. Specifically, we propose to improve teacher-student model alignment by leveraging an internal image-to-BEV alignment module. As shown in Fig. 2, the three residual blocks following the alignment module can consequently facilitate knowledge transfer via *direct distillation of most of the teacher's features* in Sec. 3.4.

IPM-based Transformer Alignment Module: Explicitly mapping images to BEV space for downstream tasks can be challenging. While this can be achieved via depth estimation and lifting [32,53], we sought to incorporate an efficient and differentiable module which can effectively attend and reason over various features while mapping from front view to BEV. A transformer-based architecture [14,65] provides a natural choice as it can be optimized end-to-end with respect to the driving task while also carefully aligning features among *spaces with arbitrary dimensionalities*. As shown in Fig. 2, we sample queries $\mathbf{Q}_{init} \in \mathbb{R}^{L_H \times L_W \times L_C}$ using a spatial parameterization of the BEV space and process them using a self-attention module [14,65]. The output

of self-attention module Q can be formulated as

$$\mathbf{Q} = \text{softmax}\left(\frac{\mathbf{Q}_{init}\mathbf{K}_{init}^T}{\sqrt{d}}\right)\mathbf{V}_{init} \qquad (2)$$

where $\mathbf{K}_{init}, \mathbf{V}_{init} \in \mathbb{R}^{L_H \times L_W \times L_C}$ provide the key-value pairs and d is the dimension of the query. Subsequently, each query point $\mathbf{q} \in \mathbf{Q}$ is mapped to its corresponding reference point $\mathbf{p} \in \mathbb{R}^3$ (homogeneous coordinates) in the image features $\mathcal{F}_{RGB} \in \mathbb{R}^{Z_H \times Z_W \times Z_C}$ using an Inverse Perspective Mapping (IPM)

$$\mathbf{p} = s\mathbf{P}_k\mathbf{R}_k(\mathbf{q} - \mathbf{t}_k) \qquad (3)$$

where s is a scale factor, $\mathbf{P}_k \in \mathbb{R}^{3 \times 3}$, $\mathbf{R}_k \in \mathbb{R}^{3 \times 3}$, and $\mathbf{t}_k \in \mathbb{R}^3$ are the kth camera intrinsics, rotation and translation relative to the center of the ego-vehicle, respectively [6]. Note that we employ a different IPM mapping for each camera view to perform the projection.

To populate the student's BEV features $\mathcal{F}_{BEV} \in \mathbb{R}^{H_B \times W_B \times C_B}$ (i.e., the output of the alignment module in Fig. 2), we leverage a deformable cross-attention mechanism based on deformable DETR [85]. This enables the network to attend to multiple regions around the reference point in the image features when transforming to BEV feature space. The reference point $\mathbf{p}$ and the learned deformable offsets can then be used as keys $\mathbf{K}$, with their corresponding features from the extracted images features the values $\mathbf{V}$, for a cross-attention module

$$\mathcal{F}_{BEV} = \text{DeformAttn}\left(\mathbf{Q}, \mathcal{F}_{RGB}, \mathbf{H}\right) \qquad (4)$$

where $\mathbf{H} = \{\mathbf{P}_k, \mathbf{R}_k, t_k\}_{k=1}^3$ contains the combined multi-view IPM parameters from Eqn. 3. In addition to the proposed alignment and teachable modules, we also incorporate prediction modules for auxiliary tasks (i.e., BEV prediction [16]), discussed in Sec. 3.4.

GRU-based Waypoint Predictor: The transformed image-to-BEV features are inputted to three residual blocks and a GRU-based conditional branch based on the command [11].

3.4. Student Loss

Even with our carefully designed architecture, the student's image-based learning task remains significantly more challenging than the teacher's. Thus, both optimization and alignment with the teacher can benefit from extensive supervision and gradual student instruction, i.e., to scaffold supervision and better match the student's ability. Our distillation process provides ample beneficial supervision for the student through three holistic mechanisms. First, we incorporate deep distillation losses, i.e., both over output (i.e., waypoint regression task) and feature maps, as shown in Fig. 2, that can more directly supervise internal layers. Second, we incorporate additional auxiliary losses that regularize the optimization process. This includes ground-truth supervised losses, e.g., defined over segmentation and control command, and a geometric loss defined over the post-alignment features that supervises for correct BEV structure. Finally, we leverage a student-paced coaching mechanism for gradually increasing the difficulty of the waypoint prediction task throughout the iterations in Sec. 3.5.

Loss Functions: Our optimization objective for guiding the distillation process is a weighted sum over both distillation and auxiliary tasks:

$$\mathcal{L}_{CaT} = \mathcal{L}_{out} + \mathcal{L}_{feat} + \mathcal{L}_{seg} + \mathcal{L}_{cmd} \qquad (5)$$

The various loss terms enable guiding holistic aspects of student learning.

Output Distillation: We leverage an L_1 loss computed over *all of the conditional branches* for each instance

$$\mathcal{L}_{out} = \sum_{c=1}^{C} \|f^s_{\boldsymbol{\theta}}(\mathbf{x}^s, c) - f^t_{\boldsymbol{\psi}}(\mathbf{x}^t, c)\|_1 \qquad (6)$$

This can be done by sampling from the teacher with different commands c (following Chen et al. [13]).

Feature Distillation Loss: Our proposed student architecture provides BEV-space feature alignment, i.e., for direct internal feature matching between the student and the teacher (the first term in our loss in Eqn. 7). This is in contrast to current methods which distill the output or a final fully-connected layer [11, 13, 71, 79]. However, while this can provide rich supervision for the student, the two models should not match entirely, i.e., due to differences in information processing for a perception-based planning task. Moreover, the BEV-based features can be highly sparse and structured, with slight feature offsets being less meaningful than obtaining a task and scene-relevant representation. To flexibly account for the student-based task and effectively represent structural information in BEV space, our feature loss is computed as

$$\mathcal{L}_{feat} = \sum_{i=1}^{3} \Big[\|\mathcal{F}^s_i(\mathbf{x}^s) - \mathcal{F}^t_i(\mathbf{x}^t)\|_2 + $$
$$\|T^s_i(\mathcal{F}^s_i(\mathbf{x}^s)) - T^t_i(\mathcal{F}^t_i(\mathbf{x}^t))\|_2 + $$
$$\lambda_{CD}\|\mathcal{F}^s_i(\mathbf{x}^s) - \mathcal{F}^t_i(\mathbf{x}^t))\|_{CD} \Big] \qquad (7)$$

where we use $\mathcal{F}_i$ to denote layers within the networks (as shown in Fig. 2, we directly distill three layers), T_i indicates the features following several convolutional layers (trained jointly with the student network model), and CD stands for a Chamfer Distance [51] (we set $\lambda_{CD} = 0.1$). The final term is computed over thresholded activations and a spatial soft-argmax [13] from each feature map.

Task-oriented Auxiliary Tasks: To ensure essential task-relevant information is preserved throughout the distillation process, we leverage regularizing supervision in the form of segmentation and command-based auxiliary tasks interlaced throughout the network [16, 49]. Our segmentation loss $\mathcal{L}_{seg}$ is a cross-entropy loss, computed over the ground-truth BEV and averaged across the segmentation prediction heads. Our command prediction loss $\mathcal{L}_{cmd}$ is binary cross-entropy supervised by the ground-truth command c, which facilitates learning task-relevant features [11, 19].

3.5. Student-paced Coaching

While Sec. 3.4 provides ample supervision on the student's internal representations, the teacher's targets may still be difficult to imitate for the student. To better consider the learning ability of the student, we propose to leverage a student-paced training mechanism which can gradually coach the student, i.e., through increasingly difficult samples. When computing the loss in Eqn. 5, we define a coach,

$$\mathcal{F}^t \leftarrow \lambda_i \mathcal{F}^s + (1 - \lambda_i)\mathcal{F}^t, \text{ if } \mathcal{L}_{CaT} > \tau_i \qquad (8)$$

which interpolates the teaching targets with the student's predictions and features. λ_i is a hyperparameter which is linearly decreased to 0 over training iteration i. The modified targets are only computed over the hard samples at each batch. Within each batch, τ_i is a threshold defining the lowest 50% of the scored samples for the loss with respect to the privileged teacher. While smoothing the targets in this manner may seem counterintuitive, i.e., compared to techniques which mine hard examples, the coaching mechanism aims to stabilize training by reducing the difficulty when the student is unable to perform the optimal action. Eqn. 8 enables adjusting the learning rate in a sample-selective manner (initially supressing difficult samples).

4. Experiments

We use the CARLA simulator (version 0.9.10.1) [21] for data generation and closed-loop evaluation of the proposed CaT framework. We also use open-loop evaluation

Table 1. **Quantitative Evaluation on the Longest6 Benchmark.** Comparison of CaT with prior methods in terms of Driving Score (DS), Route Completion (RC), and Infraction Score (IS). Additional infraction metrics (Pedestrian Collisions (Ped), Vehicle Collisions (Veh), Layout Collisions (LC), Red Light Violations (Red), Off-road Infraction (OR), Route deviation (Dev), Route Timeouts (TO), Agent Blocked (Blk)) are shown. FD refers to Feature Distillation, SH refers to Safety Hints. *-re-trained by us using the publicly available code. Mean and standard deviation are computed over three runs. CaT outperforms the state-of-the-art by a large margin in terms of DS.

Method	RGB	LiDAR	DS ↑	RC ↑	IS ↑	Ped ↓	Veh ↓	LC ↓	Red ↓	OR ↓	Dev ↓	TO ↓	Blk ↓
LAV [11]	✓	✓	48.41±3.40	80.71±0.84	0.60±0.04	**0.00**	0.50	0.19	0.07	0.20	0.07	0.01	**0.29**
TransFuser [16]	✓	✓	46.20±2.57	**83.61±1.16**	0.57±0.00	0.29	0.38	0.28	**0.04**	0.07	0.07	**0.00**	0.32
WOR [10]	✓	✗	17.36±2.95	43.46±2.99	0.54±0.06	0.05	0.64	0.15	0.84	0.15	0.89	0.04	0.45
NEAT [15]	✓	✗	24.08±3.30	59.94±0.50	0.49±0.02	0.01	0.71	0.21	0.18	0.18	**0.00**	0.02	2.83
TCP* [71]	✓	✗	42.86±0.63	61.83±4.19	0.71±0.04	0.01	0.42	0.11	**0.04**	0.08	0.01	0.01	0.75
CaT (w/o Alignment, Coaching, FD)	✓	✗	39.48±0.67	60.96±1.65	0.68±0.01	0.03	1.29	0.20	0.08	0.84	**0.00**	0.02	1.49
CaT (w/o Alignment, Coaching)	✓	✗	40.64±0.98	62.45±0.46	0.67±0.01	0.02	1.07	0.33	0.14	0.38	**0.00**	0.02	1.11
CaT (w/o Coaching, FD, SH)	✓	✗	44.10±0.40	65.84±5.55	0.72±0.03	0.01	0.26	0.05	0.11	0.15	**0.00**	**0.00**	0.66
CaT (w/o Coaching, SH)	✓	✗	49.69±2.28	81.10±0.58	0.64±0.02	0.01	0.78	0.03	0.08	0.20	**0.00**	0.02	0.51
CaT (w/o Coaching)	✓	✗	55.55±1.41	81.97±2.34	0.68±0.01	0.02	0.30	0.05	0.07	**0.06**	**0.00**	0.02	0.35
CaT	✓	✗	**58.36±2.24**	78.79±1.50	**0.77±0.02**	0.01	**0.20**	**0.02**	0.05	0.30	**0.00**	0.04	0.44
Privileged Agents:													
RL Expert (Roach) [79]	-	-	60.14±2.40	85.83±0.60	0.69±0.03	0.06	0.22	**0.00**	0.01	**0.00**	**0.00**	0.04	**0.07**
Rule-based Expert	-	-	71.96±2.13	77.46±3.11	**0.91±0.00**	**0.01**	**0.06**	**0.00**	**0.00**	**0.00**	**0.00**	0.03	0.41
Basic BEV Agent [13]	-	-	24.08±2.83	73.36±1.08	0.31±0.06	**0.01**	1.45	2.45	0.09	0.27	0.11	**0.01**	0.24
+ History and Desired Path	-	-	52.81±1.79	79.34±3.65	0.71±0.06	0.02	0.30	**0.00**	**0.00**	**0.00**	**0.00**	0.02	0.53
+ Agent Forecast	-	-	65.73±0.93	83.50±1.18	0.79±0.02	**0.01**	0.21	**0.00**	0.02	**0.00**	**0.00**	0.01	0.28
+ Entity Attention	-	-	**73.30±1.07**	**87.44±0.28**	0.83±0.02	0.05	0.11	**0.00**	0.02	**0.00**	**0.00**	0.02	0.12

on nuScenes [7, 33, 77] (ADE, FDE, and collision rate). In Carla, we leverage the Longest6 Benchmark [16], which uses the six longest routes of each town (Town01 - Town06) from the set of 76 routes provided by the official CARLA leaderboard [1] (a total of 36 routes). We note in addition to the standard cross weather and time of day generalization evaluation, the long and dense traffic conditions in Longest6 present the most challenging settings across the various CARLA-based benchmarks. To evaluate our models, we follow standard metrics and report Route Completion (RC, in terms of completed route percentage), Infraction Score (IS, a penalty factor over infractions), and Driving Score (DS, computed from the prior two) [1, 11, 16]).

4.1. Comparison with Prior Methods

CARLA Results: As shown in Table 1, using feature distillation, safety hints-based BEV for the teacher, and the proposed coaching mechanism, CaT is able to obtain a 58.36% DS. Specifically, we achieve state-of-the-art performance among all prior agents, including LiDAR-based approaches [11, 16] (by 20.6% in terms of DS, from 48.41 to 58.36, and 28.3% in terms of IS). We also note that such approaches provide strong baselines, due to various LiDAR-based safety checks that the agents perform on top of the trained policy, which we do not employ. Moreover, CaT outperforms the prior RGB-only state-of-the-art agent TCP [71] by 36.16% DS. To further validate the generalization of CaT, we discuss an additional benchmark split from TCP [71] and LAV [11] in the supplementary. We also find that removal of the alignment module degrades DS from 44.10% to 39.48%, and is shown to only slightly benefit from feature distillation (40.64% DS) indicating the effectiveness of the proposed architecture.

Table 2. **Open-Loop Evaluation on nuScenes.**

Method	ADE (m) ↓	FDE (m) ↓	Coll. (%) ↓
BEV Agent	0.33	0.52	0.49
CaT (w/o Coaching, FD, SH)	0.48	0.43	0.68
CaT	**0.41**	**0.36**	**0.27**

Real-World Evaluation: To further analyse the benefits of CaT in realistic driving settings, Table 2 shows open-loop evaluation (ADE, FDE, and Collision rate [77]) for nuScenes [7]. On the official validation split, the privileged agent performs best with an ADE of 0.33. CaT achieves 0.41 ADE, improving by 14.6% gain over an agent without distillation and coaching. Moreover, we find a significant reduction in the collision rate by 60.3% compared to the baseline. Additional details regarding nuScenes evaluation can be found in the supplementary.

4.2. Ablation Studies

Teacher Comparison: Table 1 depicts our teacher model ablation. Specifically, we find the proposed BEV channels to drastically improve the privileged agent's performance by simplifying the learning task (73.30% DS, surpassing the built-in rule-based expert of 71.96%). We note that our behavior cloning agent also significantly outperforms an RL-based expert [79] (60.14%), simply through an effective BEV design. Thus, our findings apply to real-world scenarios where interactive agent training may be unsafe. We also observe that the improved teacher benefits the student agent. Specifically, incorporating the proposed BEV-based safety hints results in student performance gains, from 49.69% to 55.55%.

Impact of Feature Distillation: Table 3 presents the benefits of our proposed multi-layer feature distillation frame-
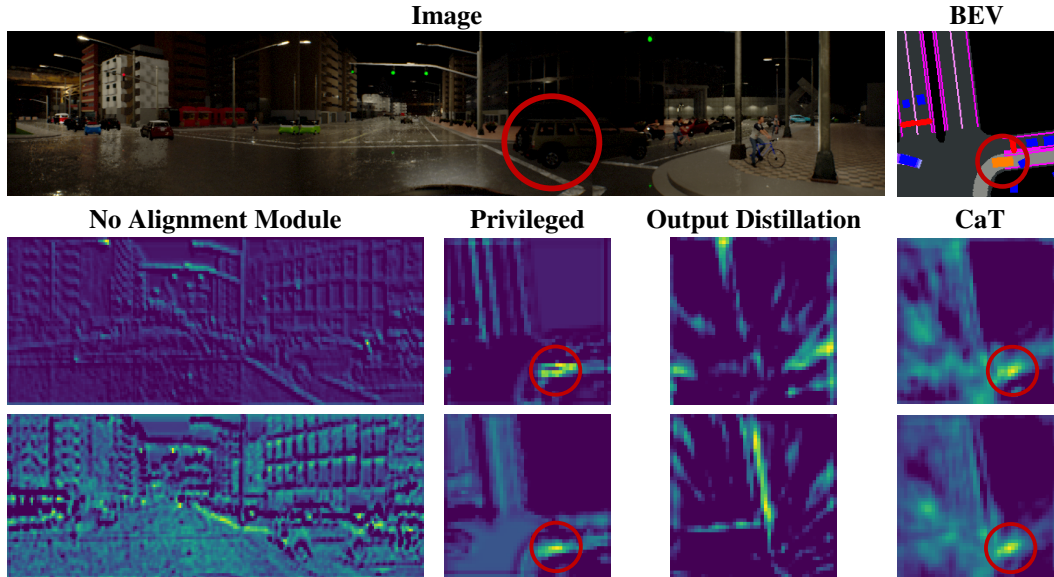
Figure 4. **Visualization of Model Input and Features.** We show the BEV ground truth (BEV) and RGB image ground truth (Image) on the top row. Next, we visualize the intermediate layer activations of a sensorimotor agent without the proposed alignment module [71] (No Alignment Module), privileged teacher agent (Privileged), student agent trained with output distillation only (Output Distillation), and the proposed agent (CaT). We demonstrate that our teacher agent with proposed knowledge distillation is able to guide the student agent in learning scene-relevant representations and attend to safety-critical entities (circled in red).

Table 3. **Ablation Study on Feature Distillation Layers.** All loss terms in Eqn. 7 benefit training with deep feature distillation.

Method	DS ↑	RC ↑	IS ↑
No Distillation	44.10	65.84	0.72
One Layer [71, 79]	45.23	69.33	0.68
Three Layers $\mathcal{L}_2$	49.31	66.92	0.78
Three Layers $\mathcal{L}_2 + \mathcal{L}_{CD}$	51.95	62.82	**0.87**
Three Layers $\mathcal{L}_{feat}$	**55.55**	**81.97**	0.68

work. We note that prior methods only distill one final fully connected layer (before the prediction head [71, 79], referred to as 'one layer' in Table 3). We note that such approaches result in minimal gains, whereas CaT is able to benefit from the extensive feature distillation (improving from 45.23% to 55.55%). Incorporating multiple types of feature distillation losses also benefits performance.

4.3. Feature Visualization

Fig. 4 depicts the learned features of the initial feature distilled layer (out of the three in Fig. 2). We visualize activations from two channels, to better understand the learned representation by the agents. Interestingly, the visualization shows how traditionally trained sensorimotor agents (i.e., without the proposed alignment module) mostly operate in image-space features in this layer despite the BEV waypoint prediction task. This validates our main hypothesis, that such agents have limited ability to learn from a BEV-based agent, even with BEV-based auxiliary prediction tasks. We

also find that output distillation provides a weak supervision for critical components of the BEV, including layout and object features. After feature distillation, the features of our CaT model are able to not only keep the driving-related semantic information i.e., drivable road and route, but also focus on safety-critical entities. More visualization examples are provided in the supplementary.

5. Conclusion

We present CaT, a novel knowledge distillation method for effectively training a sensorimotor student agent using the supervision from a privileged agent. We leverage an alignment module to better map image features to the BEV space, thus enabling extensive supervision from a BEV-based teacher over the intermediate feature learning. As both agents can be trained offline via imitation learning, our findings are directly relevant to real-world settings where on-policy interactions may be unsafe to perform. Finally, we carefully account for inherent differences between the student and the teacher using a student-paced coaching mechanism with various auxiliary supervision tasks. Through the improved knowledge distillation process, our experiments result in a state-of-the-art camera-based agent in CARLA. Our ablation into input and model design guides future directions through interpretable analysis into current limitations of end-to-end driving models.

References

[1] Carla autonomous driving leaderboard. *https://leaderboard.carla.org/*, 2022. 7

[2] Yonathan Aflalo, Asaf Noy, Ming Lin, Itamar Friedman, and Lihi Zelnik. Knapsack pruning with inner distillation. *arXiv preprint arXiv:2002.08258*, 2020. 2

[3] Gustavo Aguilar, Yuan Ling, Yu Zhang, Benjamin Yao, Xing Fan, and Chenlei Guo. Knowledge distillation from internal representations. In *AAAI*, 2020. 2

[4] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. ChauffeurNet: Learning to drive by imitating the best and synthesizing the worst. In *RSS*, 2019. 1

[5] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *ICML*, 2009. 2

[6] Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. An extension to the inverse perspective mapping to handle non-flat roads. In *IV*, 1998. 5

[7] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020. 7

[8] Raphael Chekroun, Marin Toromanoff, Sascha Hornauer, and Fabien Moutarde. Gri: General reinforced imitation and its application to vision-based autonomous driving. *arXiv preprint arXiv:2111.08575*, 2021. 2

[9] Chenyi Chen, Ari Seff, Alain L. Kornhauser, and Jianxiong Xiao. DeepDriving: Learning affordance for direct perception in autonomous driving. In *ICCV*, 2015. 2, 4

[10] Dian Chen, Vladlen Koltun, and Philipp Krähenbühl. Learning to drive from a world on rails. In *ICCV*, 2021. 2, 7

[11] Dian Chen and Philipp Krähenbühl. Learning from all vehicles. In *CVPR*, 2022. 1, 2, 3, 4, 5, 6, 7

[12] Defang Chen, Jian-Ping Mei, Yuan Zhang, Can Wang, Zhe Wang, Yan Feng, and Chun Chen. Cross-layer distillation with semantic calibration. In *AAAI*, 2021. 2

[13] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *CoRL*, 2020. 1, 2, 3, 4, 6, 7

[14] Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi, Yu Qiao, et al. Persformer: 3d lane detection via perspective transformer and the openlane benchmark. *arXiv preprint arXiv:2203.11089*, 2022. 2, 5

[15] Kashyap Chitta, Aditya Prakash, and Andreas Geiger. Neat: Neural attention fields for end-to-end autonomous driving. In *ICCV*, 2021. 7

[16] Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. *arXiv preprint arXiv:2205.15997*, 2022. 3, 4, 5, 6, 7

[17] Inseop Chung, SeongUk Park, Jangho Kim, and Nojun Kwak. Feature-map-level online adversarial knowledge distillation. In *ICML*, 2020. 2

[18] Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *ICRA*, 2018. 3, 4

[19] Felipe Codevilla, Eder Santana, Antonio M López, and Adrien Gaidon. Exploring the limitations of behavior cloning for autonomous driving. In *ICCV*, 2019. 4, 6

[20] Pim de Haan, Dinesh Jayaraman, and Sergey Levine. Causal confusion in imitation learning. In *NeurIPS*, 2019. 1

[21] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *CoRL*, 2017. 2, 3, 4, 6

[22] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology press, 2014. 4

[23] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. In *IJCV*, 2021. 1, 3

[24] Sheng Guo, Weilin Huang, Haozhi Zhang, Chenfan Zhuang, Dengke Dong, Matthew R Scott, and Dinglong Huang. Curriculumnet: Weakly supervised learning from large-scale web images. In *ECCV*, 2018. 2

[25] Saurabh Gupta, James Davidson, Sergey Levine, Rahul Sukthankar, and Jitendra Malik. Cognitive mapping and planning for visual navigation. In *CVPR*, 2017. 2

[26] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 1968. 3

[27] He He, Jason Eisner, and Hal Daume. Imitation learning by coaching. In *NeurIPS*, 2012. 1, 2, 3

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 4

[29] Byeongho Heo, Jeesoo Kim, Sangdoo Yun, Hyojin Park, Nojun Kwak, and Jin Young Choi. A comprehensive overhaul of feature distillation. In *ICCV*, 2019. 2

[30] Roger Hiemstra and Burton Sisco. *Individualizing Instruction. Making Learning Personal, Empowering, and Successful*. 1990. 2

[31] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015. 2

[32] Anthony Hu, Zak Murez, Nikhil Mohan, Sofía Dudas, Jeffrey Hawke, Vijay Badrinarayanan, Roberto Cipolla, and Alex Kendall. Fiery: Future instance prediction in bird's-eye view from surround monocular cameras. In *ICCV*, 2021. 2, 3, 5

[33] Peiyun Hu, Aaron Huang, John Dolan, David Held, and Deva Ramanan. Safe local motion planning with self-supervised freespace forecasting. In *CVPR*, 2021. 7

[34] Zanming Huang, Zhongkai Shangguan, Jimuyang Zhang, Gilad Bar, Matthew Boyd, and Eshed Ohn-Bar. ASSISTER: Assistive navigation via conditional instruction generation. In *ECCV*, 2022. 3

[35] Zehao Huang and Naiyan Wang. Like what you like: Knowledge distill via neuron selectivity transfer. *arXiv preprint arXiv:1707.01219*, 2017. 2

[36] Mingi Ji, Byeongho Heo, and Sungrae Park. Show, attend and distill: Knowledge distillation via attention-based feature matching. In *AAAI*, 2021. 2

[37] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *ICML*, 2018. 2

[38] Jangho Kim, SeongUk Park, and Nojun Kwak. Paraphrasing complex network: Network compression via factor transfer. In *NeurIPS*, 2018. 2

[39] Tom Kocmi and Ondrej Bojar. Curriculum learning and minibatch bucketing in neural machine translation. *arXiv preprint arXiv:1707.09533*, 2017. 2

[40] Animesh Koratana, Daniel Kang, Peter Bailis, and Matei Zaharia. Lit: Learned intermediate representation training for model compression. In *ICML*, 2019. 2

[41] M Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *NeurIPS*, 2010. 2

[42] Ravi Kumar Kushawaha, Saurabh Kumar, Biplab Banerjee, and Rajbabu Velmurugan. Distilling spikes: Knowledge distillation in spiking neural networks. In *ICPR*, 2021. 2

[43] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016. 1

[44] Dong Li, Jia-Bin Huang, Yali Li, Shengjin Wang, and Ming-Hsuan Yang. Weakly supervised object localization with progressive domain adaptation. In *CVPR*, 2016. 2

[45] Guohao Li, Matthias Mueller, Vincent Casser, Neil Smith, Dominik L Michels, and Bernard Ghanem. Oil: Observational imitation learning. In *RSS*, 2019. 2

[46] Yuang Liu, Wei Zhang, and Jun Wang. Source-free domain adaptation for semantic segmentation. In *CVPR*, 2021. 2

[47] Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 7(62), 2022. 1

[48] Matthias Müller, Alexey Dosovitskiy, Bernard Ghanem, and Vladlen Koltun. Driving policy transfer via modularity and abstraction. In *CoRL*, 2018. 3, 4

[49] Eshed Ohn-Bar, Aditya Prakash, Aseem Behl, Kashyap Chitta, and Andreas Geiger. Learning situational driving. In *CVPR*, 2020. 6

[50] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, Jan Peters, et al. An algorithmic perspective on imitation learning. *Foundations and Trends® in Robotics*, 2018. 2, 3

[51] Despoina Paschalidou, Angelos Katharopoulos, Andreas Geiger, and Sanja Fidler. Neural parts: Learning expressive 3D shape abstractions with invertible neural networks. In *CVPR*, 2021. 6

[52] Nikolaos Passalis and Anastasios Tefas. Learning deep representations with probabilistic knowledge transfer. In *ECCV*, 2018. 2

[53] Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D. In *ECCV*, 2020. 2, 5

[54] Dean A Pomerleau. ALVINN: An autonomous land vehicle in a neural network. In *NeurIPS*, 1989. 1, 2

[55] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550*, 2014. 2

[56] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *AISTATS*, 2011. 2, 3

[57] Avishkar Saha, Oscar Mendez, Chris Russell, and Richard Bowden. Translating images into maps. In *ICRA*, 2022. 3

[58] Axel Sauer, Nikolay Savinov, and Andreas Geiger. Conditional affordance learning for driving in urban environments. In *CoRL*, 2018. 4

[59] Miaojing Shi and Vittorio Ferrari. Weakly supervised object localization using size estimates. In *ECCV*, 2016. 2

[60] Suraj Srinivas and Francois Fleuret. Knowledge transfer with Jacobian matching. In *ICML*, 2018. 2

[61] Yuxing Tang, Xiaosong Wang, Adam P Harrison, Le Lu, Jing Xiao, and Ronald M Summers. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In *MLMI*, 2018. 2

[62] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *CVPR*, 2020. 2

[63] Radu Tudor Ionescu, Bogdan Alexe, Marius Leordeanu, Marius Popescu, Dim P Papadopoulos, and Vittorio Ferrari. How hard can it be? estimating the difficulty of visual search in an image. In *CVPR*, 2016. 2

[64] Rachel R Van Der Stuyf. Scaffolding as a teaching strategy. *Adolescent learning and development*, 2002. 2

[65] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 5

[66] Antonio Visioli. *Practical PID control*. Springer Science & Business Media, 2006. 3

[67] Tao Wang, Li Yuan, Xiaopeng Zhang, and Jiashi Feng. Distilling object detectors with fine-grained feature imitation. In *CVPR*, 2019. 2

[68] Wenhui Wang, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. Minilmv2: Multi-head self-attention relation distillation for compressing pretrained transformers. *arXiv preprint arXiv:2012.15828*, 2020. 2

[69] Wei Wang, Isaac Caswell, and Ciprian Chelba. Dynamically composing domain-data selection with clean-data selection by" co-curricular learning" for neural machine translation. *arXiv preprint arXiv:1906.01130*, 2019. 2

[70] Chuan Wen, Jierui Lin, Trevor Darrell, Dinesh Jayaraman, and Yang Gao. Fighting copycat agents in behavioral cloning from observation histories. In *NeurIPS*, 2020. 1

[71] Penghao Wu, Xiaosong Jia, Li Chen, Junchi Yan, Hongyang Li, and Yu Qiao. Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline. *arXiv preprint arXiv:2206.08129*, 2022. 1, 2, 3, 6, 7, 8

[72] Liuyu Xiang, Guiguang Ding, and Jungong Han. Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification. In *ECCV*, 2020. 2

[73] Canwen Xu, Wangchunshu Zhou, Tao Ge, Furu Wei, and Ming Zhou. Bert-of-theseus: Compressing bert by progressive module replacing. *arXiv preprint arXiv:2002.02925*, 2020. 2

[74] Jing Yang, Brais Martinez, Adrian Bulat, and Georgios Tzimiropoulos. Knowledge distillation via adaptive instance normalization. *arXiv preprint arXiv:2003.04289*, 2020. 2

[75] Junho Yim, Donggyu Joo, Jihoon Bae, and Junmo Kim. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In *CVPR*, 2017. 2

[76] Jimuyang Zhang and Eshed Ohn-Bar. Learning by watching. In *CVPR*, 2021. 2, 3

[77] Jimuyang Zhang, Ruizhao Zhu, and Eshed Ohn-Bar. SelfD: Self-learning large-scale driving policies from the web. In *CVPR*, 2022. 2, 7

[78] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, and Fang Wen. Robust mutual learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:2106.00609*, 2021. 2

[79] Zhejun Zhang, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. End-to-end urban driving by imitating a reinforcement learning coach. In *ICCV*, 2021. 1, 2, 3, 4, 5, 6, 7, 8

[80] Albert Zhao, Tong He, Yitao Liang, Haibin Huang, Guy Van den Broeck, and Stefano Soatto. Lates: Latent space distillation for teacher-student driving policy learning. *arXiv preprint arXiv:1912.02973*, 2019. 2

[81] Albert Zhao, Tong He, Yitao Liang, Haibin Huang, Guy Van den Broeck, and Stefano Soatto. Sam: Squeeze-and-mimic networks for conditional visual driving policy learning. *CoRL*, 2020. 1, 2, 3

[82] Brady Zhou and Philipp Krähenbühl. Cross-view transformers for real-time map-view semantic segmentation. In *CVPR*, 2022. 2

[83] Brady Zhou, Philipp Krähenbühl, and Vladlen Koltun. Does computer vision matter for action? *Science Robotics*, 4(30), 2019. 2

[84] Yikai Zhou, Baosong Yang, Derek F Wong, Yu Wan, and Lidia S Chao. Uncertainty-aware curriculum learning for neural machine translation. In *ACL*, 2020. 2

[85] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 5