Uncertainty-Guided Never-Ending Learning to Drive

Lei Lai Eshed Ohn-Bar Sanjay Arora† John Seon Keun Yi
Boston University †Red Hat Inc.
{leilai, eohnbar, jskyi}@bu.edu saarora@redhat.com

Abstract

We present a highly scalable self-training framework for incrementally adapting vision-based end-to-end autonomous driving policies in a semi-supervised manner, i.e., over a continual stream of incoming video data. To facilitate large-scale model training (e.g., open web or unlabeled data), we do not assume access to ground-truth labels and instead estimate pseudo-label policy targets for each video. Our framework comprises three key components: knowledge distillation, a sample purification module, and an exploration and knowledge retention mechanism. First, given sequential image frames, we pseudo-label the data and estimate uncertainty using an ensemble of inverse dynamics models. The uncertainty is used to select the most informative samples to add to an experience replay buffer. We specifically select high-uncertainty pseudo-labels to facilitate the exploration and learning of new and diverse driving skills. However, in contrast to prior work in continual learning that assumes ground-truth labeled samples, the uncertain pseudo-labels can introduce significant noise. Thus, we also pair the exploration with a label refinement module, which makes use of consistency constraints to re-label the noisy exploratory samples and effectively learn from diverse data. Trained as a complete never-ending learning system, we demonstrate state-of-the-art performance on training from domain-changing data as well as millions of images from the open web.

1. Introduction

Despite unprecedented advancements in sensors and compute since the nascent days of learning-based autonomous driving, the manner in which driving decision policies are trained today is still surprisingly narrow when compared with human learning. In a similar fashion to the classical ALVINN [54] over three decades ago, engineers and researchers may first collect a large, clean dataset of demonstrations and then utilize it to train, and subsequently deploy, the policy model [7, 14, 19, 27–29, 31–33, 55]. However, shortly after deployment, reports of various failure events, such as colliding with a jaywalker [49] or handling emergency vehicles [2], begin to emerge. Determined to resolve such scenarios, additional data may be collected in such uncertain situations, and the model is re-trained from scratch. Yet, to their dismay, the next iteration exhibits additional failures, e.g., uncertainty when encountering an unfamiliar object on the road (a puddle or a rock), or perhaps a drastically different environment, and the cycle repeats.

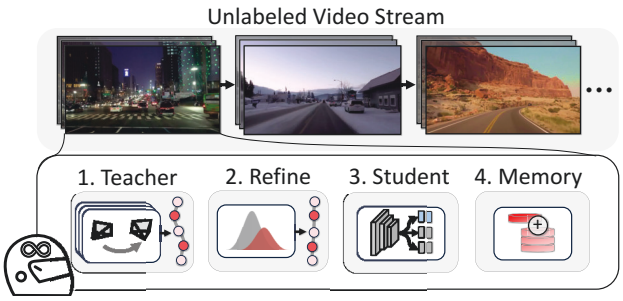

Figure 1. **Never-Ending Learning to Drive.** We present **∞-Driver**, an agent that continually learns from unlabeled incoming video data. For each video in the video stream, the agent employs an ensemble of inverse dynamics models to infer waypoint pseudo-labels and their uncertainty (higher uncertainty is depicted in red). Noisy pseudo-labels are automatically refined through consistency-based re-labeling and confidence-based filtering. Next, a driving policy student model is trained over incoming and episodic memory replay data. The memory buffer is updated to incorporate high-uncertainty samples. This maintains a diverse set of samples to retain knowledge and prevent forgetting, despite only viewing a given image once. Towards learning a generalized driving policy, our efficient framework enables highly scalable training, i.e., over millions of video frames from the web.

In contrast to such isolated policies and fixed deployment scenarios, to learn to drive safely humans integrate a continual process of acquiring, fine-tuning, and adapting knowledge based on a never-ending stream of observations that often span years and primarily comprises self-supervised experience [16, 44, 47, 52, 63]. Human drivers continuously build on what they've previously learned to handle unseen scenarios and domains, from complex merging into busy traffic and up to driving in intricate highways, tunnels, and handling bad weathers. Indeed, the likelihood of fatal crash rates for drivers is reduced by nearly a factor of three

in the following decades since first receiving a driver's licence [23, 62]. This incremental learning process is performed without retaining massive amounts of previously observed experience but only via recalling the most relevant and salient prior experiences and skills. How can we develop incremental driving policy learning frameworks that can be efficiently trained and continually adapted over massive amounts of diverse and self-supervised driving data?

Incremental policy learning is a realistic and practical learning paradigm. Yet, due to the complexity of incrementally accumulating knowledge in driving settings, e.g., unlabeled and unconstrained on-road data with out-of-distribution, non-stationary, and intricate scenes, prior work tends to instead focus on simplified or isolated subcomponents within the never-ending learning task, particularly transfer learning and domain adaptation [43, 45, 58, 66, 70]. Such approaches cannot mitigate issues with generalization, i.e., avoiding domain-specific overfitting and forgetting previous knowledge [20, 26, 64]. To improve generalization, methods may further leverage an additional pre-training stage using self-supervised strategies, e.g., from unlabeled web data [4, 9, 12, 25, 41, 68, 74, 76]. Yet, these do not scale to massive amounts of driving data available on the web as they require storage of the entire dataset in memory without performing exploration and experience selection. Towards enabling more scalable driving policies, our goal is to tackle the full continual self-supervised driving policy training. Our proposed framework enables training behavioral models from large amounts of unlabeled and unconstrained real-world on-road data with frequent out-of-distribution, non-stationary, and intricate scenes.

Contribution: We introduce ∞-Driver, a novel framework for continual self-supervised driving policy training. Our efficient framework enables scalable training from large amounts of unlabeled and unconstrained real-world on-road data with frequent out-of-distribution, non-stationary, and complex scenes (i.e., as opposed to commonly used simplistic settings on MNIST or CIFAR [5, 6, 20, 26, 34, 36, 52]). Our effective approach interfaces four highly efficient mechanisms. First, we formulate incremental policy learning as knowledge distillation from an inverse dynamics teacher model. A model ensemble produces epistemic uncertainty estimates that are used to guide subsequent steps. To leverage high-uncertainty but potentially noisy pseudo-labels, we leverage temporal consistency-based re-labeling and label filtering mechanisms. We then train a student conditional imitation learning model [18], while leveraging the epistemic uncertainty to maintain a diverse episodic memory buffer with highly informative samples. When integrated as a complete incremental learning system, our method achieves state-of-the-art performance compared to baseline algorithms in generalized, cross-dataset evaluation settings. Our code is available at https://infdriver.github.io/.

2. Related Work

Learning to Drive: Many modern autonomous driving systems [7, 14, 19, 28, 31–33, 55, 61, 73, 75] employ an end-to-end framework, where a driving policy, either a high-level plan or low-level control, is learned directly from sensor input. However, these methods require a prohibitive amount of labeled data and long interactions for training [7]. Furthermore, while some works display superior performance on closed-loop evaluations in simulation [1, 33, 61], they may fail to generalize to complex settings, multiple domains, or unseen environments in the real world. To tackle scalability and adaptability, recent works [12, 68, 74, 76, 78] utilize vast and abundant driving videos from YouTube to pre-train a visuomotor driving policy. Unlabeled and uncalibrated driving videos are used to pseudo-label the actions [12, 74, 76] or to provide feature representations catered to the driving policy [68]. Zhang et al. [74] proposes a semi-supervised driving policy that learns to map monocular images directly to ego vehicle waypoints in the BEV space. The approach utilizes an unlabeled set of YouTube videos to repeatedly train the conditional imitation learner [15, 18] and learns a generalized policy. However, pseudo-labels are obtained via sampling from the model, which can produce significant noise. Moreover, the approach does not infer ego-vehicle speed and trajectory, thus discarding useful demonstration data. In contrast, we use a visual odometry teacher model to provide pseudo-labels of speed and command. Critically, all aforementioned methods leverage inefficient and repeated visitation of the entire dataset which is stored in memory, and are thus limited in scalability. Our method only requires watching an image stream once.

Incremental Learning: As most studies leverage annotated and simple classification datasets to analyze incremental learning, only a few have considered incremental learning on a noisy data stream. Kim et al. [36] studies a noisy labeled continual learning task, however, the method assumes disjoint class scenarios where no class overlaps between task streams, which is less practical in real-world applications. Bang et al. [6] proposes a framework for online blurry continual learning on a contaminated data stream, leveraging consistency regularization to label high-noise data. Although consistency regularization can improve model generalization by enforcing the model to be consistent in its prediction of augmented data, the effectiveness in handling out-of-distribution (OOD) data is limited [59, 77]. Mirza et al. [46] leverages a domain-incremental learning approach to drive. However, the method employs a ground-truth dataset and thus cannot be learned at scale from online resources. Furthermore, their method only considers weather conditions as a variable component in driving, whereas in reality driving conditions can vary significantly,

such as traffic rules, construction styles, and moving objects. Thus, simply training on a few weather conditions from limited datasets cannot achieve a generalizable driving model. Our work employs driving videos from YouTube as the data stream for domain-incremental learning, which enables learning from vast amounts of data and diverse domains.

Episodic Memory: In contrast to transfer learning or domain adaptation, the central challenge in incremental learning to drive lies in minimizing model forgetting, i.e., retaining and developing knowledge while maintaining model generalization [20, 36, 64]. In our work, we build on recent advances in continual learning, particularly rehearsal-based approaches which have shown SOTA performance. Recently, replay buffer-based strategies, where selected training data is stored in a memory buffer and rehearsed, have emerged as an effective state-of-the-art for continual learning [5, 6, 10]. Instead of random reservoir sampling [10], we utilize an uncertainty-guided memory buffer, which maintains information from the previous data to mitigate forgetting. In contrast to prior works, we show that prioritizing high-uncertainty pseudo-labeled samples, once post-processed and smoothed using temporal consistency, can benefit incremental learning.

3. Method

We propose a novel framework that incrementally observes a stream of unlabeled, non-stationary driving data and learns a vision-based motion planning policy that can efficiently accumulate generalized driving skills[1]. Our scalable approach interfaces three highly efficient modules. First, we formulate incremental policy self-training (Sec. 3.1) as knowledge distillation from an inverse dynamics teacher model (Sec. 3.2). The teacher model is used to pseudo-label incoming image sequences and produce supervision to a sensorimotor student model. Second, to leverage high-uncertainty but potentially noisy pseudo-labels, we propose to leverage temporal consistency-based re-labeling as well as adaptive filtering mechanisms (Sec. 3.3). Finally, we propose to leverage epistemic uncertainty to guide and maintain an episodic memory buffer (Sec. 3.4). When integrated as a complete incremental learning system, these modules enable highly efficient and effective self-learning from diverse and informative pseudo-labeled samples. We describe an outline of the framework in Algorithm 1.

3.1. Problem Settings

Preliminaries: We follow standard conditional imitation learning [15, 17, 30] approaches to learn a regression function $f_{\boldsymbol{\theta}}^{\text{plan}} : \mathcal{X} \rightarrow \mathcal{Y}$, parameterized as a neural network with weights $\boldsymbol{\theta}$, to map current observations $\mathbf{x} = (\mathbf{I}, v, c) \in \mathcal{X}$, comprised of image $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$, speed $v \in \mathbb{R}$, and categorical high-level driving goal command $c \in \mathcal{C}$ [18], to a desired future ego-vehicle location output $\mathcal{Y}$. Specifically, we define $\mathcal{Y}$ as a set of K sequential ego-relative world coordinates in a top-down map view, $\mathbf{y} = \{\mathbf{y}^k\}_{k=1}^{K}$ with $\mathbf{y}^k \in \mathbb{R}^2$. In general, the waypoints prediction function,

$$\mathbf{y} = f_{\boldsymbol{\theta}}^{\text{plan}}(\mathbf{I}, v, c) \tag{1}$$

can be optimized over demonstrations, i.e., via supervised learning [15, 17, 28, 33, 54, 56, 57]. In contrast, we do not assume a single fixed labeled dataset and instead use a *continual collection of unlabeled images* extracted from driving videos, $\{\mathbf{I}_t\}_{t=1}^{\infty}$, which cannot be repeatedly re-visited in training. During training, to balance adapting to new experiences while retaining old ones, we construct two sets based on the image stream, $\mathcal{U}$, which contains the most recent observed images, and $\mathcal{B}$, which is a long-term experience replay buffer constructed via uncertainty (Sec. 3.4).

Incremental Self-Training: To train the ego-vehicle future trajectory prediction model, prior methods typically rely on high-quality datasets with expert labels [17, 27, 28, 30, 31]. However, in practice, this requirement can severely restrict the availability and diversity of the training data. Obtaining accurate waypoint labels in the real world involves a complex multi-step and multi-sensor optimization with GPS (which is unreliable in dense urban areas), LiDAR,

Algorithm 1 Uncertainty-Guided Never-Ending Learning

Input:
$\mathcal{U}$: incoming image collection
$\mathcal{G}$: initial small ground-truth dataset
$\mathcal{S}$: image dataset with pseudo-labels
$\mathcal{B}$: replay buffer
$f_{\boldsymbol{\psi}_m}^{\text{inv}}$: inverse dynamics models
$f_{\boldsymbol{\theta}}^{\text{plan}}$: motion planning policy
α, β: learning rates
procedure INVERSE DYNAMICS($\mathcal{G}, \mathcal{S}, \mathcal{U}, f_{\boldsymbol{\psi}_m}^{\text{inv}}$)
  $\boldsymbol{\psi}_m \leftarrow \boldsymbol{\psi}_m - \alpha \cdot \nabla[\mathcal{L}_{inv}]$ ▷ train on $\mathcal{G} \cup \mathcal{S}$
  $\mathcal{S}$ = reservoir sampling($\mathcal{U}$)
  $(\hat{\mathbf{y}}_t, \hat{v}_t, \hat{c}_t) \leftarrow \mathcal{T}(\{\hat{\mathbf{p}}_i\}_{i=t}^{t+h})$ ▷ Eq. 4
  $u_t = \text{std}\left(\{\hat{\mathbf{y}}_t^m\}_{m=1}^{M}\right)$
end procedure
procedure PLANNING POLICY($\mathcal{U}, \mathcal{B}, f_{\boldsymbol{\theta}}^{\text{plan}}$)
  $\mathcal{U} \leftarrow$ confidence-based filter($\mathcal{U}$) ▷ Eq. 8
  $\bar{\mathbf{y}}_t$ = sample relabeling($\hat{\mathbf{y}}_t, u_t$) ▷ Eq. 6,7
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \beta \cdot \nabla[\mathcal{L}_{plan}]$ ▷ train on $\mathcal{B} \cup \mathcal{U}$
  $\mathcal{U} \leftarrow$ low-loss filter($\mathcal{U}, \mathcal{L}(\mathbf{y}_t, \bar{\mathbf{y}}_t)$) ▷ Eq. 9
  $\mathcal{B} \leftarrow$ buffer update($\mathcal{U}, \mathcal{B}, u_t$)
end procedure

[1] In the context of continual learning to drive, we view driving within a specific domain, data distribution, or handling unseen events and maneuvers as novel skills or tasks

and HD map data [11]. Thus, instead of the widely studied supervised incremental learning setup over streaming *labeled observations* [5, 20, 52, 60], we pursue a more realistic and generalized setting over streaming unlabeled data from the web. In the following, we discuss our framework for effectively making use of such data while also retaining knowledge throughout the incremental training process. We emphasize that this setting enables leveraging diverse and unconstrained sources of observations, i.e., either when adapting the driving model over incoming frame-by-frame driving data [44] or for efficient learning from *arbitrarily large amounts of data*, e.g., from the web [68, 76]. Both use cases will be further studied as an instantiation of our general framework in Sec. 4.

3.2. Inverse Dynamics Teacher

We propose to estimate underlying decisions made by drivers in data. This enables us to leverage the unlabeled data to directly train a decision-making policy, unlike pre-training-based strategies that cannot leverage the demonstration data to learn decision-making [68, 74].

Two-Model Pseudo-Labeling Approach: We ease the difficult incremental learning from images task by formulating self-training as an *interplay between two-models*, where ego-vehicle waypoints, speed and command are first inferred from the images collection extracted from driving videos using inverse dynamics model $f^{\text{inv}}_{\boldsymbol{\psi}_m}$, and subsequently used as supervision for training the planning the policy function $f^{\text{plan}}_{\boldsymbol{\theta}}$.

We estimate the ego-motion between two consecutive images by training an inverse dynamics model $f^{\text{inv}}_{\boldsymbol{\psi}_m}$ [4, 12, 76] and compute the ego-vehicle waypoints, speed and command from a sequence of ego-motion predictions. Our model estimates pose $\mathbf{p} = [\mathbf{R}|\mathbf{t}]$ where $\mathbf{R} \in \text{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ are the rotation and translation of an image pair. Besides regressing on pose $\mathbf{p}$, we leverage a probabilistic loss, used to estimate prediction confidence and detect poor predictions. We use the matrix Fisher distribution to model the rotation distribution on $\text{SO}(3)$ [72]. The probability density function of the matrix Fisher distribution is:

$$p_{mf}(\mathbf{R}|\boldsymbol{\Phi}) = \frac{1}{F(\boldsymbol{\Phi})} \exp\left(tr(\boldsymbol{\Phi}^\top \mathbf{R})\right) \tag{2}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{3\times3}$ is the parameter of matrix Fisher distribution and $F(\boldsymbol{\Phi})$ is the normalizing constant. The advantage of introducing the probabilistic loss is that the entropy of the estimated matrix Fisher distribution is a good measurement of prediction confidence, as we will discuss in 3.3. Thus, the loss of the inverse dynamics teacher is:

$$\mathcal{L}_{inv} = \|\mathbf{p} - \hat{\mathbf{p}}\|_2^2 - \log(p_{mf}(\mathbf{R}|\hat{\boldsymbol{\Phi}})) \tag{3}$$

where $\hat{\mathbf{p}}$ and $\hat{\boldsymbol{\Phi}}$ are the predicted pose and parameter of matrix Fisher distribution, respectively.

We train $f^{\text{inv}}_{\boldsymbol{\psi}_m}$ in a semi-supervised manner. During training, we maintain a small subset of data with ground-truth $\mathcal{G}$ (e.g., nuScenes-Boston [11]) and a large image set with pseudo-labels $\mathcal{S}$. The model iteratively trains on $\mathcal{G} \cup \mathcal{S}$, pseudo-labels incoming image collection, and selects pseudo-labeled samples into $\mathcal{S}$ using reservoir sampling.

Due to the challenging unconstrained and out-of-distribution (OOD) data in our settings such as arbitrary camera settings and unseen environments, we propose to learn an *ensemble* of models and aggregate predictions from each model to improve accuracy [3, 38, 39, 50, 71]. Ensembling models effectively estimates well-calibrated *epistemic uncertainty* [35], facilitating the selection of the most informative samples for inclusion in episodic memory in Sec 3.4.

Robust Ensemble of Inverse Dynamics Models: We select a learning-based end-to-end regression model as our inverse dynamics model with extensive data augmentation [37, 65] to generalize across various camera settings (model details can be found in supplementary). We incrementally train an ensemble of $M = 5$ inverse dynamics models each regressing ego-vehicle motion and integrated to compute:

$$(\hat{\mathbf{y}}_t, \hat{v}_t, \hat{c}_t) \;=\; \mathcal{T}(\{\hat{\mathbf{p}}_i\}_{i=t}^{t+h}) \tag{4}$$

where $\hat{\mathbf{p}}_i = \frac{1}{M}\sum_{m=1}^{M} f^{\text{inv}}_{\boldsymbol{\psi}_m}(\mathbf{I}_i, \mathbf{I}_{i+1})$ are the averaged estimated ego-motion from the ensemble models, and $\hat{\mathbf{y}}_t = \{\hat{\mathbf{y}}^k\}_{k=1}^{K}$, $\hat{v}_t$, and $\hat{c}_t$ are the future waypoints, speed, and command estimations. Details of transformation $\mathcal{T}$ in Eq. 4 and length of the poses sequence h can be found in supplementary. The parameters of the ensemble models are denoted as $\{\boldsymbol{\psi}_m\}_{m=1}^{M}$. To ensure model diversity, we find the simple different random initialization to work best, compared to more complex diversity-encouraging methods [51]. In addition, we store and employ pre-aggregated predictions as training targets for continually updating each model.

Incremental Adaptation via Self-Distillation: Despite ensembling models, we observe significant noise in the sequential ego-motion predictions, resulting in inaccurate estimation of waypoints, particularly on challenging but highly informative samples such as bad weather and abrupt turns. We alleviate this issue through temporal consistency-based re-labeling and adaptive filtering mechanisms. We incrementally self-train the ensemble models over incoming data [40, 69, 74]. Moreover, we compute the disagreement $u_t = \text{std}\left(\{\hat{\mathbf{y}}_t^m\}_{m=1}^{M}\right)$ as epistemic uncertainty among the ensemble models [3, 38]. Next, we select highly uncertain samples according to the estimated epistemic uncertainty and employ a temporal consistency-based re-labeling method to refine the pseudo labels. Additionally, an adaptive filtering mechanism is applied to eliminate high-noise samples in Sec. 3.3. We also use epistemic uncertainty to construct an informative episodic memory and reduce model forgetting in Sec. 3.4.

3.3. Motion Planning Policy Student

Given pseudo-labels $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ from the inverse dynamics model in Eq. 4, we incrementally update the student model $f_{\boldsymbol{\theta}}^{\text{plan}}$ over a collection of unlabeled data $\mathcal{U}$ and a replay buffer $\mathcal{B}$ via a behavior cloning objective using an L1 loss [15]. The loss functions on current video $\mathcal{U}$ and buffer $\mathcal{B}$ are defined as $\mathcal{L}_{\mathcal{U}} = \mathbb{E}_{(\hat{\mathbf{x}},\hat{\mathbf{y}})\in\mathcal{U}}[\mathcal{L}_1(\hat{\mathbf{y}}, f_{\boldsymbol{\theta}}^{\text{plan}}(\hat{\mathbf{x}}))]$ and $\mathcal{L}_{\mathcal{B}} = \mathbb{E}_{(\hat{\mathbf{x}},\hat{\mathbf{y}})\in\mathcal{B}}[\mathcal{L}_1(\hat{\mathbf{y}}, f_{\boldsymbol{\theta}}^{\text{plan}}(\hat{\mathbf{x}}))]$ respectively. We define the weighted optimization objective as

$$\mathcal{L}_{plan} = \frac{1}{n}\mathcal{L}_{\mathcal{U}} + (1 - \frac{1}{n})\mathcal{L}_{\mathcal{B}} \tag{5}$$

where n is the number of image collections the model has seen so far. We note that the model does not revisit $\mathcal{U}$ in subsequent iterations. We set up weight in $\mathcal{L}_{plan}$ based on the number of image collections the model has been trained. Intuitively, as the buffer grows more diverse and informative, we aim to shift the model's focus increasingly toward the buffer. We find this to improve knowledge retention.

Sample Re-Labeling: Since our waypoints pseudo-label $\hat{\mathbf{y}}_t$ is generated from the average prediction from the inverse dynamics teacher, it is possible that some waypoints trajectories $\hat{\mathbf{y}}_t = \{\mathbf{y}^k\}_{k=1}^K$ are less accurate, i.e., due to the noise in the ego-motion predictions or disagreement among the inverse dynamics models. We observe that incorporating temporal information can improve the pseudo-label quality of the waypoints. For highly uncertain samples, we propose to set a time window with size l and leverage those adjacent waypoints in the window to re-estimate the waypoint $\hat{\mathbf{y}}^k$. To detect samples for re-labeling, we use the two-component Gaussian Mixture model (GMM) to fit the uncertainties of all samples in the video. Given sample $\hat{\mathbf{x}}_t$ and its uncertainty u_t, its low uncertainty probability is the posterior probability $p_{gmm}(g|u_t)$ where g is the Gaussian component with a smaller mean, i.e., smaller uncertainty. We split $\mathcal{U}$ into a high uncertainty set $\mathcal{N}$ and a low uncertainty set $\mathcal{C}$:

$$\begin{aligned} \mathcal{C} &= \{(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) \in \mathcal{U} : p_{gmm}(g|u_t) \geq \epsilon_a\} \\ \mathcal{N} &= \{(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) \in \mathcal{U} : p_{gmm}(g|u_t) < \epsilon_a\} \end{aligned} \tag{6}$$

For the high uncertainty samples $(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) \in \mathcal{N}$, we propose to apply temporal consistency smoothing to further purify each waypoint $\hat{\mathbf{y}}^k \in \hat{\mathbf{y}}_t$:

$$\bar{\mathbf{y}}^k = \frac{\sum_{i=1}^{l/2} \mathbf{y}_{-i}^k + \sum_{i=1}^{l/2} \mathbf{y}_i^k}{l} \tag{7}$$

where we denote $\bar{\mathbf{y}}^k$ as the purified waypoint, $\mathbf{y}_{-i}^k$ and $\mathbf{y}_i^k$ as adjacent waypoints preceding and succeeding $\hat{\mathbf{y}}^k$.

Noisy Sample Filter Mechanism: Unlike existing curated datasets, YouTube videos contain different kinds of noise from numerous sources, such as severe light variation, distortion, edits, etc. Such noisy images do not provide useful information and can confuse the model during training. Noise can either originate from the image (e.g., image is deprecated and unrelated to the driving task) or from the waypoints pseudo-labels that deviate significantly from the actual trajectories. We propose an adaptive filtering mechanism that contains a confidence-based filter to remove samples with potentially deprecated images and a low-loss-based filter to remove samples with potentially deprecated pseudo labels. For the confidence-based filter, we compute the entropy of the matrix Fisher distribution of each image pair $(\mathbf{I}_{t-1}, \mathbf{I}_t) \in \mathcal{U}$ as: $h(\mathbf{I}_{t-1}, \mathbf{I}_t) = H(p_{mf}(R_t|f_{\boldsymbol{\psi}_m}^{\text{inv}}(\mathbf{I}_{t-1}, \mathbf{I}_t)))$. Then, we model the entropy distribution in $\mathcal{U}$ as a normal distribution with mean m and standard deviation σ. We extract and remove the noisy samples $\mathcal{D}$ based on the three-sigma rule:

$$\mathcal{D} = \{(\mathbf{I}_{t-1}, \mathbf{I}_t) \in \mathcal{U} : h(\mathbf{I}_{t-1}, \mathbf{I}_t) > m + 3 * \sigma\} \tag{8}$$

Although the confidence-based filter can effectively remove those potentially deprecated images, it is unable to detect the quality of waypoints pseudo labels. Thus, we introduce the low-loss-based filter. The low-loss trick is commonly used to detect noisy labels [6, 42]. Once the student model completes training on the incoming data $\mathcal{U}$, before selecting samples into the buffer, we apply the low-loss-based filter to detect and remove deprecated pseudo labels by computing the loss for each sample using the trained student model. Then we fit a GMM to the loss distribution over all samples and remove the set $\mathcal{D}$ of samples with deprecated pseudo labels:

$$\mathcal{D} = \{(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) \in \mathcal{U} : p_{gmm}(g|\mathcal{L}(\mathbf{y}_t, \hat{\mathbf{y}}_t)) < \epsilon_b\} \tag{9}$$

3.4. Uncertainty-Guided Episodic Memory

The central challenge in incremental learning to drive lies in minimizing model forgetting, i.e., retaining and developing knowledge while maintaining model generalization [20, 36, 64]. Our never-ending agent training process leverages episodic memory; a long-term memory buffer of previous experiences that is rehearsed during every model update.

High-Uncertainty Experience Exemplars: Well-calibrated epistemic uncertainty is widely employed to select informative samples in Active Learning [8, 21, 48]. We adopt the same idea in constructing the informative and diverse replay buffer. Given an experience replay buffer $\mathcal{B}$ with fixed size B and an incoming image collection $\mathcal{U}$, at every iteration we sort $\mathcal{B} \cup \mathcal{U}$ by uncertainty and select the *highest-uncertainty samples* to produce an updated buffer of maximum uncertainty. This process can be interpreted as a form of *self-supervised exploration* [53], where the

model selects the most challenging and novel experiences to retain. Due to our introduced temporal-consistency re-labeling and adaptive filtering mechanisms, we demonstrate that this uncertainty-seeking strategy results in a simple yet highly effective buffer construction method, automatically accounting for noise and informativeness. We find it to outperform a range of prior sampling methods, e.g., based on random or reservoir sampling.

Implementation Details: Our student model is trained using a batch size of 96 and a learning rate of 0.001. The teacher model is trained using a batch size of 16 and a learning rate of 0.0005. Throughout most of the experiments, $\mathcal{G}$ contains 108,971 images from nuScenes-Boston. The size of $\mathcal{S}$ is set as 500,000. $f^{\text{inv}}_{\psi_m}$ incrementally trains on ten million of YouTube images (extracted from videos at 10Hz) from various weather conditions and driving scenarios. The time window size in temporal consistency-based re-labeling is set as six. Each label contains $K = 5$ sequential ego-relative world coordinates, and the interval between each waypoint is 0.5 second. Additional implementation details can be found in our supplementary.

4. Experiments

We perform cross-dataset analysis of ∞-Driver to evaluate for diverse driving behaviors, generalization, and knowledge retention.

Datasets: To incrementally learn driving behaviors online, we create a YouTube dataset containing hundreds of hours of driving videos with ten million extracted frames. The YouTube dataset includes various weather conditions (sunny, rainy, snowy, foggy, night) and different driving scenarios from all over the world. We expect our ∞-Driver to incrementally learn and *remember* driving skills under different weather conditions and driving scenarios. Moreover, we use three standard datasets, KITTI [24], nuScenes [11], and Argoverse2 [67], for generalization experiments. KITTI has 11 driving sequences with ground-truth. nuScenes has 15 hours of driving data from four locations: Boston-Seaport, Singapore-OneNorth, Singapore-Queenstown, and Singapore-HollandVillage. nuScenes driving scenes includes many dynamic objects and various weather conditions. Argoverse2 contains 1,000 driving scenes from six different US cities: Austin, Detroit, Miami, Pittsburgh, Palo Alto, and Washington, DC.

Experimental Setup: We conduct experiments on *incremental learning across cities* to test the performance of our incremental learning framework. We continuously train on KITTI, nuScenes, and Argoverse2 which are sorted by cities. The sequence of cities is Karlsruhe, One-north, Queenstown, Holland Village, Pittsburgh, Austin, Detroit, Palo Alto, Miami, and Washington, DC. There are around 430,000 images over the ten cities. During the training of the inverse dynamics model, the model trains on nuScenes-Boston in a supervised manner and incrementally trains across ten cities in a semi-supervised manner. For each city's dataset, we split 90% as a training set and 10% as an evaluation set. We set buffer sizes as 3,000, 6,000, and 9,000 to study incremental learning performance under limited buffer sizes.

Moreover, we conduct experiments on *incremental learning on open web data* to test the generalization ability of our driving policy. We collect around 10 million images from driving videos. During the training of the inverse dynamics model, the model trains on nuScenes-Boston in a supervised manner and incrementally trains on the stream of image collections in a semi-supervised manner. Different from the experiment setup above, we are unable to generate an evaluation set from YouTube images due to the lack of ground-truth labels. We aim to construct a large and diverse evaluation set that has *never been seen* during training to effectively validate the generalization ability of our ∞-Driver. We propose an evaluation set by mixing KITTI, nuScenes, and Argoverse2. However, considering the redundancy issue in the large dataset, samples in tough and rare scenarios might be neglected if we evaluate each sample in the evaluation set and average the loss. Thus, during evaluation, we partition the evaluation set into three subsets according to the steering command. Within each subset, we further categorize the data based on velocity. Our model is tested on each of these subsets, and the sample loss within each subset is averaged. Subsequently, the final average loss is computed based on a balanced loss. In addition to extensive open-loop evaluation, we incorporate additional experiments in the supplementary, including closed-loop evaluation in CARLA [22].

Baselines: We compare our method with the three closest incremental learning baselines, *Dark Experience Replay* (DER) [10], *Rainbow Memory* (RM) [5], and PuriDivER [6]. DER proposes a general continual learning method with blurry task boundaries and domain shifts. RM proposes a memory management method based on uncertainty and data augmentation. PuriDivER studies continual learning on noisy data and proposes a relabeling method to purify the noisy samples.

Metrics: We evaluate our incremental learning framework on three widely used metrics in the literature [13]. However, since we are solving a regression problem, we make corresponding revisions to standard classification metrics.

Average Loss (L): L_k measures the average performance on each image collection after learning on the k-th image collection. We denote $l_{k,j}$ as the average displacement error (ADE) that is evaluated on the hold-out evaluation set of the j-th image collection where $j \leq k$ after the model trained on the k-th image collection. We define the average ADE

Table 1. **Evaluating Incremental Learning Over Cities.** $L_{-1}(\downarrow)$, $F_{-1}(\uparrow)$, $I_{-1}(\uparrow)$ denote Average Loss, Forgetting and Intransigence after the model incrementally trains on the last city. $\downarrow$ indicates metrics that are the lower the better, $\uparrow$ indicates the higher the better. B stands for buffer size. Subscript $(_{-1})$ denotes the metrics of the last image collection.

	$B = 3{,}000$			$B = 6{,}000$			$B = 9{,}000$		
Methods	L_{-1}	F_{-1}	I_{-1}	L_{-1}	F_{-1}	I_{-1}	L_{-1}	F_{-1}	I_{-1}
DER [10]	0.76	-0.14	-0.38	0.82	-0.19	-0.48	0.77	-0.12	-0.38
Rainbow [5]	1.29	-0.43	-0.75	0.87	-0.31	-0.38	0.75	-0.18	-0.36
PuriDivER [6]	0.85	-0.16	-0.41	0.84	-0.11	-0.59	0.72	-0.08	-0.34
∞-Driver (Ours)	**0.66**	**-0.07**	**-0.35**	**0.61**	**-0.04**	**-0.28**	**0.60**	**-0.03**	**-0.30**

Table 2. **Evaluating Incremental Learning on Open Web Data.** $\bar{L}_{-1}(\downarrow)$, $\bar{F}_{-1}(\uparrow)$, ADE_{-1} $(\downarrow)$ are the revised Average Loss, Forgetting and ADE on the evaluation set after the model continuously trains on 10 million YouTube images.

Methods	$\bar{L}_{-1}$	$\bar{F}_{-1}$	ADE_{-1}
DER [10]	1.24	**-0.02**	1.18
Rainbow [5]	1.33	-0.05	1.24
PuriDivER [6]	1.25	-0.04	1.18
∞-Driver (Ours)	**1.17**	**-0.02**	**1.13**

after learning on the k-th image collection as:

$$L_k = \frac{1}{k}\sum_{j=1}^{k} l_{k,j} \tag{10}$$

Forgetting (F): Forgetting a specific image collection is defined by the difference between the peak level of the image collection knowledge acquired during past learning and the present understanding that the model holds about the image collection. We define the forgetting of the j-th image collection after the model continuously learns k image collections as $f_j^k = \min_{l\in\{j,\ldots,k-1\}} l_{l,j} - l_{k,j}, \forall j < k$. The average forgetting of previous image collections after learning the k-th image collection is defined as

$$F_k = \frac{1}{k-1}\sum_{j=1}^{k-1} f_j^k \tag{11}$$

Intransigence (I): Intransigence measures the difficulty for the model to learn on new image collection as below:

$$I_k = l_k^* - l_{k,k} \tag{12}$$

where l_k^* is the loss if we performed supervised learning on all the samples the model has seen so far, i.e., with unlimited buffer size.

However, experiments of incremental learning on open web data do not provide a hold-out evaluation set with ground-truth for each image collection. Instead, we construct a single evaluation set by mixing KITTI, nuScenes, and Argoverse2. Thus, we provide revised Average Loss $\bar{L}_k$ and Forgetting $\bar{F}_k$ to measure the incremental learning performance. We denote $\bar{l}_k$ as the ADE that is measured on the evaluation set after the model is trained on the k-th image collection, where ADE is computed as described in the experimental setup. The Average Loss is the average ADE over image collections: $\bar{L}_k = \frac{1}{k}\sum_1^k \bar{l}_i$. Here, we define the forgetting after the model continuously learns k image collections as $\bar{F}_k = \min_{l\in\{1,\ldots,k-1\}} \bar{l}_l - \bar{l}_k, \forall j < k$.

4.1. Results

Incremental Learning Performance: Table 1 presents the results of incremental learning on image collections sorted by cities with varying buffer sizes: 3,000, 6,000, and 9,000. L_{-1}, F_{-1}, I_{-1} represents the evaluations after the model trains on the last image collection. Our ∞-Driver outperforms other baselines across all metrics and exhibits a significant advantage on the Average Loss and Forgetting measures compared to other baseline models, especially on buffer sizes 3,000 and 6,000. This indicates that our filter mechanism and temporal consistency re-labeling method can effectively remove and refine noisy samples. Additionally, it demonstrates that our uncertainty-based buffer sampling method can select more informative samples into the buffer. Thus, our buffer is enriched with more informative samples and less misleading labels. It enhances the model's ability to retain previously acquired knowledge without hindering the model from learning new incoming knowledge.

Generalization Performance: To further demonstrate the generalization ability of ∞-Driver, we conduct open web experiments that continuously train on the YouTube dataset. It is important to note that all samples within the evaluation set are *entirely unseen* to ∞-Driver before evaluation. Table 2 presents the revised Average Loss, Forgetting, and ADE measured on the evaluation set after the model completes the final training phase. ∞-Driver obtains the best result on Forgetting, which indicates that the model's final

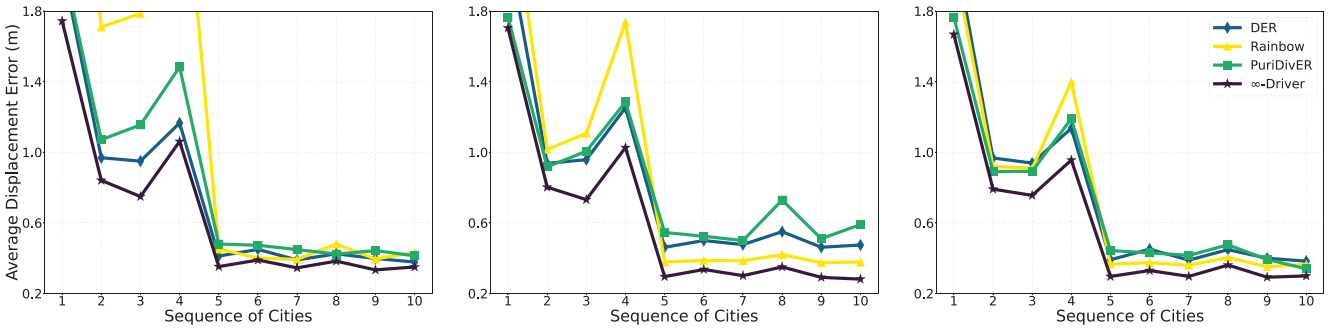

Figure 2. **Cross-City Continual Learning Experiment.** The figure depicts ADE loss of models trained with an incremental number of different cities, with buffer size 3,000, 6,000, and 9,000 (from left to right), with our proposed approach showing best performance even on harsh generalization conditions, e.g., KITTI and nuScenes (cities 1, 2, 3, and 4 in the figure).

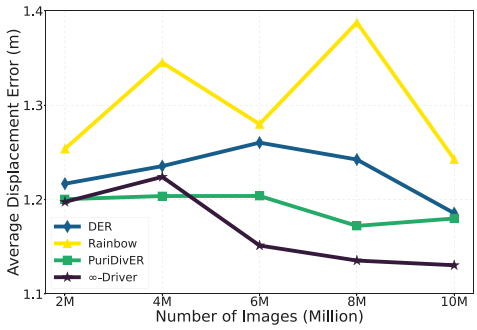

Figure 3. **Open Web Learning Experiment.** We plot ADE measured after the model has seen 2, 4, 6, 8, and 10 million images.

performance on the evaluation set essentially matches the model's best performance. The lowest Average Loss and ADE demonstrate that the model exhibits best generalization capabilities among the analyzed models.

Training Trends: Fig. 2 shows the ADE scores on each city's hold-out evaluation set. Performance is measured after the model has completed training on the last city's image collection. According to the figure, we demonstrate that after incrementally training over ten cities, our ∞-Driver consistently outperforms other baseline models in each city under all buffer settings. Moreover, Fig. 3 displays the trend of ADE score as ∞-Driver incrementally trained on YouTube data. We plot the trend by selecting the ADE scores measured after the model has seen 2, 4, 6, 8, and 10 million images. There is a significant gap between our method and other baselines. ADE curves from RM and DER fluctuate drastically over time. We believe this is because selecting samples with an interval in RM may avoid adding a few duplicate samples to the buffer, but does not ensure the sample diversity and potentially selects numerous non-informative (e.g., low epistemic uncertainty) samples into the buffer. The reservoir sampling strategy in DER ensures that each sample has an equal probability of being selected in the buffer. While simple, this approach may result in a less informative buffer. Conversely, our epistemic uncertainty-based buffer selection strategy effectively maintains an informative buffer and our temporal consistency re-labeling method and filtering mechanisms are capable of effectively purifying and eliminating noisy samples. Further ablations can be found in the supplementary.

5. Conclusion

We introduce a never-ending learning-to-drive framework that can leverage unlabeled data streams to learn a robust and generalized driving policy. ∞-Driver utilizes an inverse dynamics teacher model to provide pseudo-labels for the driving policy student to learn. Combined with our uncertainty-guided sampling model, relabeling, and filtering mechanisms, our incremental learning framework can successfully learn from continuous driving data from the web and generalize to different scenarios from unseen images. So far, ∞-Driver has been watching hundreds of videos and has yet to plateau. At the time of submission, it has visited more than 10,000 miles (estimated based on the inverse dynamics model). Our proposed approach is self-trained and does not require prohibitive supervision, e.g., labels for perception tasks. In the future, towards accessible, open-source, and large-scale navigation models, we plan to further analyze the role of ∞-Driver with larger model architectures and additional video data. In the process, we hope to uncover and address unique limitations in never-ending learning to drive at an extreme scale.

Acknowledgments: We thank the Red Hat Collaboratory (awards #2024-01-RH02, #2024-01-RH07) and NSF (IIS-2152077) for supporting this research.

References

[1] Carla autonomous driving leaderboard. *https://leaderboard.carla.org/*, 2022. 2

[2] California DMV removes cruise's driverless vehicle testing permits. https://www.government-fleet.com/10208569/cruise-updates-driverless-taxis-to-better-respond-to-emergency-vehicles, 2023. 1

[3] Taiga Abe, Estefany Kelly Buchanan, Geoff Pleiss, Richard Zemel, and John P Cunningham. Deep ensembles work, but are they necessary? *NeurIPS*, 2022. 4

[4] Bowen Baker, Ilge Akkaya, Peter Zhokov, Joost Huizinga, Jie Tang, Adrien Ecoffet, Brandon Houghton, Raul Sampedro, and Jeff Clune. Video pretraining (VPT): Learning to act by watching unlabeled online videos. *NeurIPS*, 2022. 2, 4

[5] Jihwan Bang, Heesu Kim, YoungJoon Yoo, Jung-Woo Ha, and Jonghyun Choi. Rainbow memory: Continual learning with a memory of diverse samples. In *CVPR*, 2021. 2, 3, 4, 6, 7

[6] Jihwan Bang, Hyunseo Koh, Seulki Park, Hwanjun Song, Jung-Woo Ha, and Jonghyun Choi. Online continual learning on a contaminated data stream with blurry task boundaries. In *CVPR*, 2022. 2, 3, 5, 6, 7

[7] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. ChauffeurNet: Learning to drive by imitating the best and synthesizing the worst. In *RSS*, 2019. 1, 2

[8] Lucas Berry and David Meger. Normalizing flow ensembles for rich aleatoric and epistemic uncertainty modeling. In *AAAI*, 2023. 5

[9] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. RT-2: Vision-language-action models transfer web knowledge to robotic control. *CoRL*, 2023. 2

[10] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *NeurIPS*, 2020. 3, 6, 7

[11] 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. 4, 6

[12] Matthew Chang, Arjun Gupta, and Saurabh Gupta. Semantic visual navigation by watching YouTube videos. In *NeurIPS*, 2020. 2, 4

[13] Arslan Chaudhry, Puneet K Dokania, Thalaiyasingam Ajanthan, and Philip HS Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *ECCV*, 2018. 6

[14] Dian Chen and Philipp Krähenbühl. Learning from all vehicles. In *CVPR*, 2022. 1, 2

[15] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *CoRL*, 2020. 2, 3, 5

[16] Zhiyuan Chen and Bing Liu. *Lifelong machine learning*. 2018. 1

[17] Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. *PAMI*, 2022. 3

[18] Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *ICRA*, 2018. 2, 3

[19] Daniel Dauner, Marcel Hallgarten, Andreas Geiger, and Kashyap Chitta. Parting with misconceptions about learning-based vehicle motion planning. *CoRL*, 2023. 1, 2

[20] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. Continual learning: A comparative study on how to defy forgetting in classification tasks. *arXiv preprint arXiv:1909.08383*, 2019. 2, 3, 4, 5

[21] Stefan Depeweg, Jose-Miguel Hernandez-Lobato, Finale Doshi-Velez, and Steffen Udluft. Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. In *ICML*, 2018. 5

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

[23] Insurance Institute for Highway Safety. Fatality facts 2021. https://www.iihs.org/topics/fatality-statistics/detail/older-people, 2023. 2

[24] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 6

[25] Priya Goyal, Dhruv Mahajan, Abhinav Gupta, and Ishan Misra. Scaling and benchmarking self-supervised visual representation learning. *arXiv preprint arXiv:1905.01235*, 2019. 2

[26] Hexiang Hu, Ozan Sener, Fei Sha, and Vladlen Koltun. Drinking from a firehose: Continual learning with web-scale natural language. *PAMI*, 2022. 2

[27] Peiyun Hu, Aaron Huang, John Dolan, David Held, and Deva Ramanan. Safe local motion planning with self-supervised freespace forecasting. In *CVPR*, 2021. 1, 3

[28] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *CVPR*, 2023. 2, 3

[29] INSIDEEVs. Tesla working on "final piece of the FSD AI puzzle," says elon musk. https://insideevs.com/news/680090/tesla-working-on-final-piece-fsd-ai-puzzle-musk/, 2023. 1

[30] Bernhard Jaeger, Kashyap Chitta, and Andreas Geiger. Hidden biases of end-to-end driving models. *ICCV*, 2023. 3

[31] Xiaosong Jia, Yulu Gao, Li Chen, Junchi Yan, Patrick Langechuan Liu, and Hongyang Li. Driveadapter: Breaking the coupling barrier of perception and planning in end-to-end autonomous driving. In *ICCV*, 2023. 1, 2, 3

[32] Xiaosong Jia, Penghao Wu, Li Chen, Jiangwei Xie, Conghui He, Junchi Yan, and Hongyang Li. Think twice before driving: Towards scalable decoders for end-to-end autonomous driving. In *CVPR*, 2023.

[33] Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Vad: Vectorized scene representation for efficient autonomous driving. *ICCV*, 2023. 1, 2, 3

[34] Nazmul Karim, Umar Khalid, Ashkan Esmaeili, and Nazanin Rahnavard. Cnll: A semi-supervised approach for continual noisy label learning. In *CVPR*, 2022. 2

[35] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *NeurIPS*, 2017. 4

[36] Chris Dongjoo Kim, Jinseo Jeong, Sangwoo Moon, and Gunhee Kim. Continual learning on noisy data streams via self-purified replay. In *ICCV*, 2021. 2, 3, 5

[37] Lei Lai, Zhongkai Shangguan, Jimuyang Zhang, and Eshed Ohn-Bar. XVO: Generalized visual odometry via cross-modal self-training. In *ICCV*, 2023. 4

[38] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *NeurIPS*, 2017. 4

[39] Olivier Laurent, Adrien Lafage, Enzo Tartaglione, Geoffrey Daniel, Jean-Marc Martinez, Andrei Bursuc, and Gianni Franchi. Packed-ensembles for efficient uncertainty estimation. *arXiv preprint arXiv:2210.09184*, 2022. 4

[40] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICMLW*, 2013. 4

[41] Alexander Cong Li, Ellis Langham Brown, Alexei A Efros, and Deepak Pathak. Internet explorer: Targeted representation learning on the open web. In *ICML*, 2023. 2

[42] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*, 2020. 5

[43] Shunkai Li, Xin Wang, Yingdian Cao, Fei Xue, Zike Yan, and Hongbin Zha. Self-supervised deep visual odometry with online adaptation. In *CVPR*, 2020. 2

[44] Zhiqiu Lin, Deva Ramanan, and Aayush Bansal. Streaming self-training via domain-agnostic unlabeled images. *arXiv preprint arXiv:2104.03309*, 2021. 1, 4

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

[46] M Jehanzeb Mirza, Marc Masana, Horst Possegger, and Horst Bischof. An efficient domain-incremental learning approach to drive in all weather conditions. In *CVPR*, 2022. 2

[47] Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Bishan Yang, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, et al. Never-ending learning. *Communications of the ACM*, 2018. 1

[48] Jishnu Mukhoti, Andreas Kirsch, Joost van Amersfoort, Philip HS Torr, and Yarin Gal. Deep deterministic uncertainty: A new simple baseline. In *CVPR*, 2023. 5

[49] NBC News. Self-driving uber car that hit and killed woman did not recognize that pedestrians jaywalk. https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281, 2018. 1

[50] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *NeurIPS*, 2019. 4

[51] Matteo Pagliardini, Martin Jaggi, François Fleuret, and Sai Praneeth Karimireddy. Agree to disagree: Diversity through disagreement for better transferability. *ICLR*, 2023. 4

[52] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 2019. 1, 2, 4

[53] Deepak Pathak, Dhiraj Gandhi, and Abhinav Gupta. Self-supervised exploration via disagreement. In *ICML*, 2019. 5

[54] Dean A Pomerleau. ALVINN: An autonomous land vehicle in a neural network. In *NeurIPS*, 1989. 1, 3

[55] Aditya Prakash, Kashyap Chitta, and Andreas Geiger. Multi-modal fusion transformer for end-to-end autonomous driving. In *CVPR*, 2021. 1, 2

[56] Nicholas Rhinehart, Rowan McAllister, and Sergey Levine. Deep imitative models for flexible inference, planning, and control. 2020. 3

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

[58] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. Asymmetric tri-training for unsupervised domain adaptation. In *ICML*, 2017. 2

[59] Kuniaki Saito, Donghyun Kim, and Kate Saenko. Openmatch: Open-set consistency regularization for semi-supervised learning with outliers. *arXiv preprint arXiv:2105.14148*, 2021. 2

[60] Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 2012. 4

[61] Hao Shao, Letian Wang, Ruobing Chen, Steven L. Waslander, Hongsheng Li, and Yu Liu. Reasonnet: End-to-end driving with temporal and global reasoning. In *CVPR*, 2023. 2

[62] Brian Tefft. Rates of motor vehicle crashes, injuries and deaths in relation to driver age, united states, 2014-2015. *AAA Foundation for Traffic Safety*, 2017. 2

[63] Sebastian Thrun. Lifelong learning algorithms. In *Learning to learn*. 1998. 1

[64] Gido M Van de Ven and Andreas S Tolias. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*, 2019. 2, 3, 5

[65] Wenshan Wang, Yaoyu Hu, and Sebastian Scherer. Tartanvo: A generalizable learning-based vo. In *CoRL*, 2021. 4

[66] Yuxi Wang, Junran Peng, and ZhaoXiang Zhang. Uncertainty-aware pseudo label refinery for domain adaptive semantic segmentation. In *ICCV*, 2021. 2

[67] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, et al. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *arXiv preprint arXiv:2301.00493*, 2023. 6

[68] Penghao Wu, Li Chen, Hongyang Li, Xiaosong Jia, Junchi Yan, and Yu Qiao. Policy pre-training for end-to-end autonomous driving via self-supervised geometric modeling. *ICLR*, 2023. 2, 4

[69] I Zeki Yalniz, Hervé Jégou, Kan Chen, Manohar Paluri, and Dhruv Mahajan. Billion-scale semi-supervised learning for image classification. *arXiv preprint arXiv:1905.00546*, 2019. 4

[70] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. ST3D: Self-training for unsupervised domain adaptation on 3d object detection. In *CVPR*, 2021. 2

[71] Jingkang Yang, Pengyun Wang, Dejian Zou, Zitang Zhou, Kunyuan Ding, Wenxuan Peng, Haoqi Wang, Guangyao Chen, Bo Li, Yiyou Sun, et al. OpenOOD: Benchmarking generalized out-of-distribution detection. *NeurIPS*, 2022. 4

[72] Yingda Yin, Yingcheng Cai, He Wang, and Baoquan Chen. Fishermatch: Semi-supervised rotation regression via entropy-based filtering. In *CVPR*, 2022. 4

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

[74] Jimuyang Zhang, Ruizhao Zhu, and Eshed Ohn-Bar. SelfD: self-learning large-scale driving policies from the web. In *CVPR*, 2022. 2, 4

[75] Jimuyang Zhang, Zanming Huang, and Eshed Ohn-Bar. Coaching a teachable student. In *CVPR*, 2023. 2

[76] Qihang Zhang, Zhenghao Peng, and Bolei Zhou. Learning to drive by watching youtube videos: Action-conditioned contrastive policy pretraining. In *ECCV*, 2022. 2, 4

[77] Xujiang Zhao, Killamsetty Krishnateja, Rishabh Iyer, and Feng Chen. Robust semi-supervised learning with out of distribution data. *arXiv preprint arXiv:2010.03658*, 2020. 2

[78] Ruizhao Zhu, Peng Huang, Eshed Ohn-Bar, and Venkatesh Saligrama. Learning to drive anywhere. *CoRL*, 2023. 2