

Physically Interpretable Probabilistic Domain Characterization

Anais Halin^{*1}, Sébastien Piérard^{*1}, Renaud Vandeghen¹,
Benoît Gérin², Maxime Zanella^{2,4}, Martin Colot³, Jan Held¹,
Anthony Cioppa¹, Emmanuel Jean⁵, Gianluca Bontempi³,
Saïd Mahmoudi⁴, Benoît Macq², and Marc Van Droogenbroeck¹

¹ Montefiore Institute, University of Liège (ULiège), Liège, Belgium

² Catholic University of Louvain (UCLouvain), Louvain-la-Neuve, Belgium

³ Université Libre de Bruxelles (ULB), Brussels, Belgium

⁴ University of Mons (UMons), Mons, Belgium

⁵ Multitel research & innovation centre, Mons, Belgium

Abstract. Characterizing domains is essential for models analyzing dynamic environments, as it allows them to adapt to evolving conditions or to hand the task over to backup systems when facing conditions outside their operational domain. Existing solutions typically characterize a domain by solving a regression or classification problem, which limits their applicability as they only provide a limited summarized description of the domain. In this paper, we present a novel approach to domain characterization by characterizing domains as probability distributions. Particularly, we develop a method to predict the likelihood of different weather conditions from images captured by vehicle-mounted cameras by estimating distributions of physical parameters using normalizing flows. To validate our proposed approach, we conduct experiments within the context of autonomous vehicles, focusing on predicting the distribution of weather parameters to characterize the operational domain. This domain is characterized by physical parameters (absolute characterization) and arbitrarily predefined domains (relative characterization). Finally, we evaluate whether a system can safely operate in a target domain by comparing it to multiple source domains where safety has already been established. This approach holds significant potential, as accurate weather prediction and effective domain adaptation are crucial for autonomous systems to adjust to dynamic environmental conditions.

Keywords: Domain Characterization · Distribution Prediction
· Normalizing Flows · Simulation-Based Inference · Domain Adaptation
· Weather · Autonomous Vehicles · Operational Design Domain

1 Introduction

Advances in computer vision allow widespread camera monitoring, but diverse weather conditions lead to visually different data, sometimes impacting the per-

* Equal contributions. {anais.halin,s.pierard}@uliege.be

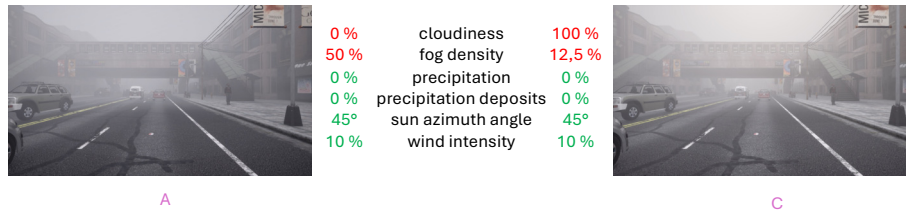


Fig. 1: The two synthetic images, A and C, generated using the *CARLA* software, appear almost identical, despite being acquired under very different weather conditions. This highlights the challenge of information loss from sensors like cameras when dealing with weather-related physical parameters. As a result, predictions that diverge from the ground truth in such ambiguous cases should not be penalized during evaluation.

formance of high-level tasks like object detection and surveillance. Current solutions often lack generalizability across multi-weather scenarios, highlighting the need for adaptive methods that can process visual data under diverse conditions.

More specifically, weather conditions significantly affect the perception capabilities of autonomous driving systems [34], particularly under harsh conditions, such as heavy rain or fog, which can compromise their ability to operate safely. Therefore, it is essential to develop reliable approaches to analyze the environment, irrespective of the weather conditions. Detecting critical circumstances such as extreme weather events allow the system to respond appropriately within and outside its *Operational Design Domain (ODD)*, defined by SAE International [42] as the “operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics”.

However, predicting exact weather conditions is challenging due to the many ambiguous cases. A single observation of the environment (*e.g.*, recorded by a vehicle-mounted camera) could be characterized by multiple sets of weather parameters. Fig. 1 presents two synthetic images, generated using the *CARLA* software [6], that appear nearly identical but were captured under different weather conditions. This shows the ambiguities inherent to characterizing the domain based on a single observation (*i.e.*, image).

Current methods predict weather conditions through regression or classification [17, 18, 26]. However, those approaches only produce a single crisp answer, lacking the insight about the ambiguities of characterizing the weather conditions. Fig. 2 illustrates how the two scenarios from Fig. 1, when processed with an intermediate set of values, result in very different visual representations.

In this work, we propose a novel solution based on a probabilistic characterization of the weather. Historically, statistical inference relying on likelihood estimation was intractable when dealing with high dimensional data [4]. Recently, one class of density estimation techniques based on neural networks called normalizing flows [20, 32, 40] has become quite popular to solve this kind of problem.

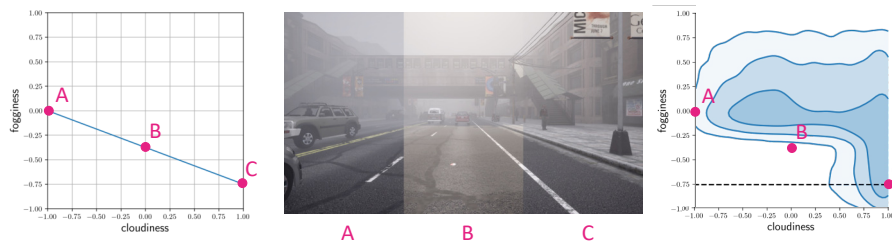


Fig. 2: Considering the images A and C from Fig. 1, generating an image B for the arithmetically averaged parameters, leads to an image very different from A and C. In other words, there are images i such that the probability $P(I = i, W = \hat{w}(i))$ of the pair (image, estimated weather) is zero when the estimated weather is the expected value of the weather knowing the image, $\hat{w}(i) = E[W|I = i]$, as done in regression. Working with distributions of weather parameters (as shown using a contour plot on the right-hand side) as proposed in this work, rather than predicting specific values for each parameter, *i.e.*, regression (as shown on the left-hand side), avoids this problem.

These models learn invertible transformations to go from complex distributions to more handy ones, and can therefore be used for modeling weather parameters from highly complex image and weather distributions.

These novel techniques allow us to express the domain of an image by a probabilistic distribution, *e.g.*, weather conditions, compared to current deterministic approaches. More specifically, we show on three consecutive tasks how various weather conditions can be predicted based (1) on a single color image acquired in front of a vehicle, (2) on a bag of color images (*absolute characterization*), and (3) how the current domain is related to arbitrarily chosen source domains (*relative characterization*).

Notions similar to *ODD* are very common in other fields where equivalent terms are used, such as *Operational Envelope* for maritime and *Operational Context* for railroad [46]. Additionally, other fields are facing domain shifts, such as in medical imagery from different acquisition devices [9]. We argue that our approach could serve a large range of practical applications. For example, determining the most suitable model for a given scenario within a fleet of lightweight AI models able to analyze the environment, as proposed in previous works [8, 30, 35], detecting significant domain transitions to collect new data for domain adaptation methods that rely on buffers or adaptable internal statistics [10, 15, 49–52] or activating adaptation mechanisms based on clustering [2, 45, 53, 54].

We summarize our contributions as follows: (1) We propose a novel probabilistic methodology to characterize domains in the case of autonomous vehicles driving in various weather conditions. (2) We demonstrate that simulation-based inference (normalizing flows) is adequate to obtain distribution for weather parameters that are used to characterize the domain and compare different backbones for features extraction. (3) Based on this weather domain characterization, we show how to characterize a new target domain as a mixture of source models.

2 Related Work

2.1 Predicting Weather Conditions from Images

Prediction of weather conditions from images was first formulated as a single-label classification task (*e.g.*, sunny, cloudy, or foggy). In 2014, Lu *et al.* [28] proposed a binary classification task between sunny and cloudy weathers, using features extracted from visual cues such as the sky, shadows, reflections, contrast, and haze. Later, Guerra *et al.* [48] proposed a multi-class dataset extending the scope of the classification task to rain, snow, and fog. Recent works focus on optimizing *Convolutional Neural Network (CNN)* to obtain strong features for common and uncommon weather conditions [27,55]. However, weather conditions can hardly be represented by crisp classification due to its continuous nature.

To reach a more realistic description of intricate relations between weather conditions, recent methods simultaneously regress several physical parameters [17, 18]. To increase interpretability, Li *et al.* [26] also assign cues of weather characteristics to each pixel. However, these methods still lack the ability to represent ambiguous scenarios (see Fig. 2). In this work, we predict the joint distribution of weather parameters by proposing a novel method based on normalizing flows.

2.2 Handling Weather Conditions for Autonomous Driving

Autonomous car driving systems need to be efficient under all weather conditions. Some methods propose to integrate a generalization step to the model to remove the environmental influences on the acquired images, by introducing a style layer inspired by images style-transfer neural networks [38] or through adversarial training [25]. Generalized features obtained from alignment of source and target domains tend to be suboptimal, as they do not consider the task. To improve the estimation, a solution is to add a domain adaptation step [24]. Jeon *et al.* [19] further improved their estimation by adding an unsupervised domain adaptation step after domain alignment.

Many studies have also highlighted the importance of a strong *ODD* definition to properly assess the ability of automatic driving systems to work in given conditions regarding weather, location, other vehicles on the road, state of the car sensors and many other environmental parameters [3,12,33]. Many strategies have been proposed to evaluate different situations according to specific evaluations of potential damage cost [23,44] and define the boundaries of the *ODD*. In this work, we characterize the domain by a probability distribution focusing on weather parameters in autonomous driving environment.

2.3 Probabilistic Modeling of Parameters Distributions

We propose to leverage recent observations in the *Simulation-Based Inference (SBI)* literature [4,47] for domain characterization. This literature has seen a rapid expansion thanks to new density estimation techniques in problems where likelihood estimation was often intractable, especially for high dimensional data.

One class of these density estimation techniques based on neural networks is normalizing flows [32]. The principle consists in transforming an arbitrarily chosen distribution (*e.g.*, a Gaussian) into the desired distribution. Different types exist, such as the *Neural Spline Flow (NSF)* type [7] that can be used in two ways: (1) to determine the value of the *Probability Density Function (PDF)* at a given point and (2) to draw samples at random. Different techniques can be used to learn them, *e.g.* the *Neural Posterior Estimation (NPE)* technique [11, 29].

There are a few techniques to analyze the performance of models predicting distributions. *Coverage Plots* [14] show, objectively and quantitatively, whether the distribution prediction models are underconfident (*i.e.*, conservative), calibrated, or overconfident. Another technique, specific for parameter distributions (*e.g.*, weather parameters) predicted from an observation (*e.g.*, an image) and widespread in the field of *SBI*, is known as *Posterior Predictive Check (PPC)* [39]. It consists in drawing parameters at random from a predicted distribution, injecting these parameters into a simulator or physical system and comparing the resulting observations with the one from which the distribution was predicted. In this work, we leverage those analysis techniques for assessing the quality of our weather characterization models.

3 The Three Fundamental Tasks Behind the Physically Interpretable Probabilistic Domain Characterization

Our experiments are organized around three different tasks, involving a prediction of the distribution of weather conditions given some images acquired by color cameras placed in front of vehicles. Before elaborating on these tasks in Sec. 3.2, 3.3, and 3.4, we briefly introduce our framework in Sec. 3.1.

3.1 Framework

Mathematical Modeling. We denote the set of all possible values for the physical parameters (*e.g.* weather conditions) by \mathbb{W} and the set of all observations of interest (*e.g.* images) by \mathbb{I} . We adopt the probability theory of Kolmogorov [21, 22] and consider a measurable space (Ω, Σ) as well as the (generalized) random variables $W : \Omega \rightarrow \mathbb{W}$ for the physical parameters and $I : \Omega \rightarrow \mathbb{I}$ for the observation. Following the mathematical modeling introduced in [35], we consider the set $\mathbb{D}_{(\Omega, \Sigma)}$ of domains d in which there is a probability measure P_d on (Ω, Σ) . We see the *ODD* of a given autonomous system as the set of domains in which it can be used safely, no matter if this has been established by design or by testing. Thus, $ODD \subseteq \mathbb{D}_{(\Omega, \Sigma)}$.

Data. All our experiments are performed on data generated using the *CARLA* software, an open-source simulator for autonomous driving research [6]. All images are acquired by a simulated camera placed in front of a vehicle called *ego vehicle*. The weather is controlled through 13 real-valued parameters (see Table 1). There are 6 parameters for which we make predictions. Thus, $\mathbb{W} \subseteq \mathbb{R}^6$.

Table 1: In *CARLA*, the weather is controlled through 13 physical parameters. This table shows the range of values that we consider in our experiments and indicates, for each of them, if they are considered in our predictions.

parameter	range	predicted
cloudiness	0 to 100%	yes
fog density	0 to 100%	yes
precipitation	0 to 100%	yes
sun azimuth angle	0° to 360°	no
sun altitude angle	-90° to 90°	yes
wind intensity	0 to 100%	yes
precipitation deposits	0 to 100%	yes

parameter	range	predicted
fog distance	fixed to 0.75	no
fog falloff	fixed to 0.1	no
mie scattering scale	fixed to 0.03	no
rayleigh scattering scale	fixed to 0.033	no
scattering intensity	fixed to 1.0	no
wetness	fixed to 0.0	no

3.2 Task I: Predicting Distributions of Physical Parameters

The first task (see Fig. 3) consists in predicting how likely various physical parameters are, jointly. As these parameters cannot, in general, be measured directly, it is necessary to estimate the likelihood, *i.e.* the distribution of plausible values, based on an indirect observation.

Case Study. We study here the particular case in which the physical parameters are relative to the weather conditions and the observations are color images $i \in \mathbb{I}$ acquired by cameras placed in front of vehicles. We aim at learning, offline, a deep learning model $i \mapsto \hat{P}_{d_R}(W|I=i)$, where d_R denotes a domain of reference in which the probability measure on (Ω, Σ) is P_{d_R} . This domain is arbitrarily chosen in such a way that P_{d_R} has a large support.

Data. We use *CARLA* to generate a dataset with 635k images and the corresponding ground-truth values for the weather parameters. The dataset (Fig. 4) is split into a learning set (*LS*) with 600k images (500k for the training set and 100k for the validation set) and a test set (*TS*) with 35k images. Letting the

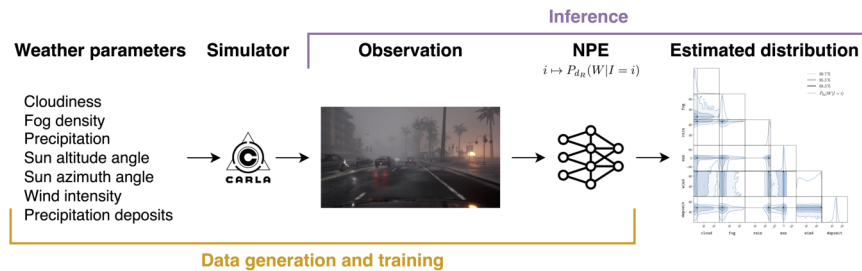


Fig. 3: Task I. The aim of the first task is, based on an image, to predict the joint distribution of the weather parameters. For this purpose, (1) we generate data using the *CARLA* software for uniformly distributed weather parameters (offline), (2) we train a *NPE* model using the learning set (*LS*) of our generated data (offline), and (3) we infer, given an image from the test set (*TS*), the estimated weather distribution (normalizing flow) and show the result on a corner plot.

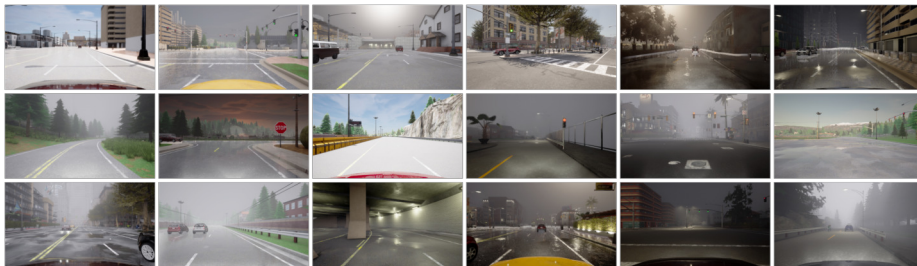


Fig. 4: Excerpt of the images in our dataset generated with the *CARLA* simulator. The ground-truth weather parameters are drawn at random for each image, following a uniform distribution with the bounds given in Table 1.

model of the ego vehicle, the map, the number of pedestrians, and the number of vehicles vary brings a touch of diversity in our data.

Method. We consider three different (frozen) backbones to extract features from the input images: *ResNet-50* [13], *DINOv2* [31], and *CLIP* [37]. We use the libraries *LAMPE* [41] and *ZUKO* [40] to learn a model (of type *NPE*) and to manipulate the normalizing flows (of type *NSF*), respectively. All predicted weather distributions are posteriors relative to the weather priors in the *LS*. Note that *LAMPE* and *ZUKO* were not developed for domain characterization, but rather for *SBI*. Also, to the best of our knowledge, these libraries have only been used once with very high dimensional input data [47].

Evaluation and Results. Five different analyses are carried out.

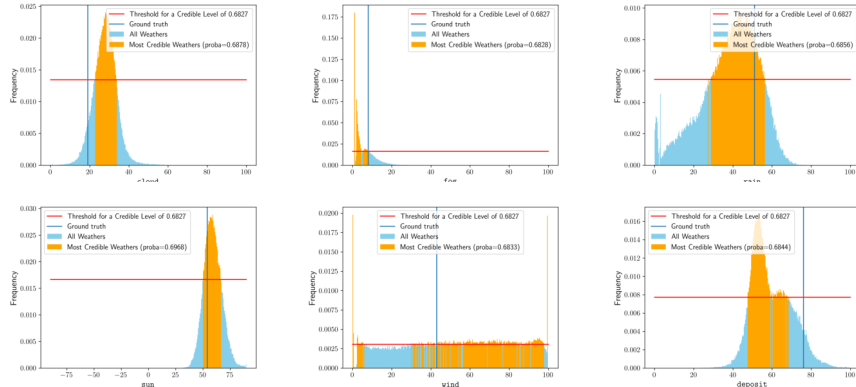
1st analysis: histograms (Fig. 5b). Due to their physical meaning, the weather parameters are easily interpretable. This paves the way to a first, subjective, evaluation. We draw samples at random out of the predicted weather distributions $\hat{P}_{dR}(W|I = i)$, for some images i arbitrarily chosen in TS and conduct a visual inspection of the histograms for the 6 marginals. The most credible weather parameters (*a.k.a.* highest density credibility sets, highest density regions, plausible sets, *etc.*) are highlighted for a credible level $l = 68.27\%$. For any input image i , these weathers are those such that the predicted *PDF* is above some threshold $t(i)$ and the predicted probability of the set is l [16].

2nd analysis: corner plots (Fig. 5c). A corner plot is a triangular array of plots. Those on the diagonal show the n marginals of a n -D distribution. Those below the diagonal depict the highest density credibility regions delimited at some arbitrarily chosen credibility levels (in this paper: 68.27%, 95.45%, and 99.73%), for each pair of parameters. We observe on the corner plots wide distributions, meaning that there is a large uncertainty for the weather parameters given an image. However, when analyzing the results, this uncertainty is explainable and

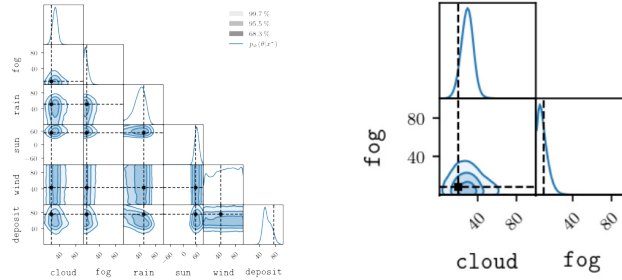


weather parameter	ground-truth value
cloudiness (cloud)	19
fog density (fog)	8
precipitation (rain)	51
sun azimuth angle	90
sun altitude angle (sun)	54
wind intensity (wind)	43
precipitation deposits (deposit)	76

(a) An input image i and the corresponding ground-truth values.



(b) Histograms for the 6 marginals of the predicted weather distribution for i .



(c) Corner plot, with the ground-truth weather pinned (left: complete, right: zoomed).



(d) Posterior predictive check.

Fig. 5: Task I: results obtained, with the model learned with the *ResNet-50* backbone on 500k learning samples, for an arbitrarily chosen input image in the *TS*.

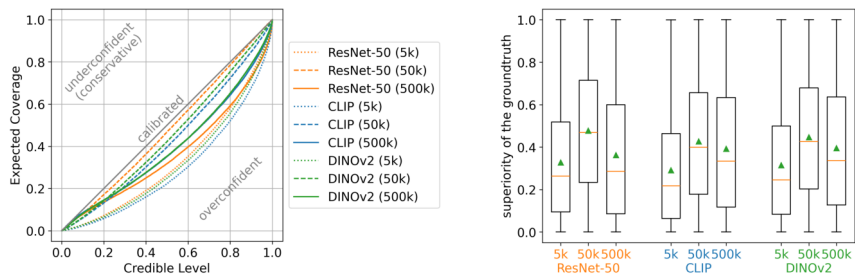


Fig. 6: Task I: comparison of our 9 models on the overall TS . Left: coverage plot. Right: box-and-whisker plots for the π statistic. Both analyses agree that the model learned with the *ResNet-50* backbone on 50k learning samples is preferable.

the following analysis shows that this uncertainty is not excessive, *i.e.*, our models are not underconfident.

3rd analysis: posterior predictive checks (Fig. 5d). Once a weather distribution is predicted for an observation (input image), it is possible to (1) draw weather parameters vectors at random from it and then (2) to inject these vectors in *CARLA*, keeping all other parameters unchanged and immobilizing the vehicles and pedestrians, to obtain new images in order to finally (3) compare those with the initial input image. One cannot expect to obtain identical images as, in *CARLA*, the traffic lights continue to run, the rain continues to fall, the plants move with the wind, and pedestrians’ poses are not perfectly frozen. Putting this aside, we observe that most retrieved images are very similar to the input image. We conclude that our models are not underconfident.

4th analysis: coverage plots (Fig. 6). Coverage plots show how the expected coverage varies with the credible level [14]. By definition, the expected coverage, at the credible level $l \in [0, 1]$, is the probability that the ground-truth weather belongs to the most credible weathers at level l . A method estimating the distribution of weathers based on an image is said underconfident (*i.e.*, conservative), calibrated, or overconfident at level l when the expected coverage at level l is, respectively, $> l$, $= l$, or $< l$. We observe that our models are all overconfident. The best calibrated model is the one that we obtained with 50k learning samples and *ResNet-50* as backbone.

5th analysis: superiority of the ground truth (Fig. 6). We also determine to what extent the ground truth is more credible than the other weathers. This is the proportion π of weathers that can be drawn at random from the normalizing flow and that are predicted as having a *PDF* value (*i.e.*, likelihood) lower or equal than the one from the ground truth. The higher π is, the better it is, but approaching 1.0 is notably very challenging. We estimated π for each image of the TS and made box-and-whisker plots for the 9 models. This analysis is complementary to the coverage plot in the sense that a model can be perfectly

calibrated while still presenting room for improvement. That being said, we are in the particular case in which this analysis leads to the same conclusion as the coverage plot: the model based on the backbone *ResNet-50* and learned from 50k samples is preferable to the others.

3.3 Task II: Absolute Domain Characterization

The second task (see Fig. 7) consists in characterizing a domain in an easy-to-interpret way. For this, we opt for a distribution of physical parameters estimated based on a bag (*a.k.a.* multiset) of observations.

Case Study. We characterize a domain $d \in \mathbb{D}_{(\Omega, \Sigma)}$ by the estimated distribution $\hat{P}_d(W)$ of weather conditions based on a real-valued bag b (multiset) of arbitrarily weighted images acquired by cameras placed in front of vehicles. In the following, we denote the weight (multiplicity) of the image $i \in b$ by $\omega(i)$. The images can either originate from a unique vehicle or from several, in case of vehicle-to-vehicle (*V2V*) communications. We want to establish $b \mapsto \hat{P}_d(W|B = b)$.

Data. We consider a bag b of 1,000 equally weighted images generated in maps already used in Task I. The weather parameters have been drawn at random from an arbitrary distribution $P_d(W)$ for which we fixed all parameters but two, fog density and precipitation, that follow a mixture of Gaussians.

Method. We use the *NPE* model developed for task I and build our solution for task II on top of it. We implement the following estimator for $P_d(W)$: $\hat{P}_d(W) = \sum_{i \in b} \omega(i) \hat{P}_{d_R}(W|I = i)$. The motivation for this estimator is threefold. (1) It

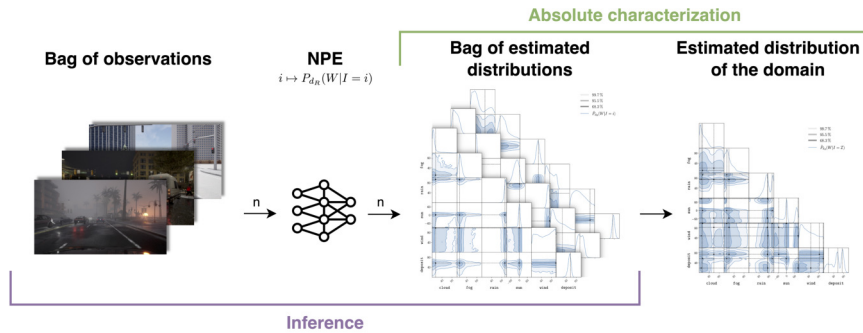


Fig. 7: Task II. The aim of the second task is to obtain an absolute characterization of a domain of interest. We process with the *NPE* each observation of a bag of observations of the domain of interest to obtain the bag of estimated distributions corresponding to the observations, then we obtain the distribution of weather parameters for the domain of interest by averaging the individual distributions of the bag.

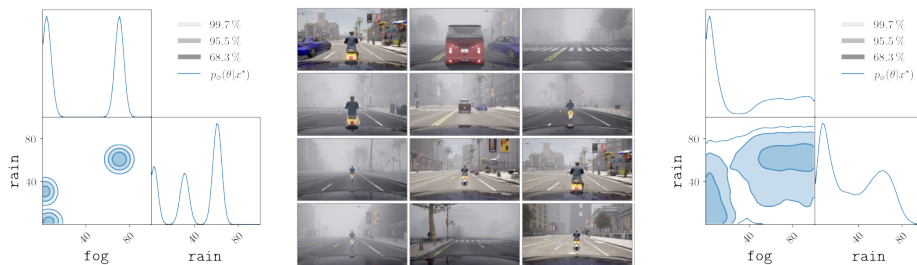


Fig. 8: Task II: example of result. Left: corner plot showing an arbitrarily chosen distribution of weather conditions, $P_d(W)$ (ground truth). Center: a bag b of images generated using *CARLA* with weather conditions drawn from $P_d(W)$. Right: corner plot showing the estimated likelihood of the weather conditions $\hat{P}_d(W)$ based on b .

is straightforward to evaluate the probability density function of $\hat{P}_d(W)$ and to draw weather conditions $w \in \mathbb{W}$ at random, following $\hat{P}_d(W)$, as we can do it with the normalizing flow $P_{d_R}(W|I = i)$. This will be valuable in our third task. (2) This estimator is useful for linear temporal filtering, *e.g.*, when one wants to weight more the recently acquired images than the old ones. (3) Finally, this estimator is fully justifiable under the assumptions that $P_d(W|I) = P_{d_R}(W|I)$ and $\omega(i) = P_d(I = i)$ as, in this case, $P_d(W) = \int_i \omega(i) P_{d_R}(W|I = i) di$.

Evaluation and Results. Our result is shown in Fig. 8. We observe that the three modes of the ground-truth weather distribution $P_d(W)$ are within the highest density regions of the prediction $\hat{P}_d(W)$. We made the same observation for many other ground-truth distributions (results omitted due to the limited space). We stress the fact that $\hat{P}_d(W)$ differs significantly from $P_d(W)$ can be explained by the inherent loss of information resulting from the use of color cameras, as already discussed in the introduction (*cf.* Fig. 1).

3.4 Task III: Relative Domain Characterization

The third task (see Fig. 9) consists in characterizing any target domain $d_T \in \mathbb{D}_{(\Omega, \Sigma)}$, relatively, w.r.t. to some arbitrarily chosen source domains $d_{S_1}, d_{S_2}, \dots, d_{S_{n_S}} \in \mathbb{D}_{(\Omega, \Sigma)}$. Our motivation for this task originates from the importance of knowing if a system implementing a given many-to-one domain adaptation method can operate safely in the target domain d_T when it is known to operate safely in the source domains $d_{S_1}, d_{S_2}, \dots, d_{S_{n_S}}$.

Case Study. We discuss the case of *Mixture Domain Adaptation (MDA)* in which the system adapts to any target domain for which $P_{d_T} = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}$, with $\sum_{k=1}^{n_S} \lambda_k = 1$ and $\lambda_k \geq 0 \forall k$ (this is the *mixture assumption*). In this framework, the *ODD* is the convex hull of $\{P_{d_{S_1}}, P_{d_{S_2}}, \dots, P_{d_{S_{n_S}}}\}$. An autonomous car implementing *MDA* is expected to drive safely when $P_{d_T} \in \text{ODD}$. So, our goal

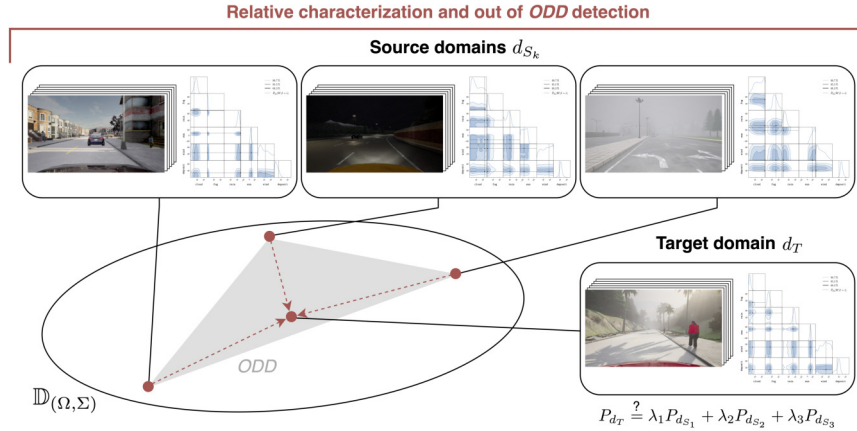


Fig. 9: Task III. The aim of the third task is twofold: (1) having a relative characterization of a target domain d_T based on source domains d_{S_k} and (2) detecting when this target domain d_T is out of the *ODD*.

in this third task is to determine if the mixture assumption holds and if so, what the values of the mixture weights λ_k are.

MDA supports many potential applications in the high-level tasks of the *Sense* pillar of the *Sense-Plan-Act* model [43]. While Mansour studied it generically [30], the application to the two-class classification task has been studied in [36], and the application to the semantic segmentation task of images acquired by vehicle-mounted cameras has been studied in [35]. These latter two works put an important emphasis on on-the-fly applicability and provide mathematically proven exact solutions. However, a critical limitation of these works is the need for the mixture weights to be known at adaptation time. Here, we remove this limitation by introducing a method that determines these weights automatically.

Data. We consider 4 subsets of the *TS* that we created for Task I: b_0 , b_1 , b_2 , and b_3 , containing 176, 27, 21, and 26 images, respectively. In the bag b_k , the distribution of the ground-truth weather parameters follows a uniform distribution on $\mathbb{W}_k \subsetneq \mathbb{W}$. These sets are such that $\mathbb{W}_0 \cap (\mathbb{W}_1 \cup \mathbb{W}_2 \cup \mathbb{W}_3) = \emptyset$.

Method. We characterize the source and target domains with the method presented for Task II, using the same weather distribution predictive model for all domains. We define the mean squared gap between the target domain and the mixture of the source domains as:

$$\begin{aligned} \delta(\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S}) &= \int_{\mathbb{W}} \left[\hat{P}_{d_T}(W = w) - \sum_{k=1}^{n_S} \hat{\lambda}_k \hat{P}_{d_{S_k}}(W = w) \right]^2 \hat{P}_{d_T}(W = w) dw \\ &\simeq \frac{1}{n_W} \sum_i^{n_W} \left[\hat{P}_{d_T}(W = w_i) - \sum_{k=1}^{n_S} \hat{\lambda}_k \hat{P}_{d_{S_k}}(W = w_i) \right]^2 \quad (1) \end{aligned}$$

with $\{w_i\}_{i=1}^{n_W} \sim \hat{P}_{d_T}(W)$. We aim at finding the values of $\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S}$ that minimize $\delta(\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S})$. This is a constrained least squares problem in which the constraints are $\sum_{k=1}^{n_S} \hat{\lambda}_k = 1$ and $\hat{\lambda}_k \geq 0 \forall k$. In order to use the *CVXPY* library [1,5] to solve our problem, we converted our original problem into a convex quadratic programming problem, with the same constraints.

Evaluation and Results. The goal of this experiment is twofold.

1. To show that it is possible to recover the mixture weights needed for the *MDA* technique presented in [35]. If the probability measure in the target domain is a mixture of the probability measures in the source domains, then we expect our characterization of the target domain to be a mixture of our characterizations for the source domains, with the same weights, as $P_{d_T} = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}} \Rightarrow P_{d_T}(W) = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}(W)$.
2. To show that it is possible to detect when the target domain is not a mixture of the source domains, which means that the target domain is out of the *ODD* for the *MDA* technique presented in [35]. If our characterization of the target domain is not a mixture of our characterizations for the source domains, then we expect that the probability measure in the target domain is not a mixture of the probability measures in the source domains, as we have $P_{d_T}(W) \neq \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}(W) \Rightarrow P_{d_T} \neq \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}$.

To achieve this goal, we perform experiments in which we characterize target domains relatively to $n_S = 3$ source domains. However, to study the robustness of our method, we let the probability measures in the target domains d_T be mixtures not only of the probability measures in the source domains, but also in some domains that are out of the *ODD*. We use the bags $b_1, b_2,$ and b_3 for the source domains $d_{S_1}, d_{S_2},$ and d_{S_3} , respectively. We also use the bag b_0 for domains $d_{\notin ODD}$ that are, by construction, guaranteed to be out of the *ODD*. Putting this into equations, we have $P_{d_T} = (1 - \eta)P_{d_{\in ODD}} + \eta P_{d_{\notin ODD}}$ with $P_{d_{\in ODD}} = \lambda_1 P_{d_{S_1}} + \lambda_2 P_{d_{S_2}} + \lambda_3 P_{d_{S_3}}$. The quantity $\eta \in [0, 1]$ is interpreted as a proportion of noise, which is the keystone to assess the robustness of our method.

In each experiment, the images in $b_1, b_2,$ and b_3 are equally weighted, whereas those in b_0 are randomly weighted following a uniform distribution. The method introduced in Task II for the absolute characterization of domains is applied for $d_{S_1}, d_{S_2}, d_{S_3},$ and d_T . We arbitrarily chose $(\lambda_1, \lambda_2, \lambda_3) = (0.2, 0.3, 0.5)$. The mixture weights are optimized on $n_W = 16$ points.

The overall experiment consists in (1) choosing a target domain at random, as described here-above, (2) computing its absolute characterization with the method of Task II, (3) executing the weight estimation algorithm, and (4) reporting both the mean square gap $\delta(\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S})$ and the Euclidean distance d_E between $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ and $(\lambda_1, \lambda_2, \lambda_3)$. This experiment has been performed 330 times (30 times for 11 values of η). The results are shown in Fig. 10. The mean square gap achievable by chance is shown (baseline). Note that the target domain belongs to the *ODD* only when $\eta = 0$. We observe that d_E is negligible when $d_T \in ODD$ and that the mean square gap is effective to detect when $d_T \notin ODD$.

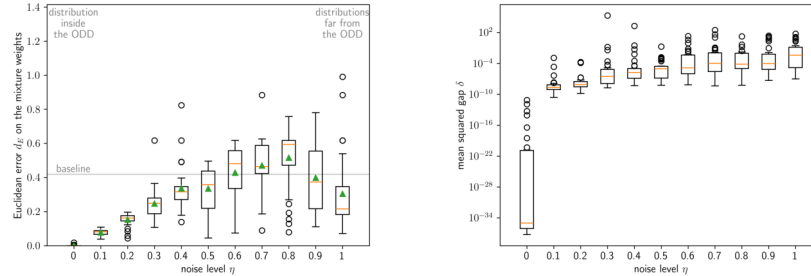


Fig. 10: Task III: results obtained when the noise level η sweeps the $[0, 1]$ interval. Left: box-and-whisker plots showing the distributions of d_E . Right: box-and-whisker plots showing the distributions of $\delta(\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S})$. These plots show (1) that we can recover the relationship between the target domain and the source domains, under the mixture assumption, when the target domain belongs to the *ODD* and (2) that the $\delta(\hat{\lambda}_1, \dots, \hat{\lambda}_{n_S})$ can be used to determine when the target domain belongs to the *ODD*.

4 Conclusion

In this work, we present a novel approach to characterize domains by estimating the distribution of physical parameters. Our method enhances interpretability, facilitates domain adaptation, and provides safeguards for systems operating outside their *ODD*. Our experiments, organized around three tasks, are performed in the particular, but very important, case of autonomous vehicles and demonstrate how to obtain an absolute characterization of a domain by predicting a distribution of weather parameters (1) using a single image acquired by a vehicle-mounted camera and (2) using a bag of images. Our experiments also demonstrate (3) how to obtain a relative characterization of a target domain based on arbitrarily chosen source domains. Our approach includes two types of domain characterization: absolute and relative. The relative characterization is particularly valuable for domain adaptation, allowing the expression of a target domain in terms of source domains and verifying whether the current domain is part of the *ODD*. This is important for autonomous driving systems as well as for other fields requiring interpretable and trustworthy domain adaptation.

Acknowledgments. This work has been made possible thanks to the *TRAIL* initiative (<https://trail.ac>). Part of it was supported by the Walloon region (Service Public de Wallonie Recherche, Belgium) under grant n°2010235 (ARIAC by DIGITALWALLONIA4.AI). J. Held and A. Cioppa are funded by the F.R.S.-FNRS (<https://www.frs-fnrs.be/en/>).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agrawal, A., Verschueren, R., Diamond, S., Boyd, S.: A rewriting system for convex optimization problems. *Journal of Control and Decision* **5**(1), 42–60 (January 2018). <https://doi.org/10.1080/23307706.2017.1397554>
2. Boudiaf, M., Mueller, R., Ayed, I.B., Bertinetto, L.: Parameter-free online test-time adaptation. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 8334–8343. Institute of Electrical and Electronics Engineers (IEEE), New Orleans, Louisiana, USA (June 2022). <https://doi.org/10.1109/CVPR52688.2022.00816>
3. Colwell, I., Phan, B., Saleem, S., Salay, R., Czarnecki, K.: An automated vehicle safety concept based on runtime restriction of the operational design domain. *2018 IEEE Intelligent Vehicles Symposium (IV)* (June 2018). <https://doi.org/10.1109/IVS.2018.8500530>
4. Cranmer, K., Brehmer, J., Louppe, G.: The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences (PNAS)* **117**(48), 30055–30062 (May 2020). <https://doi.org/10.1073/pnas.1912789117>
5. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* **17**(83), 1–5 (2016)
6. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: *Annual Conference on Robot Learning. Proceedings of Machine Learning Research*, vol. 78, pp. 1–16. Mountain View, California, USA (November 2017), <https://proceedings.mlr.press/v78/dosovitskiy17a.html>
7. Durkan, C., Bekasov, A., Murray, I., Papamakarios, G.: Neural spline flows. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 32 (2019)
8. Gérin, B., Halin, A., Cioppa, A., Henry, M., Ghanem, B., Macq, B., De Vleeschouwer, C., Van Droogenbroeck, M.: Multi-stream cellular test-time adaptation of real-time models evolving in dynamic environments. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. vol. 33, pp. 4472–4482. Institute of Electrical and Electronics Engineers (IEEE), Seattle, Washington, USA (June 2024). <https://doi.org/10.1109/CVPRW63382.2024.00450>
9. Gérin, B., Zanella, M., Wynen, M., Mahmoudi, S., Macq, B., De Vleeschouwer, C.: Exploring viability of test-time training: Application to 3D segmentation in multiple sclerosis. In: *IEEE Conference on Artificial Intelligence (CAI)*. vol. 34, pp. 557–562. Institute of Electrical and Electronics Engineers (IEEE), Singapore, Singapore (June 2024). <https://doi.org/10.1109/CAI59869.2024.00110>
10. Gong, T., Jeong, J., Kim, T., Kim, Y., Shin, J., Lee, S.J.: NOTE: Robust continual test-time adaptation against temporal correlation. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 35, pp. 27253–27266. Curran Associates, Inc. (2022), <https://openreview.net/forum?id=E9HNxrCFZPV>
11. Greenberg, D., Nonnenmacher, M., Macke, J.H.: Automatic posterior transformation for likelihood-free inference. In: *International Conference on Machine Learning (ICML)*. *Proceedings of Machine Learning Research*, vol. 97, pp. 2404–2414 (June 2019), <https://proceedings.mlr.press/v97/greenberg19a.html>
12. Gyllenhammar, M., Johansson, R., Warg, F., Chen, D., Heyn, H.M., Sanfridson, M., Söderberg, J., Thorsén, A., Ursing, S.: Towards an operational design domain that supports the safety argumentation of an automated driving system. In: *European congress on embedded real time systems (ERTS)*. pp. 1–10. Toulouse, France (January 2020)

13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. Las Vegas, Nevada, USA (June 2016). <https://doi.org/10.1109/CVPR.2016.90>
14. Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., Louppe, G.: A trust crisis in simulation-based inference? your posterior approximations can be unfaithful. arXiv **abs/2110.06581** (2021). <https://doi.org/10.48550/arXiv.2110.06581>
15. Houyon, J., Cioppa, A., Ghunaim, Y., Alfarra, M., Halin, A., Henry, M., Ghanem, B., Van Droogenbroeck, M.: Online distillation with continual learning for cyclic domain shifts. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). vol. abs 2211 16234, pp. 2437–2446. Institute of Electrical and Electronics Engineers (IEEE), Vancouver, Canada (June 2023). <https://doi.org/10.1109/CVPRW59228.2023.00242>
16. Hyndman, R.J.: Computing and graphing highest density regions. *The American Statistician* **50**(2), 120–126 (May 1996). <https://doi.org/10.1080/00031305.1996.10474359>
17. Ibrahim, M., Haworth, J., Cheng, T.: WeatherNet: Recognising weather and visual conditions from street-level images using deep residual learning. *ISPRS International Journal of Geo-Information* **8**(12), 1–18 (November 2019). <https://doi.org/10.3390/ijgi8120549>
18. Introvigne, M., Ramazzina, A., Walz, S., Scheuble, D., Bijelic, M.: Real-time environment condition classification for autonomous vehicles. In: IEEE Intelligent Vehicles Symposium (IV). Institute of Electrical and Electronics Engineers (IEEE), Jeju Island, Republic of Korea (June 2024). <https://doi.org/10.1109/iv55156.2024.10588692>
19. Jeon, M., Seo, J., Min, J.: DA-RAW: Domain adaptive object detection for real-world adverse weather conditions. In: IEEE International Conference on Robotics and Automation (ICRA). vol. 17, pp. 2013–2020. Institute of Electrical and Electronics Engineers (IEEE), Yokohama, Japan (May 2024). <https://doi.org/10.1109/ICRA57147.2024.10611219>
20. Kobyzev, I., Prince, S.J., Brubaker, M.A.: Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**(11), 3964–3979 (November 2021). <https://doi.org/10.1109/TPAMI.2020.2992934>
21. Kolmogorov, A.N.: Grundbegriffe der wahrscheinlichkeitsrechnung. In: Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik und Ihrer Grenzgebiete, vol. 2, p. 62 pages. Springer Berlin Heidelberg (1933). <https://doi.org/10.1007/978-3-642-49888-6>
22. Kolmogorov, A.N.: Foundations of the theory of probability. Chelsea Publishing Company (1950), <https://archive.org/details/foundationsofthe00kolm>
23. Lee, C.W., Nayer, N., Garcia, D.E., Agrawal, A., Liu, B.: Identifying the operational design domain for an automated driving system through assessed risk. In: IEEE Intelligent Vehicles Symposium (IV). Institute of Electrical and Electronics Engineers (IEEE) (October 2020). <https://doi.org/10.1109/IV47402.2020.9304552>
24. Lee, S., Kim, D., Kim, N., Jeong, S.G.: Drop to adapt: Learning discriminative features for unsupervised domain adaptation. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 91–100. Institute of Electrical and Electronics Engineers (IEEE) (October 2019). <https://doi.org/10.1109/ICCV.2019.00018>

25. Li, J., Xu, R., Ma, J., Zou, Q., Ma, J., Yu, H.: Domain adaptive object detection for autonomous driving under foggy weather. In: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 612–622. Institute of Electrical and Electronics Engineers (IEEE), Waikoloa, Hawaii, USA (January 2023). <https://doi.org/10.1109/WACV56688.2023.00068>
26. Li, X., Wang, Z., Lu, X.: A multi-task framework for weather recognition. In: ACM International Conference on Multimedia. pp. 1318–1326. Mountain View, California, USA (2017). <https://doi.org/10.1145/3123266.3123382>
27. Li, Z., Li, Y., Zhong, J., Chen, Y.: Multi-class weather classification based on multi-feature weighted fusion method. In: Earth and Environmental Science. Journal of Physics: Conference Series, vol. 558, pp. 1–13. IOP Publishing (August 2020). <https://doi.org/10.1088/1755-1315/558/4/042038>
28. Lu, C., Lin, D., Jia, J., Tang, C.K.: Two-class weather classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2510–2524 (December 2017). <https://doi.org/10.1109/TPAMI.2016.2640295>
29. Lueckmann, J.M., Goncalves, P.J., Bassetto, G., Öcal, K., Nonnenmacher, M., Macke, J.H.: Flexible statistical inference for mechanistic models of neural dynamics. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30, pp. 1–11. Curran Associates, Inc., Long Beach, California, USA (November 2017)
30. Mansour, Y., Mohri, M., Rostamizadeh, R.: Domain adaptation with multiple sources. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 21, pp. 1041–1048. Vancouver, Canada (December 2008), https://papers.nips.cc/paper_files/paper/2008/hash/0e65972dce68dad4d52d063967f0a705-Abstract.html
31. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: DINOv2: Learning robust visual features without supervision. arXiv **abs/2304.07193** (2023). <https://doi.org/10.48550/arXiv.2304.07193>
32. Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B.: Normalizing flows for probabilistic modeling and inference. Journal of Machine Learning Research **22**(57), 1–64 (2021), <http://jmlr.org/papers/v22/19-1028.html>
33. Pappalardo, G., Caponetto, R., Varrica, R., Cafiso, S.: Assessing the operational design domain of lane support system for automated vehicles in different weather and road conditions. Journal of Traffic and Transportation Engineering (English Edition) **9**(4), 631–644 (August 2022). <https://doi.org/10.1016/j.jtte.2021.12.002>
34. Perrels, A., Votsis, A., Nurmi, V., Pilli-Sihvola, K.: Weather conditions, weather information and car crashes. ISPRS International Journal of Geo-Information **4**(4), 1–23 (November 2015). <https://doi.org/10.3390/ijgi4042681>
35. Piérard, S., Cioppa, A., Halin, A., Vandeghen, R., Zanella, M., Macq, B., Mahmoudi, S., Van Droogenbroeck, M.: Mixture domain adaptation to improve semantic segmentation in real-world surveillance. In: IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). pp. 22–31. Institute of Electrical and Electronics Engineers (IEEE), Waikoloa, Hawaii, USA (January 2023). <https://doi.org/10.1109/WACVW58289.2023.00007>
36. Piérard, S., Marcos Alvarez, A., Lejeune, A., Van Droogenbroeck, M.: On-the-fly domain adaptation of binary classifiers. In: Belgian-Dutch Conference on Machine Learning (BENELEARN). pp. 20–28. Brussels, Belgium (June 2014)

37. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 139, pp. 8748–8763. Proceedings of Machine Learning Research (July 2021), <https://proceedings.mlr.press/v139/radford21a.html>
38. Rebut, J., Bursuc, A., Perez, P.: StyleLess layer: Improving robustness for real-world driving. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 8992–8999. Institute of Electrical and Electronics Engineers (IEEE), Prague, Czech Republic (September 2021). <https://doi.org/10.1109/IROS51168.2021.9636204>
39. Robertson, C., Long, J.A., Nathoo, F.S., Nelson, T.A., Plouffe, C.C.F.: Assessing quality of spatial models using the structural similarity index and posterior predictive checks. *Geographical Analysis* **46**(1), 53–74 (January 2014). <https://doi.org/10.1111/gean.12028>
40. Rozet, F., Divo, F., Schnake, S.: Zuko: Normalizing flows in PyTorch. Software (2022). <https://doi.org/10.5281/ZENODO.7625672>, <https://pypi.org/project/zuko>
41. Rozet, F., Miller, B.K., Delaunoy, A.: LAMPE: Likelihood-free amortized posterior estimation. Software (2021). <https://doi.org/10.5281/ZENODO.8405782>, <https://pypi.org/project/lampe>
42. SAE International: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Tech. Rep. SAE Standard J3016 202104, Society of Automobile Engineers, Warrendale, PA, USA (April 2021), https://doi.org/10.4271/J3016_202104
43. Shoker, A., Yasmin, R., Esteves-Verissimo, P.: Savvy: Trustworthy autonomous vehicles architecture. In: Symposium on Vehicles Security and Privacy (VehicleSec). pp. 1–7. San Diego, California, USA (February 2024), <https://www.ndss-symposium.org/ndss-paper/auto-draft-465/>
44. Sun, C., Deng, Z., Chu, W., Li, S., Cao, D.: Acclimatizing the operational design domain for autonomous driving systems. *IEEE Intelligent Transportation Systems Magazine* **14**(2), 10–24 (March–April 2022). <https://doi.org/10.1109/MITS.2021.3070651>
45. Tang, H., Chen, K., Jia, K.: Unsupervised domain adaptation via structurally regularized deep clustering. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8722–8732. Institute of Electrical and Electronics Engineers (IEEE), Seattle, WA, USA (June 2020). <https://doi.org/10.1109/CVPR42600.2020.00875>
46. Tonk, A., Boussif, A.: Operational design domain or operational envelope; seeking a suitable concept for autonomous railway systems. In: European Safety and Reliability Conference. pp. 2104–2111. Research Publishing Services (2022). https://doi.org/10.3850/978-981-18-5183-4_S06-08-245-cd
47. Vasist, M., Rozet, F., Absil, O., Mollière, P., Nasedkin, E., Louppe, G.: Neural posterior estimation for exoplanetary atmospheric retrieval. *Astronomy & Astrophysics* **672**, A147 (April 2023). <https://doi.org/10.1051/0004-6361/202245263>
48. Villarreal Guerra, J.C., Khanam, Z., Ehsan, S., Stolkin, R., McDonald-Maier, K.: Weather classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of convolutional neural networks. In: NASA/ESA Conference on Adaptive Hardware and Systems (AHS). pp. 305–310. Institute

- of Electrical and Electronics Engineers (IEEE), Edinburgh, UK (August 2018). <https://doi.org/10.1109/AHS.2018.8541482>
49. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv [abs/2006.10726](https://arxiv.org/abs/2006.10726) (2020). <https://doi.org/10.48550/arXiv.2006.10726>
 50. Wang, Q., Fink, O., Van Gool, L., Dai, D.: Continual test-time domain adaptation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7191–7201. Institute of Electrical and Electronics Engineers (IEEE), New Orleans, Louisiana, USA (June 2022). <https://doi.org/10.1109/CVPR52688.2022.00706>
 51. Wang, W., Zhong, Z., Wang, W., Chen, X., Ling, C., Wang, B., Sebe, N.: Dynamically instance-guided adaptation: A backward-free approach for test-time domain adaptive semantic segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 24090–24099. Institute of Electrical and Electronics Engineers (IEEE), Vancouver, Canada (June 2023). <https://doi.org/10.1109/CVPR52729.2023.02307>
 52. Yuan, L., Xie, B., Li, S.: Robust test-time adaptation in dynamic scenarios. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15922–15932. Institute of Electrical and Electronics Engineers (IEEE), Vancouver, Canada (June 2023). <https://doi.org/10.1109/CVPR52729.2023.01528>
 53. Zanella, M., Ayed, I.B.: On the test-time zero-shot generalization of vision-language models: Do we really need prompt learning? In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). vol. 33, pp. 23783–23793. Institute of Electrical and Electronics Engineers (IEEE), Seattle, Washington, USA (June 2024). <https://doi.org/10.1109/CVPR52733.2024.02245>
 54. Zanella, M., Gérin, B., Ayed, I.B.: Boosting vision-language models with transduction. arXiv [abs/2406.01837](https://arxiv.org/abs/2406.01837) (2024). <https://doi.org/10.48550/arXiv.2406.01837>
 55. Zhang, Y., Sun, J., Chen, M., Wang, Q., Yuan, Y., Ma, R.: Multi-weather classification using evolutionary algorithm on EfficientNet. In: IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). pp. 546–551. Institute of Electrical and Electronics Engineers (IEEE), Kassel, Germany (March 2021). <https://doi.org/10.1109/PerComWorkshops51409.2021.9430939>