When Cars meet Drones: Hyperbolic Federated Learning for Source-Free Domain Adaptation in Adverse Weather
Supplementary Material

Giulia Rizzoli*, Matteo Caligiuri*, Donald Shenaj, Francesco Barbato, Pietro Zanuttigh
University of Padova, Italy

This document contains the supplementary material for the paper *When Cars meet Drones: Hyperbolic Federated Learning for Source-Free Domain Adaptation in Adverse Weather*. We begin by presenting the implementation specifics of our network. Next, we introduce additional details on the employed datasets. We then proceed with ablations supporting the usage of the weather batch-normalization and hyperbolic prototypes. Finally, we show further experimental data including per-class accuracy scores and qualitative results.

1. Implementation Details

Server pretraining. We chose DeepLabV3 architecture with MobileNetV2 as in [12]. We trained our model on the source synthetic dataset using a decreasing power-law learning rate η, starting at $\eta = 5 \times 10^{-3}$ with a power of 0.9. The optimization used SGD with momentum 0.9 and no weight decay, lasting for 5 epochs with batch size 16. For the weather classifier, we implemented a 3-layer ConvNet (see Fig. 3), trained for 8 epochs using SGD optimization with a learning rate of 1×10^{-4} and a batch size of 88.
Clients Adaptation. For the target dataset, experiments were run with a fixed learning rate of $\eta = 1 \times 10^{-4}$ with SGD optimizer. The training involved 5 clients per round for a total of $R = 100$ rounds, with $\lambda_{cl} = 140$. The pseudo-label teacher model was updated at the end of each round. For the optimizer of the manifolds, we use RiemmanianAdam as in [14] with $\gamma = 0.1$ as initialization, learning rate equal to 1×10^{-3} and weight decay of 4×10^{-4}.
Training of Competitors To establish performance upper bounds, we conducted two experiments:

1. Fully Centralized Training: We performed supervised training on the target dataset for 250 epochs with a learning rate of 5×10^{-3}.

2. Federated Fine-Tuning: We implemented federated fine-tuning with full supervision on the clients for 250 rounds, representing an upper limit scenario for the federated model.

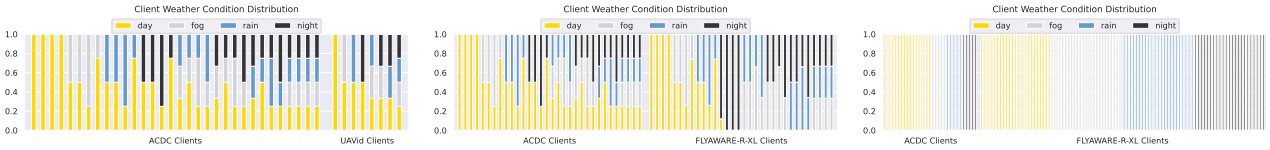
2. Data Selection and Distribution

In standard autonomous driving for semantic segmentation, a sufficient number of datasets provide a diversity of adverse weather conditions. For our study on car agents, we relied on existing datasets, adapting them to enable distributed learning in adverse weather scenarios. Here, we will first describe the technical choices adopted for this purpose. Conversely, there is a scarcity of datasets featuring adverse conditions for aerial viewpoints. To address this gap, we introduced the **FLYAWARE** aerial dataset, as it represents a novel contribution to unlock further research on this topic. Lastly, we provide additional details on weather distribution among clients.

2.1. Driving Datasets

For the synthetic source dataset for cars, we used the **SELMA** dataset [13]. It offers a comprehensive set of 27 weather and daytime conditions, resulting in a vast dataset of over 20M samples. To better align with the weather scenarios considered in real data [11], we opted not to use the standard SELMA split, but downloaded over 24k samples in the *ClearNoon, ClearNight, HardRainNoon, MidFoggyNoon* splits. In total, for this work, we employed almost 100k SELMA samples from the Desk Cam point of view that match the one used in the real-world dataset. As the real counterpart, we used the **ACDC** dataset, from which we selected the 3 domains — night, rain, and fog — that match our pretraining. As ACDC lacks images in clear weather, we supplemented it with daytime images coming from the Cityscapes dataset [1], which is the most similar to ACDC. Note that, the ACDC dataset has been built to create an adverse condition version of Cityscapes, and shares the same class-set. Cityscapes provides more samples than those in the thematic splits by ACDC. Therefore, we subsampled its training set to match the sizes. Since the goal was to select clear weather conditions, we devised an automatic way of assigning a sunlit level (Eq. 1) to each image and selected the needed images by sorting them according to this metric:

Figure 1. Cityscapes samples and corresponding *sunlit* score. Decreasing from the left, high/mid/low-score ($\sim 300/200/100$).



(a) Scenario (i) (b) Scenario (ii) (c) Scenario (iii)

Figure 2. Distribution of weather conditions across clients

$$\text{sunlit}(\mathbf{X}) = \sum_{c \in \{r,g,b\}} \frac{1}{HW} \sum_{i=0}^{W} \sum_{i=0}^{W} \mathbf{X}[i,j,c] \qquad (1)$$

As shown in **??** in the main paper, we selected a set of images to match the total count of all other conditions combined, aiming for a ratio of three clear sky images to one adverse condition. This decision was driven by the desire to retain "clear sky" as the most probable scenario in real-world contexts. In Fig. 1 we report some examples of samples with their rating (a bright one, a dark one and a mid-range one).

2.2. Aerial Dataset

A demo video of the dataset is available at https://github.com/LTTM/HyperFLAW.

FLYAWARE-S: Adverse Synthetic Dataset. Recently released, Syndrone [10] is a synthetic dataset based on the CARLA simulator [2]. While the dataset is richly annotated with 28 semantic classes, it currently lacks imagery in different weather conditions. We extended their work starting from the codebase provided in [10] by generating images in 3 adverse weather conditions (*i.e.*, rain, fog, and night), while maintaining the capability of the system to produce images from multiple viewpoints and heights. The dataset includes drone views at heights ranging from 20 to 80 meters and with angles varying from 30 to 90 degrees, all with a resolution of 1920x1080. Moreover, we also generated 3D data (depth maps and LiDAR) for future usage in multimodal architectures.

FLYAWARE-R: Real Dataset Translation. Since no adverse weather dataset for aerial vehicles is available we opted for using image translation over standard drone datasets to build the adverse weather imagery. Addressing the task of converting daytime images into adverse conditions requires the availability of images of clear weather samples like sunny or cloudy conditions, and of adverse samples occurring in specific conditions of interest, such as rainy weather. Inspired by [6] which performs adverse domain translations for autonomous driving in the context of depth estimation, we opted for the use of Generative Adversarial Networks (GANs). For the rain and night adverse conditions, we employed the ForkGAN model [17] to translate clear-day training samples from the UAVid dataset [8]. While for the fog samples, we have used a combination of Omnidata [4] and FoHIS [16] methods. In the following, we will present an in-depth explanation of how we have performed the aforementioned conversion.

Day2Night. To convert daylight into nighttime, we trained ForkGAN over $14K$ images, half of which represent daytime and the other half nighttime. All the used images were sampled from Visdrone [18], and UAVDT [3], two datasets designed for Object Detection. We have decided to sample the training data from different datasets to increase the overall number of nighttime images. This was necessary to ensure that the GAN architecture reaches good enough reconstruction performance. The training phase lasted 40 epochs. After the training, for all the nighttime samples in the UAVid dataset, we convert clear-day images into their nighttime counterparts using the pretrained model.

Day2Rain. Training the GAN models directly on drone data proved advantageous as it helped to mitigate domain shifts. In the rain case however there are no available drone datasets - not even for other tasks as in the case of night images -, therefore we had to train ForkGAN using diverse datasets that included adverse weather scenarios, like BDD100K [15], ACDC [11] and RainCityscapes [7] (which are datasets for automotive applications). For daytime to rainy conversion, we followed the same strategy applied for

ST	AG	CL	BN	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	All
				35.4	5.7	55.8	4.6	4.2	14.2	13.3	10.9	61.9	6.3	43.7	26.0	17.2	11.1	3.3	3.8	1.4	8.9	10.7	22.7
✓				43.7	22.0	57.2	6.6	4.7	15.8	9.8	13.9	50.2	5.7	68.4	24.3	11.4	34.6	8.9	4.7	3.5	16.0	14.9	26.1
✓	✓			43.9	23.2	59.2	7.1	5.9	16.2	11.1	14.8	55.3	6.1	58.5	26.0	11.9	34.2	9.2	4.3	3.0	15.8	13.5	26.6
✓	✓		✓	43.8	23.6	59.7	6.9	5.9	15.9	10.7	14.8	56.5	5.1	60.6	25.4	11.9	37.7	9.0	4.5	2.8	15.1	13.0	26.9
✓	✓	✓		46.8	22.4	59.9	7.4	6.7	17.6	17.0	22.4	56.6	7.1	62.6	25.4	10.5	35.5	9.5	6.9	3.7	16.6	25.7	28.5
✓	✓	✓	✓	47.2	22.7	60.6	7.0	6.7	17.3	16.8	22.7	57.7	5.9	64.9	25.4	10.6	36.9	9.3	6.8	3.5	16.8	25.5	29.0

Table 1. Per class IoU for the proposed approach and its ablated versions.

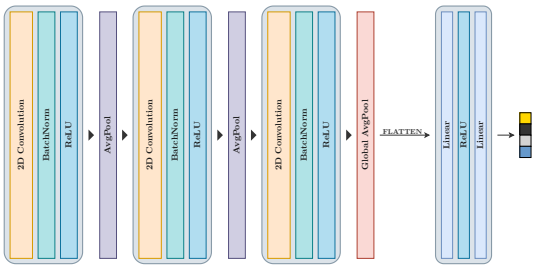


Figure 3. Weather classifier architecture model.

Modules	Clear	Night	Rain	Fog
Pretrain	25.41	13.62	26.72	26.08
$\mathcal{L}_{st}$ + AG w/o BN	29.27	13.47	29.47	27.86
$\mathcal{L}_{st}$ + AG w BN	26.81	13.50	30.19	28.86
ALL w/o BN	31.58	14.04	30.53	28.93
ALL w BN	**31.73**	**14.18**	**34.19**	**29.86**

Table 2. Effect of the different modules on the weather.

the conversion to nighttime training the ForkGAN model over $9K$ images for 40 epochs and we applied the pretrained model to convert images into their rainy counterparts.

Day2Fog. To perform daytime to foggy conversion we first estimated a depth map for each image in the UAVid dataset using Omnidata [4], then, we exploited the estimated depth as input to the FoHIS method [16] to apply fog.

2.3. Weather Heterogeneity

Scenario i: Fig. 2a illustrates the weather distribution among clients in the configuration **ACDC+ FLYAWARE-R**. We deliberately supplied clear day samples to each client to emulate a common scenario where clear daylight conditions prevail, with fewer instances of challenging weather conditions. This introduces an inherent challenge as adverse weather data samples are less prevalent. Additionally, the imbalance among clients further complicates the setting.

Scenario ii: Fig. 2b shows the weather distribution for clients in the **ACDC+FLYAWARE-R-XL** setup. While this configuration exhibits better data balance across clients, it introduces new complexities: some drone clients exclusively operate in adverse conditions, and the training and testing domains originate from different datasets (VisDrone [18] and UAVid [8], respectively). Notably, VisDrone images offer real night data, whereas we utilized domain-translated images for fog and rain using the same strategies.

Scenario iii: Fig. 2c shows the weather distribution for clients in the **ACDC+FLYAWARE-R-XL** setup in the extreme case where each client observes a single weather condition. Although the number of clients is balanced in terms of different viewpoints, operating in a single weather condition implies less training variability.

3. Impact of Weather-Batch Normalization

As shown in **??** of the main paper, the inclusion of ad-hoc batch norms (BNs) enhances performance in comparison to utilizing a non-personalized network. First of all, in Fig. 3, we show in detail the architecture of the weather classifier. Although the classifier model is simple, it achieves an accuracy of 98.96% on the source synthetic datasets used to train it and of 72.37% on the real world target ones.

We further examine in Tab. 2 the advantage of adapting the system to accommodate different weather conditions as a remarkable feature for the decision-making of autonomous driving agents. The basic self-training strategy, coupled with the proposed server-side aggregation scheme, has shown stability and an overall improvement over pretraining. However, due to the prevalence of clear-day images in most clients' samples, the network tends to learn more about this weather condition. Introducing the personalized weather BNs helps to mitigate this, improving performances in adverse weather conditions, with an mIoU increase of 3% on Rain and 1.78% on Fog. Similar results are evident when comparing the full method with and without the weather BNs, showcasing mIoU increases of 3.66% on Rain and 0.93% on Fog. As a side note, the most difficult scenario remains the night where the improvement over the pretraining is just 0.56% of mIoU. Overall, we remark on the significance of incorporating personalized weather BNs for mitigating bias towards clear day images and enhancing performance across various weather conditions.

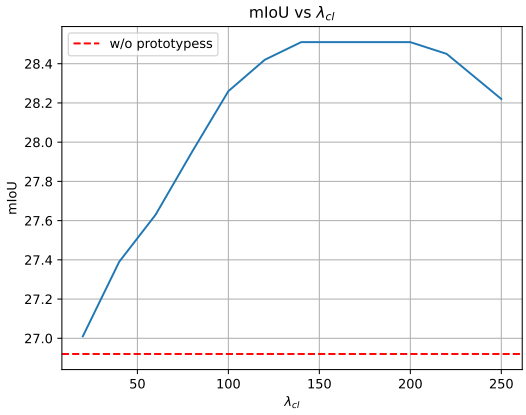


Figure 4. Tuning on λ_{cl} weight.

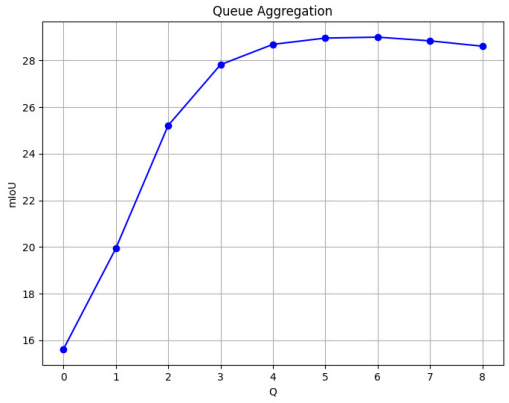


Figure 5. Tuning on the queue aggregation parameter Q.

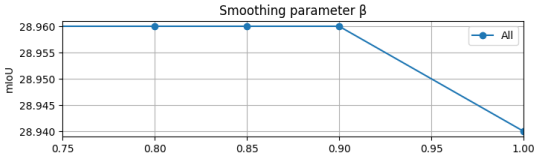


Figure 6. Tuning on the prototype smoothing parameter β.

Method	Backbone	# Params	Supervised on	Unsup. on	City mIoU
FedDrive [5]	BiseNetV2	8.2M	City	–	43.85
Fed. Oracle†	MobileNetV2	3.4M	City	–	58.16
Fed. Fine-Tune†	MobileNetV2	3.4M	GTAV → City	–	59.35
Source Only	MobileNetV2	3.4M	GTAV	–	20.23
LADD [12] †	MobileNetV2	3.4M	GTAV	City	36.49
Ours	MobileNetV2	3.4M	GTAV	City	38.21

Table 3. GTA → Cityscapes results.

As a final note, considering that modern vehicles often have access to real-time weather information through several sensors (e.g., automatic windshield wipers for rain detection, clock or optical light detection for night and fog), and the weather information might be directly obtained.

4. Additional Parameters Ablation

Fig. 4 shows the mIoU for different values of the λ_{cl} parameter. Best performances are achieved in the range 140-180, with quite stable maxima. We report the effect of the queue aggregation parameter Q, which stabilizes at value 5 (Fig. 5). Finally, the network does not exhibit sensitivity to the smoothing parameter of the prototypes β ($\beta = 0.85$) (Fig. 6).

5. Per-class Results

Tab. 1 contains a more detailed version of the results in **??** of the main paper that also shows the class-by-class accuracy. As common on the employed datasets, the mIoU is higher in the common classes (e.g., road or building) and lower in the rare and more challenging ones. However, it can be seen that by adding the various components of the model, results tend to increase consistently in most classes, even if a few challenging ones remain hard to detect. Some classes like traffic sign, car or bicycle show impressive improvements. On the other side, there are a few cases in which our approach does not improve the results on some under-represented classes such as terrain and rider.

6. Qualitative Results

Fig. 7 shows some qualitative results for car and drone samples in both clear and adverse weather. FedAvg [9] and LADD [12] appear to be overconfident on the road and car classes, respectively. While our model adeptly captures the structure of the street better than other models in both adverse and non-adverse conditions, the adverse conditions images remain a challenging aspect. For drones, the discrepancies between the predicted classes and the ground truth arise also due to the fact that the network is trained on a superset of classes beyond those present in the drone dataset. Notice that sometimes the network assigns "fine" classes that share a semantic affiliation with the respective "coarse" classes, e.g., predict terrain instead of vegetation, but on the drone dataset labeling terrain is not present and terrain samples have ground truth set to vegetation. The same thing happens for the sidewalk and road or rider and person. For this reason, we remapped the 19 classes of the full set into the 5 drone ones using the mapping proposed in [10]. The figure shows both the original prediction maps from the network and the ones obtained after remapping the 19 classes into the 5 of the drone datasets. The comparison with the ground truth shows that some predictions not matching are just due to the different types of labeling in drone and car datasets.

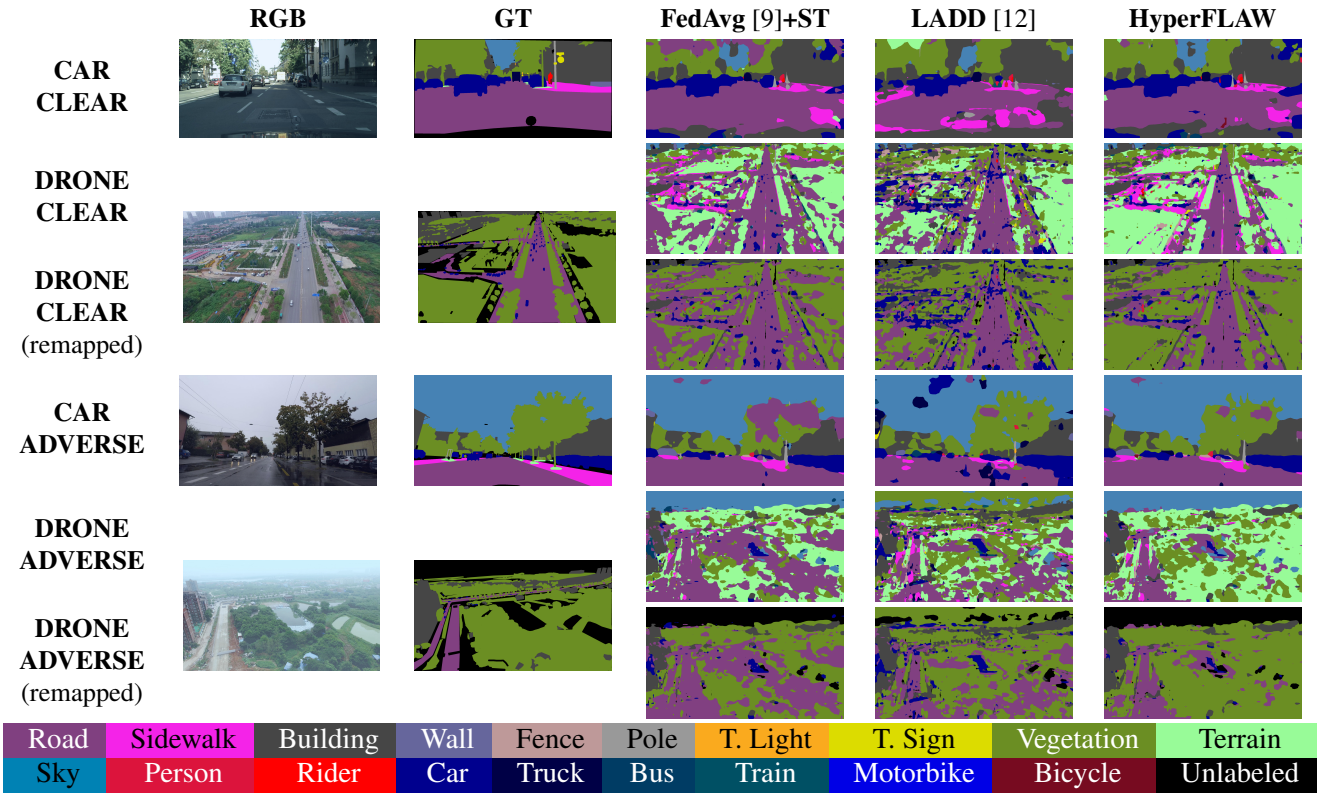


Figure 7. Qualitative results.

7. Comparison over GTA → Cityscapes

We demonstrate the advantage of our approach in transferring a more generalizable representation from synthetic to real-world urban environments, even in scenarios not involving drones. To this end, we evaluated our method on the GTA → Cityscapes domain adaptation task, with results presented in Tab. 3. We utilized the same federated split of Cityscapes as employed in previous works [5, 12] to ensure a fair comparison. Our approach outperforms previous unsupervised methods and narrows the gap with supervised techniques. Specifically, we achieve results that are only 5.6 mIoU lower than the supervised method in [5], while using a less complex model.

Supplementary References

[1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1

[2] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2

[3] Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, and Qi Tian. The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 370–386, 2018. 2

[4] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 2, 3

[5] Lidia Fantauzzo, Eros Fanì, Debora Caldarola, Antonio Tavera, Fabio Cermelli, Marco Ciccone, and Barbara Caputo. Feddrive: Generalizing federated learning to semantic segmentation in autonomous driving. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11504–11511. IEEE, 2022. 4, 5

[6] Stefano Gasperini, Nils Morbitzer, HyunJun Jung, Nassir Navab, and Federico Tombari. Robust monocular depth estimation under challenging conditions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8177–8186, 2023. 2

[7] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 8022–8031, 2019. 2

[8] Ye Lyu, George Vosselman, Gui-Song Xia, Alper Yilmaz,

and Michael Ying Yang. Uavid: A semantic segmentation dataset for uav imagery. *ISPRS journal of photogrammetry and remote sensing*, 165:108–119, 2020. 2, 3

[9] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282. PMLR, 2017. 4, 5

[10] Giulia Rizzoli, Francesco Barbato, Matteo Caligiuri, and Pietro Zanuttigh. Syndrone-multi-modal uav dataset for urban scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2210–2220, 2023. 2, 4

[11] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10765–10775, 2021. 1, 2

[12] Donald Shenaj, Eros Fanì, Marco Toldo, Debora Caldarola, Antonio Tavera, Umberto Michieli, Marco Ciccone, Pietro Zanuttigh, and Barbara Caputo. Learning across domains and devices: Style-driven source-free domain adaptation in clustered federated learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 444–454, 2023. 1, 4, 5

[13] Paolo Testolina, Francesco Barbato, Umberto Michieli, Marco Giordani, Pietro Zanuttigh, and Michele Zorzi. Selma: Semantic large-scale multimodal acquisitions in variable weather, daytime and viewpoints. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 1

[14] Max van Spengler, Erwin Berkhout, and Pascal Mettes. Poincare resnet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5419–5428, October 2023. 1

[15] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 2

[16] Ning Zhang, Lin Zhang, and Zaixi Cheng. Towards simulating foggy and hazy images and evaluating their authenticity. In *International Conference on Neural Information Processing*, pages 405–415. Springer, 2017. 2, 3

[17] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi. Forkgan: Seeing into the rainy night. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 155–170. Springer, 2020. 2

[18] Pengfei Zhu, Longyin Wen, Dawei Du, Xiao Bian, Heng Fan, Qinghua Hu, and Haibin Ling. Detection and tracking meet drones challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7380–7399, 2021. 2, 3