Supplementary Material

Julian Strohmayer Rafael Sterzinger Matthias Wödlinger Martin Kampel
Computer Vision Lab, TU Wien
Vienna, Austria
{firstname.lastname}@tuwien.ac.at

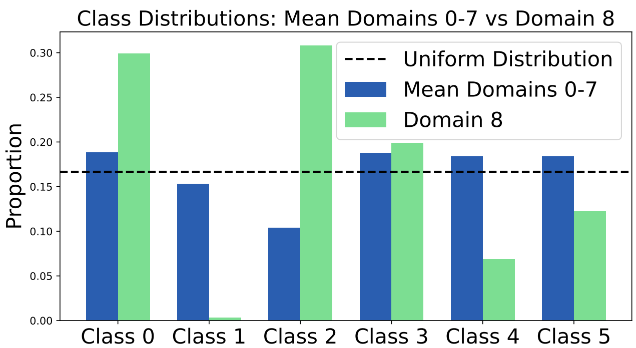

Figure 1. Comparison of class distributions between the mean distribution of domains 0–7 and the highly skewed distribution of domain 8. The figure illustrates the deviation of domain 8 from the uniform distribution observed in domains 0–7, highlighting its class imbalance.

1. Training Hyperparameters

Table 2 summarizes the hyperparameters for conventional training (W) and domain-adversarial training (DAT) (W +DAT [1], W_{DAT}). Parameters include the optimizer, learning rate, weight decay, batch size, and the domain-adversarial loss weights α, β, as well as the gradient reversal layer (GRL) scaling parameter γ.

Table 3 presents the hyperparameters for models utilizing test-time adaptation (TTA). Parameters include the optimizer, learning rate, batch size, loss function, use of random weight resetting (R), and the corresponding reset probability (p).

Table 1 describes the augmentations applied during conventional training and DAT, including amplitude perturbation and circular rotation, as well as pixel-wise and column-wise dropout with mean-replacement [3].

2. Widar3.0-G6D D8 Failue Analysis

As highlighted in our cross-domain adaptation experiment (*Section 4.3*), domain D8 causes a significant drop in performance during TTA. To investigate this issue, we perform a PCA on domains D0–D8, as visualized in Figure 2. We observe that domain D8 (right) diverges considerably from the nominal feature space defined by domains D1–D7 (left), exhibiting low intra-class cohesion and poor class separability. In contrast, a well-behaved domain such as D0 (middle) demonstrates clear and consistent clusters of activities.

Moreover, D8 shows a highly skewed class distribution and distinctly different data statistics compared to domains D0–D7. Figure 1 illustrates this discrepancy by comparing the mean class distribution of domains D0–D7 to that of domain D8, emphasizing the severe class imbalance and its potential negative impact on model adaptation.

Augmentation	p	m
Amplitude Perturbation	0.5	0.01
Circular Rotation	0.5	0.50
Pixel-wise Dropout	0.001	-
Column-wise Dropout	0.001	-

Table 1. Augmentations applied during conventional and domain-adversarial training (DAT), along with their probabilities (p) and magnitudes (m). For amplitude perturbation and circular rotation, the magnitude is uniformly sampled from $[-m, m]$. Circular rotation is conducted along the temporal axis. Pixel-wise and column-wise dropout use mean-replacement [3], setting the respective values to the mean of the CSI amplitude spectrogram instead of zero.

References

[1] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*, pages 289–304, 2018. 1, 2

[2] Wei Lin, Muhammad Jehanzeb Mirza, Mateusz Kozinski, Horst Possegger, Hilde Kuehne, and Horst Bischof. Video Test-Time Adaptation for Action Recognition. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22952–22961, Vancouver, BC, Canada, 2023. IEEE. 2

[3] Julian Strohmayer and Martin Kampel. Wifi CSI-based long-

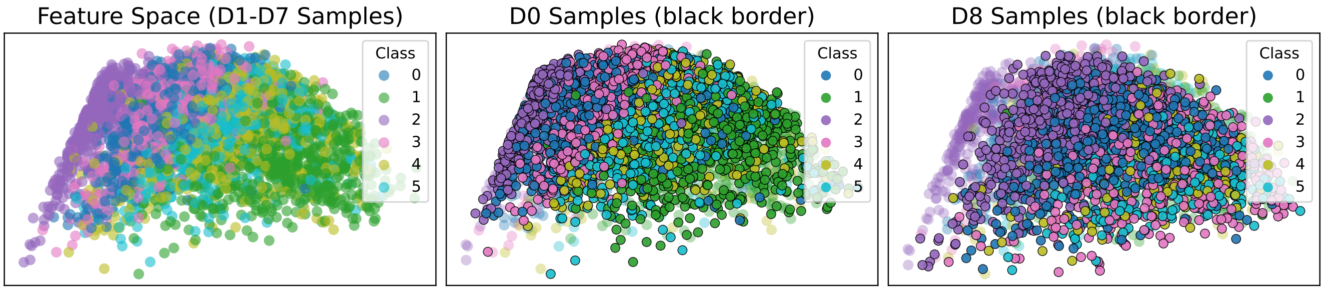

Figure 2. Comparison of feature spaces for domains D1–D7, D0, and D8, highlighting the low intra-class cohesion and poor separability of D8 samples.

Model	Optimizer	learning rate	weight decay	loss	batch size	α	β	γ
W (baseline)	AdamW	0.001	0.001	CE	32	-	-	-
W +DAT [1]	AdamW	0.001	0.001	$\mathcal{L}$	8	0.3	0.2	8.0
W_{DAT}	AdamW	0.001	0.001	$\mathcal{L}$	8	0.3	0.2	8.0

Table 2. Hyperparameter settings for conventional training (W) and domain-adversarial training (DAT) (W +DAT [1], W_{DAT}). α and β represent the weights for the domain-adversarial loss $\mathcal{L}$, and γ denotes the gradient reversal layer (GRL) scaling parameter.

Model	Optimizer	learning rate	loss	batch size	R	p
W +ViTTA [2]	SGD	0.0001	ℓ_1	1	-	-
W_{DAT} +ViTTA [2]	SGD	0.00001	ℓ_1	1	-	-
W +DAT [1] +ViTTA [2]	SGD	0.00001	ℓ_1	1	-	-
W_{TTA} +DAT [1]	SGD	0.001	ℓ_2	1	-	-
W_{TTA}	SGD	0.001	ℓ_2	1	-	-
W_{DATTA}	SGD	0.0001	ℓ_2	1	-	-
	SGD	0.0001	ℓ_2	1	✓	0.0001

Table 3. Hyperparameter settings for models trained with test-time adaptation (TTA). Column R indicates the use of random weight resetting, and p represents the reset probability.

range person localization using directional antennas. In *The Second Tiny Papers Track at ICLR 2024*, 2024. 1