

A. Supplementary material

A.1. Priors of the source domains

In this section, we provide further details about the priors of each source domain used in our experiments, we provide the histograms with respect to the semantic classes for all domains (each corresponding to a dataset) in Figure 9. As can be seen, there is a huge class imbalance with some of the classes. Some of them are especially underrepresented, like two wheels on the RainSnow, MTID/Drone, GRAM-RTM/M-30-HD and UT/Sherbrooke datasets.

A.2. Details about the networks

We use the TinyNet network for both the source and the domain discriminator models. This is a segmentation architecture first introduced in [40] for soccer player segmentation and later refined in [35]. Interestingly, due to its lightweight architecture, this network can be used for real-time inference on embedded devices, making it a great choice for real-world surveillance. Compared to the original architecture, we adapt it for the 720p input image size (compared to 1080p in the original publication). For that, only the values of the pooling layers applied in the pyramid pooling have to be adjusted. The architecture of the adapted network can be found in Figure 10.

A.3. Details about the domain discriminator model

Figure 11 shows an example of multi-domain images that we generated to train our domain discriminator model. As can be seen, four patches cropped from four randomly drawn images are combined to create training images, so that the domain discriminator model learns to recognise different domains within a same image. This is indeed an interesting feature since a target domain could combine some elements from different source domains (for example, the road of the target domain could look like the one from one source domain but the weather and lighting conditions may be those of another source domain).

A.4. Interpretation of posteriors

The interpretation given to the posteriors depends on the priors. In particular, $P(H | E) = P(H)$ when the evidence provides no information about the hypotheses. In fact, the priors can be seen as a reference, the posteriors being expressed w.r.t. it. Consider P' and P'' , two probability measures sharing common likelihoods $P'(E | H) = P''(E | H)$. The posteriors $P'(H | E)$ and $P''(H | E)$ convey the same information, but are expressed w.r.t. different references $P'(H)$ and $P''(H)$. The change of reference is termed *prior shift* or *target shift*. Assuming non-zero priors, it is given by [3, 36]

$$P''(H | E) \propto \frac{P''(H)}{P'(H)} P'(H | E)$$

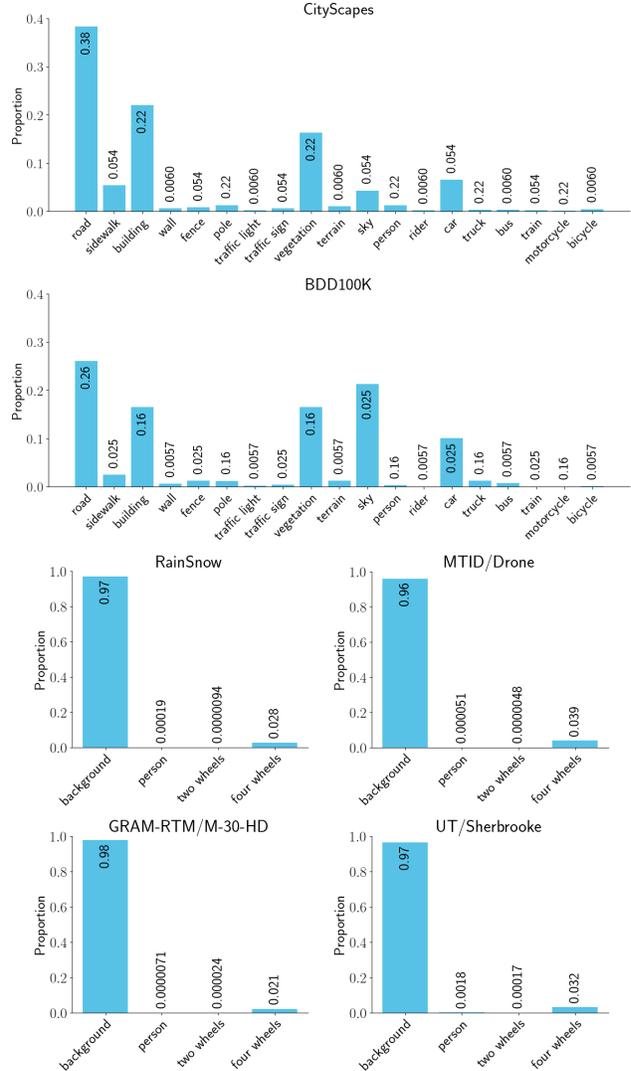


Figure 9. Histograms giving the priors, *i.e.* the proportion of each semantic class, for each domain (from top to bottom and from left to right: CityScapes, BDD100K, RainSnow, MTID/Drone, GRAM-RTM/M-30-HD and UT/Sherbrooke). All numbers are given up to 2 significant figures.

and, when \mathbb{H} forms a partition of Ω , by

$$P''(H | E) = \frac{\frac{P''(H)}{P'(H)} P'(H | E)}{\sum_{H \in \mathbb{H}} \frac{P''(H)}{P'(H)} P'(H | E)}$$

A.5. Validation of predicted posteriors

As stated in Section 4.1, it is important that the source and domain discriminator models output trustworthy posteriors. To verify that, we perform two tests: (1) we compare the class-by-class average of the posteriors estimated by our algorithm to the priors of the dataset—in particular, its test set (shown for one source model in Figure 12), and

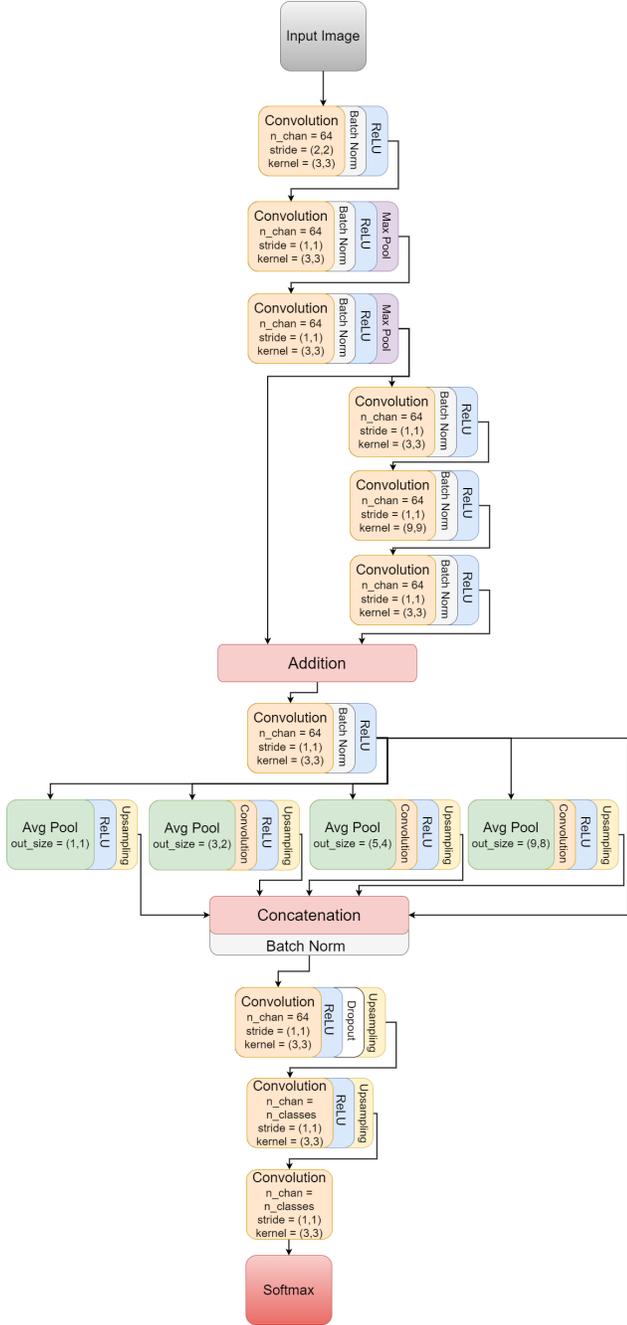


Figure 10. TinyNet architecture adapted to 720p images and an arbitrary number of output classes (either the semantic classes or the number of source domains).

(2) we verify that our model is well calibrated using calibration plots (shown for one source model and one semantic class in Figure 13).

Since both the average of the posteriors is comparable to the priors of the dataset and the calibration plots are close to the $y = x$ line, we are confident that the outputs of our algorithm correspond to posteriors. We perform these tests



Figure 11. Example of multi-domain image used to train our domain discriminator model. The image is obtained with a mosaic transform for images of the CityScapes and BDD100K datasets.

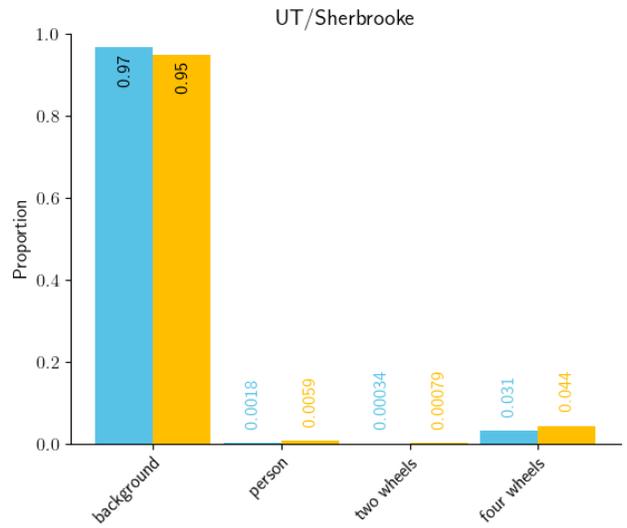


Figure 12. Histogram for the UT/Sherbrooke dataset comparing the priors of the test set (in blue) and the average of the posteriors obtained with our algorithm for each class (in orange). All numbers are given to 2 significant figures.

for all source and discriminator models.

A.6. Extra results

In this section, we provide extra results for both of our experiments. First, we provide the same graph as the one depicted in Figure 6 for the 3 remaining segmentation scores (*i.e.* balanced accuracy, mean IoU, and balanced mean IoU) in Figure 14. As can be seen, for all scores, our algorithm outperforms all other heuristics for each target domain.

Next, we provide the performance of our algorithm for our second experiment with 4 source domains for different λ_k in Table 1 for the accuracy and mean IOU, and in Table 2 for the balanced accuracy and balanced mean IOU. We compare our algorithm with the four source models (one model for each source domain), the random selection

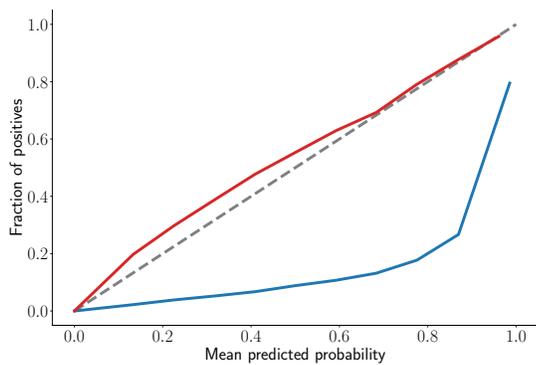


Figure 13. Calibration plot for the UT/Sherbrooke source model **before the target shift (in blue)** and **after the target shift (in red)** for the "four wheels" class. The dotted line corresponds to the perfect case where the output of the model correspond to the posteriors.

of source models, and the linear combination of posteriors. Note that the histograms presented in Figure 8 represent the column-by-column average values of these tables.

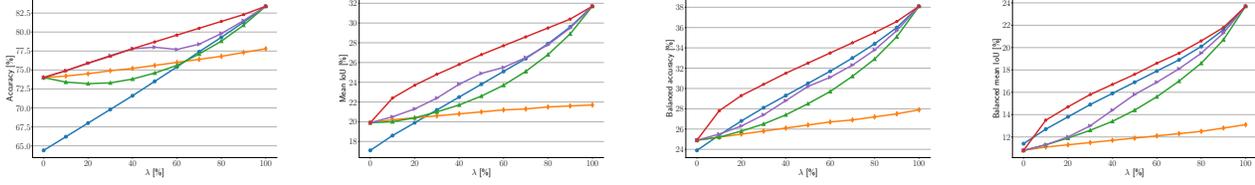


Figure 14. Results of our first experiment (all decisions made with the MAP strategy) showing from left to right: the accuracy, mean IoU, balanced accuracy, and balanced mean IoU. We compare the **source model trained on CityScapes (in blue)**, the **source model trained on BDD100K (in orange)**, the **random selection of source models (in green)**, the **linear combination of posteriors (in purple)**, and **our algorithm (in red)**. $\lambda = 0$ (resp. $= 100$) corresponds to BDD100K (resp. CityScapes) as target domain.

Table 1. **Results of our second experiment (all decisions made with the MAP strategy) for various target domains obtained from 4 source domains for various λ_k for the accuracy/mean IoU scores. We compare the different source models (SM) trained on their source domain, the random selection of source models, the linear combination of posteriors, and our algorithm.**

Parameters				Methods							
λ_1	λ_2	λ_3	λ_4	SM ₁ RainSnow	SM ₂ MTID/Drone	SM ₃ GRAM-RTM/M-30-HD	SM ₄ UT/Sherbrooke	Random selection of source models	Linear combination of posteriors	Ours	
1	0	0	0	97.9/35.1	96.4/26.3	96.0/27.6	96.1/28.3	97.9/35.1	97.9/35.1	97.9/35.1	
0	1	0	0	96.8/33.3	98.2/40.4	95.5/31.3	95.0/29.4	98.2/40.4	98.2/40.4	98.2/40.4	
0	0	1	0	98.2/34.6	97.8/31.8	99.5/55.7	96.1/28.2	99.5/55.7	99.5/55.7	99.5/55.7	
0	0	0	1	96.2/32.9	95.7/27.4	96.5/32.3	99.0/67.1	99.0/67.1	99.0/67.1	99.0/67.1	
½	½	0	0	97.4/ 35.0	97.2/33.7	95.7/29.6	95.5/28.9	97.3/34.0	97.7 /34.1	97.0/31.7	
½	0	½	0	98.0/35.3	97.2/28.8	97.8/38.0	96.1/28.4	97.9/36.9	98.5 /36.4	98.0/ 39.3	
½	0	0	½	97.1/33.5	96.0/27.0	96.2/30.0	97.5/52.6	97.3/42.5	98.2 /42.8	98.2 / 57.4	
0	½	½	0	97.4/33.7	98.0/36.8	97.6/ 43.1	95.6/28.9	97.8/39.7	98.4 /39.2	97.4/33.7	
0	½	0	½	96.5/33.6	96.9/34.1	96.0/31.9	97.1/54.6	97.0/43.0	97.8/42.8	98.0 / 60.4	
0	0	½	½	97.2/33.7	96.8/29.6	98.1/38.3	97.5/54.1	97.8/46.8	98.7 /43.5	98.5/ 62.0	
⅓	⅓	⅓	0	97.6/34.9	97.4/33.2	97.1/ 36.4	95.8/28.8	97.3/34.8	98.0 /34.5	97.1/31.3	
⅓	⅓	0	⅓	97.0/33.8	96.7/31.7	96.0/30.5	96.7/48.4	96.8/37.3	97.6/35.4	97.7 / 55.3	
⅓	0	⅓	⅓	97.4/33.8	96.7/28.4	97.4/34.4	97.0/47.9	97.3/38.7	98.2 /36.4	98.0/ 56.3	
0	⅓	⅓	⅓	97.0/33.7	97.2/33.5	97.3/35.8	96.7/49.5	97.1/39.3	97.7/35.5	97.8 / 57.3	
¼	¼	¼	¼	97.2/33.9	97.0/31.7	96.9/33.7	96.5/45.4	96.9/35.9	97.8 /34.1	97.6/ 53.9	

Table 2. Results of our second experiment (all decisions made with the MAP strategy) for various target domains obtained from 4 source domains for various λ_k for the balanced accuracy/balanced mean IoU scores. We compare the different source models (SM) trained on their source domain, the random selection of source models, the linear combination of posteriors, and our algorithm.

Parameters				Methods							
λ_1	λ_2	λ_3	λ_4	SM ₁ RainSnow	SM ₂ MTID/Drone	SM ₃ GRAM-RTM/M-30-HD	SM ₄ UT/Sherbrooke	Random selection of source models	Linear combination of posteriors	Ours	
1	0	0	0	38.7/21.0	28.0/9.5	30.8/12.2	31.9/13.4	38.7/21.0	38.7/21.0	38.7/21.0	
0	1	0	0	37.0/18.8	46.7/29.6	37.0/18.6	34.0/15.6	46.7/29.6	46.7/29.6	46.7/29.6	
0	0	1	0	39.1/18.8	35.2/15.7	58.8/39.0	33.8/15.4	58.8/39.0	58.8/39.0	58.8/39.0	
0	0	0	1	38.7/19.7	30.1/11.9	36.7/17.1	75.9/61.1	75.9/61.1	75.9/61.1	75.9/61.1	
$\frac{1}{2}$	$\frac{1}{2}$	0	0	38.5/20.8	38.1/19.4	34.2/15.5	33.1/14.5	38.3/20.1	36.0/17.7	34.2/15.9	
$\frac{1}{2}$	0	$\frac{1}{2}$	0	39.2/20.4	31.3/12.8	43.1/25.2	32.8/14.3	41.1/22.8	37.7/19.2	43.4/25.3	
$\frac{1}{2}$	0	0	$\frac{1}{2}$	38.0/19.0	29.3/10.9	33.9/14.7	62.2/45.4	50.1/32.0	45.1/27.5	66.5/50.4	
0	$\frac{1}{2}$	$\frac{1}{2}$	0	37.6/18.0	41.7/22.2	48.3/30.3	33.7/15.1	45.0/26.2	41.3/22.6	37.6/18.3	
0	$\frac{1}{2}$	0	$\frac{1}{2}$	38.5/19.3	38.6/20.1	36.9/17.2	63.4/46.7	51.0/33.0	45.3/27.8	70.1/54.5	
0	0	$\frac{1}{2}$	$\frac{1}{2}$	39.0/19.8	32.3/13.9	42.4/22.7	64.6/48.2	53.5/35.2	45.0/27.1	74.0/58.8	
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	38.5/19.7	37.2/18.2	41.7/23.6	33.2/14.7	39.1/20.5	36.2/17.5	34.3/15.6	
$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	38.1/19.0	35.5/17.0	35.1/15.7	56.5/39.1	43.3/24.9	37.4/19.1	63.6/47.3	
$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{3}$	38.3/19.2	30.9/12.5	38.8/19.4	56.8/39.7	44.6/26.0	37.9/19.5	66.1/49.9	
0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	38.4/19.1	37.6/19.1	40.6/20.9	57.8/40.7	45.3/26.7	38.2/19.9	66.3/49.9	
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	38.1/18.9	35.3/16.8	38.4/19.0	53.0/35.5	41.2/22.5	35.8/17.4	62.0/45.1	