Supplementary Material

A. SPP Ablation Study

In order to study the effect of the SPP block in the network, we analyze the contributions from each branch. This is done by taking an already trained model and removing all pooling branches except one, *i.e*. only the features from one branch are used in the decoder.

As seen in Table 1, the network is quite robust to these changes. However, there is a consistent decline in perfomance as the average pooling kernel size increases. We theorize that this is due to the smaller kernels providing additional local context, allowing for fine-grained matching.

Features	Seasonal AUC	Cross-season AUC
Baseline	**99.82**	**96.56**
$k = 4$	<u>99.81</u>	<u>96.27</u>
$k = 8$	99.80	96.25
$k = 16$	99.66	95.55
$k = 32$	99.74	94.86

Table 1: Within season and cross-season AUCs using a single SPP branch with kernel size k. The baseline was taken as the best performing model from Table 1, DVF with $\alpha = 0.2$.

B. Cross-Seasonal Correspondences

Here we explore the use of the recently proposed Cross-Seasonal Correspondences dataset [2], which aims to provide 2D correspondences betweeen images of the same location obtained at different times. The dataset provides correspondences for two sub-datasets: CMU and RobotCar Seasons.

CMU Seasons. Trajectories from multiple separate runs were aligned by performing bundle adjustment on a mix of SIFT and manually annotated correspondences. Correspondences are then obtained from 3D geometric matching between dense pointclouds from each trajectory. The obtained correspondences are of better quality than those in RobotCar Seasons. However, this dataset does not contain nighttime trajectories or other significant appearance changes.

RobotCar Seasons. Similarly to CMU, the reference trajectory is iteratively refined by triangulating 3D points and performing bundle adjustment. Additional trajectories are then aligned to this reference using 3D points obtained from LiDAR. In this case, correspondences were obtained from the ICP aligned LiDAR pointclouds from different runs. In contrast to CMU Seasons, this dataset provides greater seasonal variation, but worse correspondences.

It is worth noting that the correspondences provided by this dataset are still far from perfect. Figure 1 provides an example visalization for a random training pair in RobotCar Seasons. It can be seen how, even in the reference image, the LiDAR points are missaligned, *e.g*. those corresponding to the car and the building edge. Dynamic objects, such as cars and pedestrians, also cause false correspondences.

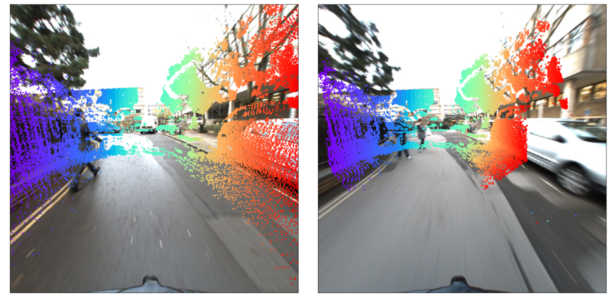

Figure 1: Sample RobotCar Seasons correpondences [2]

Despite these issues, we evaluate two cases using the procedure from [3]. Table 2 provides results when evaluating on the CMU Seasons correspondences. In this case, the correspondences are of better quality, but all the data was collected during daytime. As such, it is not surprising that D2-Net [1] slighly outperforms the proposed approach, since it was trained on supervised daytime data.

On the other hand, Table 3 corresponds to a cross-seasonal case, matching from reference-overcast to nighttime images in RobotCar Seasons. Once again, this highlights the benefits of our approach, which is invariant to large illumination changes.

Method	μ_+	Global perf.		Local perf.	
		AUC	μ_-	AUC	μ_-
SAND G	0.4747	82.79	0.8747	62.07	0.5716
SAND GL	0.7255	86.37	1.2929	73.90	1.0566
SAND L	0.5765	82.09	0.9433	<u>76.08</u>	0.8843
D2-Net	0.8182	**96.86**	1.1361	**83.56**	1.0234
DVF - $\alpha = 0$	**0.0305**	87.30	0.254	61.70	0.0381
DVF - $\alpha = 0.2$	0.0952	91.18	0.4276	61.38	0.1155
DVF - $\alpha = 0.4$	<u>0.0325</u>	87.03	0.2619	57.72	0.0388
DVF - $\alpha = 0.6$	0.1665	<u>91.37</u>	0.5744	58.02	0.1911
DVF - $\alpha = 0.8$	0.1098	90.67	0.4307	61.09	0.1297
DVF - $\alpha = 1$	0.0382	87.94	0.2818	60.34	0.0466

Table 2: Evaluation on CMU Seasons correspondences

Method	μ_+	Global perf.		Local perf.	
		AUC	μ_-	AUC	μ_-
SAND G	1.3459	49.76	1.3342	49.94	1.345
SAND GL	1.4751	49.13	1.4723	49.33	1.4676
SAND L	0.9377	45.78	0.8957	48.25	0.92
D2-Net	1.0943	77.45	1.1503	**59.67**	1.1149
DVF - $\alpha = 0$	**0.0564**	79.90	0.2179	51.09	0.0569
DVF - $\alpha = 0.2$	0.1688	74.61	0.3909	50.33	0.1696
DVF - $\alpha = 0.4$	<u>0.0589</u>	**84.51**	0.2256	<u>51.20</u>	0.0594
DVF - $\alpha = 0.6$	0.2642	<u>83.91</u>	0.5737	50.74	0.2668
DVF - $\alpha = 0.8$	0.2069	77.80	0.4245	50.79	0.2087
DVF - $\alpha = 1$	0.066	83.59	0.2368	51.06	0.0665

Table 3: Evaluation on RobotCar Seasons nighttime correspondences

References

[1] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. *CVPR*, may 2019.

[2] M. Larsson, E. Stenborg, L. Hammarstrand, M. Pollefeys, T. Sattler, and F. Kahl. A cross-season correspondence dataset for robust semantic segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2019-June, pages 9524–9534. IEEE Computer Society, jun 2019.

[3] J. Spencer, R. Bowden, and S. Hadfield. Scale-Adaptive Neural Dense Features : Learning via Hierarchical Context Aggregation. *CVPR*, 2019.