Supplementary Material:

Efficient Neighbourhood Consensus Networks via Submanifold Sparse Convolutions

Anonymous ECCV submission

Paper ID 909

In this supplementary material we present additional qualitative results on the HPatches Sequences (Sec. 1), InLoc (Sec. 2) and Aachen Day-Night (Sec. 3) benchmarks, and insights about the way Sparse-NCNet operates (Sec. 4).

1 HPatches Sequences benchmark

Mean matching accuracy. The Mean Matching Accuracy (MMA) metric is used in the HPatches Sequences benchmark to assess the fraction of correct matches under different tolerance thresholds. It is computed in the following way:

$$\mathrm{MMA}\big(\{(p_i^A, p_i^B)\}_{i=1}^N; t\big) = \frac{\sum_{i=1}^N \mathbb{1}_{>0}\big(t - \|\mathcal{T}_H(p_i^A) - p_i^B\|\big)}{N}, \tag{1}$$

where $\{(p_i^A, p_i^B)\}_{i=1}^N$ is the set of matches to be evaluated, $\mathcal{T}_H(p_i^A)$ is the warped point p_i^A using the ground-truth homography H, $\mathbb{1}_{>0}$ is the indicator function for positive numbers, and t is the chosen tolerance threshold (in pixels).

Additional qualitative results are presented in Figures 1 and 2. We compare the MMA of Sparse-NCNet with the state-of-the-art methods SuperPoint [1], D2-Net [2] and R2D2 [5], which are trainable methods for joint detection and description on local features. The correctly matched points are shown in green, while the incorrectly matched ones are shown in red, for a threshold value $t = 3$ pixels. For the proposed Sparse-NCNet, results are presented for two different numbers of matches, 2000 and 6000. Results show that our method produces the largest fraction of correct matches, even when considering as many as 6000 correspondences. In particular, note that our method is able to produce a large amount of correct correspondences even under strong illumination changes, as shown in Fig. 2. Furthermore, note that the nature of the correspondences produced by Sparse-NCNet is different from those of local feature methods. While local feature methods can only produce correspondences on the detected points, which are the local extrema of a particular feature detection function, our method produces densely packed sets of correspondences. This results from Sparse-NCNet's propagation of information in local neighbourhoods, as discussed in Sec. 4 of this supplementary material.

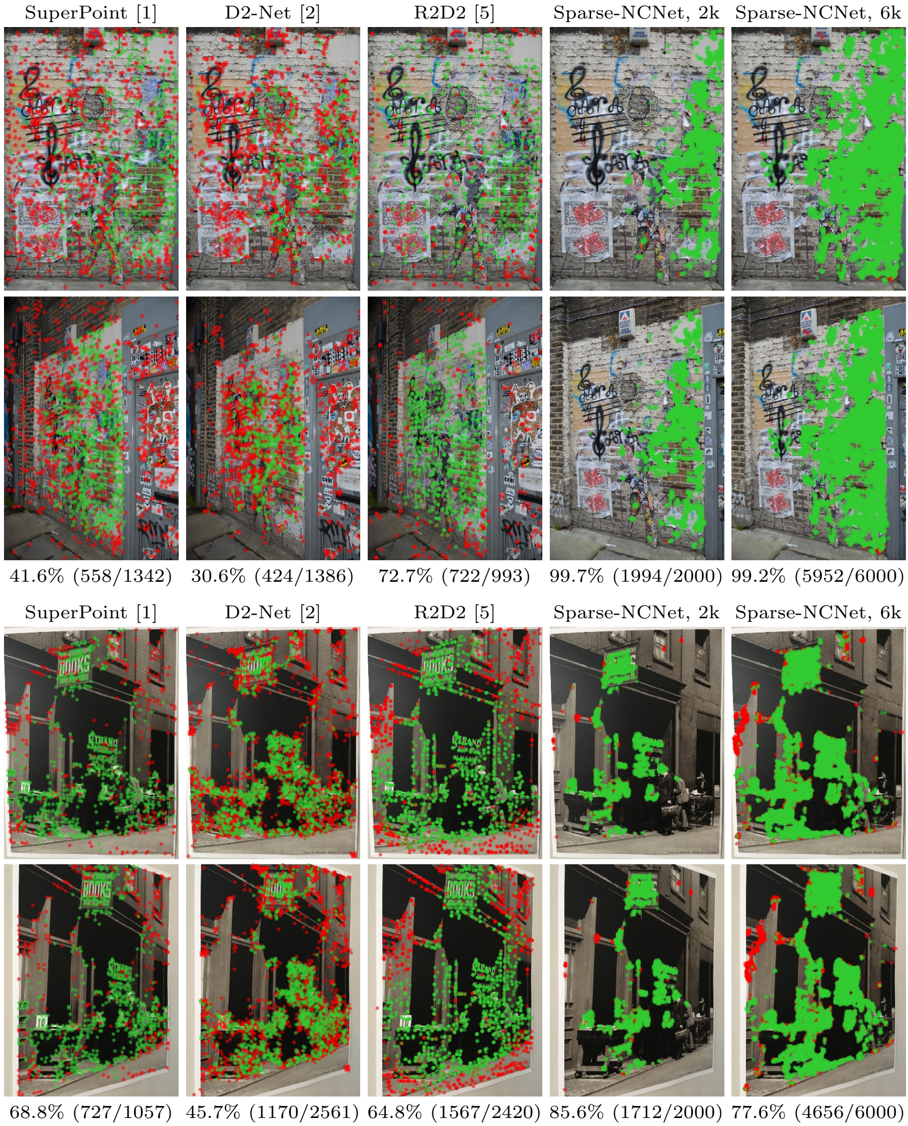


Fig. 1: **HPatches qualitative results (viewpoint).** We present the results of Sparse-NCNet, along with several state-of-the-art methods. The correct correspondences are shown in green, and the incorrect ones in red for a threshold $t = 3$px. Below each pair we indicate the fraction of correct matches (both in percentage and absolute values). Our method is presented for both the top 2K matches and the top 6K matches, and it obtains the largest fraction of correct matches for both cases. Examples are from the *viewpoint* sequences.

Fig. 2: **HPatches qualitative results (illumination).** We present the results of Sparse-NCNet, along with several state-of-the-art methods. The correct correspondences are shown in green, and the incorrect ones in red for a threshold $t = 3$px. Below each pair we indicate the fraction of correct matches (both in percentage and absolute values). Our method is presented for both the top 2K and top 6K matches, and it obtains the largest fraction of correct matches for both cases. Examples are from the *illumination* sequences.

2 InLoc benchmark

We present additional qualitative results from the InLoc indoor localisation benchmark [6] in Fig. 3, where the task is to estimate the 6-dof pose of a query image within a large university building. Each image pair is composed of a query image (top row) captured with a cell-phone and a database image (middle row), captured several months earlier with a 3D scanner. Note that the illumination conditions in the two types of images are different. Furthermore, because of the time difference between both images, some objects may have been displaced (*e.g.* furniture) and some aspects of the scene may have changed (*e.g.* wall decoration). For ease of visualisation, we overlay only the top 500 correspondences for each image pair, which appear in green. These correspondences have not been geometrically verified, and therefore contain a certain fraction of incorrect matches. Note however, that most matches are coherent and the few incorrect outliers are likely to be removed when running RANSAC [3] within the PnP pose solver [4], therefore obtaining a good pose estimate. Also note how Sparse-NCNet is able to obtain correspondences in low textured areas such as walls or ceilings, or on repetitive patterns such as carpets.

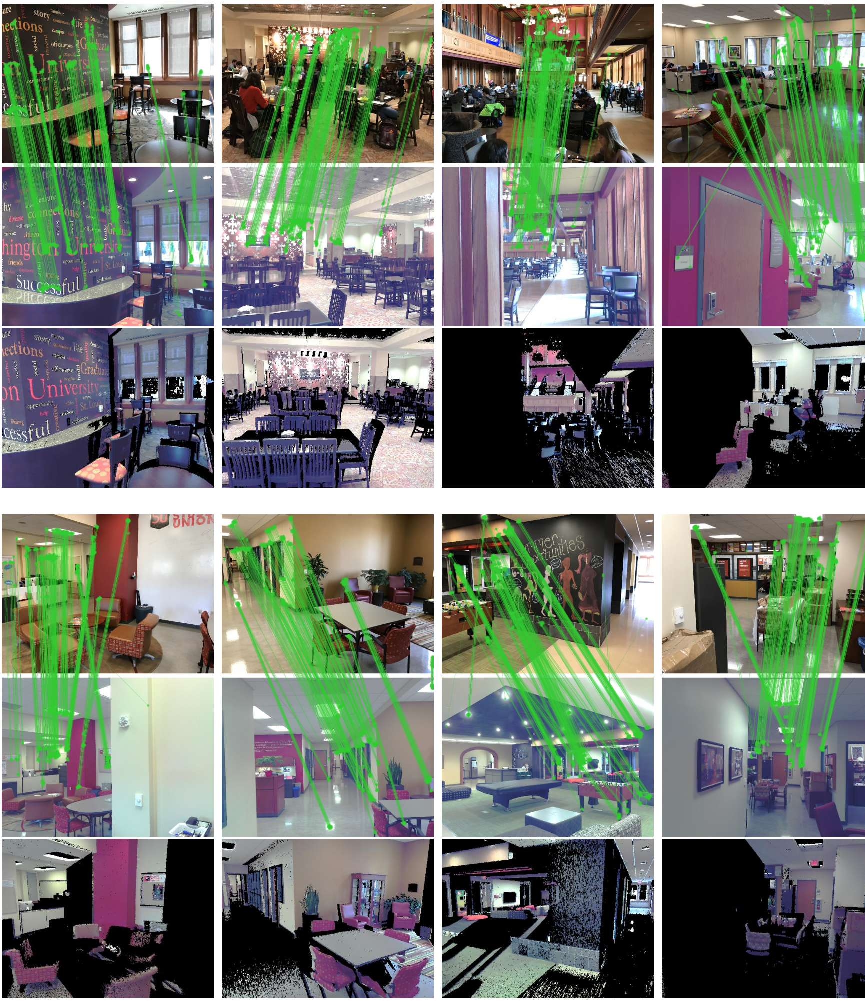


Fig. 3: **InLoc qualitative results.** For each image pair, we show the top 500 matches produced by Sparse-NCNet between the query image (top row) and database image (middle row). In addition we show the rendered scene from the estimated query 6-dof pose (bottom row), obtained by running RANSAC+PnP[3, 4] on our matches. Note these rendered images are well aligned with the query images, demonstrating that the estimated poses have low translation and rotation errors.

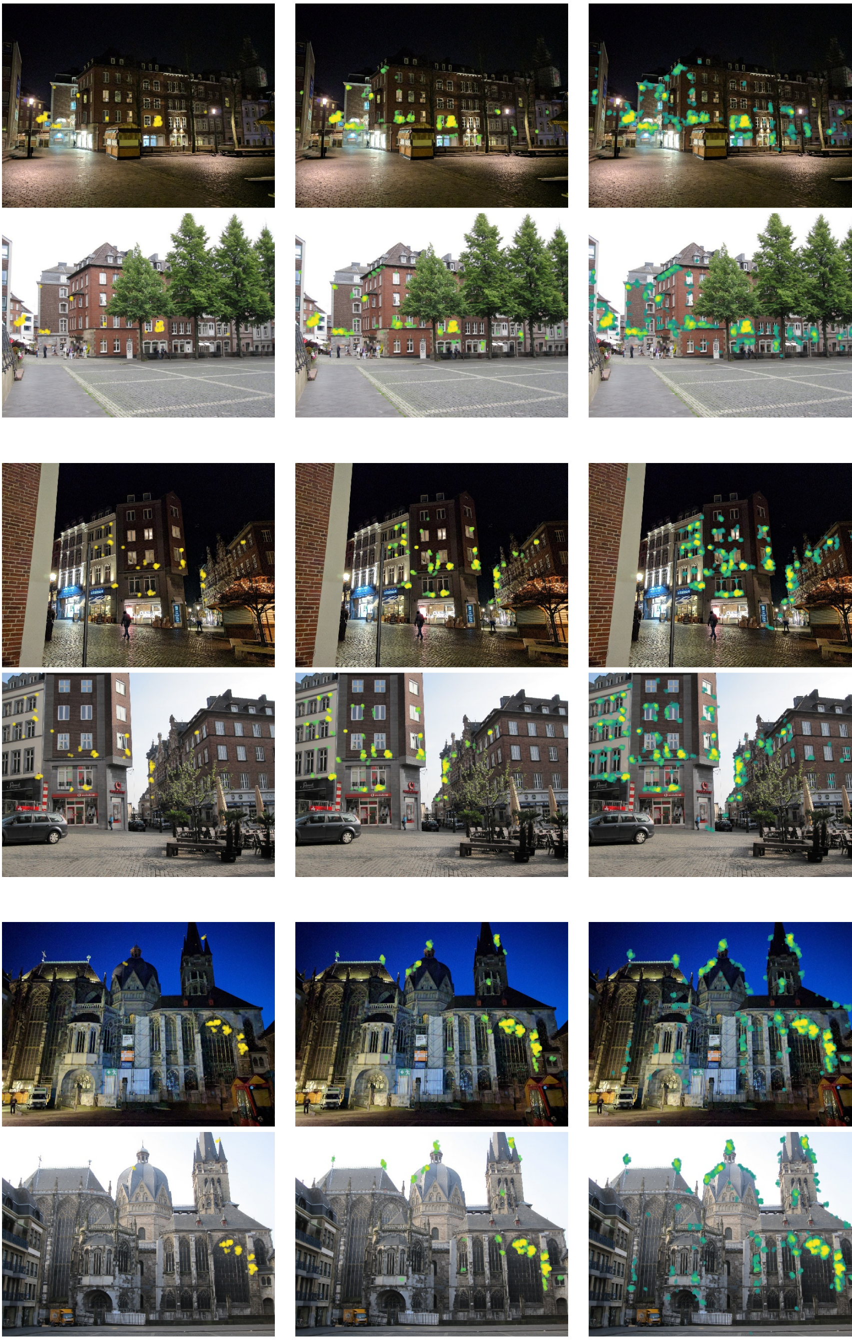
3 Aachen day-night benchmark

Additional qualitative results from the Aachen-day benchmark are shown in Fig. 4. We show several image pairs composed of night query images (top) and their top matching database images (bottom), according to the average matching score of Sparse-NCNet. For each image pair, we overlay the top 500 correspondences obtained with Sparse-NCNet. Note that these correspondences were not geometrically verified by any means. Nevertheless, as seen in Fig. 4, most correspondences are coherent and seem to be correct, despite the strong changes in illumination between night and day images.

Fig. 4: **Aachen day-night results.** We show the top 500 correspondences obtained by Sparse-NCNet between the night query image (top) and the database day image (bottom). Note that the large majority of matches are correct, despite the strong illumination changes.

4 Insights about Sparse-NCNet

In this section we provide additional insights about the way Sparse-NCNet operates, which differs from traditional local feature detection and matching methods. In Fig. 5 we plot the top N matches produced by Sparse-NCNet for different values of N: 100 (left column), 400 (middle column) and 1600 (right column). By comparing the middle column (showing the top 400 matches) with the left column (showing the top 100), we can observe that many of the additional 300 matches are close to the initial 100 matches. A similar effect is observed when comparing the right column (top 1600 matches) with the middle column (top 400 matches). This could be attributed to the fact that Sparse-NCNet propagates information from the strongest matches to their neighbours. In this sense, strong matches, which are typically non-ambiguous ones, can help in matching their neighbouring features, which might not be so discriminative.



$N = 100$ $N = 400$ $N = 1600$

Fig. 5: **Insights about Sparse-NCNet.** We show the top N matches between each pair of images for different values of N. The strength of the match is shown by color (the more yellow the stronger). Please note how new matches tend to appear close to high scoring matches, demonstrating the propagation of information in Sparse-NCNet.

References

1. DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-Supervised Interest Point Detection and Description. In: CVPR Workshops (2018)
2. Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T.: D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. In: Proc. CVPR (2019)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
4. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. IEEE PAMI **25**(8), 930–943 (2003)
5. Revaud, J., Weinzaepfel, P., de Souza, C.R., Humenberger, M.: R2D2: repeatable and reliable detector and descriptor. In: NeurIPS (2019)
6. Taira, H., Okutomi, M., Sattler, T., Cimpoi, M., Pollefeys, M., Sivic, J., Pajdla, T., Torii, A.: InLoc: Indoor visual localization with dense matching and view synthesis. In: Proc. CVPR (2018)