

# SUPPLEMENTARY MATERIALS

Paper-ID [183]

We have added the following experiments and results based on the reviewers' comments:

- Search for appropriate  $\beta$  values to eliminate background texture in Fourier transform-based domain augmentation. This experiment will be added to part E of Section IV. Experiments.
- Comparison to the DISP (GAN-based local feature learning method). This experiment will be added to part D of section IV. Experiments.
- Comparisons of keypoints matching results with baseline method using Aachen Day-Night actual image datasets.

## A. Experiment on selecting $\beta$ values

As demonstrated by Equation 9,  $\beta=0$  will produce the same result as the original source image. In contrast, when  $\beta$  equals 1, the result's amplitude is replaced by that of the target image. Experiments were conducted in order to determine a suitable  $\beta$  value. First, we tested images with varying  $\beta$  values and observed that when  $\beta$  was greater than or equal to 0.03, image artefacts became obvious, as shown in Fig. 1. Subsequently, we sampled numerous  $\beta$  values (0.01, 0.015, 0.02, and 0.025) less than 0.03 to train the network (baseline with Fourier transform-based domain augmentation). We use the open-source code of ASLFeat (CVPR) as the baseline. The performance of models trained with  $\beta$  values 0.01, 0.015, and 0.02 are comparable, the performance of the model trained with  $\beta$  value 0.025 was lower. The test results on Image Matching task with HPatches are shown in Table I. For each iteration in the training process, one of the three  $\beta$  values (0.01, 0.015, or 0.02) was chosen at random. This  $\beta$  selection strategy is employed for all subsequent experiments.

TABLE I. EVALUATION RESULTS OF LOCAL FEATURES LEARNED WITH VARIOUS  $\beta$  VALUE ON HPATCHES

$\beta$ value	Precision	Recall	MMA	Homo.
baseline	72.27	63.63	70.83	73.51
0.01	<u>73.20</u>	<b>71.87</b>	71.22	74.22
0.015	73.16	<u>71.73</u>	71.19	<b>74.34</b>
0.02	<b>73.26</b>	71.70	<u>71.24</u>	74.28
0.025	73.08	71.66	71.15	73.94

TABLE II. RESULTS ON VISUAL LOCALIZATION FOR DIFFERENT DISTANCE AND ORIENTATION THRESHOLDS

Method	day			night		
	0.25m 2°	0.5m 5°	5m 10°	0.25m 2°	0.5m 5°	5m 10°
NV+DISP	79.2	87.4	93.9	62.2	72.4	81.6
NV+SADGFeat (data)	80.4	86.8	94.2	63.8	71.8	83.0
NV+SADGFeat (data&align)	82.7	88.4	97.5	68.6	73.4	85.8

## B. Comparison with GAN based learning method

Since the GAN-based local feature learning method DISP [30] mentioned in our paper is not open-source, we conducted experiments using the same experimental pipeline as DISP and compared the results on the Aachen Day Night dataset to those of DISP [30]. We employ the same efficient hierarchical localization concept proposed in [62]. Using NetVLAD (Abbreviated NV) [63] as a global image descriptor, we first locate candidate images by matching the query against images in the database. The 6-DoF camera



Figure 1. Examples of images after Fourier transform-based domain augmentation with different  $\beta$  values. When  $\beta$  value is greater than or equal to 0.03 the image artefacts became obvious.



Figure 2. Examples for local feature matching for image pairs taken from Aachen Day-Night datasets. The left column displays the baseline's matching results, while the right column displays that of ours

pose is then calculated by matching the query image's local features to those of the retrieved candidates. We compare it to two versions of our proposed features, one of which employs only Fourier transform-based data augmentation during representational learning and the other of which employs domain augmentation and feature distribution alignment strategies. The results of the comparison are shown in Table II.

As observed, the proposed local feature only with Fourier transform-based domain augmentation for representational learning is comparable to that of DISP, indicating that Fourier transform-based domain augmentation can also increase domain diversity without requiring additional time-consuming GAN network training. The addition of a domain feature alignment strategy results in further improvement, proving the validity of the proposed strategy.

### C. Comparison of Keypoints Matching on real images

In the visual localization task with the Aachen Day-Night datasets, we illustrate local feature matching for image pairs in Fig. 2. It can be seen that our proposed local feature gets more matching keypoints pairs than that of the baseline under environment changing conditions. It is extremely advantageous for the long-term localization of robots in the actual world.

