

1 **Paper 1932.** We thank the reviewers for their work and feedback. We first address the general comments
2 A-D related to the main contributions from **R1**, **R2**, **R4** and then the specific ones.

3 **A. R1, R2.** Many works have empirically shown that L2 normalisation improves the matching performance
4 of local features, however, our work is the first to reveal and give insight into why L2 normalisation and its
5 gradients are important in the descriptor training and testing. Furthermore, based on these insights, we take
6 advantage of the underlying mechanism of L2 normalization and further improve the matching performance
7 by proposing a novel loss function. Therefore, we agree with **R1** that investigating this problem from the
8 perspective of the gradients is our strongest and novel contribution.

9 **B. R1.C3, R2, R4.** Following our gradient analysis, we design a new architecture that applies L2
10 normalisation in the intermediate feature maps. Thus, our work is the first local descriptor that proposes and
11 displays the benefit of using L2 normalisation within the model and not just on the output as in [21,22]. The
12 new components, *i.e.*, the loss function and the network architecture with FRNs, result from the presented
13 analysis and establish a new state-of-the-art in several benchmarks when used together. Also, see our response
14 to **R4**.

15 **C. R1, R2.** Previous methods have explored different directions such as data sampling (mining), optimisation
16 of descriptor space distribution, or heuristic designs of loss functions. In contrast, we tackle it from a novel
17 perspective by making better use of the gradients, which is validated by the superior results.

18 **D. R1, R2.** We believe that our work will have a wide impact in other areas, where feature embedding and
19 matching also rely on L2 normalisation but so far lacked theoretical support, *e.g.* in person re-identification,
20 or face recognition.

21 **R1.C1 Effect of L2-regularisation term (R_{L2}).** The test set of HPatches contains 2M samples (1M
22 positives and 1 M negatives), thus given this large number of samples a 0.39 MAP increase is not marginal
23 (See response to **R4**). Moreover, we further investigate the matching MAP increase brought by R_{L2} in
24 HPatches and find an increase of 0.57 and 0.21 for illumination and viewpoint, respectively. It indicates that
25 our regularisation on the descriptor norm makes the network more robust to illumination variations in the
26 input.

27 **R1.C2 Hyper-parameter α .** Despite the sensitivity, comparing Fig. 5(a) and Fig. 4 (matching MAP
28 in the middle), we can see that all choices of α perform either on par or better than the previous SOTA
29 methods (HardNet, SOSNet). Moreover, such peakedness in the MAP curve by varying α indicates that there
30 is an optimal value of α that can well capture the distribution of the training set.

31 **R4. Architecture and loss.** The improvements brought by individual components are discussed in lines
32 235-241. The gain from the new similarity (s_H) is up to +1.87 MAP. The gain from the new architecture
33 (FRN+TLU) is +1.5 MAP. Note that both new components benefit each other and lead to higher gains when
34 used together.

35 New results will be added to Table 3 to further expose the improvements, namely HardNet+FRN: 51.89 (+1.33
36 from the original HardNet+BN) and SOSNet+FRN: 52.12 (+1.10, from the original SOSNet+BN). Note
37 that the increase of HyNet from BN to FRN is +1.93, indicating that our new loss is more compatible with
38 the FRN layer and that both together take better advantage of the gradients.

39 The baseline L2Net+proposed loss is indeed ($BN + s_H + R_{L2}$) in Table 3. Under the evaluation protocol
40 of the HPatches dataset with a large number of test samples (1M positives and 1M negatives), 1.0 MAP
41 is significant. The improvement of the previous SOTA SOSNet over HardNet was 0.96 MAP. Finally, we
42 acknowledge that the second order regularisation term (SOSR) from SOSNet is beneficial. Table 1 and Fig. 1
43 in supplementary material show that SOSR further improves the matching MAP of HyNet by 0.15.

44 **R4. L161 (L2 normalisation within the architecture).** The analysis in Sec. 2.2 shows the benefits of
45 L2 normalisation. We implement it per feature map by the FRN layer, hence, no further L2 normalisation is
46 needed inside the architecture. We will clarify in Sec. 2.2 where and how the normalisation occurs in the
47 network.

48 **R4. Similarity measure.** We mentioned on page 3 footnote that L2 is a distance metric (or inverse
49 similarity) but we agree about possible confusion and will change the notation.

50 **R1. Color coding.** We will highlight the top results with color.