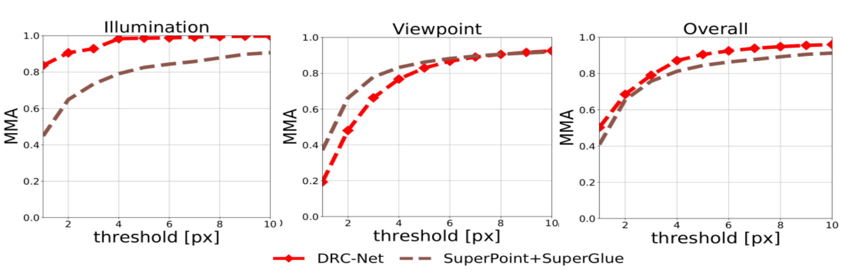
We appreciate all reviewers for their valuable comments and confirming the simplicity of our design and the repeatability of our experiments. We address the main concerns below.

R1,R2,R3: SuperGlue & OANet First of all, the Aachen Day-Night updated its ground truth after our submission. We now re-evaluate our method against all baselines and include the SuperPoint+SuperGlue (SP+SG) and ASLFeat+OANet (upper left table). We also compared with SP+SG on HPatches using the standard MMA metric (the bottom right figures) and the overall MMA of DRC-Net is significantly higher. On Aachen, DRC-Net performs comparably well with SP+SG. By adding Orthogonal Loss [21], the accuracy under low error threshold is improved. We believe injecting SuperPoint into our framework will further boost DRC-Net. Note that SP+SG and ASLFeat+OANet were published in CVPR20 (after our submission). DRC-Net was SOTA during submission. **R1,R2: Comparison on InLoc in table** The plots of InLoc intended to emphasise the robustness and stability of all methods, and we also provide a comparison in table (lower left table). **R1,R2: MMA for fair comparison** We believe it is fair because we follow the identical evaluation protocol as [5,6]. As described in [6], mutual NN is applied on other description-matching methods to obtain about 1k matches to ensure a comparable number of matches. Therefore, MMA is evaluated on nearly an identical number of matches for fair comparison. **R1,R2: Experiments on more datasets** We follow the suggestion to evaluate on Aachen v1.1 and the results are 71.2/86.9/97.9. We will include more results in the final version. **R1: Negative scores and softmax** The ReLU layers are employed in neighbourhood consensus module, hence it is guaranteed the output scores at the coarse level are non-negative, thus adding softmax becomes optional. We choose to switch off softmax as we found softmax slows down the training convergence, possibly because of reduced gradient after softmax. **R2,R3: Novelty** DRC-Net is inspired by NCNet but significantly different. DRC-Net tackles the scalability issue of dense matching with the subtle design of dual-resolution feature framework, which can effectively make use of feature maps of different resolutions, substantially outperforming all neighbourhood consensus based methods. **R2: Same training principle as Sparse-NCNet** Our training principle is different from Sparse-NCNet. Sparse-NCNet is supervised by image level annotations, while DRC-Net is supervised by sparse keypoint annotations. The training losses are different as well, which will be clarified in the final version. **R2: Why 1024 channels** The use of 1024 channels is inherited from [5,6]. We also find that using 256 channels in our model can provide comparable (slightly inferior) accuracy which has been reported in Fig. 4 in supplementary. **R2: Insignificance of reporting performance over large error band** The performance over large error band represents the stability and robustness. It is a common practice to plot up to 2m for InLoc [24,5,6] and 10 pixel for HPatches [6,7,8]. This is meaningful because the relative errors of 10 pixel are about 1% in HPatches and less than 10% at 1m for InLoc. Our method is superior than baselines in these circumstances. See lower left table for details. **R2,R4: Notation and "mask"** (i', j') is for coarse-level and (i, j) is for fine-level coordinates. We will further clarify. We use "mask" to indicate that some fine-resolution scores would be zeroed by coarse-resolution scores since the ReLU layers enable zeros in 4D tensor. This can be an analogue to binary mask. We follow the notation convention used in [5,6,21] to use $ijkl$ to index a correlation score. **R2,R4: Why train on MegaDepth, not on IVD** Training on MegaDepth and testing on the localisation datasets for establishing correspondence has been successfully adopted in literatures (See [7,8,40] and S2DNet by Germain et al 2020) for MegaDepth contains rich viewpoint and illumination variations. Testing on other datasets with standard 3D reconstruction pipelines allows fair comparison with a large number of baselines. IVD is a popular alternative, however, IVD lacks of the sparse pixel-wise annotation required to train DRC-Net. **R3,R5: Runtime and memory** We follow the suggestion and evaluate the runtime/GPU memory on a fixed feature map size 200×150 for three methods, the average processing time per image pair are 2.05s/0.82s/4.15s by DRC-Net/Sparse-NCNet/NCNet with GPU memory cost of 1232MB/680MB/7868MB respectively. All three methods are evaluated on a GTX 1080Ti GPU. **R3,R4: Performance in illumination** Please refer to sect. B supplementary. It will be included in the main paper in the final version. **R3: Qualitative results and failure** Please refer to sect. C in supp. Failure cases will be included and discussed in the final version. **R4: Including non-isotropic filters** We have tried to include similar adaptive module as [21] into our framework, but no obvious gain is observed possibly because the only feasible non-isotropic filters is small and hence inadequate to deal with strong perspective scale variation. **R5: NC module configuration** We use the same configuration as NCNet (as mentioned in line 206). It will be further clarified. **R1,R2,R5: Typos and figures** Thanks, all will be fixed.

Method	Correctly localized queries (%) 0.25m, 2°	0.5m, 5°	5m, 10°
SP+SG	**79.6**	**90.8**	**100.0**
Sparse-NCNet	76.5	84.7	<u>98.0</u>
ASLFeat+OANet	77.6	<u>89.8</u>	**100.0**
DRC-Net w/o orth.	<u>78.6</u>	88.8	**100.0**
DRC-Net w/ orth.	**79.6**	88.8	**100.0**

Method	Correctly localized queries (%) 0.25m	0.5m	1m
IL+DRC-Net	43.2	<u>67.5</u>	**83.0**
IL+DRC-Net (Type(e))	**48.0**	**68.1**	**83.0**
IL+D2-Net	43.2	61.1	74.2
IL+Sparse-NCNet	<u>45.4</u>	66.2	<u>79.9</u>



Upper left table: Aachen Day-Night. Lower left table: InLoc. Three figs on the right: DRC-Net vs SP+SG on HPatches.