Table 1: Ablations on Megadepth-1500. Relative pose estimation, measured in AUC (higher is better). The last section of the table contain ablations described in Section A.

Method ↓	AUC →	@5°	@10°	@20°
DeDoDe-B [2]		49.4	65.5	77.7
Steerers-B-C4-Perm [1]		51	67	79
AffEqui-B		46.1	62.4	74.8
AffSteer-B		**52.7**	**68.9**	**81.0**
DeDoDe-G [2]		52.8	69.7	82.0
Steerers-G-C4-Perm [1]		52	69	81
AffEqui-G		50.6	67.2	80.1
AffSteer-G		**53.7**	**70.0**	**82.1**
DeDoDe-B + affine augmentation w/o steerer		40.8	57.0	70.3
AffEqui-B with pretraining		47.5	63.4	75.6
AffSteer-B without pretraining		47.2	63.7	76.2
AffEqui-G with pretraining		51.7	68.0	80.3
AffSteer-G without pretraining		50.7	67.6	80.1

A Ablation study on MegaDepth-1500

We present a couple of ablations on the MegaDepth-1500 test set [3, 5] in Table 1. First of all, we train DeDoDe-B on MegaDepth with the same affine augmentations as we train *AffEqui*, but this time without a steerer. We find that this deteriorates results by a large margin compared to the baseline DeDoDe-B without a steerer. This corroborates the finding in [1] that training with large augmentations requires the addition of a steerer. Secondly, we train *AffEqui* with homography pretraining and *AffSteer* without homography pretraining and find as explained in the main text that pretraining matters most for *AffSteer*.

B Experiment details

We use the same hyperparameters as in DeDoDe [2], with the exception of matching parameters. Since we use negative L2-distance instead of cosine similarity for measuring description similarity, the similarity of two descriptions is unbounded from below. For this reason, we found that lowering the inverse temperature from 20 to 5 was necessary in order to not exaggerate low similarities.

For consistency with [1], we use their DeDoDe-C4 detector in our experiments. An exception from this is the experiment on AIMS, where we use the DeDoDe-SO2 detector as do [1]. For the datasets where there are no reported numbers in [2] (*i.e.* Hpatches and WxBS), we use the DeDoDe-C4 detector for the DeDoDe baseline as well.

When training *AffSteer*, we pretrain for 50k iterations and finetune for 50k iterations, yielding the same total number 100k of training iterations as for *AffEqui* and DeDoDe.

B.1 Runtime

Computing similarities multiple (three) times does incur additional runtime—matching for 10k DeDoDe keypoints: 40ms, for AffSteer: 60ms—but this is low compared to the time for detection and description: 240ms for size B and 610ms for G. (On an RTX3080 Ti GPU.)

C Using the pipeline to train XFeat

XFeat [4] is a recent keypoint detector/descriptor which was developed for minimizing computational cost. To show that the end-to-end pipeline for training upright descriptors that we have proposed does not only work for the DeDoDe architecture, we evaluate using the pipeline for training the descriptor part of XFeat. *I.e.*, we pre-train with affine steering and finetune using the max similarity method. We evaluate with 30k DeDoDe keypoints on MegaDepth. For the baseline, we also use 30k DeDoDe keypoints (giving slightly higher scores than in the original XFeat paper's sparse results, which only used 4k keypoints). We use LO-RANSAC as in the XFeat paper. As seen in Table 2, a substantial improvement is obtained by training using our pipeline.

Table 2: Training the XFeat descriptor in our pipeline improves it significantly.

	AUC@5°	AUC@10°	AUC@20°
XFeat	43.2	57.9	69.1
AffXFeat (ours)	48.6	62.3	72.8

References

1. Bökman, G., Edstedt, J., Felsberg, M., Kahl, F.: Steerers: A framework for rotation equivariant keypoint descriptors. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2024)
2. Edstedt, J., Bökman, G., Wadenbäck, M., Felsberg, M.: DeDoDe: Detect, Don't Describe – Describe, Don't Detect for Local Feature Matching. In: 2024 International Conference on 3D Vision (3DV). IEEE (2024)
3. Li, Z., Snavely, N.: Megadepth: Learning single-view depth prediction from internet photos. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR). pp. 2041–2050 (2018)

4. Potje, G., Cadar, F., Araujo, A., Martins, R., Nascimento, E.R.: Xfeat: Accelerated features for lightweight image matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2682–2691 (2024)
5. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8922–8931 (2021)