A. Appendix

A.1. Hyperparameter Ablations

In equation 10, we defined detector-weighted similarity matrices as:

$$\bar{\mathbf{S}}_{ij}^{ST} = \left(\frac{\mathbf{w}_i^S}{\tau_s}\right)\mathbf{S}_{ij}^{ST}\left(\frac{\mathbf{w}_j^T}{\tau_t}\right) \qquad \bar{\mathbf{S}}_{ij}^{TT} = \left(\frac{\mathbf{w}_i^T}{\tau_t}\right)\mathbf{S}_{ij}^{TT}\left(\frac{\mathbf{w}_j^T}{\tau_t}\right)$$

$\tau_\mathbf{s}$		1	0.5	0.1	0.5
$\tau_\mathbf{t}$		1	0.5	0.1	0.1
HPatches	$\epsilon = 1$	0.59	**0.60**	0.57	0.58
	$\epsilon = 3$	0.82	**0.84**	082	0.81

Table 3. Ablation study of the temperature parameters $\tau_\mathbf{s}$ and $\tau_\mathbf{t}$ used in the Joint Detector–Descriptor Distillation loss.

This formulation allows us to control the influence of the student and teacher detector confidences in the joint similarity space $\bar{\mathbf{S}}$. Table 3 investigates the impact of the temperature parameters $\tau_\mathbf{s}$ and $\tau_\mathbf{t}$ on performance. We observe that decreasing the temperature values (which corresponds to increasing the influence of the detector confidences) from 1.0 to 0.5 generally improves performance. However, performance begins to decline when the temperature is decreased further (e.g., below 0.5). Furthermore, the use of different temperature values for the student ($\tau_\mathbf{s}$) and teacher ($\tau_\mathbf{t}$) confidences does not appear to offer any significant benefit.

In Equation 13, we defined our total training loss as a weighted combination of the geometric matching loss ($\mathcal{L}_{\text{match}}$) and the joint detector–descriptor distillation loss ($\mathcal{L}_{\text{KD}}$), balanced by the factor λ_{KD}. Table 4 presents an ablation study on the impact of this weighing factor. When $\lambda_{\text{KD}} = \mathbf{0}$, only the matching loss is utilized, yielding results identical to the $\mathcal{L}_{\text{match}}$-only case reported in Table 2. As λ_{KD} is increased, performance steadily improves, reaching maximum performance at $\lambda_{\text{KD}} = \mathbf{2}$. Further increasing λ_{KD} shifts the balance toward the distillation loss, leading to results that approach those of the $\mathcal{L}_{\text{KD}}$-only case presented in Table 2.

λ_{KD}	**HPatches** HEA		**ScanNet** RP-AUC	
	($\epsilon = 1$)	($\epsilon = 3$)	@10°	@20°
0	0.53	0.70	21.6	35.8
1	0.54	0.79	29.5	45.0
2	**0.59**	**0.83**	**31.5**	**48.5**
4	0.57	0.81	30.0	47.0

Table 4. Analyzing the impact of λ_{KD} on HPatches and ScanNet Datasets. We report Homography Estimation Accuracy (HEA) for HPatches and Relative Pose Prediction AUC (RP-AUC) for ScanNet.

Param	**HPatches**		**HPatches**	
Residual?			✓	
	($\epsilon = 1$)	($\epsilon = 3$)	($\epsilon = 1$)	($\epsilon = 3$)
0.13M	0.60	0.84	0.60	0.83
0.8M	0.59	0.83	0.59	0.83

Table 5. Analyzing the impact of adding residual connections.

A.2. Exploring Different Architectures

While the main paper presented results for four model sizes (0.12M, 0.08M, 0.06M, and 0.04M parameters), Figure 6 provides an extended analysis incorporating a wider range of model complexities, including much smaller (0.02M and 0.005M) and larger (0.25M and 0.5M) models. We observe that for model sizes of $\mathbf{0.25}$M parameters and above, the performance of both the symmetric and Asymmetric pipelines closely approximates that of the full Teacher model, with the Asymmetric approach exhibiting a slight advantage. As the parameter count is reduced, the Asymmetric pipeline proves significantly more robust, retaining performance much better than its symmetric counterpart. Specifically, the Asymmetric approach maintains performance close to the Teacher's down to $\mathbf{0.04}$M parameters, but then begins to show a sharp decline at $\mathbf{0.02}$M parameters, though it still outperforms the symmetric pipeline.

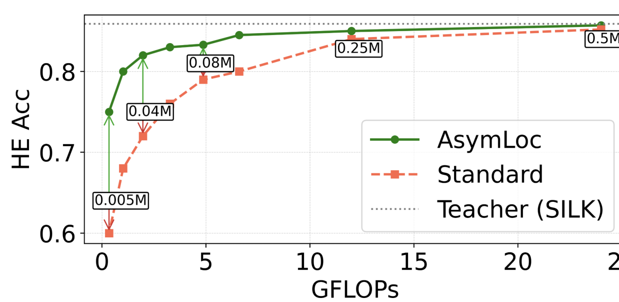

Figure 6. Homography estimation accuracy on HPatches with a wide range of model sizes. Here we use SILK as the teacher.

Table 5 presents an analysis of the impact of incorporating residual connections into our pipeline. We find that the addition of these connections offers no major performance advantage, likely due to the fact that our pipeline utilizes only small CNN, which do not typically suffer from the vanishing gradient issues that residual connections are designed to mitigate in deeper architectures.

A.3. Training Details

Our pipeline was trained using the $\mathbf{Adam}$ optimizer with a base learning rate of $\mathbf{0.001}$ and standard momentum settings ($\beta_1 = \mathbf{0.9}, \beta_2 = \mathbf{0.999}$). To enhance robustness, we employed a comprehensive suite of data augmentation techniques (adopted from [20]). This suite included color and exposure manipulations such as Random Gamma and Hue, Saturation, and Value shifts, various blurring effects including standard Blur and Motion Blur, and Gaussian Noise injection. Furthermore, we applied Random Brightness and Contrast adjustments to broaden the model's tolerance to varying lighting conditions.

A.4. Additional Baseline Ablations

Table 1 presents our main results, demonstrating that asymmetric matching outperforms standard symmetric matching. We further compare our method against other asymmetric baselines, showing that our formulation yields superior performance. Due to space constraints, the main analysis focuses on the 0.13M parameter models. For completeness, we report results for additional model capacities (0.08M and 0.06M) on HPatches and ScanNet in Tables 6 and 7.

		(0.08M)		**(0.06M)**	
Asym?	**Technique**	($\epsilon = 1$)	($\epsilon = 3$)	($\epsilon = 1$)	($\epsilon = 3$)
✗	Standard	0.55	0.79	0.52	0.76
✗	Naive Distillation	0.54	0.76	0.50	0.77
✓	AML	0.56	0.81	0.53	0.78
✓	RKD	0.55	0.80	0.52	0.77
✓	CSD	0.56	**0.83**	0.55	0.80
✓	D3still	0.56	0.80	0.54	0.80
✓	Ours	**0.59**	**0.83**	**0.58**	**0.83**

Table 6. **Homography estimation accuracy on HPatches (0.08M and 0.06M).**

		(0.08M)		**(0.06M)**	
Asym?	**Technique**	@10°	@20°	@10°	@20°
✗	Standard	27.6	44.6	24.2	38.9
✗	Naive Distillation	26.3	44.8	24.2	39.9
✓	AML	28.3	44.9	26.7	40.9
✓	RKD	28.6	45.1	25.8	42.9
✓	CSD	29.5	46.1	28.6	44.4
✓	D3still	28.9	46.5	28.6	45.3
✓	Ours	**31.5**	**48.5**	**31.0**	**47.4**

Table 7. **Relative pose estimation accuracy on ScanNet (0.08M and 0.06M).**

A.5. Speed vs Accuracy

We report additional latency analysis in Figure 7, plotting FPS vs. Homography Estimation Accuracy on HPatches using SILK as the teacher model. This analysis is done on an NVIDIA RTX A5000 GPU.

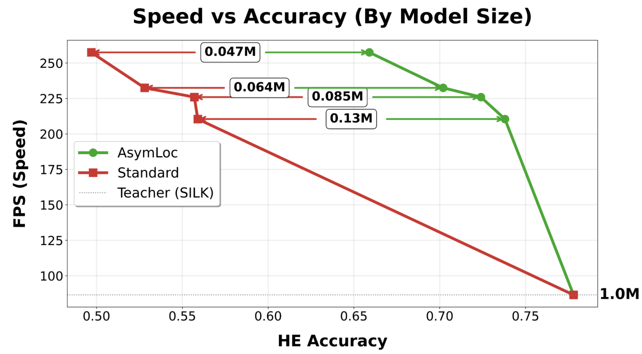

Figure 7. FPS Comparison on HPatches with SILK Teacher

A.6. Additional Results with XFeat

XFeat is trained on both COCO homography (like us) as well as MegaDepth-v1. For thoroughness, we evaluate two asymmetric configurations: (1) **XFeat-Mega**, using the official pre-trained weights as the teacher and training the student models on both COCO and MegaDepthv1; (2) **XFeat-COCO**, training XFeat teacher and students solely on COCO homography. We show results on both ScanNet as well as MegaDepth-1500. Results indicate that the asymmetry consistently outperforms the symmetric baseline by a wide margin, and that asymmetry nears the oracle teacher performance.

	Method	**Asym?**	**# params** (online)	**# params** (offline)	**ScanNet** Rel. Pose AUC **@10°**	**@20°**	**MegaDepth-1500** Rel. Pose AUC **@10°**	**@20°**
XFeat (Mega)	**Backbone (0.17M)**							
	Standard	✗	0.17M	0.17M	25.2	39.3	66.5	72.4
	AsymLoc (Ours)	✓	0.17M	1.5M	30.2	44.4	71.3	78.6
	Oracle Teacher Performance							
	XFeat (Teacher) (Pre-trained)	✗	1.5M	1.5M	32.2	46.8	75.2	81.9
XFeat (COCO)	**Backbone (0.17M)**							
	Standard	✗	0.17M	0.17M	24.0	38.5	53.7	65.9
	AsymLoc (Ours)	✓	0.17M	1.5M	29.2	43.2	60.1	71.1
	Oracle Teacher Performance							
	XFeat (Teacher)	✗	1.5M	1.5M	31.3	46.6	65.1	73.7

Table 8. Results with XFeat on ScanNet and MegaDepth