Image fidelity and geometric accuracy in paired restoration–matching pipelines

Congshuai Dong, Wenxue Zhang, Yiying Zhang, Hao Zhang and Yiteng Teng

School of Science, Shandong Jianzhu University, Jinan 250101, China
Correspondence to Congshuai Dong dong_cs@qq.com

S1 Paired observations and controls

The original experiment used 60 HPatches pairs from 12 sequences (Balntas et al., 2017). Six sequences were selected at evenly spaced positions in each of the sorted illumination and viewpoint lists. Within each sequence, the first image was paired with each of the five remaining views. On MegaDepth (Li and Snavely, 2018), we took the first 15 non-overlapping pairs in sorted order in each of scenes 0015 and 0022. Selection used no matching outcomes.

We converted both views to grayscale and limited the longer side to 384 pixels. Gaussian noise with standard deviation $\sigma \in \{10, 25, 50\}$ in 8-bit units was added before clipping and rounding. The noise realizations used seeds 20261024, 20261025 and 20261026. Interval estimation uses separate, fixed resampling seeds recorded in the analysis scripts; the values given below apply to the specified intervals. Each restorer received the same corrupted arrays. *Raw* passes these arrays directly to the matcher. The clean control passes the resized images without corruption or restoration. Table S1 gives their absolute success rates.

Table S1 Geometric success (%) before restoration. Each noisy entry averages three realizations per pair. NN denotes mutual nearest-neighbour matching and LG denotes SIFT–LightGlue

Dataset	Matcher	Clean	Raw, $\sigma = 10$	Raw, $\sigma = 25$	Raw, $\sigma = 50$
HPatches	NN	80.0	66.1	49.4	23.9
	LG	90.0	85.0	70.6	48.3
MegaDepth	NN	46.7	21.1	8.9	2.2
	LG	63.3	44.4	36.7	12.2

Restoration and matching. The Gaussian filter used a 5×5 kernel with spatial standard deviation 0.8 px. Non-local means (NLM) used $h = 0.8\sigma$, a 7×7 template and a 21×21 search window. We used the pretrained blind grayscale DnCNN without fine-tuning. Both matchers saw the same RootSIFT features, capped at 1,024 keypoints per image. NN required a ratio below 0.8 in both directions. LightGlue used a match threshold of 0.1 and depth and width confidence settings of 0.95 and 0.99.

Geometric endpoints. HPatches success required a mean error below 3 resized pixels over the four projected corners. MegaDepth success required both the rotation error and the sign-invariant translation-direction error to be below 10°. Failed fits, including poses with zero cheirality-consistent inliers, counted as unsuccessful.

Additional pipeline controls. BM3D receives the known noise level. ORB changes detection, binary description and matching together. Table S2 reports BM3D success and correspondence precision on the original RootSIFT grid, together with DnCNN success under the ORB pipeline. ORB clean success was 58.3 [43.3, 75.0]% on HPatches and 13.3 [3.3, 26.7]% on MegaDepth. These extensions use the original pairs and remain outside the original multiplicity family. Their cluster intervals use 2,000 draws; ORB clean-control intervals use seed 20261030.

Verification records. A separate package contains measurements, result tables, analysis and plotting scripts, and implementation records. The corresponding author provides it on request; benchmark images and pretrained models follow their providers' terms.

S2 Local analysis and pipeline controls

Image-domain error. Let x contain m noise-free pixel values and let $y = x + n$, with $\mathbb{E}[n \mid x] = 0$ and $\mathrm{Cov}(n \mid x) = \sigma^2 I$. Before clipping or quantization, a fixed linear filter W has error $Wy - x = (W - I)x + Wn$. Expanding the square and taking the conditional expectation gives

$$\frac{1}{m}\mathbb{E}[\|Wy - x\|_2^2 \mid x] = \frac{\|(W - I)x\|_2^2}{m} + \frac{\sigma^2}{m}\|W\|_F^2. \tag{S1}$$

The cross term vanishes. A wider averaging kernel can lower the noise term while increasing bias at an edge. Which effect dominates depends on the image and the noise level. For a realized squared error $u > 0$, PSNR is $10 \log_{10}(255^2/u)$ when u is the per-pixel MSE. We averaged PSNR over views. Equation (S1) concerns expected MSE and assumes a fixed filter, so it supplies no quantitative prediction for NLM or DnCNN.

A perturbed corner. Hold correspondence identities, inlier membership and normalization maps fixed. A normalized direct linear transform estimates the unit homography vector $h \in \mathbb{R}^9$ as the eigenvector of $B = A^\top A$ corresponding to its smallest eigenvalue λ_0, assumed to have multiplicity one. Perturbing the point coordinates changes the design matrix by δA. Differentiating $Bh = \lambda_0 h$ and $h^\top h = 1$, then projecting orthogonally to h, gives to first order

$$\delta h = -(B - \lambda_0 I)^\dagger (I - hh^\top)(A^\top \delta A + \delta A^\top A)h. \tag{S2}$$

Here † is the Moore–Penrose pseudoinverse. The inverse on that orthogonal subspace depends on the gap between the two smallest eigenvalues; the approximation requires perturbations small relative to that gap. Nearly degenerate point configurations can therefore amplify a small coordinate perturbation.

Let K_c map the stacked correspondence perturbation η to the linearized displacement of a projected corner c, including denormalization and division by the last homogeneous coordinate. For mean b and covariance Σ_η,

$$\mathbb{E}\|\delta u_c\|_2^2 \simeq \|K_c b\|_2^2 + \mathrm{tr}(K_c \Sigma_\eta K_c^\top). \tag{S3}$$

Here K_c is a fixed local Jacobian, and displacement is relative to the unperturbed fit, not to reference geometry. The same pixel MSE can accompany different K_c, biases and point covariances. The homogeneous projection denominator must stay away from zero. A denoiser may also remove keypoints or change which matches survive robust fitting. Those changes lie outside the expansion.

Evaluation coordinates. If S_0, S_1 resize the two views, then $u_1 \sim Hu_0$ implies $H_e = S_1 H S_0^{-1}$, and $K_{e,i} = S_i K_i$ for camera intrinsics. For features at an enlarged scale s, an output pixel centre q maps back to $p = (q + 1/2)/s - 1/2$ on each axis. The inverse map corrects coordinates. It leaves the detector responses obtained on the enlarged image intact.

Table S2 Changes from each pipeline's Raw control (pp). BM3D uses RootSIFT with NN or LightGlue (LG). DnCNN uses ORB with brute-force Hamming matching. Brackets give unadjusted 95% paired percentile intervals from resampling HPatches sequences or pairs within each fixed MegaDepth scene

		BM3D success		BM3D precision		DnCNN success
Dataset	σ	NN	LG	NN	LG	ORB
HPatches	10	+4.4 [−0.6, 8.9]	+5.6 [1.1, 11.1]	+1.8 [0.3, 3.5]	+6.4 [3.9, 9.2]	+8.9 [3.9, 14.4]
	25	+9.4 [3.3, 15.6]	+3.3 [−3.9, 11.7]	−0.2 [−3.4, 2.4]	+13.3 [10.6, 15.9]	+7.8 [3.3, 11.7]
	50	+14.4 [7.2, 23.3]	+18.9 [12.2, 26.1]	−3.6 [−6.6, −0.5]	+11.9 [9.5, 14.4]	+6.7 [0.0, 15.6]
MegaDepth	10	+11.1 [−1.1, 22.2]	+3.3 [−8.9, 14.4]	+0.7 [−1.7, 3.3]	+3.9 [1.6, 6.5]	+3.3 [−1.1, 10.0]
	25	+14.4 [4.4, 25.6]	+0.0 [−11.1, 11.1]	−4.4 [−8.7, −0.6]	+7.8 [4.6, 10.8]	+4.4 [−3.3, 11.1]
	50	+11.1 [4.4, 17.8]	+20.0 [11.1, 28.9]	−2.5 [−8.4, 2.9]	+13.7 [11.4, 16.2]	+2.2 [−1.1, 6.7]

S3 Gaussian filtering versus non-local means

Table S3 differences NLM and Gaussian measurements on shared images, seeds and matchers before resampling. Separate intervals against Raw do not give this direct interval.

Table S3 NLM minus Gaussian filtering on the original pairs. The PSNR difference is shared by the two matchers. Success differences are in percentage points (pp). Brackets give unadjusted 95% paired percentile intervals

Dataset	σ	PSNR (dB)	NN success (pp)	LG success (pp)
HPatches	10	+1.15 [0.25, 2.19]	−2.78 [−8.33, 1.67]	−2.78 [−7.78, 1.11]
	25	+0.79 [0.33, 1.19]	−7.22 [−15.56, 0.00]	−1.67 [−4.44, 0.57]
	50	+1.86 [1.19, 2.51]	+7.78 [1.11, 15.00]	−1.67 [−9.44, 6.11]
MegaDepth	10	+2.00 [1.74, 2.28]	−10.00 [−20.03, 0.00]	−3.33 [−13.33, 6.67]
	25	+0.57 [0.39, 0.75]	+5.56 [−4.44, 15.56]	−6.67 [−18.89, 4.44]
	50	+1.58 [1.38, 1.77]	−1.11 [−7.78, 4.44]	−1.11 [−10.00, 8.89]

NLM had the higher mean PSNR in every stratum. Its mean success was lower in 10 of 12 strata, with zero included in all 10 corresponding success intervals. The upper limits for HPatches NN at $\sigma = 25$ and MegaDepth NN at $\sigma = 10$ are exactly zero. The exceptions were HPatches with NN at $\sigma = 50$ and MegaDepth with NN at $\sigma = 25$. These are repeated uses of the same image pairs.

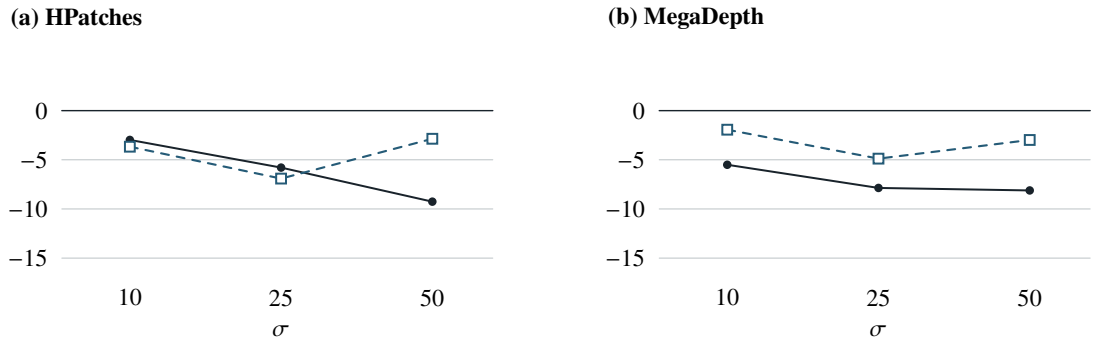

Fig. S1 Correspondence precision, NLM minus Gaussian filtering (pp). Circles and solid lines denote NN, open squares and dashed lines denote LG. Each point averages the original 60 HPatches or 30 MegaDepth pairs after averaging seeds. Lines connect observed noise levels. The plot shows sample means

With Gaussian filtering as the comparator, NLM lost precision in every cell of Fig. S1. Table S3a instead compares NLM with Raw. Mean changes across the six dataset–severity combinations were −6.1 pp for NN and +3.3 pp for LG. Precision is the fraction of matches within 3 px of the reference transfer or epipolar constraint, with zero for an empty set; epipolar consistency does not identify the physical point.

Table S3a NLM minus Raw correspondence precision (pp). Point estimates average the 60 HPatches or 30 MegaDepth pairs after averaging the three realizations within each pair. Brackets give unadjusted 95% percentile intervals from 2,000 resamples of HPatches sequences or pairs within fixed MegaDepth scenes (seed 20261029)

Dataset	σ	NN	LightGlue
HPatches	10	−1.1 [−1.8, −0.3]	+1.8 [0.8, 2.7]
	25	−4.7 [−8.5, −1.6]	+3.2 [1.2, 4.9]
	50	−8.3 [−13.3, −3.5]	+4.8 [2.7, 7.0]
MegaDepth	10	−1.1 [−2.9, 0.9]	+2.5 [0.8, 4.2]
	25	−9.6 [−13.9, −5.9]	+1.0 [−2.5, 4.1]
	50	−12.2 [−17.7, −7.4]	+6.3 [4.2, 8.5]

Selection cost. Among the three denoisers, PSNR selected a highest-success arm in 9 of 12 strata, counting ties. Mean/maximum in-sample regret was 0.42/2.22 pp for PSNR and 1.76/12.22 pp for precision.

S4 Effects against the degraded control

We averaged repeated corruptions within each pair before differencing restored and Raw success. The 95% percentile intervals use 2,000 resamples: HPatches sequences, or pairs within each fixed MegaDepth scene with equal scene weights. They describe the sampled design.

Table S4 Restored minus Raw success (pp). Brackets give unadjusted paired cluster intervals. HPatches p_s enumerates all 4,096 sign vectors of the 12 sequence-mean effects; $p_{s,H}$ applies Holm to all 18 HPatches contrasts, including the retrospective focus. MegaDepth intervals resample pairs within two fixed scenes. Flipping whole scenes gives four sign patterns and a minimum two-sided p of 0.5

			HPatches			MegaDepth
σ	Match	Restorer	Effect [95% CI]	p_s	$p_{s,H}$	Effect [95% CI]
10	NN	Gaussian	+6.7 [1.7, 12.2]	0.0625	0.9375	+14.4 [4.4, 24.4]
		NLM	+3.9 [−1.1, 9.4]	0.3438	1.0000	+4.4 [−2.2, 12.2]
		DnCNN	+5.0 [0.6, 10.0]	0.0938	1.0000	+13.3 [2.2, 24.4]
	LG	Gaussian	+2.8 [0.0, 5.6]	0.1875	1.0000	−2.2 [−14.4, 7.8]
		NLM	0.0 [−5.0, 5.0]	1.0000	1.0000	−5.6 [−15.6, 4.4]
		DnCNN	+3.3 [0.6, 6.7]	0.1250	1.0000	−2.2 [−12.2, 7.8]
25	NN	Gaussian	+5.6 [0.0, 13.3]	0.1797	1.0000	+6.7 [−1.1, 14.5]
		NLM	−1.7 [−7.8, 4.4]	0.7305	1.0000	+12.2 [1.1, 24.4]
		DnCNN	+8.3 [2.8, 13.9]	0.0234	0.3750	+10.0 [3.3, 17.8]
	LG	Gaussian	+3.3 [0.6, 6.7]	0.1562	1.0000	−3.3 [−16.7, 10.0]
		NLM	+1.7 [−2.8, 6.1]	0.6406	1.0000	−10.0 [−20.0, 0.0]
		DnCNN	+6.7 [1.1, 12.2]	0.0703	0.9844	+3.3 [−8.9, 16.7]
50	NN	Gaussian	−2.2 [−6.1, 1.1]	0.4062	1.0000	+5.6 [−1.1, 12.2]
		NLM	+5.6 [−2.2, 12.8]	0.2305	1.0000	+4.4 [0.0, 10.0]
		DnCNN	+10.0 [3.3, 17.8]	0.0156	0.2656	+7.8 [2.2, 14.4]
	LG	Gaussian	+5.0 [−2.8, 11.1]	0.2578	1.0000	+5.6 [−3.3, 14.4]
		NLM	+3.3 [−1.1, 8.3]	0.2852	1.0000	+4.4 [−4.4, 13.3]
		DnCNN	+12.2 [6.7, 17.2]	0.0029	0.0527	+16.7 [6.7, 27.8]

The sign-flip test keeps pairs and seeds together within each sequence. Enumeration gives an exact test under joint sign symmetry of the sequence effects under the null. Deterministic sequence selection provides no design-based guarantee of that assumption. No HPatches contrast survives Holm at 0.05. Severe-noise DnCNN–LG gives $p_s = 0.0029$ and $p_{s,H} = 0.0527$. For MegaDepth NLM–LG at $\sigma = 25$, the upper limit for the success change rounds to zero. The percentile intervals and sign-flip tests use different constructions. Post hoc testing leaves contrast selection and transfer unresolved.

The earlier exact trial-level McNemar values ignored shared pairs and images. For the retrospectively selected HPatches DnCNN–LG contrast at $\sigma = 50$, that value was 0.0002. Of 35 other comparisons, HPatches DnCNN–NN and MegaDepth DnCNN–LG at $\sigma = 50$ passed the trial-level Holm cutoff (0.0181 and 0.0096). We retain them as historical summaries, rather than independent-trial inference.

Model and margin sensitivities. Random-intercept fits to seed-averaged effects included restorer, matcher, severity and all interactions. Pair variances were 41.29 and 95.85 pp^2 for HPatches and MegaDepth. Fixed-design/pair/residual shares were 2.73/8.62/88.65% and 5.80/11.17/83.03%. Among the 33 secondary comparisons with trial-level Holm-adjusted values greater than 0.05, 0, 0 and 1 had 90% intervals inside ±1, ±2 and ±5 pp. These post hoc margins establish no equivalence.

S5 Pose thresholds, scenes and image size

Table S5 separates endpoint thresholds, scene composition and image size for the severe-noise DnCNN–LightGlue comparison.

Table S5 DnCNN–LightGlue versus Raw at $\sigma = 50$. Raw and DnCNN columns give absolute percentages, or normalized AUC on a 0–100 scale. The final column is the paired difference on the corresponding scale with its unadjusted 95% percentile interval

Endpoint or subset	Raw	DnCNN	Difference [95% CI]
MegaDepth, both scenes at 384 px			
Success at 5°	4.4	11.1	+6.7 [1.1, 12.2]
Success at 10°	12.2	28.9	+16.7 [6.7, 27.8]
Success at 20°	33.3	50.0	+16.7 [4.4, 28.9]
Normalized AUC to 20°	14.1	26.1	+12.0 [4.5, 20.3]
MegaDepth by scene, 384 px, success at 10°			
Scene 0015	13.3	24.4	+11.1 [0.0, 26.7]
Scene 0022	11.1	33.3	+22.2 [6.7, 37.8]
Longer side of 512 px			
HPatches (< 3 px)	41.1	64.4	+23.3 [15.0, 32.8]
HPatches (< 4 px)	49.4	70.6	+21.1 [15.6, 26.7]
MegaDepth (< 10°)	18.9	38.9	+20.0 [12.2, 28.9]

The normalized AUC integrates empirical recall by trapezoids to 20°, holds recall flat after the last included error, and divides by 20°. Failures remain in the denominator. Scene intervals resample 15 pairs within the specified scene.

Resolution. The 512-px run shares noisy arrays across restorers. Rescoring its fitted homographies at 4 px scales the success cutoff by 512/384; the 3-px estimator threshold stays fixed. This checks endpoint scale only. Clean NN/LG success was 85.0/90.0% on HPatches (3 px) and 40.0/60.0% on MegaDepth (10°). HPatches clean LG success at 4 px was 93.3%. The 4-px effect interval uses 2,000 sequence-bootstrap draws with seed 1049177346, derived deterministically from its analysis identifier.

Other degradations. Low-light applies $I^{1/0.4}$, Poisson noise (peaks 120/60/30), Gaussian read noise (1/2/4 grey levels) and 8-bit quantization. JPEG uses qualities 20/30/40; repeated seeds give identical inputs. NLM uses nominal $\sigma = 25$ in both groups. Their clean NN/LG success was 80.0/90.0% on HPatches and 46.7/63.3% on MegaDepth. Table S6 gives every contrast; low-light PSNR includes the imposed exposure change.

Table S6 NLM minus Gaussian filtering under other degradations. P is PSNR (dB); NN and LG are success differences (pp). These descriptive pair-mean contrasts retain every stratum (60 HPatches or 30 MegaDepth pairs) and carry no interval or multiplicity claim

		HPatches			MegaDepth		
Group	Level	P	NN	LG	P	NN	LG
Low-light	1	−0.15	−15.00	−15.00	−0.11	−7.78	−11.11
Low-light	2	−0.17	−11.11	−10.00	−0.13	−1.11	−5.56
Low-light	3	−0.27	−8.89	−0.56	−0.33	0.00	−10.00
JPEG	20	−1.80	−6.67	−11.67	−1.25	−26.67	−16.67
JPEG	30	−2.17	−18.33	−18.33	−1.42	−13.33	−10.00
JPEG	40	−2.36	−13.33	−15.00	−1.48	−13.33	−3.33

S6 Recorded image pairs and feature sampling

Figure S2 shows a viewpoint pair and a scene-0022 pair, the first eligible records in those subsets. Each crop uses the same location across restorers. Table S7 reports the separate Real-ESRGAN comparison on all original pairs.

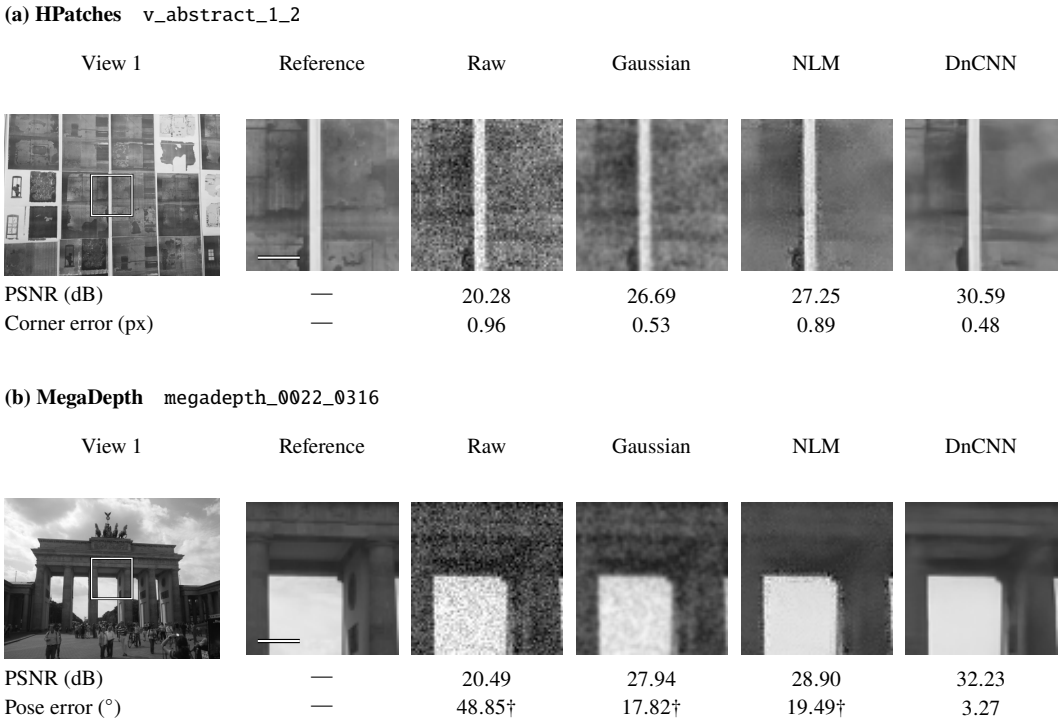

Fig. S2 First-view crops at $\sigma = 25$, seed 20261024, with LightGlue. Boxes locate the 72×72 px crops. Bars span 20 evaluation pixels. PSNR averages both views. A dagger marks failure against reference geometry (Balntas et al., 2017; Li and Snavely, 2018)

On the HPatches pair, NLM gave higher PSNR than Gaussian filtering but larger corner error (0.89 versus 0.53 px). On the MegaDepth pair, only DnCNN brought the error below 10°.

Table S7 Compact Real-ESRGAN minus Raw. The PSNR change on the reduced image is shared by both matchers and feature rasters. Each dataset and noise level therefore has one PSNR estimate and interval, using the enlarged NN run's resampling stream. Geometric success is evaluated with features extracted at 4× enlargement or after reduction to native evaluation size. Brackets give unadjusted 95% percentile intervals over HPatches sequences or pairs within fixed MegaDepth scenes

Dataset	σ	ΔPSNR (dB)	Match	Enlarged (pp)	Native (pp)
HPatches	10	+1.08 [−1.04, 2.87]	NN	−8.9 [−17.8, −2.2]	−2.8 [−6.1, 0.6]
			LG	−12.8 [−24.4, −2.2]	−2.8 [−8.9, 2.2]
	25	+1.58 [0.91, 2.19]	NN	−28.9 [−46.7, −14.4]	−5.0 [−11.1, 1.1]
			LG	−46.7 [−62.8, −32.2]	−2.2 [−7.8, 3.3]
	50	+1.69 [1.46, 1.93]	NN	−23.9 [−36.7, −11.1]	−1.7 [−6.1, 2.2]
			LG	−48.3 [−66.1, −29.4]	+3.9 [−0.6, 8.3]
MegaDepth	10	+0.73 [0.29, 1.13]	NN	−6.7 [−16.7, 1.1]	−1.1 [−8.9, 6.7]
			LG	−11.1 [−20.0, −2.2]	−12.2 [−25.6, 2.2]
	25	+2.17 [2.01, 2.35]	NN	−2.2 [−10.0, 6.7]	+1.1 [−5.6, 6.7]
			LG	−18.9 [−32.2, −4.4]	−10.0 [−21.1, 1.1]
	50	+2.15 [2.03, 2.29]	NN	−2.2 [−5.6, 0.0]	+8.9 [3.3, 14.4]
			LG	−10.0 [−17.8, −3.3]	+4.4 [−3.3, 13.3]

S7 Transfer to unused HPatches sequences

We froze the arm choices from the original HPatches $\sigma = 25$ grid before processing 12 additional sequences. They shared no image or sequence with the original subset. Six illumination and six viewpoint sequences contributed five pairs each. Three realizations were averaged within each pair. The unadjusted 95% intervals in Tables S8 and S9 use 10,000 draws of sequences within each type, with equal weight for the two types.

Table S8 Controls and restoration effects on the 60 follow-up pairs. Upper rows give absolute success or differences from Raw. The lower group compares NLM directly with Gaussian filtering. Brackets enclose sequence-stratified percentile intervals

Measurement	NN	LightGlue
Clean success (%)	90.0 [80.0, 98.3]	93.3 [83.3, 100.0]
Degraded Raw success (%)	52.8 [36.7, 68.3]	78.3 [61.1, 91.7]
Gaussian vs Raw, success (pp)	+8.33 [3.33, 13.89]	+6.67 [0.00, 17.22]
NLM vs Raw, success (pp)	+10.56 [5.56, 15.56]	+1.67 [−1.67, 5.56]
DnCNN vs Raw, success (pp)	+11.67 [5.00, 17.78]	+7.22 [1.11, 13.89]
NLM minus Gaussian filtering		
ΔPSNR (dB)	+0.52 [−0.04, 1.14]	+0.52 [−0.04, 1.14]
ΔSSIM	+0.089 [0.050, 0.129]	+0.089 [0.050, 0.129]
Precision (pp)	−4.26 [−5.70, −2.82]	−7.12 [−9.17, −5.10]
Success (pp)	+2.22 [−3.33, 7.78]	−5.00 [−13.33, 0.56]

PSNR and SSIM are independent of the matcher; identical values are repeated to align the columns. NLM gave a higher mean PSNR and lower correspondence precision than Gaussian filtering with both matchers. Its success difference changed sign from the original HPatches estimate for NN and remained negative for LightGlue. Both success intervals include zero.

Table S9 Success of arms chosen on the original sequences. Regret is the difference from the best mean success among the three denoisers on the follow-up sample, with that best arm reselected in each bootstrap draw

Match	Selection score	Frozen arm	Success (%)	Regret (pp)
NN	PSNR	DnCNN	64.44 [46.67, 80.00]	0.00 [0.00, 3.33]
	Precision	Gaussian	61.11 [43.89, 77.22]	3.33 [0.00, 8.89]
LG	PSNR	DnCNN	85.56 [72.22, 96.11]	0.00 [0.00, 5.56]
	Precision	DnCNN	85.56 [72.22, 96.11]	0.00 [0.00, 5.56]

The paired difference between the PSNR and precision choices was +3.33 [−2.78, 8.33] pp with NN. For LightGlue both rules selected DnCNN, so the difference is zero by construction. DnCNN had the highest observed success with each matcher on these sequences. This follow-up covers one noise level within HPatches and adds no new MegaDepth scene.

Dataset references

V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk. HPatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3852–3861, 2017. doi: 10.1109/cvpr.2017.410.

Z. Li and N. Snavely. MegaDepth: Learning Single-View Depth Prediction from Internet Photos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2041–2050, 2018. doi: 10.1109/cvpr.2018.00218.