Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Image fidelity and geometric accuracy in paired restoration–matching pipelines

Congshuai Dong

dong_cs@qq.com

Shandong Jianzhu University

Wenxue Zhang

Shandong Jianzhu University

Yiying Zhang

Shandong Jianzhu University

Hao Zhang

Shandong Jianzhu University

Yiteng Teng

Shandong Jianzhu University

Research Article

Keywords: Image restoration, Image denoising, Feature matching, Homography estimation, Geometric estimation, Task-aware evaluation

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11744730/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Image fidelity and geometric accuracy in paired restoration–matching pipelines

Congshuai Dong[1*], Wenxue Zhang[1], Yiying Zhang[1], Hao Zhang[1], Yiteng Teng[1]

[1*]School of Science, Shandong Jianzhu University, Jinan, 250101, China.

*Corresponding author(s). E-mail(s): dong_cs@qq.com;

Abstract

Image fidelity and geometric success can rank denoisers differently even when the highest-fidelity choice performs well on geometry. We tested three denoisers against the same corrupted inputs using one detector–descriptor family and two matchers, on 60 HPatches pairs and 30 MegaDepth pairs from 2 scenes. Non-local means (NLM) gave higher peak signal-to-noise ratio (PSNR) than Gaussian filtering in all 12 Gaussian-noise strata, but fewer successful geometric fits in 10. In each of these reversal strata, the paired interval for the success difference included zero. Choosing the highest-PSNR denoiser in each stratum incurred a mean in-sample loss of 0.42 percentage points relative to the highest observed geometric success, with a maximum of 2.22. At $\boldsymbol{\sigma = 25}$, the PSNR choice made on the original HPatches sample, DnCNN, also attained the highest mean success with both matchers on 60 pairs from unused sequences. The NLM–Gaussian success difference changed direction for nearest-neighbor matching on these sequences. Precision was associated with homography success, while changes in mean correspondence precision depended on the matcher, the noise level and the comparator. A model using precision and PSNR explained little of the variation in success. Reporting geometric effects alongside fidelity shows both where rankings disagree and how much that disagreement matters for the choice of denoiser.

Keywords: Image restoration, Image denoising, Feature matching, Homography estimation, Geometric estimation, Task-aware evaluation

1 Introduction

Image restoration and geometric estimation optimize different outputs. A denoiser can reduce pixelwise error while changing the locations, descriptors or spatial distribution of the points used to estimate a homography or camera pose. The practical question is whether inserting a particular restoration operation improves the geometric pipeline on its intended inputs. Neither a visually cleaner image nor a higher peak signal-to-noise ratio (PSNR) settles that question.

Evidence is difficult to transfer across pipeline compositions. Restoration benchmarks measure image fidelity, whereas application studies may report matching stability, recognition accuracy or panorama appearance. Summers et al. [1], for example, assess underwater enhancement through frame-matching measures and a complete SLAM system. Here we measure transformation error against reference geometry under controlled corruptions. Shared input arrays, a no-restoration control, consistent coordinates and an explicit failure denominator isolate the effect of inserting a particular restoration implementation.

We address three linked questions: whether denoisers reverse rank across image and geometric endpoints; how much success is lost by choosing an arm on fidelity or correspondence precision; and whether choices made on the original sample transfer to unused sequences. The study combines paired effect estimates with a local perturbation analysis and a sequence-disjoint follow-up. Each comparison fixes the feature extractor and geometric estimator for a given matcher, and unsuccessful fits remain in the success denominator (Fig. 1).

2 Related work

Non-local means (NLM) and transform-domain collaborative filtering exploit repeated structure [2, 3]; DnCNN learns a noise-removal mapping [4]. These designs leave open how restoration affects feature localization or matching ambiguity. Perceptual quality and distortion are also distinct objectives [5]. Our compact Real-ESRGAN experiment concerns one released adversarial super-resolution composition [6]; its findings are specific to that implementation.

Sparse matching separates detection, description, assignment and robust estimation. SIFT and RootSIFT provide a fixed detection/description stage [7, 8]; LightGlue changes the assignment stage using learned contextual reasoning [9]. Sharing extracted features isolates a matcher contrast within that family. Detector-free and dense matching change the representation and correspondence process more broadly [10, 11]. Robust fitting limits its outlier influence but cannot make a poorly distributed constraint set informative [12–14].

Related models alter the information available during restoration: SIR-DIFF shares information across views [15], UniFuse handles degradation and misalignment in multimodal medical fusion [16], and T-PhISH-Net introduces temporal context for underwater enhancement [17]. Task-inspired underwater enhancement also changes training to support perception [18]. Those settings differ from independently restoring each view before a fixed estimator. We distinguish image fidelity, reference correspondence consistency and transformation error, so improvement in one is not taken as evidence of improvement in another.

3 Paired evaluation protocol

3.1 Controls and coordinates

In Fig. 1, D_ϕ denotes the corruption rule with settings ϕ. For view j, $y_j = D_\phi(x_j; \epsilon_j)$ uses the image-specific noise realization ϵ_j. Each restoration arm receives the same pair y_0, y_1. Writing g_a for the restorer and h_ψ for the downstream pipeline with a fixed matcher, the two estimates are

$$\widehat{G}_a = h_\psi(g_a(y_0), g_a(y_1)), \quad \widehat{G}_0 = h_\psi(y_0, y_1). \quad (1)$$

Raw denotes the identity arm, not the clean image. A clean control is reported separately to expose low-baseline regimes. Input hashes, shared random streams and frozen downstream settings define the pairing. Failed estimates remain in all success denominators. A ranking stratum fixes dataset, severity and matcher; an arm is an implementation rather than an architecture family.

Feature and evaluation frames must remain explicitly related. With resize maps S_i, homographies and intrinsics transform as

$$H_e = S_1 H S_0^{-1}, \qquad K_{e,i} = S_i K_i. \qquad (2)$$

For enlargement by s, pixel-center coordinates return by $p = (q + 1/2)/s - 1/2$ in each axis. Recorded round-trip and estimator checks use known geometry. This coordinate correction does not undo changes in detected features caused by enlargement; extracting features after downsampling is a separate sensitivity analysis.

Image error and geometric sensitivity weight perturbations differently. Online Resource 1, Section S2 derives a fixed-filter risk decomposition and a local homography perturbation analysis. Their assumptions exclude changes in detected correspondences or inlier support; they motivate separate endpoints rather than identify a smoothing mechanism.

3.2 Effects, uncertainty and selection cost

For pair i, arm a and repeated corruption z, let s_{iaz} denote binary success. The paired sample

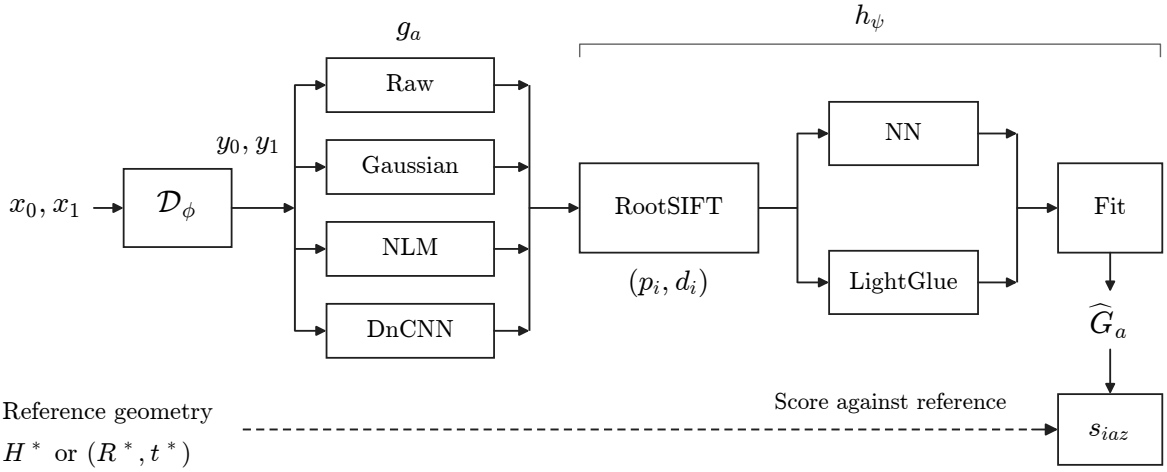

Fig. 1 Paired evaluation in the primary Gaussian grid. Raw passes the corrupted pair directly to the feature extractor. For each arm, nearest-neighbor matching (NN) and LightGlue use the same RootSIFT features. The fitted homography or relative pose is scored against reference geometry. A failed fit receives zero success. Table 1 separates this main grid from the extensions and follow-up

effect is

$$d_{ia} = \frac{1}{Z}\sum_z (s_{iaz} - s_{i0z}),$$
$$\widehat{\Delta}_a = \sum_i w_i d_{ia}, \qquad \sum_i w_i = 1. \tag{3}$$

HPatches uncertainty resamples sequences, retaining dependent pairs and repeated configurations together. MegaDepth resamples pairs within each fixed scene and equally weights scene means. We report sample statistics with percentile intervals from 2,000 fixed-seed draws [19]; these intervals do not estimate variability across new scenes.

For HPatches, we averaged each paired contrast within sequence and enumerated all 4096 sign vectors for the sequence means. The two-sided test compares the absolute observed mean with its sign-flip distribution. Validity requires joint sign symmetry under the null, as would follow from independent sequence contrasts each symmetric about zero. The fixed selection of sequences does not establish this assumption. We adjusted the resulting values by Holm across all 18 HPatches restorer–severity–matcher contrasts, including the retrospective focus. Online Resource 1, Section S4 reports this sensitivity analysis alongside variance and equivalence-margin checks.

For a score v selecting arm $\widehat{a}_v(t) = \arg\max_a v_{at}$ in stratum t, the observed regret is

$$R_v(t) = \max_a \widehat{\Delta}_{at} - \widehat{\Delta}_{\widehat{a}_v(t),t}. \tag{4}$$

Selection uses the three-denoiser grid and excludes Raw and the separately added arms. The comparison is in-sample: the best arm is defined by the observed geometric means in the same grid. Transfer is evaluated separately using arm choices fixed before processing the follow-up sequences.

4 Experimental setup

4.1 Data, corruption and restoration

The HPatches subset [20] contains 60 pairs from 12 sequences (6 illumination and 6 viewpoint). MegaDepth [21] uses 15 image-disjoint pairs from each of scenes 0015 and 0022 in the public MegaDepth-1500 release. Online Resource 1, Section S1 describes the deterministic subset selection. The reported scores refer to these subsets. A frozen follow-up adds 60 pairs from 12 non-overlapping HPatches sequences at $\sigma = 25$, testing original-sample arm choices unchanged (Online Resource 1, Section S7).

Grayscale images are resized without upsampling to maximum side 384 pixels. Gaussian corruption uses $\sigma \in \{10, 25, 50\}$ in 8-bit units, then clipping and rounding. Recorded seeds and

Table 1 Measured design blocks and their roles. Only the main Gaussian grid enters the original endpoint families and the HPatches sequence-level Holm analysis; extensions retain their own controls

Block	Role	Coverage
Gaussian	Main grid	Raw, Gaussian, NLM, DnCNN; RootSIFT–NN/LightGlue
Adversarial	Extension	Compact Real-ESRGAN; enlarged/native features
Low-light, JPEG	Extension	Fixed nominal NLM noise level
Resolution	Extension	Gaussian grid at 512 px
BM3D, ORB	Extension	Denoiser and feature-pipeline checks
Follow-up	Transfer	New HPatches sequences; fixed choices at $\sigma = 25$

image-path/severity hashes fix each noisy array, including shared images. Resized originals are the fidelity references and may contain acquisition artifacts.

The original restorers are a fixed Gaussian filter, OpenCV fast NLM and the pretrained blind grayscale DnCNN checkpoint from KAIR. NLM uses $h = 0.8\sigma$ and receives the synthetic noise level; DnCNN is blind. The separate BM3D extension also uses known noise. No model is trained or fine-tuned. Online Resource 1, Section S1 gives the filter and matcher settings.

The compact `realesr-general-x4v3` release uses default denoising-strength interpolation 0.5. Grayscale input is replicated into 3 channels and output returned to quantized grayscale. Its primary mode extracts features on the 4-times enlarged output before returning coordinates to the evaluation frame. PSNR and structural similarity (SSIM) [22] use its Lanczos-downsampled output. SSIM is recorded alongside PSNR; paired follow-up values are reported in Online Resource 1, Section S7. A native-raster sensitivity extracts features from that same downsampled image, reusing the inference and noisy-array hashes.

4.2 Matching, endpoints and analysis

Both core matchers consume the same RootSIFT features. Figure 2 reports success relative to Raw at each noise level. NN uses mutual matches and bidirectional ratio threshold 0.8; pretrained SIFT–LightGlue uses match threshold 0.1. Extractor and confidence settings are held fixed (Online Resource 1, Section S1). The public SIFT–LightGlue recipe excludes the evaluated MegaDepth scenes from gradient training but includes them in validation. Pose results therefore concern scenes exposed during model development.

HPatches uses USAC_MAGSAC homography fitting with a 3-pixel threshold. Success requires mean reference transfer error at the four source corners below 3 resized pixels. MegaDepth uses essential-matrix RANSAC in calibrated coordinates, retaining the candidate with the largest cheirality-consistent inlier set. Success requires the larger of rotation error and sign-invariant translation-direction error to be below 10 degrees. Estimator seeds are shared across matched runs. Failed fits, including zero-inlier poses, receive infinite error and zero success.

Correspondence precision is the fraction passing a reference consistency threshold, with zero for an empty match set. HPatches uses symmetric homography transfer below 3 pixels; MegaDepth uses symmetric epipolar distance below 3 pixels. The latter is a consistency measure, not verification of physical correspondence. Other diagnostics are match count, inlier count and normalized convex-hull support area.

We chose HPatches DnCNN–LightGlue at $\sigma = 50$ for closer examination after inspecting the original grid. It is included in the HPatches sequence-level family. An earlier trial-level analysis excluded this one contrast and adjusted the remaining 35 across both datasets. Repeated trials shared pairs and images. Their McNemar values appear as historical summaries in Online Resource 1, Section S4. Both families exclude the extension arms. We also regressed seed-averaged Δsuccess on ΔPSNR and Δprecision, separately by dataset and in a pooled model, with cluster intervals from 2000 draws. These coefficients describe associations across configurations. The

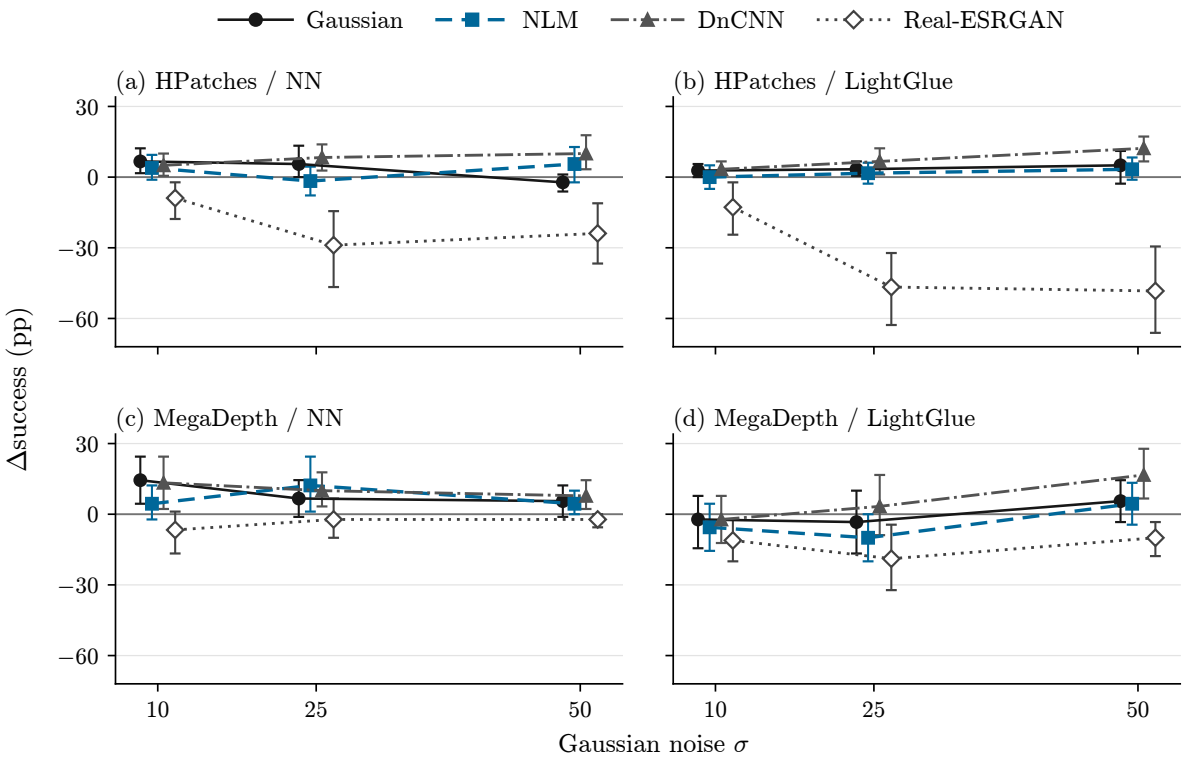

Fig. 2 Success change relative to Raw by noise level and matcher. Points show paired sample effects on 60 HPatches or 30 MegaDepth pairs. Bars show unadjusted 95% percentile intervals obtained by resampling HPatches sequences or pairs within each fixed MegaDepth scene. Repeated seeds stay together. All panels use the same vertical scale, with small horizontal offsets for legibility. Real-ESRGAN lies outside the original Holm family

pooled fit combines the two datasets' different precision definitions.

Synthetic low-light combines $I^{1/0.4}$, Poisson/read noise and quantization; JPEG uses qualities 20, 30 and 40. NLM retains nominal $\sigma = 25$ throughout these groups. Repeated JPEG seeds do not create independent compressed inputs. Online Resource 1, Sections S1 and S5 report clean controls, per-scene and threshold results, and the 512-pixel rerun. ORB changes detection, description and matching together, rather than isolating a matcher effect.

5 Results

5.1 Quality gain versus geometric change

NLM exceeded Gaussian filtering in PSNR in all 12 Gaussian-noise strata but had lower observed success in 10. These are point-estimate rankings: the directly paired NLM-minus-Gaussian success intervals include zero in every reversal stratum (Online Resource 1, Table S3). The strata reuse pairs across restorers, severities and matchers. Table 2 gives selected effects against Raw; Online Resource 1 reports controls in Table S1, direct denoiser contrasts in Table S3 and original-grid effects against Raw in Table S4.

At severe noise, DnCNN–LightGlue had higher observed success than Raw on both datasets (Table 2, Fig. 2). The HPatches DnCNN–LightGlue contrast had the smallest sequence-level p value ($p = 0.00293$). For this first-ranked contrast, Holm multiplies the unrounded value by the family size of 18, giving $p_H = 0.0527$. No HPatches contrast met the adjusted 0.05 threshold.

5.2 Matcher interaction

The estimated restoration effects had opposite signs for the two matchers in 7 of 18 dataset–severity–restorer combinations.

Table 2 Selected success changes relative to Raw. HP and MD denote HPatches and MegaDepth, and LG denotes LightGlue. Moderate and severe noise correspond to $\sigma = 25$ and $\sigma = 50$. Brackets give unadjusted 95% paired cluster percentile intervals. Online Resource 1 gives the controls (Table S1) and original-grid effects (Table S4). The BM3D (Table S2) and Real-ESRGAN (Table S7) extensions lie outside the original Holm family

Comparison	Success change (pp)	Scope
DnCNN, HP–LG, severe	$+12.2\,[+6.7, +17.2]$	Retrospective focus (Sect. 5.1)
DnCNN, MD–LG, severe	$+16.7\,[+6.7, +27.8]$	Conditional on the fixed scenes
BM3D, HP–LG, severe	$+18.9\,[+12.2, +26.1]$	Separate exploratory denoiser extension
BM3D, MD–LG, severe	$+20.0\,[+11.1, +28.9]$	Separate exploratory denoiser extension
NLM, MD–LG, moderate	$-10.0\,[-20.0, 0.0]$	Fidelity and success disagree
Real-ESRGAN, HP–LG, severe	$-48.3\,[-66.1, -29.4]$	Features on enlarged output
Real-ESRGAN native, same condition	$+3.9\,[-0.6, 8.3]$	Same images, features at evaluation size

For MegaDepth NLM at $\sigma = 25$, success changed by $+12.2\,[1.1, 24.4]$ pp with NN and $-10.0\,[-20.0, 0.0]$ pp with LightGlue relative to each matcher's Raw control (Online Resource 1, Table S4). The LightGlue interval's upper limit rounds to zero. The paired LightGlue-minus-NN interaction was $-22.2\,[-36.7, -7.8]$ pp. These unadjusted intervals condition on the two sampled scenes.

5.3 Correspondence precision as a diagnostic

After adjustment for PSNR, the HPatches precision coefficient was $0.343\,[0.202, 0.567]$ pp per pp. The corresponding MegaDepth interval spanned zero. Together, the predictors accounted for about 3.0% of the variation on HPatches and 0.3% on MegaDepth (Table 3).

Condition-level contrasts separate the precision changes by matcher and noise level. Relative to Raw, NLM changed mean precision by $-1.4\,[-2.5, -0.4]$ pp when strata were weighted equally. Mean precision fell with NN and rose with LightGlue on both datasets at each noise level, although some intervals included zero (Online Resource 1, Table S3a). With Gaussian filtering as the comparator, NLM had lower precision with both matchers throughout (Online Resource 1, Section S3).

BM3D also showed a matcher-dependent pattern relative to Raw. At $\sigma = 25$ and 50, its mean precision fell with NN and rose with LightGlue on both datasets. At $\sigma = 10$, both matchers gained precision on average in each dataset (Online Resource 1, Table S2). The sign of the observed change thus depends on the noise level as well as the matcher and the choice of control.

5.4 Compact adversarial restoration and feature scale

With features extracted on its enlarged output, compact Real-ESRGAN increased mean PSNR and reduced observed success in all 12 strata (Fig. 2). The largest observed loss occurred on HPatches with LightGlue at $\sigma = 50$. Relative to Raw, PSNR rose by $+1.69\,[1.46, 1.93]$ dB and success changed by $-48.3\,[-66.1, -29.4]$ pp. Extracting features after reducing the same images to evaluation size gave $+3.9\,[-0.6, 8.3]$ pp (Table 2). The point estimate changed sign, but the native-raster interval includes zero. Across the full grid (Online Resource 1, Table S7), native-raster Real-ESRGAN gained fidelity and lost success in 8 of 12 strata. DnCNN showed this pattern in 1 of 12 strata. These counts describe the tested implementations.

5.5 Degradation, resolution and additional pipelines

The Gaussian-versus-NLM reversal did not persist under synthetic low-light or JPEG: both PSNR and success favored Gaussian filtering, allowing success ties. These are synthetic low-light and compression tests; low-light PSNR also includes the imposed radiometric change (Online Resource 1, Section S5). With the longer side limited to 512 pixels, severe-noise DnCNN–LightGlue changed success by $+23.3\,[+15.0, +32.8]$ pp on HPatches and $+20.0\,[+12.2, +28.9]$ pp on MegaDepth. The HPatches rerun retains the original pixel cutoff. Rescoring the same errors at a proportionally enlarged cutoff of 4 pixels gives $+21.1\,[+15.6, +26.7]$ pp. Both definitions retain a

Table 3 Descriptive regression of paired success change on PSNR and precision changes with an intercept. Coefficients are pp per dB for PSNR and pp per pp for precision; brackets are 95% cluster-bootstrap intervals. Row counts include repeated configurations for the same pairs. HPatches resamples sequences; MegaDepth resamples pairs within the fixed scenes. The pooled model combines different precision definitions

Sample	Pairs / rows	Intercept (pp)	$\beta_{\Delta\text{PSNR}}$ [CI]	$\beta_{\Delta P}$ [CI]	R^2
Pooled	90 / 1620	1.324	0.375 [−0.005, 0.816]	0.194 [0.070, 0.327]	0.010
HPatches	60 / 1080	1.193	0.294 [−0.023, 0.662]	0.343 [0.202, 0.567]	0.030
MegaDepth	30 / 540	0.695	0.579 [−0.734, 1.904]	−0.035 [−0.237, 0.162]	0.003



Fig. 3 Recorded outputs for (a) HPatches i_ajuntament_1_2 and (b) MegaDepth megadepth_0015_0619, at $\sigma = 25$, seed 20261024 and LightGlue. Boxed crops span 72 pixels; bars mark 20 evaluation pixels. PSNR uses both full views; error is mean corner error (px) or pose error (degrees). A dagger marks failure. All panels use the recorded grayscale range without display contrast or gamma adjustment

positive effect. Online Resource 1, Section S5 compares the cutoffs, angular thresholds, empirical-recall AUC and fixed scenes.

At $\sigma = 50$ with LightGlue, BM3D used the known noise level and DnCNN operated blind. Their estimated success changes from Raw were +18.9 versus +12.2 pp on HPatches and +20.0 versus +16.7 pp on MegaDepth (Table 2). ORB clean success was 58.3% on HPatches and 13.3% on MegaDepth. The low baseline and broad intervals limit ORB to a separate-pipeline check (Online Resource 1, Section S2).

Figure 3 compares recorded outputs and common crops for the first pair listed for each dataset in the fixed evaluation manifest, `i_ajuntament_1_2` and `megadepth_0015_0619`.

The MegaDepth example retains NLM's higher PSNR but failed pose estimate. Online Resource 1, Fig. S2 instead shows a viewpoint pair and a scene-0022 pair. All displayed examples come from the recorded experimental subsets.

5.6 Selection regret

PSNR selected an arm with the highest geometric success in 9 of 12 strata, including ties. Among the original three denoisers, mean regret was 0.42 pp for PSNR and 1.76 pp for precision. The corresponding maxima were 2.22 and 12.22 pp. Both scores require reference information: a clean image for PSNR and known geometry for precision.

Sequence-disjoint validation.

On the new HPatches sequences, the NLM-minus-Gaussian success difference was $+2.22\,[-3.33, +7.78]$ pp for NN and $-5.00\,[-13.33, +0.56]$ pp for LightGlue. With NN, the original PSNR choice (DnCNN) exceeded the precision choice (Gaussian) by $+3.33\,[-2.78, +8.33]$ pp. With LightGlue both criteria chose DnCNN. DnCNN attained the highest new-sample mean with both matchers (Online Resource 1, Section S7).

6 Discussion

The Gaussian-grid results suggest a trade-off between fidelity and correspondence quality, but its mechanism remains unresolved. Denoisers differ in more than smoothing strength, which was neither controlled independently nor measured here. The local analysis in Online Resource 1, Section S2 explains how image perturbations can affect geometric error when matches and inlier support are fixed. Restoration can change both, so the derivation provides conditional intuition rather than a causal explanation of the measured ranks.

On HPatches with NN at $\sigma = 50$ (Online Resource 1, Table S2), BM3D changed precision by $-3.6\,[-6.6, -0.5]$ pp while success rose by $14.4\,[7.2, 23.3]$ pp. A fraction of consistent matches omits their number, arrangement and influence on the fitted transformation. The uncertain pose association and greater precision-based regret further limit a monotone interpretation. Conditioning on precision, itself affected by restoration, does not establish mediation.

Figure 4 shows a pair selected because the NLM fit failed. The marked corner is displaced by 7.28 pixels, and the mean error over all four corners is 3.53 pixels. NLM had slightly higher full-pair match precision than Gaussian filtering on this pair. The fraction of consistent matches alone therefore does not account for this failure.

The original grid covers selected HPatches sequences, two MegaDepth scenes and synthetic degradations. LightGlue used the pose scenes for validation. Testing additional scenes outside model-development splits would help establish how far the pose results transfer. The sequence-disjoint HPatches follow-up covered only $\sigma = 25$; additional sequences across noise levels would extend the evidence for homography estimation. Gaussian noise levels were available to NLM and BM3D, while DnCNN operated blind. Comparing blind and noise-informed variants under matched settings would help separate this advantage from differences among restorers. The ORB extension covers selected conditions, and the compact Real-ESRGAN experiment leaves diffusion and flow models untested.

7 Conclusion

On the original grid, the highest-PSNR denoiser fell short of the best observed geometric success by 0.42 pp on average and by at most 2.22 pp. This small selection cost coexisted with different rankings for NLM and Gaussian filtering in 10 of 12 strata. The original PSNR choice, DnCNN, also attained the highest mean success with both matchers on 60 pairs from 12 previously unused HPatches sequences at $\sigma = 25$. The direction of mean precision changes varied with matcher, noise level and control. Precision remained associated with homography success, although the model combining precision and PSNR explained little of the variation across pairs and configurations. For a new matching pipeline, evaluating promising restorers against the same degraded inputs gives the geometric effect and its uncertainty for the deployment setting.

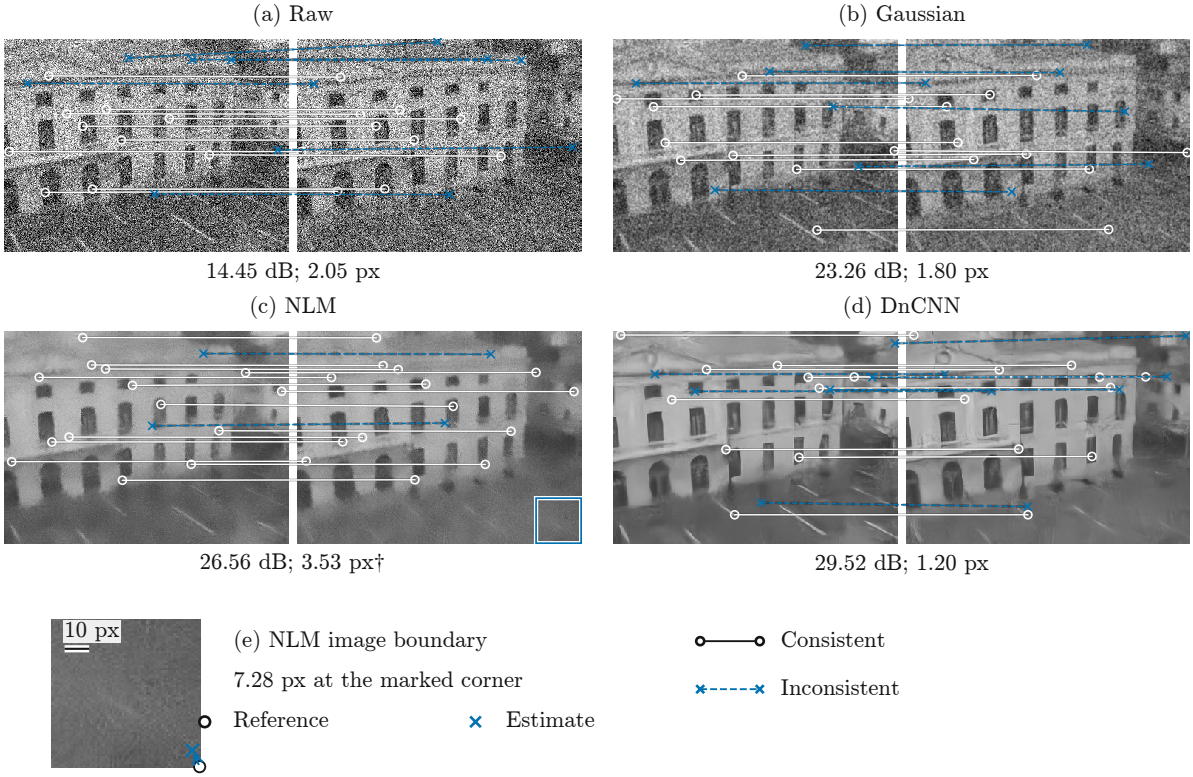

Fig. 4 LightGlue matches on i_ajuntament_1_2, $\sigma = 50$, seed 20261024. Each arm displays 16 matches sampled at uniformly spaced source-coordinate ranks. Open circles joined by solid lines mark matches with symmetric reference-transfer error below 3 pixels. Crosses joined by dashed lines mark the remaining matches. Labels give full-pair PSNR and mean corner error (†: failure). The NLM inset separates the marked corner displacement from the four-corner mean

Supplementary information. Online Resource 1 contains core controls, local derivations, paired effects, sequence-level statistical sensitivity, threshold and degradation checks, recorded image examples, and the sequence-disjoint follow-up.

Acknowledgements. The authors thank the maintainers of the HPatches and MegaDepth benchmarks and the developers of the open-source implementations used in this study.

Statements and Declarations

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. Not applicable. The study used public benchmark images and involved no participant recruitment or animal experiments.

Consent to participate. Not applicable.

Consent for publication. Not applicable. The article reports no identifiable participant information.

Author contributions. Congshuai Dong: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – original draft, Visualization, Project administration. **Wenxue Zhang:** Data curation, Validation, Investigation, Writing – review & editing. **Yiying Zhang:** Writing – review & editing, Validation. **Hao Zhang:** Software, Formal analysis, Visualization. **Yiteng Teng:** Validation, Resources, Investigation.

Data and code availability. Recorded measurements, result tables, analysis and plotting scripts, and implementation records form a reproducibility package described in Online Resource 1, Section S1, and are provided by the corresponding author for verification on request. Online Resource 1 reports the supporting results. Access to benchmark images and pretrained models follows the original providers' terms.

Declaration of generative AI and AI-assisted technologies. AI-assisted tools were used to revise the manuscript text and to draft plotting and analysis code under the authors' supervision. Experimental images and measurements derive from the recorded evaluation pipeline. The authors are responsible for verifying the scientific content and for the final decision to submit.

References

[1] Summers, J.M., Jones, M.W., Seale, C.: Impact of Underwater Image Enhancement on Feature Matching. Sensors **25**(22), 6966 (2025) https://doi.org/10.3390/s25226966

[2] Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 60–65 (2005). https://doi.org/10.1109/CVPR.2005.38

[3] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007) https://doi.org/10.1109/tip.2007.901238

[4] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017) https://doi.org/10.1109/tip.2017.2662206

[5] Blau, Y., Michaeli, T.: The Perception-Distortion Tradeoff. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6228–6237 (2018). https://doi.org/10.1109/cvpr.2018.00652

[6] Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, pp. 1905–1914 (2021). https://openaccess.thecvf.com/content/ICCV2021W/AIM/html/Wang_Real-ESRGAN_Training_Real-World_Blind_Super-Resolution_With_Pure_Synthetic_Data_ICCVW_2021_paper.html

[7] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004) https://doi.org/10.1023/b:visi.0000029664.99615.94

[8] Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2911–2918 (2012). https://doi.org/10.1109/CVPR.2012.6248018

[9] Lindenberger, P., Sarlin, P.E., Pollefeys, M.: LightGlue: Local Feature Matching at Light Speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 17581–17592 (2023). https://doi.org/10.1109/iccv51070.2023.01616

[10] 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 (CVPR), pp. 8918–8927 (2021). https://doi.org/10.1109/cvpr46437.2021.00881

[11] Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: RoMa: Robust Dense Feature Matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 19790–19800 (2024). https://doi.org/10.1109/cvpr52733.2024.01871

[12] Fischler, M.A., Bolles, R.C.: Random sample consensus. Communications of the ACM **24**(6), 381–395 (1981) https://doi.org/10.1145/358669.358692

[13] Barath, D., Noskova, J., Matas, J.: Marginalizing Sample Consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 8420–8432 (2022) https://doi.org/10.1109/tpami.2021.3103562

[14] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004). https://doi.org/10.1017/cbo9780511811685

[15] Mao, Y., Wang, B., Kulkarni, N., Park, J.J.: SIR-DIFF: Sparse Image Sets Restoration with Multi-View Diffusion Model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21620–21630 (2025). https://doi.org/10.1109/cvpr52734.2025.02014

[16] Su, D., Zhang, Y., Li, H., Li, J., Liu, Y.: UniFuse: A Unified All-In-One Framework for Multi-Modal Medical Image Fusion Under Diverse Degradations and Misalignments. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14238–14247 (2025). https://doi.org/10.1109/iccv51701.2025.01321

[17] Weinbach, B.C.: T-PhISH-Net: Temporally consistent underwater image enhancement via a transformer-based extension of PhISH-Net. Machine Vision and Applications **37**, 135 (2026) https://doi.org/10.1007/s00138-026-01903-z

[18] Lin, B., Gao, F., Yu, Y., Dong, J., Du, Q.: Downstream Task-Inspired Underwater Image Enhancement: A Perception-Aware Study From Dataset Construction to Network Design. IEEE Transactions on Image Processing **35**, 2873–2888 (2026) https://doi.org/10.1109/tip.2026.3671595

[19] Efron, B.: Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics **7**(1), 1–26 (1979) https://doi.org/10.1214/aos/1176344552

[20] Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: 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), pp. 3852–3861 (2017). https://doi.org/10.1109/cvpr.2017.410

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

[22] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004) https://doi.org/10.1109/tip.2003.819861

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf
- ESM1.pdf