DUALU: MODELING DUAL UNCERTAINTIES FOR CORRESPONDENCE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Correspondence learning aims to estimate geometric models from putative correspondences contaminated by severe outliers and is fundamental to many computer vision tasks. Existing methods have substantially improved outlier rejection by predicting inlier scores for each correspondence. However, under repeated structures or insufficient geometric context, outliers may exhibit patterns similar to inliers and consequently receive high scores, while even correctly identified inliers may still have different localization errors and therefore provide geometric measurements of varying accuracy. An inlier score alone does not quantify the reliability of the prediction or the geometric error of an inlier. To address this limitation, we propose DualU, a correspondence learning method that explicitly models dual uncertainties (i.e., the reliability of inlier predictions and the geometric errors of inliers), and uses them to guide geometric model estimation. Extensive experiments demonstrate that DualU outperforms state-of-the-art methods across multiple tasks and exhibits stronger generalization ability.

1 INTRODUCTION

Recovering a geometric model from image correspondences is a fundamental problem in computer vision that underlies relative pose estimation, visual localization, image registration, and large-scale reconstruction systems (Hartley & Zisserman, 2003; Ma et al., 2021; Sarlin et al., 2019; Schonberger & Frahm, 2016). Given a set of putative correspondences between two images, correspondence learning aims to identify inliers and use them to recover the underlying geometric model. In practice, weak textures, ambiguous scenarios, and appearance changes often produce unreliable correspondences, with outliers frequently substantially outnumbering inliers (Ma et al., 2021). Accurate geometric model estimation therefore requires both effective outlier rejection and appropriate use of the retained correspondences.

Classical robust estimators address this problem through hypothesize-and-verify sampling, where RANSAC (Fischler & Bolles, 1981) and its variants (Torr & Zisserman, 2000; Chum et al., 2003; Raguram et al., 2012; Barath et al., 2019; 2020) repeatedly fit models to minimal subsets and select the one with the largest geometric consensus. However, their sampling cost grows rapidly as the inlier ratio decreases. Learning-based methods instead predict an inlier score for each correspondence by exploiting the context among putative correspondences (Yi et al., 2018), and subsequent work has steadily strengthened this contextual reasoning through hierarchical filtering, progressive pruning, neighborhood consistency, and motion-field modeling (Zhang et al., 2019; Zhao et al., 2021; Liu & Yang, 2023; Zhang et al., 2024; 2025; Le & Ma, 2026). Across these methods, the predicted inlier scores are commonly used to select or weight correspondences during model estimation.

Despite these advances, an inlier score alone does not fully characterize a correspondence for geometric estimation. Under repeated structures or weak geometric context, an outlier can be indistinguishable from inliers and thus receive a high score; conversely, a correctly identified inlier may still carry a large localization error and provide an inaccurate geometric measurement. In other words, an inlier score reflects neither the reliability of the prediction itself nor the geometric error of the correspondence as an inlier. Accurate geometric estimation therefore requires explicitly modeling both aspects.

To address this limitation, we propose DualU, a correspondence learning method that explicitly models dual uncertainties: *inlier uncertainty* characterizes the reliability of the predicted inlier probabil-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ity, and *residual uncertainty* captures the geometric error of the correspondence as an inlier. DualU combines the two estimates into correspondence weights for a differentiable geometric solver, so that unreliable predictions and geometrically inaccurate inliers are suppressed during model estimation.

Our contributions are threefold. (i) We identify two complementary forms of uncertainty overlooked in correspondence learning. (ii) We introduce DualU, a correspondence learning method that explicitly models inlier uncertainty and residual uncertainty and uses them to construct correspondence weights for a differentiable geometric solver. (iii) Extensive experiments demonstrate that DualU outperforms state-of-the-art methods across multiple geometric estimation tasks.

2 Related Work

2.1 Traditional Methods

Traditional geometric estimation methods reject outliers through sampling or handcrafted consistency assumptions. RANSAC (Fischler & Bolles, 1981) and its variants (Torr & Zisserman, 2000; Chum et al., 2003; Raguram et al., 2012; Barath et al., 2019; 2020) generate hypotheses from minimal subsets and select the model with the largest consensus, with successive improvements in sampling, scoring, and noise-scale marginalization. Other methods identify consistent correspondences through local motion statistics, neighborhood preservation, or smooth vector fields (Bian et al., 2017; Ma et al., 2019; 2014). However, sampling becomes inefficient at low inlier ratios, and handcrafted consistency assumptions often fail under repeated structures or large viewpoint changes, which has motivated learning correspondence consistency directly from data.

2.2 Correspondence Learning

Correspondence learning predicts an inlier score for each correspondence by exploiting the relationships among putative matches. PointCN (Yi et al., 2018) introduces a permutation-equivariant architecture for joint inlier classification and model estimation, and subsequent work strengthens contextual reasoning through order-aware filtering, attentive normalization, progressive consensus pruning, and neighborhood consistency (Zhang et al., 2019; Sun et al., 2020; Zhao et al., 2021; Liu & Yang, 2023). More recent methods model correspondence relations as motion fields, either regularizing an implicit field or decomposing it into locally coherent or expert-specialized components (Zhang & Ma, 2023a; Zhang et al., 2024; 2025; Le & Ma, 2026). Despite their strong outlier rejection performance, these methods describe each correspondence with a single inlier score without explicitly modeling uncertainty.

2.3 Uncertainty Estimation

Uncertainty estimation has been widely studied in vision tasks to provide information beyond point predictions. Representative approaches distinguish model uncertainty from observation noise (Kendall & Gal, 2017) or place distributions over class probabilities to quantify predictive support (Sensoy et al., 2018). In correspondence-related tasks, uncertainty has been modeled as conditional error distributions or covariance estimates for dense matching and feature positions (Truong et al., 2021; 2023; Muhle et al., 2022; 2023; Edstedt et al., 2026; Cai et al., 2025). However, uncertainty remains underexplored in correspondence learning task, where most methods rely on deterministic inlier scores.

3 Theoretical Background

Given a set of putative correspondences $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^{N}$, correspondence learning aims to recover a geometric model $\mathbf{z} \in \mathbb{R}^D$ (e.g., the fundamental matrix or the homography) by casting the problem as point-wise inlier–outlier classification coupled with differentiable model estimation. A network predicts an inlier probability $q_i = P(y_i = 1 \mid \mathcal{C})$ for each correspondence, where $y_i \in \{0, 1\}$ is the inlier label, and the prediction is used as the weight in a geometric model estimation solver:

$$\widehat{\mathbf{z}} = \underset{\|\mathbf{z}\|_2=1}{\arg\min} \sum_{i=1}^{N} q_i \|\mathbf{M}_i \mathbf{z}\|_2^2. \tag{1}$$

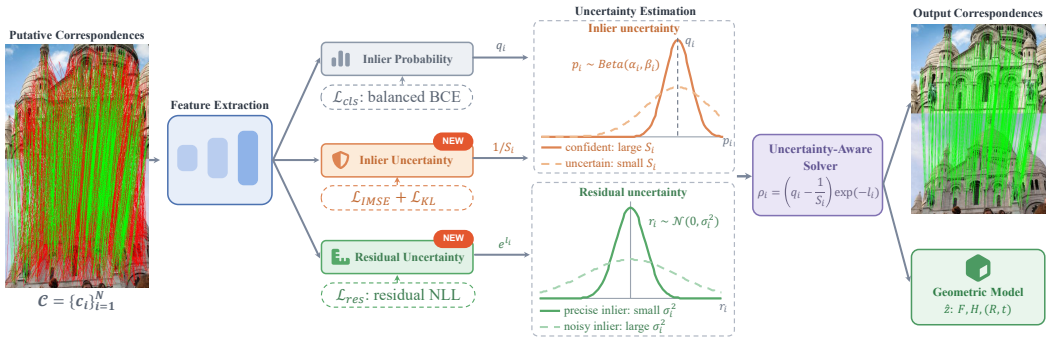

Figure 1: Overview of DualU. Given putative correspondences, the network predicts an inlier probability q_i together with an evidential-strength S_i and a residual log-variance ℓ_i for each correspondence. The two uncertainty estimates are combined with q_i into a correspondence weight that guides a differentiable geometric solver, and all predictions are optimized jointly through the classification, residual, and geometric losses.

Here, $\mathbf{M}_i$ is the data matrix induced by $\mathbf{c}_i$, so that $\|\mathbf{M}_i\mathbf{z}\|_2$ is the algebraic residual of the correspondence, and the unit-norm constraint avoids the trivial solution $\mathbf{z} = \mathbf{0}$. The construction of $\mathbf{M}_i$ for different tasks is detailed in Appendix A. Equivalently, with the weighted normal matrix $\mathbf{A}(\mathbf{q}) = \sum_{i=1}^N q_i \mathbf{M}_i^\top \mathbf{M}_i$, where $\mathbf{q} = (q_1, \ldots, q_N)^\top$, the solution is the unit eigenvector associated with the smallest eigenvalue of $\mathbf{A}(\mathbf{q})$. The classifier is supervised with a class-balanced binary cross-entropy loss $\mathcal{L}_{\text{bce}}$ (detailed in Appendix C), and the differentiable eigensolver further allows a geometric loss $\mathcal{L}_{\text{geo}}$ to constrain the estimated model.

In this formulation, q_i serves simultaneously as a classification result and as the weight assigned to a correspondence during model estimation. However, q_i alone quantifies neither the reliability of the inlier prediction nor the geometric error of the correspondence as an inlier, which motivates the dual-uncertainty modeling introduced in the following section.

4 Method

We model the uncertainty of each observation from two aspects. *Inlier uncertainty* arises because a correspondence classifier may assign a high inlier probability without sufficient support from the observed context. *Residual uncertainty* arises because even correctly identified inliers exhibit residuals of different scales and therefore contribute differently to the final model estimation.

As illustrated in Fig. 1, DualU builds on a conventional correspondence learning backbone that extracts a contextual representation $\mathbf{h}_i$ and predicts an inlier probability q_i for each observation, and introduces two additional outputs: an evidential-strength parameter S_i and a residual log-variance ℓ_i. The former models inlier uncertainty and the latter models residual uncertainty. During model estimation, DualU combines q_i with the two uncertainty estimates into each observation's weight in the differentiable solver. The following subsections introduce the modeling and supervision of the two uncertainties (Secs. 4.1 and 4.2), followed by the uncertainty-aware weighting and the overall training objective (Sec. 4.3).

4.1 Inlier Uncertainty

To quantify the reliability of the inlier prediction, we follow evidential deep learning (Sensoy et al., 2018) and place a Beta distribution over the latent inlier probability, $p_i \sim \text{Beta}(\alpha_i, \beta_i)$, whose mean coincides with the predicted inlier probability:

$$\mathbb{E}[p_i] = \frac{\alpha_i}{\alpha_i + \beta_i} = q_i. \tag{2}$$

Let $S_i = \alpha_i + \beta_i$ denote the total strength, so that the Beta distribution is fully specified by (q_i, S_i) through $\alpha_i = q_i S_i$ and $\beta_i = (1 - q_i)S_i$. For a fixed inlier probability, the strength controls how concentrated the distribution is around its mean. A larger S_i indicates that the prediction is supported by

stronger evidence, while a smaller S_i expresses greater uncertainty about the probability itself. Note that the strength S_i cannot be estimated freely. Evidential modeling views the Beta parameters as a uniform prior $\mathrm{Beta}(1,1)$ augmented with non-negative evidence for each class, so the parameters must satisfy $\alpha_i, \beta_i \geq 1$. This imposes a lower bound on the strength (derivation in Appendix B):

$$S_i \geq S_i^{\min} = \max\left(\frac{1}{q_i}, \frac{1}{1-q_i}\right). \tag{3}$$

We therefore parameterize the strength as $S_i = S_i^{\min} + \mathrm{softplus}(c_i)$, which maps the unconstrained network output c_i to a valid strength for any $q_i \in (0,1)$. The inlier evidence $\alpha_i - 1$ measures the support collected for the inlier hypothesis. Normalizing it by the total strength yields the belief mass assigned to the inlier class, which we term the positive belief mass (derivation in Appendix B):

$$b_i^+ = \frac{\alpha_i - 1}{S_i} = q_i - \frac{1}{S_i}. \tag{4}$$

Compared with the raw probability q_i, the belief mass decreases when a prediction is supported by little evidence. However, the constraints $(\alpha_i, \beta_i \geq 1)$ imply $\max(0, 2q_i - 1) \leq b_i^+ \leq q_i$. Thus, the positive belief remains close to q_i when q_i is close to one. In this case, the strength parameter has only a limited effect on the correspondence weight. The mean q_i and strength S_i are learned jointly using the integrated mean-squared error of evidential classification (Sensoy et al., 2018). For the two-class target $(y_i, 1-y_i)$,

$$\mathcal{L}_{\mathrm{IMSE},i} = \mathbb{E}_{p_i}\left[\|(y_i, 1-y_i) - (p_i, 1-p_i)\|_2^2\right] = \underbrace{2(y_i - q_i)^2}_{\mathcal{L}_{\mathrm{err},i}} + \underbrace{\frac{2q_i(1-q_i)}{S_i+1}}_{\mathcal{L}_{\mathrm{var},i}}. \tag{5}$$

The first term fits the mean inlier probability, while the second minimizes the variance of the Beta distribution. However, the variance term alone is insufficient, since it can be minimized by simply increasing S_i even when the predicted probability is wrong, i.e., the network can become confident about incorrect predictions. To prevent this, we remove the evidence of the ground-truth class by setting $\widetilde{\alpha}_i = y_i + (1-y_i)\alpha_i$ and $\widetilde{\beta}_i = (1-y_i) + y_i\beta_i$. The evidence corresponds to misleading support for the incorrect class and is penalized through its divergence from the uniform prior:

$$\mathcal{L}_{\mathrm{KL}} = \frac{1}{N}\sum_{i=1}^{N} D_{\mathrm{KL}}\left[\mathrm{Beta}(\widetilde{\alpha}_i, \widetilde{\beta}_i) \,\middle\|\, \mathrm{Beta}(1,1)\right]. \tag{6}$$

This divergence admits a closed form given in Appendix B. The overall classification loss is

$$\mathcal{L}_{\mathrm{cls}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_{\mathrm{IMSE},i} + \lambda_{\mathrm{KL}}(t)\mathcal{L}_{\mathrm{KL}}, \tag{7}$$

where $\lambda_{\mathrm{KL}}(t)$ is linearly annealed over the first T_{KL} iterations to avoid penalizing uncertain predictions during early training (details in Appendix C).

4.2 Residual Uncertainty

To quantify the geometric precision of each inlier, we model its algebraic residual as a zero-mean heteroscedastic Gaussian. Let $\mathbf{z}^{\mathrm{gt}}$ denote the ground-truth model and $r_i^{\mathrm{gt}} = \mathbf{M}_i\mathbf{z}^{\mathrm{gt}}$ the corresponding algebraic residual. We assume:

$$r_i^{\mathrm{gt}} \mid y_i = 1 \sim \mathcal{N}\left(0, \sigma_i^2\right), \tag{8}$$

where σ_i^2 is the observation-dependent residual variance. The network predicts the log-variance $\ell_i = \log\sigma_i^2$ and recovers the variance through $\sigma_i^2 = \exp(\ell_i)$, which guarantees positivity while keeping the network output unconstrained. Under this model, the residual likelihood of an inlier is

$$p(r_i \mid \mathbf{z}, y_i = 1, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{\|\mathbf{M}_i\mathbf{z}\|_2^2}{2\sigma_i^2}\right], \tag{9}$$

and ℓ_i is supervised by the negative log-likelihood over inliers:

$$\mathcal{L}_{\mathrm{res}} = \frac{1}{2N_+}\sum_{i=1}^{N} y_i\left[\exp(-\ell_i)\|r_i^{\mathrm{gt}}\|_2^2 + \ell_i\right], \tag{10}$$

where $N_+ = \sum_i y_i$ is the number of inliers. The first term scales the squared residual by the predicted precision, while the second prevents the variance from growing without bound.

4.3 UNCERTAINTY-AWARE SOLVER

To make the two uncertainty estimates guide model fitting, we construct correspondence weights using the predicted inlier probability and the two uncertainty estimates, inducing an uncertainty-aware estimator.

Under the Gaussian model in Eq. 8, an inlier contributes its quadratic term scaled by the precision σ_i^{-2}. Since $\mathbb{E}[y_i \mid \mathcal{C}] = q_i$, the expected quadratic contribution of each observation is

$$\mathbb{E}\left[\frac{y_i}{2\sigma_i^2}\|\mathbf{M}_i\mathbf{z}\|_2^2 \,\middle|\, \mathcal{C}\right] = \frac{q_i}{2\sigma_i^2}\|\mathbf{M}_i\mathbf{z}\|_2^2, \tag{11}$$

which suggests the weight q_i/σ_i^2. Since q_i is the mean of the Beta distribution, it ignores the strength S_i, so predictions with the same probability but different evidence receive identical weights. We therefore replace q_i with the positive belief mass b_i^+ in Eq. 4 and define

$$\rho_i = \frac{b_i^+}{\sigma_i^2} = \left(q_i - \frac{1}{S_i}\right)\exp(-\ell_i). \tag{12}$$

The normalized weight $w_i = \rho_i / \sum_j \rho_j$ gives the uncertainty-aware estimator

$$\widehat{\mathbf{z}} = \underset{\|\mathbf{z}\|_2=1}{\arg\min} \sum_{i=1}^{N} w_i \|\mathbf{M}_i\mathbf{z}\|_2^2, \qquad \mathbf{A}_{\text{unc}} = \sum_{i=1}^{N} w_i \mathbf{M}_i^\top \mathbf{M}_i, \tag{13}$$

whose solution is the unit eigenvector associated with the smallest eigenvalue of $\mathbf{A}_{\text{unc}}$. The belief term suppresses observations with insufficient inlier evidence, and the precision term downweights inliers with large predicted residuals. The objective for the final uncertainty head is $\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_{\text{res}}\mathcal{L}_{\text{res}} + \lambda_{\text{geo}}\mathcal{L}_{\text{geo}}$, where $\mathcal{L}_{\text{cls}}$ is the classification loss in Eq. 7, which replaces the balanced binary cross-entropy in Sec. 3 only for the final uncertainty head, and $\mathcal{L}_{\text{geo}}$ is the geometric loss, as described in Sec. 3.

5 EXPERIMENTS

5.1 IMPLEMENTATION DETAILS

DualU adopts DeMatch++ (Zhang et al., 2025) as the feature extraction backbone and retains its architecture and training settings. We train DualU for 500,000 iterations using the Adam optimizer (Kingma & Ba, 2014) with a batch size of 32 and an initial learning rate of 10^{-4}. The geometric loss weight λ_{geo} is set to 0 for the first 20,000 iterations and to 0.5 thereafter, and the residual negative log-likelihood is weighted by $\lambda_{\text{res}} = 10^{-3}$. The KL regularizer in Eq. 7 uses a maximum weight of 10^{-3} with a linear warm-up of 20,000 iterations. Correspondence labels are generated from the ground-truth model with a symmetric epipolar distance threshold of 10^{-4}, and keypoint coordinates are normalized by the camera intrinsics. All models are trained and evaluated on a single NVIDIA RTX 5090 GPU.

5.2 RELATIVE POSE ESTIMATION

Relative pose estimation recovers the relative rotation and translation between two cameras observing the same scene from different viewpoints.

Datasets. We conduct experiments on the outdoor YFCC100M dataset (Thomee et al., 2016) and the indoor SUN3D dataset (Xiao et al., 2013). Putative correspondences are generated by SIFT (Lowe, 2004) with nearest-neighbor matching, keeping at most 2,000 correspondences per image pair. YFCC100M contains 72 sequences, of which 68 are used for training and validation and the remaining four serve as unknown scenes. SUN3D consists of RGB-D sequences sampled every 10 frames, with 239 sequences for training and validation and 15 reserved for testing. Each dataset thus provides two settings: known scenes for estimation accuracy and unknown scenes for generalization to new environments.

Evaluation. Following previous methods (Zhang et al., 2019; 2025), we estimate the essential matrix using the weighted eight-point algorithm and decompose it to recover the relative rotation and

Table 1: Quantitative results of relative pose estimation on YFCC100M and SUN3D with weighted eight-point algorithm. AUC/mAP are reported. The best results are in **bold** and the second-best results are underlined.

Dataset	YFCC100M				SUN3D			
Method	Known Scene		Unknown Scene		Known Scene		Unknown Scene	
	5°	20°	5°	20°	5°	20°	5°	20°
OANet	14.16 / 33.04	47.85 / 56.81	15.94 / 39.32	57.08 / 67.67	8.18 / 21.42	40.11 / 48.77	5.98 / 16.37	34.33 / 42.22
T-Net	18.20 / 41.46	54.48 / 64.14	20.70 / 48.40	62.94 / 73.92	9.03 / 23.63	42.08 / 51.04	6.62 / 18.04	35.66 / 43.65
MSA-Net	15.69 / 36.04	49.68 / 58.81	20.95 / 48.70	62.31 / 73.07	7.24 / 19.58	38.54 / 47.03	6.07 / 16.77	34.32 / 42.10
MS2DG-Net	15.36 / 36.46	52.18 / 61.94	18.84 / 46.88	63.27 / 74.84	8.51 / 22.57	41.81 / 50.89	6.24 / 17.19	35.22 / 43.27
PGFNet	14.17 / 33.54	47.57 / 56.73	20.48 / 46.70	61.62 / 72.26	8.64 / 22.63	40.55 / 49.26	7.14 / 19.02	36.61 / 44.70
ConvMatch++	20.98 / 45.87	58.02 / 68.03	26.78 / 57.05	67.88 / 78.79	10.56 / 26.98	45.26 / 54.54	8.78 / 22.66	40.43 / 49.00
UMatch	21.80 / 46.84	58.79 / 68.76	30.90 / 60.32	69.75 / 80.28	9.62 / 25.01	42.87 / 51.87	8.05 / 21.40	38.70 / 47.12
NCMNet	25.02 / 50.24	61.31 / 71.27	32.40 / 62.65	71.82 / 82.30	9.96 / 24.99	42.88 / 51.87	7.92 / 20.41	38.16 / 46.42
DeMatch	23.78 / 49.39	60.49 / 70.54	30.97 / 61.88	70.38 / 81.07	11.82 / 29.58	46.97 / 56.43	9.33 / 23.93	41.44 / 50.23
DeMo	24.64 / 50.61	62.77 / 72.99	32.66 / 64.83	72.72 / 83.53	12.15 / 30.11	47.79 / 57.33	9.34 / 23.97	41.72 / 50.47
DeMatch++	26.33 / 53.46	63.49 / 73.66	34.33 / 65.92	72.96 / 83.45	12.91 / 31.23	48.52 / 58.07	9.87 / 25.26	42.24 / 51.13
GeoMoE	25.98 / 52.93	63.75 / 74.03	34.21 / 66.10	73.60 / 84.24	11.66 / 29.06	46.58 / 55.95	9.98 / 25.13	42.67 / 51.51
DualU (ours)	**30.77 / 59.55**	**67.53 / 77.71**	**39.45 / 69.85**	**75.56 / 85.45**	**13.18 / 31.51**	**48.91 / 58.49**	**10.28 / 25.29**	**42.77 / 51.55**

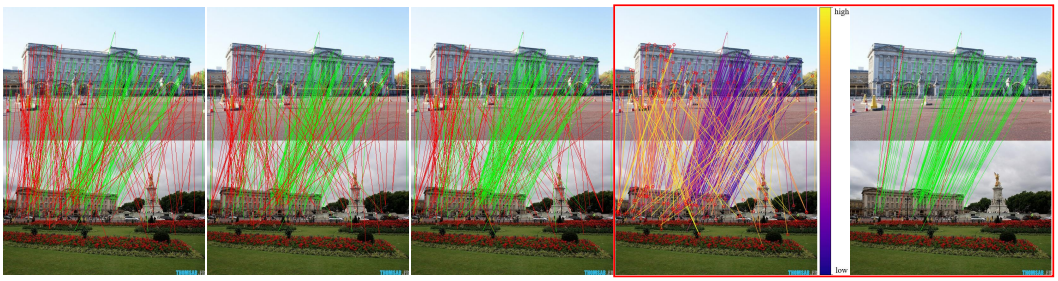

Figure 2: Qualitative comparison on YFCC100M. Green and red denote inliers and outliers, and yellow indicates high predicted uncertainty.

translation. We evaluate the maximum of the rotation and translation angular errors and report the area under the cumulative error curve (AUC) and mean average precision (mAP) at 5° and 20°. We compare DualU with representative learning-based methods, including OANet (Zhang et al., 2019), T-Net (Xiao et al., 2024), MSA-Net (Zheng et al., 2022), MS2DG-Net (Dai et al., 2022), PGFNet (Liu et al., 2023), ConvMatch++ (Zhang & Ma, 2023b), UMatch (Li et al., 2024), NCM-Net (Liu & Yang, 2023), DeMatch (Zhang et al., 2024), DeMo (Lu et al., 2025), DeMatch++ (Zhang et al., 2025), and GeoMoE (Le & Ma, 2026).

Results. DualU achieves the best AUC and mAP under all evaluation settings. The gains are most pronounced at the strict 5° threshold on YFCC100M, indicating more accurate geometric estimates. Since DualU shares the backbone with DeMatch++, the improvements over DeMatch++ are directly attributable to the proposed dual-uncertainty modeling. As shown in Fig. 2, the predicted uncertainty identifies the correspondences that DeMatch++ wrongly retains, which are then suppressed in the final estimate. Ablation studies and computational-cost comparisons are provided in Sec. 5.5.

5.3 Homography Estimation

Homography estimation recovers the projective transformation between two images of a planar scene or two images captured under pure camera rotation. In contrast to the epipolar geometry, this task evaluates the ability of DualU to generalize to a different geometric model.

Datasets. We evaluate on the HPatches benchmark (Balntas et al., 2017), which contains 116 scenes with 696 images in total, where 57 scenes exhibit illumination changes and the rest exhibit viewpoint changes. Each scene provides one reference image, five target images, and the ground-truth homographies. We detect up to 4,000 SIFT keypoints per image and generate putative correspondences by nearest-neighbor matching.

Table 2: Results of homography estimation on HPatches dataset.

Method	@1px	@3px	@5px	@7px	@10px
OANet Zhang et al. (2019)	16.38	39.31	51.38	58.45	66.21
UMatch Li et al. (2024)	15.69	41.38	55.34	60.69	67.93
ConvMatch Zhang & Ma (2023a)	20.00	47.41	61.03	66.72	73.62
DeMatch Zhang et al. (2024)	21.55	44.83	55.69	63.97	71.55
DeMatch++ Zhang et al. (2025)	25.00	49.66	62.07	67.24	72.93
GeoMoE Le & Ma (2026)	22.07	51.38	65.01	71.72	75.52
DualU (ours)	**31.72**	**63.45**	**72.93**	**76.72**	**80.52**

Table 3: Results of visual localization on the Aachen Day-Night v1.1 dataset.

Method	Day	Night
	(0.25m, 2°) / (0.5m, 5°) / (5.0m, 10°)	
RANSAC Fischler & Bolles (1981)	82.4 / 89.1 / 92.7	24.6 / 28.8 / 40.3
OANet Zhang et al. (2019)	85.4 / 93.2 / 96.5	45.0 / 55.5 / 74.3
UMatch Li et al. (2024)	85.7 / 92.6 / 96.6	50.8 / 62.3 / 77.5
ConvMatch Zhang & Ma (2023a)	85.6 / 93.3 / 96.5	51.8 / 64.9 / 81.2
DeMatch Zhang et al. (2024)	**85.9** / 93.2 / 96.5	53.4 / 64.4 / 79.6
DeMatch++ Zhang et al. (2025)	85.3 / **93.4** / 97.0	50.8 / 63.9 / 83.2
GeoMoE Le & Ma (2026)	85.7 / 93.2 / 97.1	52.9 / 64.4 / 82.2
DualU (ours)	85.0 / 92.5 / **97.6**	**55.5** / **70.2** / **84.3**

Evaluation. Following previous methods (DeTone et al., 2018; Zhang & Ma, 2023a), we measure the average corner error between the estimated and ground-truth homographies. The metric is computed by the percentage of image pairs whose error falls below a given pixel threshold. We report the accuracy at thresholds of 1, 3, 5, 7, and 10 pixels.

Results. As shown in Table 2, DualU outperforms all compared methods across all evaluation thresholds. Its advantage is particularly pronounced under the stricter thresholds. Moreover, because DualU is trained using epipolar geometry but evaluated here through homography estimation, these results demonstrate that its learned uncertainty estimates transfer effectively across different geometric models.

5.4 Visual Localization

Visual localization estimates the 6-DoF pose of a query image within a reference scene map. As a downstream task, it evaluates whether the correspondences selected and weighted by a correspondence learning method remain effective within a complete localization pipeline.

Dataset. We evaluate on the Aachen Day-Night v1.1 benchmark (Zhang et al., 2021), which contains 6,697 reference images, 824 daytime query images, and 191 nighttime query images captured using mobile phones. We extract up to 2,000 SIFT keypoints from each image and generate initial correspondences through mutual nearest-neighbor matching.

Evaluation. We integrate all compared methods into the HLoc pipeline (Sarlin et al., 2019) and use COLMAP (Schonberger & Frahm, 2016) to triangulate the reference model and recover query poses. For a fair comparison, all learning-based methods are trained on YFCC100M using 2,000 SIFT correspondences. We report the percentage of correctly localized queries under three translation and rotation error thresholds: $(0.25\,\mathrm{m}, 2^\circ)$, $(0.5\,\mathrm{m}, 5^\circ)$, and $(5\,\mathrm{m}, 10^\circ)$.

Results. As shown in Table 3, DualU is competitive with the state-of-the-art methods during the day and achieves the best result at the most relaxed threshold. Its advantage becomes clear at night, where it outperforms all compared methods under all three thresholds. Since nighttime illumination changes increase both the ambiguity of correspondence prediction and the variation of geometric errors, this improvement supports the value of modeling both uncertainties for robust localization under severe appearance changes.

Table 4: Generalization ability test using AUC/mAP@20°. The best results are in **bold** and the second-best results are underlined.

Method	YFCC100M			SUN3D			
	RootSIFT	SuperPoint	Xfeat	SIFT	RootSIFT	SuperPoint	Xfeat
OANet	57.73 / 68.54	35.64 / 43.71	17.99 / 22.87	10.31 / 13.64	10.75 / 14.19	10.09 / 13.25	7.45 / 9.98
UMatch	69.98 / 80.54	40.01 / 48.34	19.83 / 24.86	19.30 / 24.69	19.52 / 24.92	10.43 / 13.67	7.83 / 10.48
NCMNet	72.88 / 83.12	45.39 / 55.09	24.08 / 29.94	15.12 / 19.33	15.77 / 20.10	12.53 / 16.31	9.60 / 12.73
DeMatch	70.93 / 81.42	43.40 / 52.62	27.34 / 33.82	18.03 / 23.03	18.38 / 23.48	13.64 / 17.62	9.33 / 12.40
DeMatch++	73.41 / 83.95	48.68 / 58.24	29.39 / 35.94	20.90 / 26.59	21.15 / 26.82	13.23 / 17.15	9.34 / 12.33
GeoMoE	74.00 / 84.79	37.87 / 46.25	19.75 / 24.96	20.68 / 26.24	21.17 / 26.83	13.03 / 16.87	8.13 / 10.87
DualU (ours)	**76.02 / 85.86**	**57.65 / 68.16**	**41.23 / 49.86**	**21.33 / 27.17**	**21.83 / 27.77**	**17.64 / 22.46**	**10.65 / 14.03**

Table 5: Computational cost comparison.

Method	Time (ms)↓	FLOPs (G)↓	#Param (M)↓	AUC@5°↑
OANet Zhang et al. (2019)	9.771	1.841	2.473	15.94
UMatch Li et al. (2024)	37.635	3.741	7.765	30.90
ConvMatch Zhang & Ma (2023a)	18.780	3.783	7.494	26.78
NCMNet Liu & Yang (2023)	43.332	8.717	4.485	32.40
DeMatch Zhang et al. (2024)	23.869	2.346	5.853	30.97
DeMo Lu et al. (2025)	34.294	3.612	4.520	32.66
DeMatch++ Zhang et al. (2025)	29.471	3.726	5.986	34.33
GeoMoE Le & Ma (2026)	42.947	3.106	5.552	34.21
DualU (ours)	29.537	3.765	5.996	39.45

5.5 Model Analysis

Generalization Ability. The quality of putative correspondences depends on the feature descriptor, and the scene distribution varies across datasets. To evaluate robustness to these shifts, we train a single DualU model on YFCC100M with SIFT correspondences and evaluate it on YFCC100M and SUN3D with SIFT (Lowe, 2004), RootSIFT (Arandjelović & Zisserman, 2012), SuperPoint (DeTone et al., 2018), and XFeat (Potje et al., 2024). All methods use the weighted eight-point algorithm. As shown in Table 4, DualU achieves the best AUC and mAP at 20° in all settings. Its advantage is most pronounced when transferring from SIFT to SuperPoint and XFeat on YFCC100M, and it also show better performance on SUN3D, where scene and descriptor distributions shift simultaneously. These results indicate that modeling dual uncertainties reduces sensitivity to the correspondence distribution observed during training.

Computational Cost. Table 5 compares the average inference time per image, floating-point operations (FLOPs), the number of model parameters (#Params), and AUC at 5° on the unknown scenes of YFCC100M. Compared with DeMatch++, DualU introduces only approximately 0.07 ms of additional inference time, 0.04 G FLOPs, and 0.01 M parameters, owing to the lightweight uncertainty heads and the shared closed-form eigensolver. Despite having nearly identical computational cost to DeMatch++, DualU achieves a substantial improvement in pose estimation accuracy. DualU therefore provides a favorable balance between estimation accuracy and computational efficiency.

Ablation. We evaluate the contribution of each component of DualU on the unknown scenes of YFCC100M and report AUC at 5°, 10°, and 20° in Table 6. Experiment (i) uses DeMatch++ with the conventional probability-weighted solver as the baseline. Experiments (ii) and (iii) remove inlier uncertainty and residual uncertainty, respectively. Both variants outperform the baseline, demonstrating that each uncertainty contributes independently to geometric estimation. The variant retaining residual uncertainty produces a larger improvement, particularly under the strictest threshold, indicating the importance of accounting for differences in the geometric precision of inliers. Experiment (iv) removes the KL regularizer in Eq. 7, leading to lower performance in most thresholds and confirming its role in suppressing unsupported evidence. It is worth noting that several methods (Zhao et al., 2021; Liu & Yang, 2023; Miao et al., 2024) predict an additional weight beyond the inlier probability to guide the weighted solver, which also brings performance gains. Experiment (v) therefore replaces the proposed uncertainty-based weight with a freely learned per-correspondence weight. Although this variant improves upon the baseline, it remains inferior to the DualU, showing that the improvement cannot be attributed solely to the additional weighting capacity. Finally, the complete DualU model in experiment (vi) achieves the best performance at most threshold,

Table 6: Ablation studies of DualU on YFCC100M dataset.

Method	YFCC100M		
	AUC@5°	AUC@10°	AUC@20°
(i) Baseline	34.33	56.00	72.96
(ii) w/o inlier uncertainty	38.54	59.38	75.19
(iii) w/o residual uncertainty	35.81	57.57	74.18
(iv) w/o. KL	38.42	59.70	75.64
(v) Baseline w. free weight	37.61	58.11	74.38
(vi) DualU (ours)	39.45	60.06	75.56

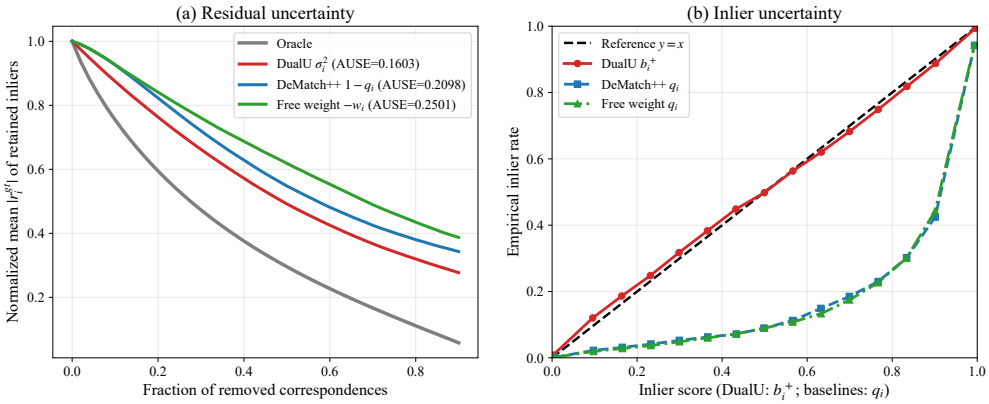

Figure 3: Effectiveness of the two uncertainty estimates on the unknown scenes of YFCC100M. (a) Sparsification plot of the residual uncertainty. (b) Empirical inlier rate versus the predicted inlier score.

demonstrating that inlier uncertainty and residual uncertainty provide complementary information for geometric model estimation.

Effectiveness. We directly evaluate the effectiveness of the two uncertainty estimates and whether they provide information that is useful to the geometric solver. For residual uncertainty, we use sparsification plots and the Area Under the Sparsification Error curve (AUSE) (Ilg et al., 2018; Truong et al., 2021; 2023). We rank ground-truth inliers by σ_i^2, progressively remove uncertain correspondences, and measure the normalized mean residual of those retained. As shown in Figure 3(a), DualU produces the sparsification curve closest to the oracle and the lowest AUSE among the compared scores, showing that σ_i^2 effectively ranks the geometric precision of inliers.

For inlier uncertainty, we assess whether the predicted score reflects correspondence correctness. We group correspondences by their predicted scores and measure the fraction of ground-truth inliers in each group, referred to as the empirical inlier rate. DualU uses the positive belief b_i^+, whereas DeMatch++ and the free-weight variant (v) in Table 6 use their respective inlier probabilities q_i. Figure 3(b) shows that groups with larger b_i^+ contain a higher proportion of true inliers, demonstrating that the uncertainty-adjusted belief is informative of correspondence correctness.

6 Conclusion

In this work, we presented DualU, a correspondence learning method that explicitly models two complementary uncertainties. Inlier uncertainty captures the reliability of predicted inlier probabilities, while residual uncertainty reflects the geometric precision of inliers. The two uncertainty estimates are combined with the predicted inlier probability to weight correspondences in a differentiable geometric solver, which suppresses unreliable and geometrically imprecise correspondences. Experiments on multiple tasks show that DualU outperforms state-of-the-art methods and generalizes better to unseen scenes and feature descriptors.

AI USE STATEMENT

In this work, we used generative AI tools for designing or providing feedback on research methodology or experiments, implementing methods, assisting with translation, cleaning and reformatting datasets, supporting qualitative and thematic data analysis, and interpreting results. We have not used generative AI tools for generating synthetic data sets or helping develop theoretical models or conceptual frameworks, and formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, and proposing or refining hypotheses are not applicable to this work. Additionally, we used generative AI tools for creating or editing software code, editing the paper to improve readability, and formatting references. We have reviewed all AI-assisted work. We checked LLM-generated research ideas for potential plagiarism through a manual literature survey, and LLM-generated code was verified and tested for correctness by two authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All experiments are conducted on publicly available datasets, including YFCC100M, SUN3D, HPatches, and Aachen Day-Night v1.1, with standard preprocessing and evaluation protocols described in Sec. 5. Implementation details, including training hyperparameters and network settings, are reported in Sec. 5 and Sec. C. Source code and pretrained models will be released publicly upon acceptance.

REFERENCES

Relja Arandjelović and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2911–2918. IEEE, 2012.

Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5173–5182, 2017.

Daniel Barath, Jiri Matas, and Jana Noskova. Magsac: marginalizing sample consensus. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10197–10205, 2019.

Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiri Matas. Magsac++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1304–1312, 2020.

JiaWang Bian, Wen-Yan Lin, Yasuyuki Matsushita, Sai-Kit Yeung, Tan-Dat Nguyen, and Ming-Ming Cheng. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4181–4190, 2017.

Lile Cai, Chuan-Sheng Foo, Xun Xu, Zaiwang Gu, Jun Cheng, and Xulei Yang. Evidential learning-based certainty estimation for robust dense feature matching. In *International Conference on Learning Representations*, 2025.

Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Joint pattern recognition symposium*, pp. 236–243. Springer, 2003.

Luanyuan Dai, Yizhang Liu, Jiayi Ma, Lifang Wei, Taotao Lai, Changcai Yang, and Riqing Chen. Ms2dg-net: Progressive correspondence learning via multiple sparse semantics dynamic graph. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8973–8982, 2022.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 224–236, 2018.

Johan Edstedt, David Nordström, Yushan Zhang, Georg Bökman, Jonathan Astermark, Viktor Larsson, Anders Heyden, Fredrik Kahl, Mårten Wadenbäck, and Michael Felsberg. RoMa v2: Harder better faster denser feature matching. In *European Conference on Computer Vision*, 2026.

Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24 (6):381–395, 1981.

Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

Eddy Ilg, Ozgun Cicek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 652–667, 2018.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Jiajun Le and Jiayi Ma. Geomoe: Divide-and-conquer motion field modeling with mixture-of-experts for two-view geometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 5845–5853, 2026.

Zizhuo Li, Shihua Zhang, and Jiayi Ma. U-match: Exploring hierarchy-aware local context for two-view correspondence learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Xin Liu and Jufeng Yang. Progressive neighbor consistency mining for correspondence pruning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9527–9537, 2023.

Xin Liu, Guobao Xiao, Riqing Chen, and Jiayi Ma. Pgfnet: Preference-guided filtering network for two-view correspondence learning. *IEEE Transactions on Image Processing*, 32:1367–1378, 2023.

David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

Yifan Lu, Jiajun Le, Zizhuo Li, Yixuan Yuan, and Jiayi Ma. Deep motion field consensus with learnable kernels for two-view correspondence learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5829–5837, 2025.

Jiayi Ma, Ji Zhao, Jinwen Tian, Alan L Yuille, and Zhuowen Tu. Robust point matching via vector field consensus. *IEEE Transactions on Image Processing*, 23(4):1706–1721, 2014.

Jiayi Ma, Ji Zhao, Junjun Jiang, Huabing Zhou, and Xiaojie Guo. Locality preserving matching. *International Journal of Computer Vision*, 127(5):512–531, 2019.

Jiayi Ma, Xingyu Jiang, Aoxiang Fan, Junjun Jiang, and Junchi Yan. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision*, 129(1):23–79, 2021.

Xiangyang Miao, Guobao Xiao, Shiping Wang, and Jun Yu. Bclnet: Bilateral consensus learning for two-view correspondence pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4225–4232, 2024.

Dominik Muhle, Lukas Koestler, Nikolaus Demmel, Florian Bernard, and Daniel Cremers. The probabilistic normal epipolar constraint for frame-to-frame rotation optimization under uncertain feature positions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1819–1828, 2022.

Dominik Muhle, Lukas Koestler, Krishna Murthy Jatavallabhula, and Daniel Cremers. Learning correspondence uncertainty via differentiable nonlinear least squares. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13102–13112, 2023.

Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2682–2691, 2024.

Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: A universal framework for random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):2022–2038, 2012.

Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12716–12725, 2019.

Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113, 2016.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.

Weiwei Sun, Wei Jiang, Eduard Trulls, Andrea Tagliasacchi, and Kwang Moo Yi. Acne: Attentive context normalization for robust permutation-equivariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11286–11295, 2020.

Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73, 2016.

Philip HS Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer vision and image understanding*, 78(1):138–156, 2000.

Prune Truong, Martin Danelljan, Luc Van Gool, and Radu Timofte. Learning accurate dense correspondences and when to trust them. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5714–5724, 2021.

Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. PDC-Net+: Enhanced probabilistic dense correspondence network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10247–10266, 2023. doi: 10.1109/TPAMI.2023.3249225.

Guobao Xiao, Xin Liu, Zhen Zhong, Xiaoqin Zhang, Jiayi Ma, and Haibin Ling. T-net++: Effective permutation-equivariance network for two-view correspondence pruning. *IEEE transactions on pattern analysis and machine intelligence*, 46(12):10629–10644, 2024.

Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1625–1632, 2013.

Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2666–2674, 2018.

Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. Learning two-view correspondences and geometry using order-aware network. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5845–5854, 2019.

Shihua Zhang and Jiayi Ma. Convmatch: Rethinking network design for two-view correspondence learning. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 37, pp. 3472–3479, 2023a.

Shihua Zhang and Jiayi Ma. Convmatch: Rethinking network design for two-view correspondence learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):2920–2935, 2023b.

Shihua Zhang, Zizhuo Li, Yuan Gao, and Jiayi Ma. Dematch: Deep decomposition of motion field for two-view correspondence learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 20278–20287, June 2024.

Shihua Zhang, Zizhuo Li, and Jiayi Ma. Dematch++: Two-view correspondence learning via deep motion field decomposition and respective local-context aggregation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Zichao Zhang, Torsten Sattler, and Davide Scaramuzza. Reference pose generation for long-term visual localization via learned features and view synthesis. *International Journal of Computer Vision*, 129(4):821–844, 2021.

Chen Zhao, Yixiao Ge, Feng Zhu, Rui Zhao, Hongsheng Li, and Mathieu Salzmann. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6464–6473, 2021.

Linxin Zheng, Guobao Xiao, Ziwei Shi, Shiping Wang, and Jiayi Ma. Msa-net: Establishing reliable correspondences by multiscale attention network. *IEEE Transactions on Image Processing*, 31: 4598–4608, 2022.

A Problem Formulations for Geometric Estimation Tasks

In this section, we provide details on how to construct $\mathbf{M}_i$ for the algebraic formulation used in Eq. 1, covering hyperplane fitting, fundamental matrix estimation, and homography estimation.

A.1 Hyperplane Fitting

For hyperplane fitting, each observation is a point $\mathbf{c}_i = \mathbf{x}_i \in \mathbb{R}^d$ in a d-dimensional space. The estimated hyperplane can be written as

$$\mathbf{a}^\mathsf{T}\mathbf{x} + c = 0, \tag{14}$$

where $\mathbf{a} \in \mathbb{R}^d$ denotes the normal vector and $c \in \mathbb{R}$ is the offset. For a point lying on the hyperplane, the above equation induces a linear constraint:

$$\mathbf{M}_i\mathbf{z} = 0, \tag{15}$$

where

$$\mathbf{M}_i = \left(\mathbf{x}_i^\mathsf{T}, 1\right), \qquad \mathbf{z} = \left(\mathbf{a}^\mathsf{T}, c\right)^\mathsf{T}. \tag{16}$$

Therefore, hyperplane fitting can be regarded as a direct high-dimensional generalization of 2D line fitting, where the goal is to estimate the normal vector and bias term from a set of noisy observations with outliers.

A.2 Fundamental Matrix

In fundamental matrix estimation, each observation $\mathbf{c}_i = (\mathbf{x}_i, \mathbf{x}_i')$ denotes a pair of 2D image correspondences, where

$$\mathbf{x}_i = (\xi_i, \eta_i, 1)^\mathsf{T}, \qquad \mathbf{x}_i' = (\xi_i', \eta_i', 1)^\mathsf{T} \tag{17}$$

denote the homogeneous coordinates of feature points detected from two different images. The epipolar constraint associated with a fundamental matrix $\mathbf{F} \in \mathbb{R}^{3\times3}$ is expressed as

$$\mathbf{x}_i'^\mathsf{T}\mathbf{F}\mathbf{x}_i = 0. \tag{18}$$

The fundamental matrix $\mathbf{F}$ has rank two and seven degrees of freedom. If we stack the entries of $\mathbf{F}$ in row-major order to form a nine-dimensional vector $\mathbf{z}$, the epipolar constraint can be equivalently rewritten as

$$\mathbf{M}_i\mathbf{z} = 0, \tag{19}$$

where

$$\mathbf{M}_i = \big(\xi_i'\xi_i,\ \xi_i'\eta_i,\ \xi_i',\ \eta_i'\xi_i,\ \eta_i'\eta_i, \eta_i',\ \xi_i,\ \eta_i,\ 1\big), \tag{20}$$

and

$$\mathbf{z} = (z_1, z_2, \ldots, z_9)^\mathsf{T} \tag{21}$$

contains the entries of $\mathbf{F}$.

A.3 Homography Transformation

For homography estimation, each observation $\mathbf{c}_i = (\mathbf{x}_i, \mathbf{x}_i')$ represents a 2D image correspondence. The homogeneous coordinates are defined as

$$\mathbf{x}_i = (\xi_i, \eta_i, 1)^\mathsf{T}, \qquad \mathbf{x}_i' = (\xi_i', \eta_i', 1)^\mathsf{T}. \tag{22}$$

The geometric relationship between the two views is formulated by a homography matrix $\mathbf{H} \in \mathbb{R}^{3\times3}$, which satisfies

$$\mathbf{x}_i' \times \mathbf{H}\mathbf{x}_i = \mathbf{0}. \tag{23}$$

The above equation provides two independent linear constraints. Each constraint induces a scalar algebraic residual and can therefore be treated as a separate observation, whose data matrix is the corresponding row of

$$\begin{bmatrix} \mathbf{0}^\mathsf{T} & -\mathbf{x}_i^\mathsf{T} & \eta_i'\mathbf{x}_i^\mathsf{T} \\ \mathbf{x}_i^\mathsf{T} & \mathbf{0}^\mathsf{T} & -\xi_i'\mathbf{x}_i^\mathsf{T} \end{bmatrix}, \tag{24}$$

and the parameter vector is defined as

$$\mathbf{z} = (z_1, z_2, \ldots, z_9)^\mathsf{T}, \tag{25}$$

which contains the entries of $\mathbf{H}$ arranged in row-major order.

In the absence of outliers, DLT (Hartley & Zisserman, 2003) typically approximates the true solution closely. However, real-world data invariably contain outliers. Consequently, relying on DLT can lead to significantly erroneous estimation, as the method is highly sensitive to even a single outlier.

B Derivations of Inlier Uncertainty

This section collects the derivations and analytical properties of the inlier-uncertainty model in Sec. 4.1. We first derive the lower bound on the evidential strength, then give the derivation of the positive belief mass together with its subjective-logic interpretation, analyze how the strength controls the dispersion of the latent inlier probability, and finally derive the closed form of the KL regularizer.

Lower bound on the evidential strength. Evidential modeling views the Beta parameters as a uniform prior $\mathrm{Beta}(1,1)$ augmented with non-negative evidence for each class. Writing $\alpha_i = e_i^+ + 1$ and $\beta_i = e_i^- + 1$ with $e_i^+, e_i^- \geq 0$, valid evidence requires $\alpha_i, \beta_i \geq 1$. Substituting the mean–strength relations $\alpha_i = q_i S_i$ and $\beta_i = (1-q_i)S_i$ from Eq. 2 gives

$$\alpha_i = q_i S_i \geq 1 \implies S_i \geq \frac{1}{q_i}, \qquad \beta_i = (1-q_i)S_i \geq 1 \implies S_i \geq \frac{1}{1-q_i}, \tag{26}$$

which together yield $S_i \geq S_i^{\min} = \max\big(1/q_i,\ 1/(1-q_i)\big)$ as stated in Eq. 3. The parameterization $S_i = S_i^{\min} + \mathrm{softplus}(c_i)$ in Sec. 4.1 satisfies this constraint by construction.

Derivation of the positive belief mass. The main text defines the positive belief mass as $b_i^+ = (\alpha_i - 1)/S_i$, i.e., the inlier evidence normalized by the total strength. Substituting $\alpha_i = q_i S_i$ from Eq. 2 gives the closed form in Eq. 4:

$$b_i^+ = \frac{\alpha_i - 1}{S_i} = q_i - \frac{1}{S_i}. \tag{27}$$

Under subjective logic, the uncertainty mass of a binary classification is $u_i = 2/S_i$, so the belief mass can be equivalently written as $b_i^+ = q_i - u_i/2$, i.e., the predicted probability discounted by half of the uncommitted uncertainty.

Beta variance and its relation to the uncertainty mass. This part formalizes the effect of the strength stated in Sec. 4.1. The mean and variance of the Beta distribution are

$$\mathbb{E}[p_i] = q_i, \qquad \mathrm{Var}(p_i) = \frac{q_i(1-q_i)}{S_i+1}, \tag{28}$$

where the variance follows from the standard Beta variance $\mathrm{Var}(p_i) = \alpha_i\beta_i/[(\alpha_i+\beta_i)^2(\alpha_i+\beta_i+1)]$ after substituting $\alpha_i = q_iS_i$ and $\beta_i = (1-q_i)S_i$. For a fixed q_i, increasing S_i decreases $\mathrm{Var}(p_i)$, while b_i^+ approaches q_i. The uncertainty mass u_i and the variance $\mathrm{Var}(p_i)$ are related but distinct: the former represents uncommitted belief under subjective logic, whereas the latter measures the dispersion of the latent inlier probability. Both are deterministic functions of the Beta parameters rather than additional network outputs, and neither should be interpreted as a posterior over the network parameters.

Closed form of the Beta KL divergence. We derive the closed form of the point-wise divergence in Eq. 6. For two Beta distributions, the KL divergence is

$$\begin{aligned} D_{\mathrm{KL}}\big[\,\mathrm{Beta}(\widetilde{\alpha}_i,\widetilde{\beta}_i)\,\|\,\mathrm{Beta}(1,1)\big] = \log\frac{B(1,1)}{B(\widetilde{\alpha}_i,\widetilde{\beta}_i)} \\ + (\widetilde{\alpha}_i-1)\big[\psi(\widetilde{\alpha}_i)-\psi(\widetilde{S}_i)\big] + (\widetilde{\beta}_i-1)\big[\psi(\widetilde{\beta}_i)-\psi(\widetilde{S}_i)\big], \end{aligned} \tag{29}$$

where $B(\cdot,\cdot)$ is the Beta function, $\psi(\cdot)$ is the digamma function, and

$$\widetilde{S}_i = \widetilde{\alpha}_i + \widetilde{\beta}_i. \tag{30}$$

Since $B(1,1) = 1$ and $B(\widetilde{\alpha}_i,\widetilde{\beta}_i) = \Gamma(\widetilde{\alpha}_i)\Gamma(\widetilde{\beta}_i)/\Gamma(\widetilde{S}_i)$, the point-wise divergence has the closed form

$$\mathcal{D}_i = \log\Gamma(\widetilde{S}_i) - \log\Gamma(\widetilde{\alpha}_i) - \log\Gamma(\widetilde{\beta}_i) + (\widetilde{\alpha}_i-1)[\psi(\widetilde{\alpha}_i)-\psi(\widetilde{S}_i)] + (\widetilde{\beta}_i-1)[\psi(\widetilde{\beta}_i)-\psi(\widetilde{S}_i)]. \tag{31}$$

All terms are differentiable with respect to $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$, so the KL regularizer in Eq. 6 can be optimized end to end.

C Training Details

Balanced binary cross-entropy. Conventional correspondence learning supervises the inlier classifier using a class-balanced binary cross-entropy:

$$\mathcal{L}_{\mathrm{bce}} = -\frac{1}{2N_+}\sum_{i=1}^{N} y_i \log q_i - \frac{1}{2N_-}\sum_{i=1}^{N}(1-y_i)\log(1-q_i), \tag{32}$$

where $N_+ = \sum_{i=1}^{N} y_i$ and $N_- = N - N_+$. The per-class normalization prevents the loss from being dominated by the majority class, which is typically the outliers. DualU replaces this point-wise supervision with the evidential classification loss in Eq. 7.

KL annealing. The KL weight in Eq. 7 is linearly annealed over the first T_{KL} iterations to avoid penalizing uncertain predictions during early training:

$$\lambda_{\mathrm{KL}}(t) = \lambda_{\mathrm{KL}}^{\max}\min\left(1, \frac{t}{T_{\mathrm{KL}}}\right), \tag{33}$$

where t is the training iteration. The values of $\lambda_{\mathrm{KL}}^{\max}$ and T_{KL} are reported in the experimental setup.

D Visualization Results

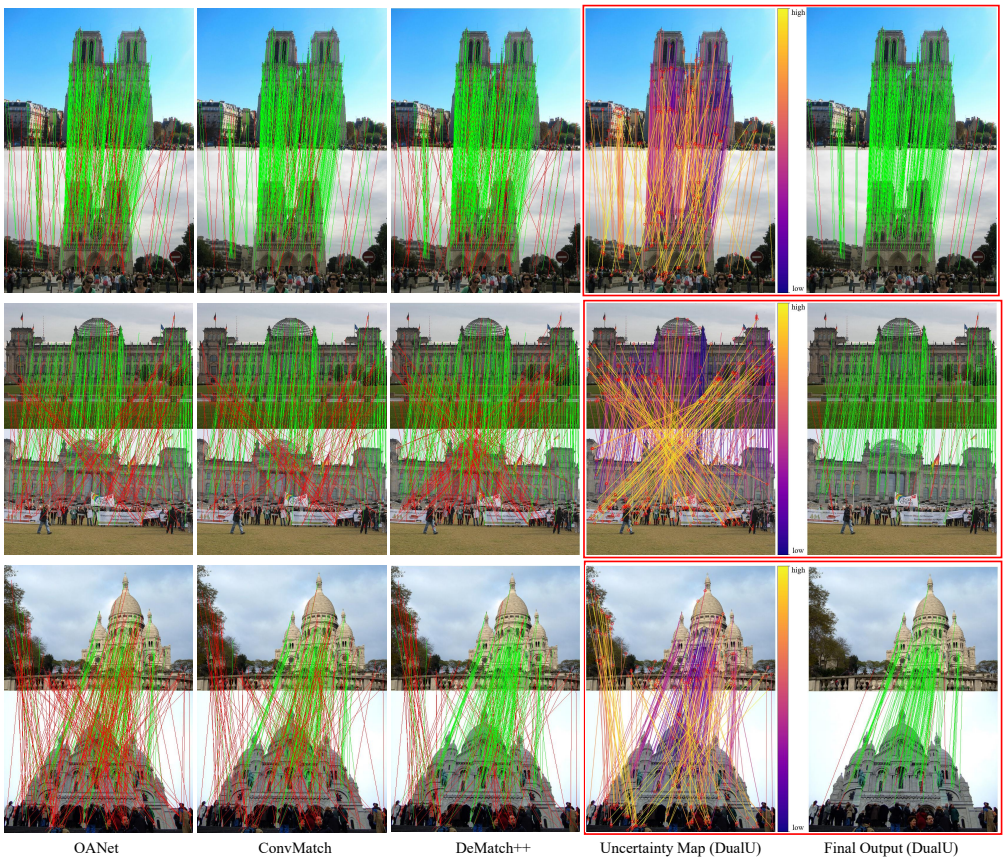

Figure 4: Qualitative comparison on YFCC100M. Green and red denote inliers and outliers, and yellow indicates high predicted uncertainty.