

Using outlier elimination to assess learning-based correspondence matching methods

Xintao Ding^{a,b,c}, Yonglong Luo^{a,b,c}, Biao Jie^{a,b,c}, Qingde Li^{d,*} and Yongqiang Cheng^{e,*}

^aSchool of Computer and Information, Anhui Normal University, Wuhu, 241002, Anhui, China

^bAnhui Province Key Laboratory of Network and Information Security, Wuhu, 241002, Anhui, China

^cAnhui Engineering Research Center of Medical Big Data Intelligent System, Wuhu, 241002, Anhui, China

^dSchool of Computer Science, University of Hull, Hull, HU67RX, UK

^eFaculty of Technology, University of Sunderland, Sunderland, SD60DD, UK

ARTICLE INFO

Keywords:

Correspondence matching
Multi-scale
Outer neighborhood
Poisoning attacks
Registration assessment

ABSTRACT


Recently, deep learning (DL) technology has been widely used in correspondence matching. The learning-based models are usually trained on benign image pairs with partial overlaps. Since DL model is usually data-dependent, non-overlapping images may be used as poison samples to fool the model and produce false registrations. In this study, we propose an outlier elimination-based assessment method (OEAM) to assess the registrations of learning-based correspondence matching method on partially overlapping and non-overlapping images. OEAM first eliminates outliers based on spatial paradox. Then OEAM implements registration assessment in two streams using the obtained core correspondence set. If the cardinality of the core set is sufficiently small, the input registration is assessed as a low-quality registration. Otherwise, it is assessed to be of high quality, and OEAM improves its registration performance using the core set. OEAM is a post-processing technique imposed on learning-based method. The comparison experiments are implemented on outdoor (YFCC100M) and indoor (SUN3D) datasets using four deep learning-based methods. The experimental results on registrations of partially overlapping images show that OEAM can reliably infer low-quality registrations and improve performance on high-quality registrations. The experiments on registrations of non-overlapping images demonstrate that learning-based methods are vulnerable to poisoning attacks launched by non-overlapping images, and OEAM is robust against poisoning attacks crafted by non-overlapping images.

1. Introduction

Correspondence matching, which aims to find the true matches from initial correspondences contaminated by numerous false matches (also called mismatches), is a fundamental problem in computer vision and pattern recognition [1, 2]. It has been widely used in many computer vision tasks, such as image registration [3, 4], geometry baseline estimation [5], structure understanding and motion estimation [6, 7, 8], and point cloud registration [9, 10].

In recent years, learning-based approaches have been proposed for correspondence matching, such as learning to find good correspondences (LFGC) [11], order-aware network (OANet) [12], SuperGlue [13], and local feature matching with Transformers (LoFTR) [14]. Traditionally, deep learning (DL) models are trained on benign image pairs with partial overlaps [11, 12, 13, 14]. The study on the model response of non-overlapping image pairs is insufficient. In addition, learning-based methods usually depend on hardware. The method application on a handheld device may have to submit its request to an artificial intelligence (AI) service system. As a servo system, the learning-based model should respond to various inputs, such as non-rigid in vivo images [15], object instances of a common category [16], and non-overlapping images. For two non-overlapping images, most of the matches between them are false inliers. However, many learning-based methods usually incorrectly output inliers. To show the vulnerability of the learning-based method on non-overlapping images, we build non-overlapping image pairs on YFCC100M [17] for demonstration. The non-overlapping image pair (I_0, I_1) for registration is built using images on four subsets buckingham, notre, reichstag, and sacre coeur, denoted as (X_0, X_1, X_2, X_3) , where $I_0 \in X_i, I_1 \in X_j, i \neq j, i, j = 0, 1, 2, 3$. We run OANet to

*Corresponding authors.

 xintaoding@163.com (X. Ding); ylluo@ustc.edu.cn (Y. Luo); jbiao@nuaa.edu.cn (B. Jie); Q.Li@hull.ac.uk (Q. Li); yongqiang.cheng@sunderland.ac.uk (Y. Cheng)

ORCID(s):

show the demonstration (Fig. 1). Fig. 1a shows that OANet usually unreasonably produces a great number of inliers on non-overlapping image pairs. It seems that non-overlapping images can be used as poison samples to fool OANet and produce false registration. Learning-based methods may be vulnerable to poisoning attacks. Furthermore, the experiment results on mean average precision (mAP) in [12] show that OANet sometimes outputs false matches on registrations of benign samples with partial overlaps. It shows that performance assessment of learning-based methods on benign and poison samples is necessary for AI applications.

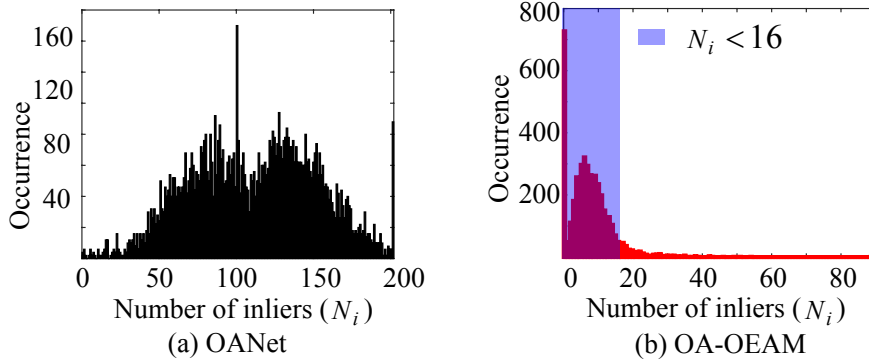


Fig. 1: The inlier number histograms of non-overlapping images on the YFCC100M dataset: (a) OANet, (b) OA-OEAM.

In recent decades, various robust methods have been proposed to explore inliers from contaminated mismatches. Generally, correspondence matching methods can be roughly categorized into three classes: (1) RANSAC-based methods, (2) outlier elimination-based methods, and (3) learning-based methods. RANSAC-based methods usually launch contributions in three streams: sampling strategy [18], model verification [19], and termination criterion [20]. There are also some studies that integrate several technologies to improve RANSAC, such as graph-cut RANSAC [21] and VSAC [22]. RANSAC-based methods are typically implemented using a loop structure [21]. They usually reach the maximum iteration count on non-overlapping image registration and consequently are inefficient against the poisoning attacks. Because the topology property of the neighborhood is useful for data refinement [23], several studies have attempted to eliminate outliers for correspondence matching. Although sampling strategy-based RANSAC methods aim to find reliable points, many of them use descriptor distance to improve the hit rate of inliers for sampling. For outlier elimination methods, they usually employ spatial clustering [24] or neighborhood coherence [25, 26] techniques to explore inliers rather than use a fitting model to check inliers. Outlier elimination-based methods usually show superiority in time efficiency and can be used for multi-structure geometry models [25]. The learning-based correspondence matching methods can be roughly summarized into two categories: detector-based methods and detector-free methods. The detector-based correspondence matching methods [12, 13] usually detect local features to produce initial matches. In contrast, the detector-free methods [14] remove the feature detector phase and directly produce dense descriptors or dense feature matches. For learning-based methods, transferring their models between two different datasets is usually unsuitable. In addition, they may be vulnerable to poisoning attacks.

In this study, we propose an outlier elimination-based assessment method (OEAM) to assess deep learning-based registration methods. The assessment contains two tasks. One is to reject low-quality image registrations that bring significant model errors. The other is to improve performance on accepted registrations with high quality. We apply OEAM on OANet, and the results on the poison samples are shown as OA-OEAM in Fig. 1b. OEAM is efficient against poisoning attacks initiated by non-overlapping images. In Fig. 1b, OA-OEAM rejects 92.35% poison samples since the resulting inlier numbers are less than 16. OEAM is advantageous for identifying false registrations.

OEAM is a post-processing technique imposed on learning-based methods. It takes correspondences generated by learning-based methods as input. OEAM contains two modules, i.e., mismatch detection and registration assessment. The procedure of mismatch detection is implemented in four steps. For a pair of images with putative correspondences, the image scale is first adjusted using the coordinates of the putative correspondences, so that the two images are registered on the same scale. Second, we construct an image pyramid to eliminate “one-to-many” mismatches. After that, OEAM uses the outer neighborhood of keypoints to generate feature descriptors and detect mismatches, in which the pixels in the outer neighborhood belong to the ring region next to the local circular neighborhood of the keypoints. Finally, OEAM detects spatially intersected mismatches. After eliminating mismatches in the detection module, a core

correspondence set is obtained. The module of the assessment is implemented in two streams. If the cardinality of the core set is sufficiently small, the input registration is inferred as a low-quality registration. Otherwise, it is considered to be of high quality. For accepted high-quality registration, OEAM uses the core set to improve the inlier precision and model accuracy. The main contributions of this article are summarized as follows.

(1) We study poisoning attacks on learning-based correspondence matching methods. Learning-based methods are usually trained on samples composed of partially overlapping images. Non-overlapping images can be used as poison samples to fool DL models and produce false registrations. Learning-based methods are vulnerable to poisoning attacks. Poisoning attacks may be a potential application deficiency for learning-based correspondence matching methods.

(2) We propose OEAM to assess learning-based correspondence matching method. OEAM produces the core set based on outlier elimination to infer low-quality registration and improve performance on high-quality registration.

(3) We implement comparison experiments on registrations of benign samples with partial overlaps and poison samples without overlap. For registrations of benign samples, OEAM can reliably infer low-quality registrations and improve performance on high-quality registrations. For registrations of poison samples, OEAM is robust against poisoning attacks crafted by non-overlapping images.

The rest of this article is organized as follows. Related works are introduced in Section 2. Section 3 provides our proposed method. The comparison experiments are presented in Section 4. Finally, conclusions are summarized in Section 5.

2. Related works

2.1. RANSAC-based methods

RANSAC method is typically implemented using a hypothesize-and-test framework [21]. First, it randomly selects the minimal set (MS) of the input matches and fits the model. Subsequently, the input matches are classified into inliers (i.e., matches that fit the model) and outliers (i.e., matches that cannot fit the model) using a tolerance error, and a consensus set (CS) composed of inliers is produced. The two aforementioned steps are repeated until the selected MS only contains true inliers.

RANSAC-based methods usually launch contributions in three streams: sampling strategy [18], model verification [19], and termination criterion [20]. In [18], a feature-matching score was used to sample the MS for model fitting. To accelerate model verification, Chum and Matas [19] added a preliminary test module to reject a massive model with limited support. Because the termination criterion of the conventional RANSAC method, which is perturbed by observation noise, may cause the iterations to terminate before a sound hypothesis is found, Imre and Hilton [20] developed a top- n criterion based on the order statistics of RANSAC to ensure accurate results. In addition, some studies have attempted to integrate several technologies to improve RANSAC, such as VSAC [22]. RANSAC-based methods usually take advantage of geometric constraints. However, they may be inefficient in processing excessive correspondences with a low inlier ratio [27].

2.2. Outlier elimination methods

Outlier elimination is demonstrated to be an efficient technology for data reduction and correspondence matching in many studies. Cai et al. [23] combined nearest neighbor distance and influence space to describe the possibility of outliers. Ren et al. [24] designed a spatial clustering method to eliminate outliers for aerial image registration. Since true correspondences typically have more similar neighbors than false correspondences, Bian et al. [28] proposed grid-based motion statistics (GMS) to count the number of similar neighbors for correspondence matching. Ma et al. [25] proposed locality preserving matching (LPM) to investigate potential true matches using the local neighborhood structures. In [26], affine transformation-based local consistency is employed for outlier filtering. Based on the neighborhood distribution of the feature points, global deformation is estimated to eliminate mismatches in [29]. For two-view geometry analysis, neighborhood coherence is often used to detect outliers. The outlier elimination-based method is usually time-efficient. In this study, we attempt to explore the spatial paradox of the matches to assess learning-based correspondence matching methods.

2.3. Learning-based correspondence matching

In recent years, learning-based approaches have been proposed for correspondence matching. The learning-based correspondence matching methods can roughly be summarized into two categories: detector-based methods and detector-free methods. The detector-based correspondence matching methods usually detect local features [30, 31, 32]

to produce initial matches. Taking the scale-invariant feature transform (SIFT) correspondence [30] composed of two two-dimensional (2D) points as input, LFGC [11] designed a 12-layer residual network (ResNet) using weight-sharing perceptron and context normalization for correspondence classification. Zhang et al. [12] proposed OANet to infer the probabilities of SIFT correspondences as inliers and regress the relative pose encoded by an essential matrix. In [33], the authors designed an attention mechanism-based network with SIFT initial matches for two-view correspondence matching. In [34], a classical learning method for correspondence matching is designed using weighted support vector regression with a quadratic insensitive loss. The method uses rotation-invariant shape context descriptor [31] to produce initial correspondences. Moreover, some studies have designed graph neural networks for correspondence matching, such as SuperGlue [13] and sparse graph attention network [35]. SuperGlue uses SuperPoint [32] to produce initial correspondences. Combining ResNet block, edge-sparsified graph attention block, and Transformer, Liao et al. [35] proposed a sparse graph attention network for feature matching tasks. Yang et al. [36] proposed dynamic attention-based detector and descriptor using derivable loss for image registration. Zhao et al. [37] developed a light-weight network termed the seed matching and filtering network to obtain sufficient high-quality correspondences.

The detector-free methods remove the feature detector phase and directly produce dense descriptors or dense feature matches. Rocco et al. [38] designed a neighborhood consensus network to directly learn the dense correspondences in an end-to-end manner. In [14], the self and cross attention layers involved in Transformer are designed to obtain feature descriptors for correspondence matching. Although learning-based methods sound efficient on experimental datasets, the experiments in [11, 12] show that DL models may produce mismatches inevitably. In particular, DL models are trained on partially overlapping image pairs. Since the models are usually data-dependent, non-overlapping images may be used as poison samples to fool DL models and produce false registrations.

3. Proposed method

Since learning-based correspondence matching methods may produce false registrations, we propose OEAM to assess the learning-based methods. Our proposed OEAM contains two modules: (I) mismatch elimination, and (II) registration assessment. The pipeline is shown in Fig. 2. The procedure of module (I) is implemented in four steps. Since the input images may be registered at different scales (Fig. 2a), OEAM first adjusts one of the image sizes so that the registration images stand at the same scale, as shown in Fig. 2b. Second, OEAM eliminates “one-to-many” mismatches. In detail, OEAM downsamples the images to build pyramids. If two correspondences are “one-to-many” matches, i.e., they have the same pixel coordinates on a layer of one pyramid but have different pixel coordinates on the corresponding layer of the other pyramid, then there is at least one mismatch. Fig. 2c shows “one-to-many” mismatches on different layers. Fig. 2d shows matches after “one-to-many” mismatch elimination. OEAM then detects globally dissimilar correspondences, as shown in Fig. 2e. For two keypoints of a true match, their local and outer neighborhoods are both expected to be similar, in which the pixels in the outer neighborhood belong to the ring region next to the local neighborhood of the keypoint. If the outer neighborhoods of the matched keypoints are dissimilar, the correspondence is detected as a mismatch. Finally, OEAM detects spatially intersected mismatches. In detail, we force the keypoints in an image to rotate around the image center until the putative correspondences have a minimum number of intersections. If one correspondence intersects with multiple others, it is inferred to be a mismatch (Fig. 2f). After eliminating mismatches in the detection module, a correspondence core set is obtained, as shown in Fig. 2g. Module (II) is implemented in two streams. If the cardinality of the resulting core set is sufficiently small, the input registration is inferred to be of low quality (Fig. 2j(1)). Otherwise, it is considered to be of high quality. For accepted registration, we first use the resulting core correspondences to fit a geometry model, as shown in Fig. 2h. Subsequently, the initial correspondences are classified into inliers and outliers using the model and a tolerance error, and a CS consisting of inliers is produced (Fig. 2i). Then, the output correspondences are the intersection of the CS and putative correspondence set (Fig. 2j(2)), and the output registration model is fitted using the output correspondences (Fig. 2j(3)). Fig. 2j shows the output of OEAM.

3.1. Mismatch detection

3.1.1. “One-to-many” mismatch detection

Let I_0 and I_1 be two images for registration. $A = \{a_1, a_2, \dots, a_N\}$ is the set of initial matches between I_0 and I_1 , where $a_i = (x_i, x'_i)$ is the i -th match, $x_i = (u_i^0, v_i^0) \in I_0$ is the keypoint in I_0 , and $x'_i = (u_i^1, v_i^1) \in I_1$ is the keypoint in I_1 . Let the putative correspondence produced by learning-based method be $M = \{c_1, c_2, \dots, c_K\} \subseteq A$, where $c_i = (z_i, q_i)$ is the i -th match of M , $z_i = (x_i^0, y_i^0) \in I_0$ is the keypoint in I_0 , $q_i = (x_i^1, y_i^1) \in I_1$ is the keypoint in

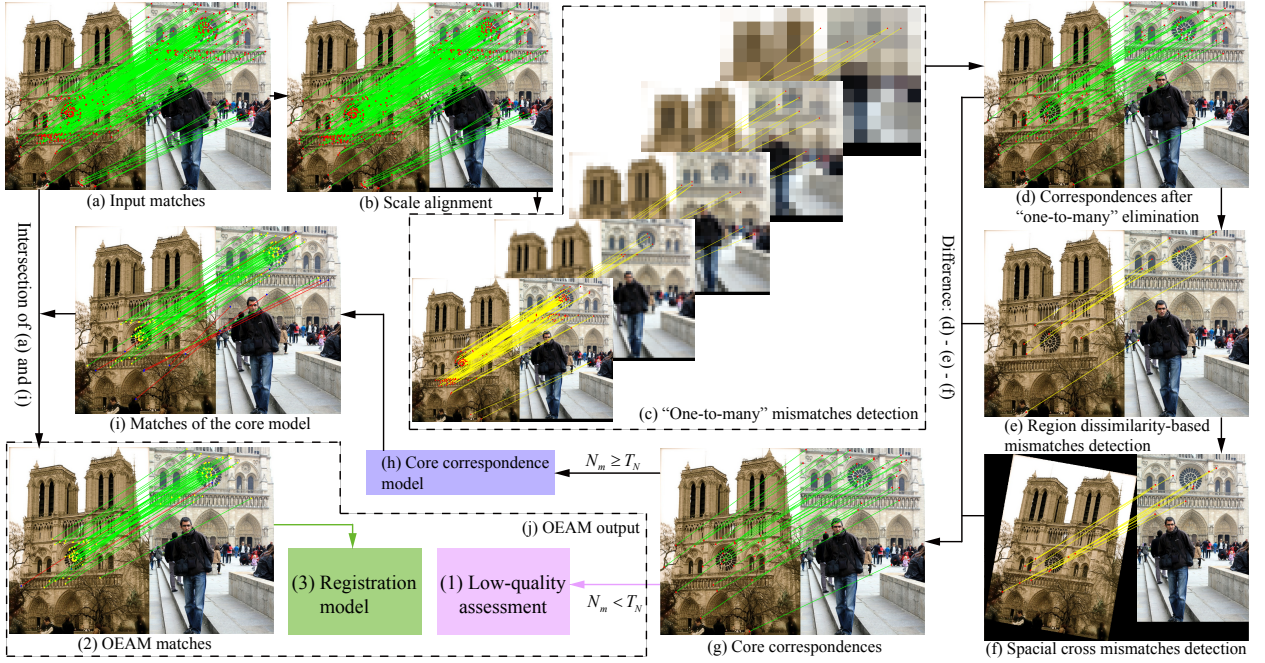


Fig. 2: The flowchart of our proposed OEAM.

I_1 . Before mismatch detection, the scale of I_0 is adjusted using the K correspondences in M so that I_0 and I_1 are at the same scale. The scale s is estimated in Eq. (1).

$$s = \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^K \sqrt{(x_i^1 - x_j^1)^2 + (y_i^1 - y_j^1)^2}}{\sum_{i=1}^{K-1} \sum_{j=i+1}^K \sqrt{(x_i^0 - x_j^0)^2 + (y_i^0 - y_j^0)^2}}. \quad (1)$$

Then, correspondences after scale adjustment are shown as $\tilde{M} = \{m_1, m_2, \dots, m_K\}$, where $m_i = (p_i, q_i)$, $p_i = sz_i = (sx_i^0, sy_i^0)$.

Let the matches pyramid of \tilde{M} be $\Phi = \bigcup_{k=0}^{d_M-1} M_k$, where $M_k = \{m_1^k, m_2^k, \dots, m_K^k\}$, $m_i^k = (p_i/2^k, q_i/2^k)$. Fig. 3 shows the two pyramids on \tilde{M} . The lines in Fig. 3 show mismatches between the pyramids. They have the same pixel coordinates on a certain layer of one pyramids, but have different pixel coordinates on the corresponding layer of the other pyramid. Let $S_1 = \{m_i \in \tilde{M} \mid \| [p_i/2^k] - [p_j/2^k] \|_\infty > 1, [q_i/2^k] = [q_j/2^k]\}$, $S_2 = \{m_i \in \tilde{M} \mid [p_i/2^k] = [p_j/2^k], \| [q_i/2^k] - [q_j/2^k] \|_\infty > 1\}$, $i, j = 1, 2, \dots, K$, $k = 0, 1, \dots, d_M - 1$, $[x]$ is the least-integer function of x . Then, m_i is detected as a ‘‘one-to-many’’ mismatch if there exist j and k so that $m_i \in S$, as shown in Eq. (2).

$$S = S_1 \cup S_2, \quad (2)$$

3.1.2. Dissimilar mismatch detection

Let the resulting correspondences after ‘‘one-to-many’’ mismatch elimination be $M_d = \tilde{M} - S$, as shown in Fig. 2d. Generally, the local and outer neighborhoods of the two keypoints of a correspondence are both expected to be similar. However, M_d may contain mismatches involving dissimilar features in the outer neighborhood.

Let $m = (P_0, P_1)$ be a correspondence in M_d . Fig. 4 shows similarity-based mismatch detection. In Fig. 4, the outer neighborhoods of P_0 and P_1 are denoted as $O(P_0)$ and $O(P_1)$, respectively. The white regions in Fig. 4 show local neighborhoods. The colored areas show outer neighborhoods, which are divided into three ring regions dyed in

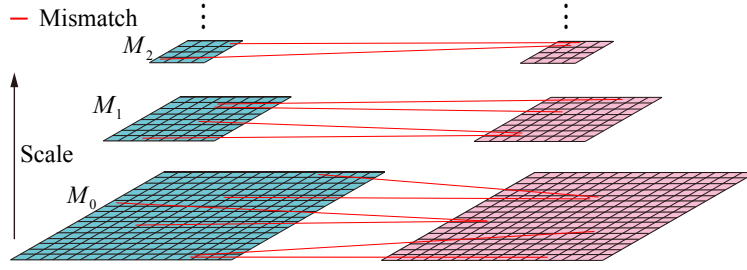


Fig. 3: “One-to-many” mismatch detection based on multi-scale.

different colors. If m is a true match, then there exists at least one ring region of P_0 that is similar to one of P_1 , i.e., $O(P_0)$ and $O(P_1)$ are similar to a certain extent.

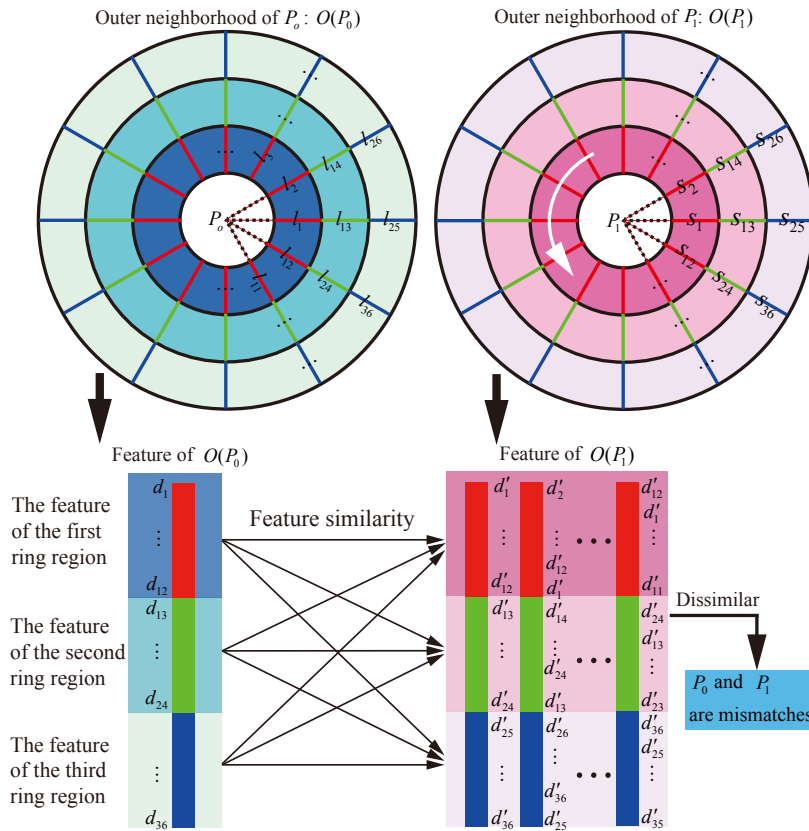


Fig. 4: Schematic diagram of mismatch detection based on the similarity of the outer neighborhoods.

To assess the similarity of $O(P_0)$ and $O(P_1)$, we build features in the outer neighborhoods. First, the outer neighborhoods are discretely sampled in n directions, and we obtain pixel segments, as shown l_j and s_j ($j = 1, 2, \dots, 3n$) in Fig. 4. Let d_j be the average deviation produced by the pixels on l_j and P_0 , as shown in Eq. (3).

$$d_j = \frac{1}{L_s} \sum_{k=1}^{L_s} (g_{jk} - g_0)^2, \quad (3)$$

where g_0 is the gray level of P_0 , g_{jk} is the gray level of the k -th pixel on l_j , L_s is the number of pixels on l_j . Then, features of the three ring regions of P_0 are built as $D_i = (d_{(i-1)n+1}, d_{(i-1)n+2}, \dots, d_{in})$, $i = 1, 2, 3$. Similarly, we have d'_j

on s_j , as shown in Eq. (4).

$$d'_j = \frac{1}{L_s} \sum_{k=1}^{L_s} (g'_{jk} - g_1)^2, \quad (4)$$

where g_1 is the gray level of P_1 , g'_{jk} is the gray level of the k -th pixel on s_j . Take the rotation into account, the features of $O(P_1)$ are built in n directions, i.e., $S_i^k = (d'_{(i-1)n+k}, \dots, d'_{(i-1)n+n}, d'_{(i-1)n+1}, \dots, d'_{(i-1)n+k-1})$, $i = 1, 2, 3$, $k = 1, 2, \dots, n$.

The Euclidean distance of the outer neighborhoods $O(P_0)$ and $O(P_1)$ is evaluated in Eq. (5), and the Chebyshev distance is evaluated in Eq. (6).

$$d_c = \min_{i,j,k} \|D_i - S_j^k\|_2. \quad (5)$$

$$d_\infty = \min_{i,j,k} \|D_i - S_j^k\|_\infty. \quad (6)$$

To offer immunity to light change, the match m is detected as a mismatch if both the d_c and d_∞ are sufficiently large. As shown in Eq. (7), the match $m \in M_e$ is discriminated as a mismatch.

$$M_e = \{m | m \in M_d, d_c > T_c, d_\infty/d_c > T_s\}, \quad (7)$$

where T_c and T_s are two thresholds.

3.1.3. Intersected mismatch detection

Without loss of generality, let I_0 and I_1 be the left and right images for registration, respectively. If the translation is the only pose change between I_0 and I_1 , true matches between the two images are approximately parallel and do not intersect each other. In this case, if a match intersects with multiple other matches, it may be a mismatch. However, rotation is practically inevitable for intersected mismatch detection. To this end, we first force the keypoints in I_0 to rotate around the center of the image until the correspondences after rotation have the minimum number of intersections. Second, we count the intersection number for each correspondence. For a given threshold T_3 , if a correspondence has a number of intersections that is greater than T_3 , it is inferred to be a mismatch.

Let the matches after dissimilar detection be $M_{de} = M_d - M_e = \{m_1, m_2, \dots, m_L\}$. Let the width and height of the left image I_0 be W and H , respectively. Then, the center coordinate of I_0 is $O = (x_o, y_o) = ([W/2], [H/2])$. If the coordinate of $p_i = (sx_i^0, sy_i^0)$ is rotated counterclockwise around the scaled coordinate of O by $k\Delta_\theta$, its rotated coordinate $\tilde{p}_i = (\tilde{x}_i, \tilde{y}_i)$ is shown in Eq. (8).

$$\begin{cases} \tilde{x}_i = sx_o + r \cos(\theta_0 + k\Delta_\theta) \\ \tilde{y}_i = sy_o + r \sin(\theta_0 + k\Delta_\theta) \end{cases}, \quad (8)$$

where $r = \sqrt{\Delta_x^2 + \Delta_y^2}$, $\Delta_x = s(x_i^0 - x_o)$, $\Delta_y = s(y_i^0 - y_o)$; $k = 0, 1, \dots, \pi/\Delta_\theta$; Δ_θ is the step width of the rotation; θ_0 is determined in Eq. (9).

$$\theta_0 = \begin{cases} \arctan(\Delta_y/\Delta_x), & \Delta_x \geq 0 \\ \pi + \arctan(\Delta_y/\Delta_x), & \Delta_x < 0 \end{cases}. \quad (9)$$

Let the matches of M_{de} after rotation be $\tilde{M}_{de} = \{\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_L\}$, where $\tilde{m}_i = (\tilde{p}_i, q_i)$. In order to obtain the rotation angle so that the putative correspondences have the minimum number of intersections, it is necessary to judge the intersection between the matches \tilde{m}_i and \tilde{m}_j . First, the line equation l_i of \tilde{m}_i is shown in Eq. (10).

$$y = k_i x + b_i, \quad (10)$$

where $k_i = (y_i^1 - \tilde{y}_i)/(x_i^1 + sW - \tilde{x}_i)$, $b_i = \tilde{y}_i - k_i \tilde{x}_i$. Let $\tilde{m}_j = (\tilde{p}_j, q_j)$, $\tilde{p}_j = (\tilde{x}_j, \tilde{y}_j)$, $q_j = (x_j^1, y_j^1)$. Similarly, the line equation l_j of \tilde{m}_j is shown in Eq. (11).

$$y = k_j x + b_j, \quad (11)$$

where $k_j = (y_j^1 - \tilde{y}_j)/(x_j^1 + sW - \tilde{x}_j)$, $b_j = \tilde{y}_j - k_j \tilde{x}_j$. If l_i and l_j are intersected, the two ends of l_i are respectively located at the two sides of l_j , meanwhile, the two ends of l_j are respectively located at the two sides of l_i , i.e.,

$$\begin{cases} (y_j^1 - k_i(x_j^1 + sW) - b_i)(\tilde{y}_j - k_i \tilde{x}_j - b_i) < 0 \\ (y_i^1 - k_j(x_i^1 + sW) - b_j)(\tilde{y}_i - k_j \tilde{x}_i - b_j) < 0 \end{cases} \quad (12)$$

After the scaled image I_0 is rotated to an optimal degree, the intersected mismatch set M_f can be detected using Eqs. (8)-(12). The intersected mismatch detection method is shown in Algorithm 1. Algorithm 1 first obtains an optimal rotation parameter k_0 with the minimum number of intersections, and then it explores intersected mismatches using the resulting k_0 .

Algorithm 1 Intersected mismatch detection method.

Input: The width W and height H of image I_0 , matches set M_{de} , and parameters Δ_θ and T_3

Output: Intersected mismatch set M_f

```

1: Initialization: Rotation parameter  $k_0 = 0$ , intersection number  $n_c = 10^6$ ,  $\tilde{M}_{de} = \emptyset$ ,  $M_f = \emptyset$ 
2: Obtain the center of  $I_0$ :  $(x_o, y_o) = ([W/2], [H/2])$ 
3: for  $k = 0, 1, \dots, \pi/\Delta_\theta$  do // Rotate the keypoints in scaled  $I_0$  around the center of it
4:    $n_d = 0$ 
5:   for  $i = 1 : L$  do // For  $m_i = (p_i, q_i) \in M_{de}$ 
6:     Obtain the rotation coordinate  $\tilde{p}_i$  using Eq. (8)
7:     Obtain the line equation  $l_i$  using Eq. (10)
8:     for  $j = 1 : L, j \neq i$  do
9:       Obtain  $\tilde{p}_j$  using Eq. (8)
10:      Obtain  $l_j$  using Eq. (11)
11:      if  $l_i$  and  $l_j$  satisfy Eq. (12) then
12:         $n_d = n_d + 1$  // Count intersections on  $M_{de}$ 
13:      end if
14:    end for
15:  end for
16:  if  $n_d < n_c$  then // Update the optimal rotation angle  $k_0 \Delta_\theta$ 
17:    Update  $n_c$  and  $k_0$ :  $n_c = n_d, k_0 = k$ 
18:  end if
19: end for
20: for  $i = 1 : L$  do // Run over the matches set  $M_{de}$  to obtain the rotated keypoint coordinates in scaled  $I_0$ 
21:   Obtain  $\tilde{m}_i = (\tilde{p}_i, q_i)$  using  $k = k_0$  and Eq. (8)
22:    $\tilde{M}_{de} = \tilde{M}_{de} \cup \{\tilde{m}_i\}$ 
23: end for
24: for  $i = 1 : L$  do
25:   for  $j = 1 : L, j \neq i$  do
26:     if  $\tilde{m}_i$  and  $\tilde{m}_j$  satisfy Eq. (12) for  $k = k_0$  then //  $l_i$  and  $l_j$  are intersected
27:        $N_i = N_i + 1$  // Count the intersection number of the  $i$ -th match
28:     end if
29:   end for
30:   if  $N_i > T_3$  then
31:      $M_f = M_f \cup \{m_i\}$  //  $m_i$  is a mismatch
32:   end if
33: end for

```

3.2. Registration assessment

After applying Algorithm 1 on M_{de} , the scaled core set $\tilde{M}_c = M_{de} - M_f = \{m_{k_1}, m_{k_2}, \dots, m_{k_{N_m}}\}$ is obtained. Correspondingly, let $M_c = \{c_{k_1}, c_{k_2}, \dots, c_{k_{N_m}}\}$ be the resulting core set, as shown in Fig. 2g. The registration

assessment of OEAM involves two tasks. One is to reject low-quality registrations. The other is to improve performance on accepted registrations.

Usually, a registration in low quality implies a high proportion of mismatches. If the size of the core set obtained by OEAM is relatively small, the quality of its input registration is hardly high. In this study, if the number of matches in the core set is less than a given threshold T_N , i.e., $N_m < T_N$, the input registration is inferred to be of low quality. Otherwise, OEAM accepts the registration to be of high quality and improves performance on the accepted registration. In detail, the matches in the core set are first used to fit a registration model F . Second, if $a_i \in A$ satisfies Eq. (13) with a given Sampson distance ϵ_S , which provides a first-order approximation of the reprojection error [39], a_i is inferred to be an inlier related to F .

$$\frac{\tilde{x}_i'^T F \tilde{x}_i}{(F \tilde{x}_i)_1^2 + (F \tilde{x}_i)_2^2 + (F^T \tilde{x}_i')_1^2 + (F^T \tilde{x}_i')_2^2} \leq \epsilon_S, \quad (13)$$

where \tilde{x}_i and \tilde{x}_i' are homogeneous coordinates of x_i and x_i' with respect to $a_i = (x_i, x_i') \in A$, $i = 1, 2, \dots, N$, respectively; $(F \tilde{x}_i)_1$ and $(F \tilde{x}_i)_2$ are the first and second elements of $F \tilde{x}_i$, respectively; $(F^T \tilde{x}_i')_1$ and $(F^T \tilde{x}_i')_2$ are similar notations. Let B be the inlier set inferred by Eq. (13). Then, the output matches are the intersection of B and input correspondences M , i.e., $M_o = B \cap M$. Finally, the output registration model F_o is the fitting model of M_o . The registration assessment is shown in Algorithm 2, which mainly contains two modules, i.e., assessments on low-quality registrations and performance improvements on high-quality registrations.

Algorithm 2 Registration assessment method.

Input: Core set M_c , initial correspondence set A , putative correspondence set M , and parameters T_N and ϵ_S

Output: Correspondence set M_o and registration model F_o

- 1: **if** $\text{Card}(M_c) < T_N$ **then**
 - 2: The input registration is in low quality
 - 3: **end if**
 - 4: **if** $\text{Card}(M_c) \geq T_N$ **then**
 - 5: Fitting a model F using M_c
 - 6: Obtain the inlier set B using Eq. (13) fed by A and ϵ_S
 - 7: $M_o = B \cap M$
 - 8: Fitting the output model F_o using M_o
 - 9: **end if**
-

4. Experiments

In this section, we implement OEAM on four deep learning-based methods, i.e., LFGC [11], OANet [12], SuperGlue [13], and LoFTR [14], to implement assessment experiments. To show the advantage of our proposed method, RANSAC [40], VSAC [22], GMS [28], and LPM [25] are employed as benchmarks for comparison. The experiments were run on a computer with an Intel Core i-3 4160 CPU and a GeForce GTX 1080 GPU. The deep learning-based experiments were implemented in TensorFlow-GPU-1.12.0 and Pytorch1.1. Our method was compiled using Microsoft Visual Studio 2015 Ultimate Edition (VS2015) equipped with the CUDA toolkit 8.0.

4.1. Experiment settings

4.1.1. Datasets and metrics

We conduct experiments on outdoor YFCC100M [17] and indoor SUN3D [41] datasets. Yahoo's YFCC100M dataset contains 100 million images from the internet. The authors generated 72 three-dimensional reconstructions of tourist landmarks from a subset of the dataset. Following [12], four sequences, including buckingham, notre, reichstag, and sacre coeur, are used as unknown scenes to test generalization ability. Every sequence contains 1000 image pairs for testing. The SUN3D dataset is used to test image pairs for indoor scenes, which is an RGBD video dataset with camera poses computed by generalized bundle adjustment. Following [12], 15 sequences are used as unknown scenes for testing. The test image pairs on SUN3D have a total of 14872.

We use the provided rotation and translation to estimate the ground truth pose. After the coordinates of the putative correspondences are normalized, the Sampson distance is set to 10^{-3} to select ground truth inliers based on the implementation presented in OANet [12]. We use the recall, precision, and area under the curve (AUC) as the evaluation metrics. Recall is the ratio of the number of true inliers identified to the total number of true inliers in the putative set. Precision is the ratio of the number of true inliers identified to the number of detected inliers. Following [11], the AUC is approximated by mAP. In detail, we obtain the predicted rotation and translation using the estimated pose model. The pose error is the maximum of the angular errors in rotation and translation, and the angular error is induced by the similarity between the estimated and ground truth vectors. We report the mAP using thresholds 5° , 10° , and 20° .

4.1.2. Experiment details

The comparison experiments are implemented on registrations of partially overlapping images in Section 4.2 and non-overlapping images in Section 4.3. The registration experiments of overlapping images are implemented on outdoor and indoor datasets. We use the image pairs proposed in [12] to implement the registrations of overlapping images. The registration experiments of non-overlapping images are implemented on the outdoor dataset YFCC100M. As aforementioned, X_0, X_1, X_2 and X_3 are the subsets of buckingham, notre, reichstag, and sacre coeur, respectively. Let $p_i^k = (I_{i1}^k, I_{i2}^k)$ be the i -th overlapping image pair for registration in X_k , $k = 0, 1, 2, 3$, $i = 1, 2, \dots, 1000$. $P_k = \{p_1^k, p_2^k, \dots, p_{1000}^k\}$ is the list of 1000 overlapping image pairs of X_k . The i -th pair of non-overlapping images is designed between X_k and X_l , and it is shown by $q_i^{kl} = (I_{i1}^{kl}, I_{i2}^{kl})$, where $k = 0, 1, 2, 3$, $l \in \{0, 1, 2, 3\}$ and $l \equiv (k + 1) \pmod{4}$, i.e., l and $k + 1$ have the same remainder modulo 4. The non-overlapping image pairs between X_k and X_l are shown by $Q_{kl} = \{q_1^{kl}, q_2^{kl}, \dots, q_{1000}^{kl}\}$. As a result, we constructed 4000 non-overlapping image pairs on YFCC100M for experiments.

For the baselines, LFGC, OANet, and SuperGlue are detector-based methods. LFGC is structured by 12 ResNet layers, where each layer is a PointCN ResNet block that contains two sequential blocks consisting of a perceptron, a context normalization layer, a batch normalization layer, and a ReLU. OANet is mainly composed of 18 sequential blocks, i.e., 6 PointCN ResNet blocks, 6 order-aware filtering blocks, and 6 PointCN ResNet blocks. Each order-aware block is composed of a spatial correlation layer inserted in the middle of the PointCN ResNet block. SuperGlue is made up of two major components: an attentional graph neural network and an optimal matching layer. As for LoFTR, it is a detector-free method. LoFTR mainly consists of four modules: local feature convolutional architecture, coarse-level local feature Transform, matching module, and coarse-to-fine module. For the experiments on LFGC and OANet, the SIFT method is employed to produce 2,000 initial correspondences. The feature-matching threshold for SIFT, defined as the ratio of the distances between the best and second-best matches, was set to 1.0. For SuperGlue, SuperPoint [32] is employed to produce initial correspondences. The confidence threshold for SuperPoint is set to 0.2. As a detector-free method, LoFTR does not explore inliers from initial correspondences. Since the initial ground truth inliers are not provided for LoFTR, the recall of LoFTR is unavailable. In this study, we use the produced inliers and ground truth model to produce precision and mAP for evaluation. The results of LoFTR are reported with a maximum limitation of 1,000 output matches [14].

In this study, RANSAC, VSAC, GMS, and LPM were employed as benchmarks for comparison, where RANSAC and VSAC are RANSAC-based correspondence matching methods; GMS and LPM are outlier elimination-based methods. For the epipolar geometry model involved in RANSAC and VSAC, the essential matrix was employed to investigate inliers. In detail, a five-point algorithm was used for model fitting. Together with camera intrinsic parameters, the Sampson distance 2 was delivered to RANSAC and VSAC for correspondence coordinate normalization and inlier selection. The maximum number of iterations for RANSAC was set to 10^4 . The experiments on VSAC, GMS, and LPM were carried out using default parameters except for the termination length used in VSAC, which was set to 1 in this study. All the models used for pose estimation were reestimated using the resulting inliers. For RANSAC, GMS, and OEAM, the resulting inliers were used to fit a fundamental matrix model. Their precisions were calculated using the inliers filtered by the fitting model with the Sampson distance 2.0 on the original image for comparison.

4.1.3. Parameter settings

Several parameters require to be set for OEAM. For convenience, we apply OEAM on OANet to tune parameters. The ablation experiments were evaluated by mAP using the threshold of 20° . Since the image size on the datasets is usually no more than $1,024 \times 1,024$, a pyramid with 8 layers can induce a sufficiently small layer in size of no more than 4×4 for ‘‘one-to-many’’ mismatch detection. Therefore, the number of pyramid layers d_M in Section 3.1.1 was set to 8. In the module of dissimilar mismatch detection, the outer neighborhood is divided into three rings. Based on the

size and orientation of SIFT descriptor, the ring width L_s was set to 16 pixels, and each ring was sampled equidistantly in $n = 12$ radial directions. T_s and T_c are the two parameters to be assigned. The ablation studies of the thresholds T_s and T_c were conducted on the notre part of the YFCC100M dataset. For T_s , three levels of 0.25, 0.30, and 0.35 were selected for the experiments. For T_c , five levels 0.05, 0.10, 0.15, 0.20, and 0.25 were employed. Fig. 5 shows the resulting mAP on the two thresholds.

0.25	40.35	41.55	42.59	43.00	43.25
T_s 0.30	42.63	42.95	43.32	43.96	44.11
0.35	43.18	43.45	43.63	43.39	43.69
	0.05	0.10	0.15	0.20	0.25
			T_c		

Fig. 5: Ablation studies of T_s and T_c on the notre dataset using the metrics mAP (%).

Generally, if the similarity parameters T_c and T_s are small, true matches may be incorrectly rejected. On the contrary, if T_c and T_s are large, the rejection ability of mismatch may not meet our expectations. As shown in Fig. 5, OA-OEAM achieves the greatest mAP of 44.11% on $T_c = 0.25$ and $T_s = 0.30$. Meanwhile, it rejects 21.35% mismatches. Although OA-OEAM obtains a smaller mAP of 43.96% on $T_c = 0.20$ and $T_s = 0.30$, a greater mismatch rejection rate of 25.44% is achieved on $T_c = 0.20$ and $T_s = 0.30$. Taking the mismatch rejection into account, $T_c = 0.20$ and $T_s = 0.30$ are employed for OEAM.

In Section 3.1.3, the threshold Δ_θ is the step width of rotation. A small Δ_θ corresponds to a fine tune of the rotation. In this study, Δ_θ was set to a small value of $\pi/10$. The threshold T_3 is used to check the intersected mismatches. Since a small T_3 is helpful for mismatch rejection, T_3 was set to 1. In the registration assessment module, the input registration is inferred to be of low quality if the number of matches in the core set is less than a given threshold T_N . Let the rejection accuracy be the ratio of the number of false registrations rejected by the model with errors greater than 20° to the number of low-quality registrations rejected by OEAM. Table 1 shows the resulting average precision, rejection accuracy, and their average value of OA-OEAM on T_N using the notre dataset. As shown in Table 1, the largest average of 84.53% is achieved at $T_N = 16$. Therefore, $T_N = 16$ is chosen for OEAM.

Table 1

Average precision, rejection accuracy, and their average value of OA-OEAM on the notre dataset using the parameter T_N (%).

T_N	8	10	12	14	16	18	20
Average precision	81.36	82.36	84.08	86.02	87.00	88.48	89.90
Rejection accuracy	85.62	85.71	84.47	81.98	81.05	80.03	76.34
Average	83.49	84.03	84.28	84.00	84.53	84.25	83.12

4.2. Experiments on overlapping images

In this section, we assess learning-based correspondence matching methods on partially overlapping images. On the one hand, OEAM rejects low-quality registrations. On the other hand, OEAM improves precision and mAP on high-quality registrations. The overlapping image experiments were implemented on indoor and outdoor datasets. To show the performance of our proposed OEAM, four deep learning-based methods, including LFGC [11], OANet [12], SuperGlue [13], and LoFTR [14] are employed as baselines for comparison. After applying OEAM to the outputs of the four learning-based methods, LF-OEAM, OA-OEAM, SG-OEAM, and Lo-OEAM are our proposed methods for performance demonstration. Since RANSAC, VSAC, GMS, and LPM can be used to refine the results of correspondence matching, they are employed as benchmarks for comparison. The benchmarks are also implemented on the output matches of the learning-based methods, as shown by LF-RANSAC, OA-VSAC, SG-GMS, etc in this section.

4.2.1. Experiments on outdoor dataset

In this section, we implement comparison experiments on outdoor scenes using benchmarks. After using $T_N = 16$ to discriminate low-quality registration, LF-OEAM, OA-OEAM, SG-OEAM, and Lo-OEAM accept 3206, 3570, 3952,

and 3919 samples as high-quality registrations on YFCC100M, respectively. Fig. 6 shows the resulting recall (R) and precision (P) of the inliers on the high-quality registrations from top to bottom. Figs. 6a - 6d show the results on LFGC, OANet, SuperGlue, and LoFTR, respectively. Since the initial ground truth inliers are unavailable for LoFTR, its recall is not provided in Fig. 6. We check the produced LoFTR inliers using the ground truth geometry model to produce its precision. The horizontal coordinates in Fig. 6 show the cumulative distribution of the accepted registrations. A point (x, y) in the graphs indicates that $x \times 100$ percent of the image pairs have a precision or recall of no more than y [25].

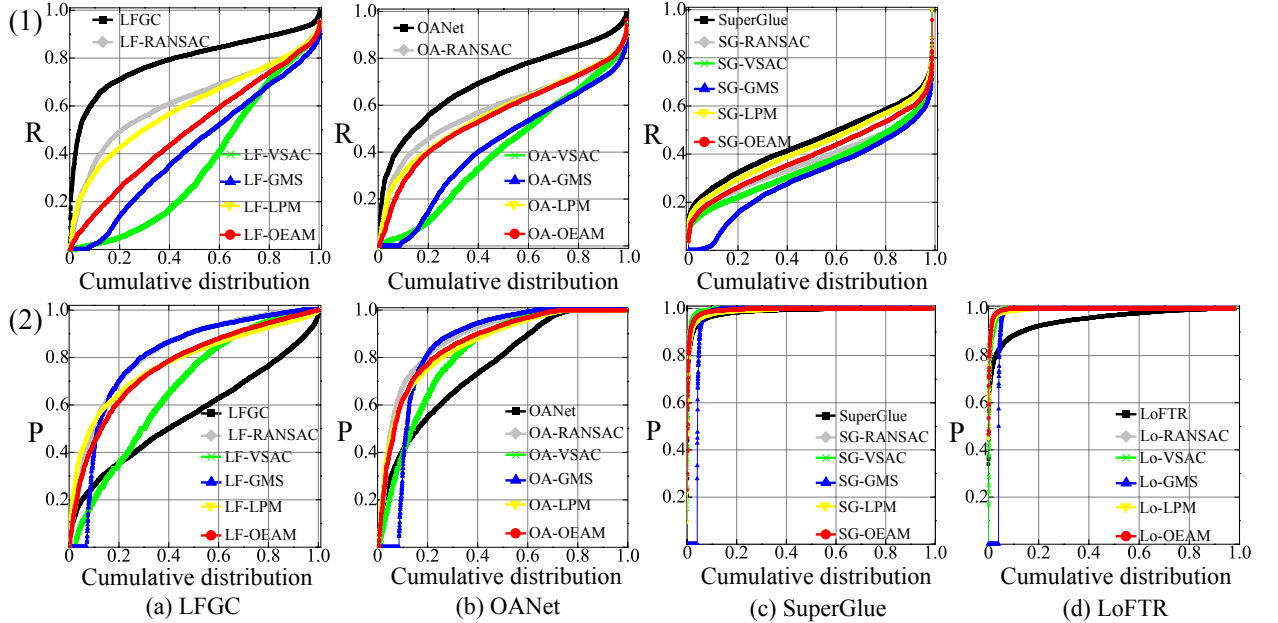


Fig. 6: The recall and precision of the high-quality registrations on the outdoor (YFCC100M) scenes.

As shown in Fig. 6, the area under the black line in Fig. 6(2) is smaller than those of the others. OEAM substantially improves the precisions of high-quality registrations. Although RANSAC, VSAC, GMS, LPM, and OEAM perform poorly in terms of recall in Fig. 6(1), they can improve precision. For pose estimation, five true inliers are required to estimate a ground truth geometry model. Compared to recall, precision may be a stringent metric for performance assessment. Generally, a higher level of precision corresponds to a more accurate model. In Fig. 6a(2), the average precisions of LFGC, LF-RANSAC, LF-VSAC, LF-GMS, LF-LPM, and LF-OEAM are 55.83%, 82.10%, 66.75%, 79.70%, 77.61%, and 76.82%, respectively. The average precisions of OANet, OA-RANSAC, OA-VSAC, OA-GMS, OA-LPM, and OA-OEAM are 75.76%, 88.19%, 81.27%, 84.06%, 85.44%, and 86.18% (Fig. 6b(2)), respectively. The average precisions of SuperGlue, SG-RANSAC, SG-VSAC, SG-GMS, SG-LPM, and SG-OEAM are 98.52%, 99.49%, 99.45%, 95.39%, 98.92%, and 99.28% (Fig. 6c(2)), respectively. As shown in Fig. 6d(2), the average precisions of LoFTR, Lo-RANSAC, Lo-VSAC, Lo-GMS, Lo-LPM, and Lo-OEAM are 95.16%, 99.71%, 99.39%, 95.66%, 99.50%, and 99.66%, respectively. The average precision improvements of RANSAC, VSAC, GMS, LPM, and OEAM on the four learning-based methods are 11.05%, 5.40%, 7.39%, 9.05%, and 9.17%, respectively. Although RANSAC achieves the best precision performance on the outdoor scenes, it usually produces a considerable number of false matches on non-overlapping image registration, as shown in Section 4.3. In particular, RANSAC usually reaches the maximum iteration count on non-overlapping image registration and therefore is time-consuming. LF-GMS outperforms LF-OEAM in Fig. 6a(2). However, the average precision improvement of OEAM on the four benchmarks is greater than that of GMS. Furthermore, the GMS ratio of $P \leq 5\%$ is the highest, which induces the largest precision deviation. A large precision deviation may be detrimental to model estimation. As shown in Table 2, GMS produced the poorest pose estimation. Since the precision of SuperGlue is greater than 98%, the performance improvement of OEAM is limited. In any case, OEAM is advantageous for precision improvement on high-quality registrations of outdoor scenes.

To evaluate model accuracy, we conducted pose estimation of the high-quality samples on the outdoor dataset and listed the resulting mAP (%) using the pose thresholds 5° , 10° , and 20° in Table 2. RANSAC achieves the best

mAP performance in Table 2. OEAM generally outperforms VSAC, GMS, and LPM on LFGC, OANet, and LoFTR in terms of mAP. The pose estimation seems to be inversely proportional to precision deviation. As shown in Fig. 6, the precision deviation of GMS is the largest. Correspondingly, the pose estimation of GMS is the poorest in Table 2. The average mAPs of learning-based methods, RANSAC, VSAC, GMS, LPM, and OEAM in Table 2 are 40.02%, 51.82%, 31.75%, 28.40%, 40.25%, and 41.42%, respectively. OEAM exhibits overall advantages in model estimation over VSAC, GMS, and LPM. Although RANSAC achieves the best precision and mAP on high-quality registrations, it exhibits a poor capability to reject low-quality registrations, as shown in Section 4.3. With the exception of model estimation on OANet, OEAM improves the mAP performance of LFGC, SuperGlue, and LoFTR. The average mAP improvement of OEAM on the four baselines is 1.40%. Overall, OEAM effectively improves mAP of pose estimation. Although RANSAC is inferior to learning-based methods in recall (Fig. 6(1)), it improves the mAP of the learning-based methods in Table 2. It seems that a reliable pose estimation is feasible even if a low recall is produced.

Table 2

Pose estimation of the high-quality registrations on the outdoor (YFCC100M) scenes using the metrics mAP (%).

Method	LFGC			OANet			SuperGlue			LoFTR		
	5°	10°	20°	5°	10°	20°	5°	10°	20°	5°	10°	20°
Learning-based	8.48	16.39	29.29	35.00	48.51	63.00	30.51	40.20	52.34	42.78	51.40	62.32
RANSAC	33.50	44.34	57.99	41.23	53.25	66.42	40.44	50.59	62.28	47.89	56.74	67.15
VSAC	14.10	19.28	27.90	25.66	32.72	42.35	34.44	45.37	58.57	20.08	25.81	34.73
GMS	6.49	11.37	19.35	18.80	25.00	34.70	21.38	29.26	40.19	36.06	44.03	54.13
LPM	9.42	16.97	28.73	30.90	42.51	56.86	34.36	44.45	56.66	44.93	53.18	64.05
OEAM	16.07	22.55	33.41	30.98	43.07	57.82	33.34	42.98	55.04	44.12	53.28	64.33

To demonstrate that the registrations rejected by OEAM are true low-quality registrations, we present the cumulative precisions of LFGC, OANet, SuperGlue, and LoFTR on the rejected registrations in five equal intervals in Fig. 7. The precision produced by the learning-based method on the rejected examples is shown by ground truth in Fig. 7. The average precision of every column is displayed at the top of the column. The gray, blue, green, pink, and purple columns show the occupation of the precision in $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, and $[0.8, 1]$, respectively. The rejection numbers in Figs. 7a - 7d are 794, 430, 48, and 81, respectively.

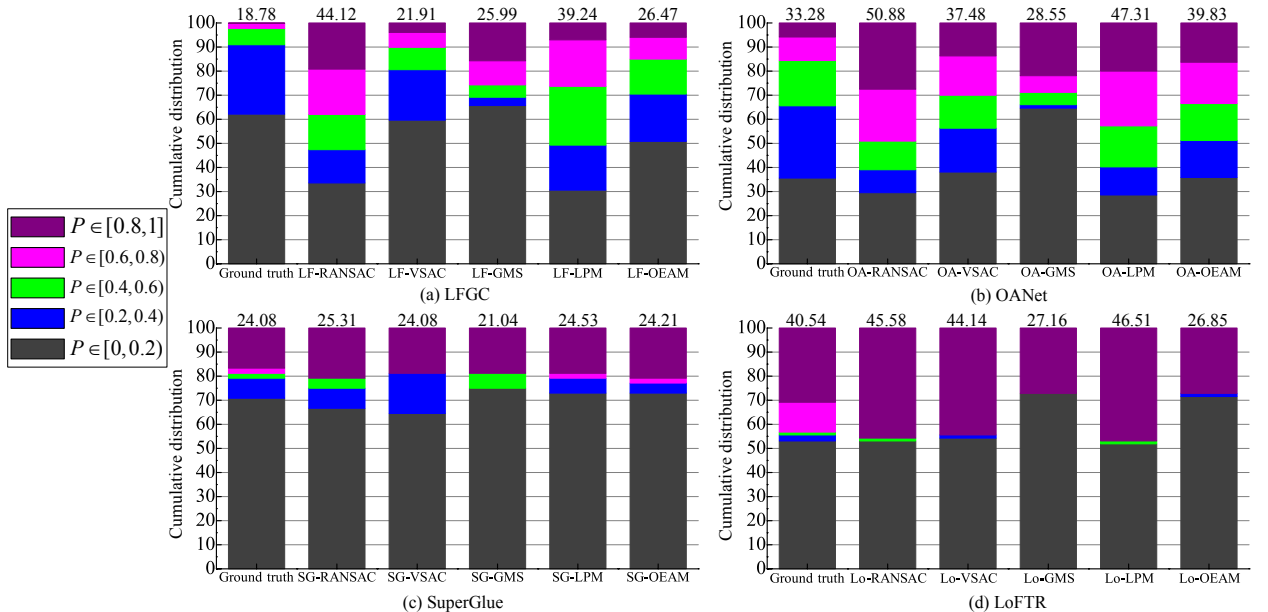


Fig. 7: Cumulative precision of rejection examples on the outdoor (YFCC100M) scenes. (a) LFGC. (b) OANet. (c) SuperGlue. (d) LoFTR. The cumulative precision produced by learning-based method on the rejection examples is shown by ground truth. The number at the top of the column shows the average precision (%) of the method.

As shown in Fig. 7, the precisions of the rejected samples are low, and most of them are truly low-quality registrations that warrant rejection. The average precisions of the rejected registrations on LFGC, OANet, SuperGlue, and LoFTR are 18.78%, 33.28%, 24.08%, and 40.54%, respectively. As aforementioned, the precisions of LFGC, OANet, SuperGlue, and LoFTR on accepted registrations are 55.83%, 75.76%, 98.52%, and 95.16%, respectively. The precisions of the rejected registrations are considerably lower than those of the accepted registrations. In addition, the ratio of the precision that is less than 80% for LFGC, OANet, SuperGlue, and LoFTR are 99.87%, 94.19%, 83.33%, and 69.14%, respectively. On average, 95.64% of the rejected samples contain at least 20% mismatches. Furthermore, the maximum average precisions in Figs. 7a - 7d are 44.12%, 50.88%, 25.31%, and 46.51%, respectively. On average, RANSAC shows the highest precision on rejected samples. For truly low-quality registration, high precision usually induces a detrimental effect on rejection. It suggests that RANSAC may not perform well in rejecting false registration. Although all the benchmarks contribute to precision improvement, most of their precisions on the rejected samples are constrained to a low level. Overall, the registrations rejected by OEAM are truly low-quality. In other words, OEAM identifies low-quality registrations.

Pose estimation of the learning-based registrations rejected by OEAM is listed in Table 3. The average mAPs of LFGC, OANet, SuperGlue, and LoFTR on the three thresholds in Table 3 are 5.61%, 21.96%, 4.69%, and 13.70%, respectively. Correspondingly, the average mAPs of the accepted high-quality registrations in Table 2 are 18.05%, 48.84%, 41.02%, and 52.17%, respectively. Compared with the accepted samples, the pose errors of the rejected samples are quite large and induce the low mAPs in Table 3. Although the learning-based methods produce correspondences on the samples rejected by OEAM, the pose estimated using these correspondences deviates significantly from ground truth values. The pose estimation results demonstrate the rejected samples are truly low-quality registrations with large pose errors.

Table 3

Pose estimation of the rejected samples on the outdoor (YFCC100M) scenes using the metrics mAP (%).

Method	5°	10°	20°
LFGC	1.51	4.03	11.30
OANet	11.63	20.70	33.55
SuperGlue	2.08	5.21	6.77
LoFTR	8.05	13.22	19.83

Experiments on the overlapping images of the outdoor scenes show that RANSAC performs best on high-quality samples. OEAM outperforms VSAC, GMS, and LPM on the outdoor scenes. Although RANSAC takes advantage of overlapping image registration, it performs poorly in non-overlapping image registration. Overall, OEAM is advantageous for inlier precision and pose estimation on accepted high-quality registrations. Furthermore, OEAM can identify truly low-quality registrations.

4.2.2. Experiments on indoor dataset

In this section, we implement baseline experiments on indoor scenes. After using $T_N = 16$ to discriminate low-quality registrations, LF-OEAM, OA-OEAM, SG-OEAM, and Lo-OEAM accept 12769, 13435, 14282, and 12094 samples as high-quality registrations on indoor scenes. We first evaluate the improvement of high-quality registrations using precision and mAP. Fig. 8 shows the recall and precision of the inliers generated by the accepted registrations. Since LoFTR is a detector-free image matching method, its recall is not provided in Fig. 8a.

OEAM is advantageous for the performance improvement of high-quality registrations on indoor scenes. In Fig. 8a, the average recalls of LFGC, LF-OEAM, OANet, OA-OEAM, SuperGlue, SG-OEAM are 79.08%, 48.12%, 60.45%, 44.15%, 71.11%, and 52.39%, respectively. Compared with the three baselines, LF-OEAM, OA-OEAM, and SG-OEAM perform poorly in terms of recall. However, the experiments on outdoor scenes in Section 4.2.1 demonstrate that a low recall does not prevent a corresponding matching method from achieving a high mAP in model estimation. The applications of OEAM on LFGC, OANet, SuperGlue, and LoFTR can improve the precisions of the baselines. The average precisions of LFGC, OANet, SuperGlue, and LoFTR on high-quality registrations of the indoor dataset are 52.76%, 61.20%, 87.44%, and 93.91% respectively. The average precisions of LF-OEAM, OA-OEAM, SG-OEAM, and Lo-OEAM, are 68.60%, 69.22%, 89.91%, and 94.43%, respectively. The precision improvements of LF-OEAM, OA-OEAM, SG-OEAM, and LoFTR-OEAM are 15.84%, 8.03%, 2.46%, and 0.52%, respectively. The insignificant

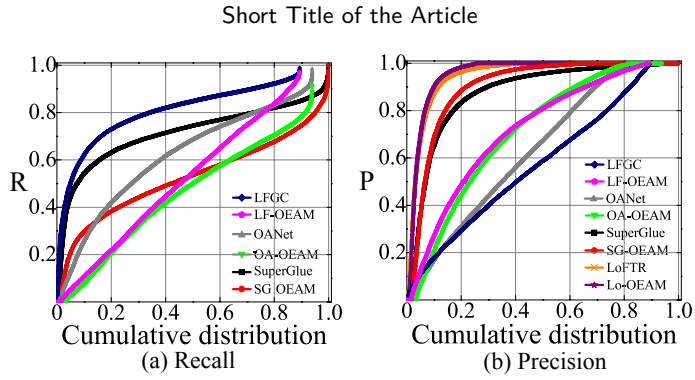


Fig. 8: The recall and precision of the high-quality registrations on the indoor (SUN3D) scenes.

precision improvement on LoFTR is partially due to the high precision of the baseline LoFTR. Overall, OEAM plays a significant role in improving the precision of the high-quality registrations on indoor scenes.

Pose estimations of the high-quality samples on the indoor scenes are listed in Table 4. OEAM is superior in terms of pose estimation. Although the mAP of Lo-OEAM is inferior to that of the baseline LoFTR, the first six rows in Table 4 demonstrate that OEAM outperforms baselines. The mAPs of LFGC, OANet, SuperGlue, and LoFTR on the threshold of 20° are 12.32%, 33.97%, 27.65%, and 23.34%, respectively. The mAPs of LF-OEAM, OA-OEAM, SG-OEAM, and Lo-OEAM are 17.07%, 36.89%, 27.78%, and 21.79%, respectively. The average mAPs of the four learning-based methods and OEAM are 24.32% and 25.88%, respectively. Overall, OEAM is efficient in improving the mAP of the indoor scenes.

Table 4

Pose estimation of the high-quality registrations on the indoor (SUN3D) scenes using the metrics mAP (%).

Method	5°	10°	20°
LFGC	2.07	5.11	12.32
LF-OEAM	4.33	8.59	17.07
OANet	12.70	21.61	33.97
OA-OEAM	13.97	23.37	36.89
SuperGlue	10.15	16.86	27.65
SG-OEAM	9.71	16.58	27.78
LoFTR	7.20	12.98	23.34
Lo-OEAM	6.49	10.72	21.79

To demonstrate the reasonability of the rejection on the low-quality registrations, we show precision distributions of the four learning-based methods on rejected registrations in Table 5. The number of low-quality registrations discriminated on LFGC, OANet, SuperGlue, and LoFTR are 2103, 1437, 590, and 2778, respectively. As shown the third column in Table 5, most precisions of the rejected samples fall within the interval $[0, 20\%)$. The proportions of the precisions that are lower than 80% for LFGC, OANet, SuperGlue, and LoFTR are 100%, 70.01%, 81.69%, and 54.64%, respectively. It seems that all the rejections produced by OEAM on LFGC are reasonable. Since 45.36% of the rejected registrations on LoFTR carry precisions that are greater than 80%, false positives may be inevitable. It may be partly due to the attention mechanism involved in the LoFTR model. The self-attention and cross-attention techniques used in the LoFTR module may have eliminated paradoxical correspondences. The average precisions of the rejected registrations are 9.01%, 37.11%, 33.91%, and 50.45%, respectively. The average precisions of the accepted high-quality registrations are 52.76%, 61.20%, 87.44%, and 93.91% respectively. Compared to the precision achieved on high-quality samples, the registrations rejected by OEAM are in low precision. Overall, OEAM can identify low-precision registrations on indoor scenes.

Experiments on outdoor and indoor scenes show that OEAM is efficient in inferring low-quality registrations for LFGC, OANet, SuperGlue, and LoFTR. Meanwhile, OEAM is beneficial to improving the precision and mAP of high-quality registrations. The comparison experiments on outdoor scenes suggest that RANSAC shows the best performance on high-quality registrations, but it performs poorly in rejecting false registrations. In Section 4.2.1,

Table 5

Precision (%) distribution of the rejected samples on the indoor (SUN3D) scenes.

Method	Rejection number	[0, 20)	[20, 40)	[40, 60)	[60, 80)	[80, 100]	Average
LFGC	2103	88.45	10.46	1.00	0.00	0	9.01
OANet	1437	57.13	8.56	2.64	1.67	29.99	37.11
SuperGlue	590	50.85	7.97	10.00	12.88	18.31	33.91
LoFTR	2778	42.69	3.49	3.20	5.26	45.36	50.45

OEAM is ranked second in terms of high-quality registrations. Compared with RANSAC, VSAC, GMS, and LPM, OEAM shows overall advantages in registration assessment.

4.3. Experiments on non-overlapping images

In this section, we assess learning-based correspondence matching methods on non-overlapping images. As shown in Section 4.1.2, we constructed 4000 non-overlapping images on the outdoor dataset YFCC100M for registration experiments. The experimental details are the same as those on the overlapping images. Fig. 9 shows the inlier number histograms of RANSAC, VSAC, GMS, LPM, and OEAM relative to LFGC, OANet, SuperGlue, and LoFTR on non-overlapping images of the outdoor (YFCC100M) scenes. The first row in Fig. 9 (Fig. 9(1)) shows the results of the learning-based methods. The horizontal coordinate in Fig. 9 shows the number of resulting inliers. The vertical coordinate shows the occurrence counts of the inlier number.

Poisoning attacks may be a potential application deficiency of the learning-based method. As shown in Fig. 9(1), LFGC, OANet, SuperGlue, and LoFTR averagely produce 97.9, 111.4, 15.8, and 122.3 inliers, respectively. Since the image pairs used for registration are composed of non-overlapping images, almost all correspondences are false matches. For correspondence matching method, it would be best not to output matches on non-overlapping images. However, the learning-based methods usually output a large number of false matches. It suggests that learning-based methods usually produce false registrations on non-overlapping images. Traditionally, registration is usually performed on benign registration samples with partial overlaps and is not designed for online service. However, learning-based methods depend on hardware. The method application on the handheld device may have to submit its request to the online service system. If the models are served as online systems, non-overlapping images can be used for poisoning attacks that fool the model to produce false registrations. Learning-based methods are vulnerable to non-overlapping image registrations. Assessment of learning-based registration methods may be necessary for model applications.

Compared with RANSAC, VSAC, GMS, and LPM, OEAM exhibits high performance in identifying false registrations. As shown in Figs 9(2) - 9(6), the resulting inlier numbers of RANSAC, VSAC, GMS, LPM, and OEAM are all smaller than those of learning-based methods. However, it does not mean that they are all robust to the results of learning-based methods. The average inlier numbers of LF-RANSAC, OA-RANSAC, SG-RANSAC, and Lo-RANSAC are 22.3, 26.1, 6.5, and 33.1, respectively. If we use the number of inliers $N_i \geq 16$ to determine the registration between a pair of images, LF-RANSAC, OA-RANSAC, SG-RANSAC, and Lo-RANSAC produce 3598, 3694, 205, and 3939 registrations, respectively. The proportion of false registrations is quite large. Admittedly, raising the threshold can alleviate the false registrations on non-overlapping images. However, it may inevitably reject true registrations on overlapping images. Taking Lo-RANSAC as an example, if the registration is determined by $N_i \geq 50$, it produces 180 false registrations on non-overlapping images and rejects 2604 registrations on overlapping images. Due to the high performance of LoFTR in registering overlapping images, it suggests that most of the 2604 rejections on the overlapping images are false rejections. Overall, RANSAC is not robust to poisoning attacks on learning-based methods. Similarly, as shown in Fig. 9(5), LPM is inefficient in identifying false registration. For VSAC, $N_i \geq 16$ brings 410, 690, 22, and 862 registrations on LFGC, OANet, SuperGlue, and LoFTR, respectively. GMS produces 52, 104, 4, and 1061 registrations for $N_i \geq 16$ on LFGC, OANet, SuperGlue, and LoFTR, respectively. As shown in Fig. 9(6), OEAM produces 187, 306, 51, and 1178 registrations on LFGC, OANet, SuperGlue, and LoFTR, respectively. On average, VSAC, GMS, and OEAM produce 496, 305, and 431 registrations, respectively, on the registration of 4000 non-overlapping images. Although VSAC and GMS show competitive performance in rejecting false registration, they are inferior in pose estimations in Table 2. Overall, OEAM shows high performance in rejecting false registration.

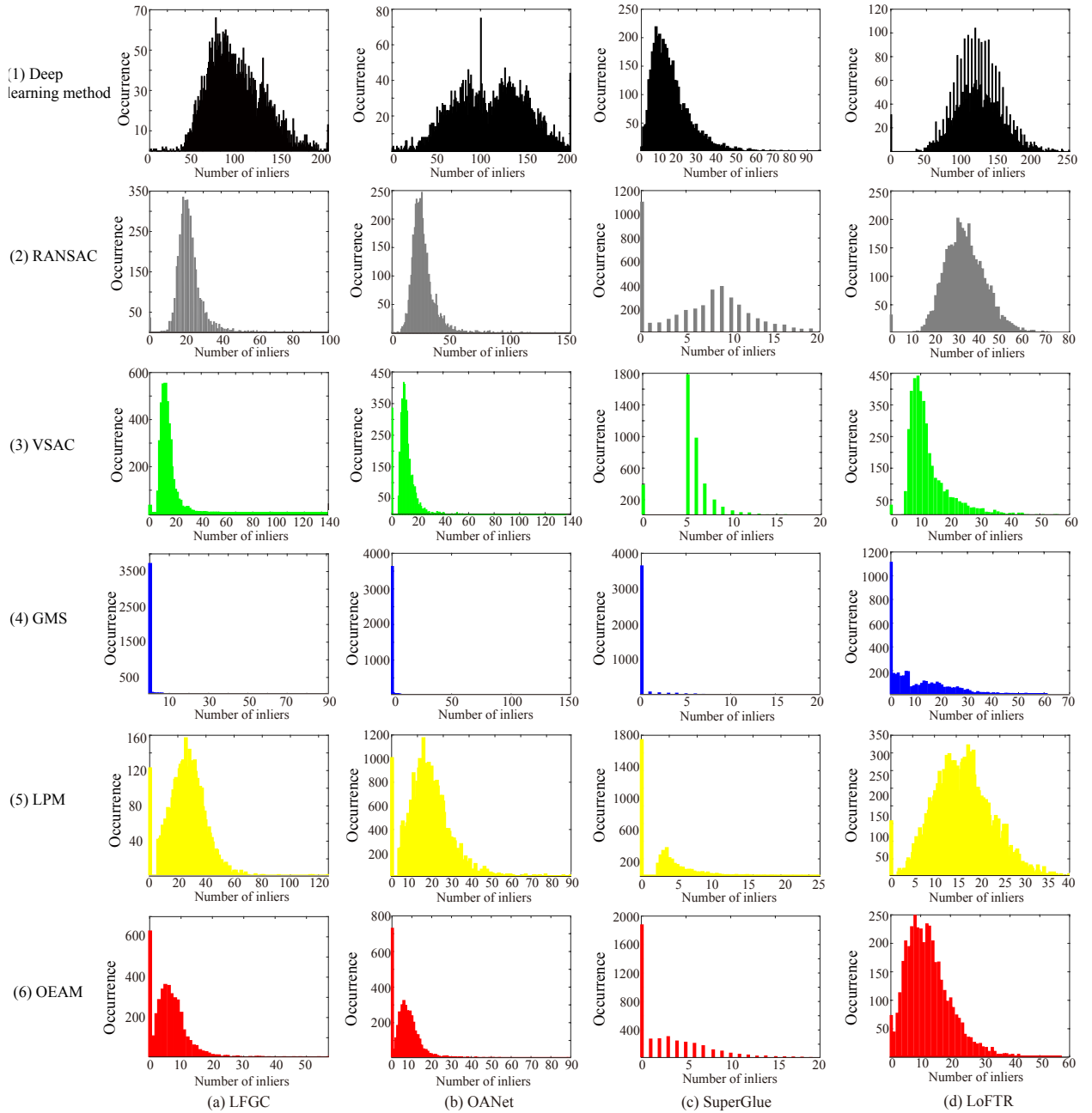


Fig. 9: The inlier number histograms of RANSAC, VSAC, GMS, LPM, and OEAM relative to LFGC, OANet, SuperGlue, and LoFTR on non-overlapping images of the outdoor (YFCC100M) scenes.

5. Conclusion

In this study, we demonstrate poisoning attacks may be a potential application deficiency of learning-based correspondence matching methods. Learning-based methods usually output a large number of false matches on registrations of non-overlapping images. Since learning-based methods rely on hardware, their applications may depend on online service systems. If learning-based registration models serve as online systems, non-overlapping images can be used as poisoning samples to fool the models. The experiments in Section 4.3 demonstrate that learning-based methods

are vulnerable to poisoning attacks launched by non-overlapping images. Assessment of learning-based registration methods may be necessary for model applications.

We propose a method, OEAM, to assess learning-based registration methods. The assessment contains two tasks. One is to infer whether image registration is of low quality. The other is to improve registration performance. OEAM first eliminates outliers based on the spatial paradox. Then, it implements a registration assessment in two streams using the obtained core correspondence set. If the cardinality of the resulting core set is sufficiently small, the input registration is inferred to be of low quality. Otherwise, it is assessed to be of high quality. When OEAM meets a high-quality assessment, it improves the inlier precision and the model mAP using the core set. The registration experiments on partially overlapping images show that OEAM can reliably infer low-quality registration and improve the performance of high-quality registration. Furthermore, experiments on registrations of non-overlapping images show that OEAM correctly rejects most of the false registrations. It shows that OEAM can not only effectively improve the performance on benign samples with partial overlaps, but also effectively reject poison samples without overlap. OEAM is robust to poisoning attacks crafted by non-overlapping images. The comparison experiments show that traditional correspondence matching methods may be incapable of rejecting poison samples without overlaps. The robustness of learning-based correspondence matching methods is an interesting research topic.

One demerit of OEAM is that the rejection ability of non-overlapping image registrations launched by LoFTR may not live up to our expectation. It may be partially due to the self-attention and cross-attention techniques used in LoFTR module, which may have eliminated paradoxical correspondences so that the efficiency of the outlier elimination used in OEAM is partially discounted.

Declaration of competing interest

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.

Acknowledgments

This work was supported in part by Anhui Provincial Natural Science Foundation [1808085MF171]; in part by the National Natural Science Foundation of China [62272006, 61972439, 61976006]; in part by Innovate UK via the Knowledge Transfer Partnership scheme (KTP) [KTP012053].

References

- [1] Jiayi Ma, Yang Wang, Aoxiang Fan, Guobao Xiao, and Riqing Chen. Correspondence attention Transformer: A context-sensitive network for two-view correspondence learning. *IEEE Transactions on Multimedia*, 25:3509–3524, 2023.
- [2] Lijie Yang, Qian Huang, Xing Li, and Yang Yuan. Dynamic-scale grid structure with weighted-scoring strategy for fast feature matching. *Applied Intelligence*, 52:10576–10590, 2022.
- [3] Shuhua Ma, Peikai Guo, Hairong You, Ping He, Guanglin Li, and Heng Li. An image matching optimization algorithm based on pixel shift clustering RANSAC. *Information Sciences*, 562:452–474, 2021.
- [4] Lina Wang, Huaidan Liang, Zhongshi Wang, Rui Xu, and Guangfeng Shi. CFOG-like image registration algorithm based on 3D-structural feature descriptor for suburban optical and SAR. *Optik*, 272:170158, 2023.
- [5] Seyed-Mahdi Nasiri, Reshad Hosseini, and Hadi Moradi. Multiple-solutions RANSAC for finding axes of symmetry in fragments of objects. *Pattern Recognition*, 131:108805, 2022.
- [6] Zhonghua Wan, Caihua Xiong, Wenbin Chen, Hanyuan Zhang, and Shiqian Wu. Pupil-contour-based gaze estimation with real pupil axes for head-mounted eye tracking. *IEEE Transactions on Industrial Informatics*, 18(6):3640–3650, 2021.
- [7] Zhiqiang Yan, Hongyuan Wang, Liuchuanjiang Ze, Qianhao Ning, and Yinxi Lu. A pose estimation method of space non-cooperative target based on ORBFPFH SLAM. *Optik*, 286:171025, 2023.
- [8] Yunpeng Sun and Xiaoli Li. Feature extraction and matching combined with depth information in visual simultaneous localization and mapping. *International Journal of Advanced Robotic Systems*, 20(2):1–15, 2023.
- [9] Enwen Hu and Lei Sun. DANIEL: A fast and robust consensus maximization method for point cloud registration with high outlier ratios. *Information Sciences*, 614:563–579, 2022.
- [10] Weixing Peng, Yaonan Wang, Hui Zhang, Yihong Cao, Jiawen Zhao, and Yiming Jiang. Deep correspondence matching based robust point cloud registration of profiled parts. *IEEE Transactions on Industrial Informatics*, 2023.
- [11] 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*, pages 2666–2674, 2018.
- [12] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Hongkai Chen, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. OANet: Learning two-view correspondences and geometry using order-aware network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3011–3122, 2022.

- [13] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4938–4947, 2020.
- [14] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with Transformers. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8922–8931, 2021.
- [15] Xishi Huang, John Moore, Gerard Guiraudon, Douglas L Jones, Daniel Bainbridge, Jing Ren, and Terry M Peters. Dynamic 2D ultrasound and 3D CT image registration of the beating heart. *IEEE Transactions on Medical Imaging*, 28(8):1179–1189, 2009.
- [16] Kui Jia, Tsung-Han Chan, Zinan Zeng, Shenghua Gao, Gang Wang, Tianzhu Zhang, and Yi Ma. ROML: A robust feature correspondence approach for matching objects in a set of images. *International Journal of Computer Vision*, 117:173–197, 2016.
- [17] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Lijia Li. YFCC100M: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73, 2016.
- [18] Mehran Fotouhi, Hamid Hekmatian, Mohammad Amin Kashani-Nezhad, and Shohreh Kasaei. SC-RANSAC: Spatial consistency on RANSAC. *Multimedia Tools and Applications*, 78(7):9429–9461, 2019.
- [19] Ondřej Chum and Jiří Matas. Optimal randomized RANSAC. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1472–1482, 2008.
- [20] Evren İmre and Adrian Hilton. Order statistics of RANSAC and their practical application. *International Journal of Computer Vision*, 111(3):276–297, 2015.
- [21] Daniel Barath and Jiří Matas. Graph-Cut RANSAC: Local optimization on spatially coherent structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4961–4974, 2022.
- [22] Maksym Ivashchkin, Daniel Barath, and Jiří Matas. VSAC: Efficient and accurate estimator for H and F. In *Proceedings of the IEEE international conference on computer vision*, pages 15243–15252, 2021.
- [23] Jianghui Cai, Yuqing Yang, Haifeng Yang, Xujun Zhao, and Jing Hao. ARIS: A noise insensitive data pre-processing scheme for data reduction using influence space. *ACM Transactions on Knowledge Discovery from Data*, 16(6):110, 2022.
- [24] Kan Ren, Yunfei Ye, Guohua Gu, and Qian Chen. Feature matching based on spatial clustering for aerial image registration with large view differences. *Optik*, 259:169033, 2022.
- [25] Jiayi Ma, Ji Zhao, Junjun Jiang, Huabing Zhou, and Xiaojie Guo. Locality preserving matching. *International Journal of Computer Vision*, 127(5):512–531, 2019.
- [26] Luca Cavalli, Viktor Larsson, Martin Ralf Oswald, Torsten Sattler, and Marc Pollefeys. Handcrafted outlier detection revisited. In *European conference on computer vision*, pages 770–787. Springer, 2020.
- [27] Xintao Ding, Boquan Li, Wen Zhou, and Cheng Zhao. Core sample consensus method for two-view correspondence matching. *Multimedia Tools and Applications*, 2023.
- [28] Jiawang Bian, Wenyan Lin, Yun Liu, Le Zhang, Saikit Yeung, Mingming Cheng, and Ian Reid. GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence. *International Journal of Computer Vision*, 128:1580–1593, 2020.
- [29] Weiqing Wang, Yongrong Sun, Zhong Liu, Zhantian Qin, Can Wang, and Jinchang Qin. Image matching via the local neighborhood for low inlier ratio. *Journal of Electronic Imaging*, 31(2):023039, 2022.
- [30] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [31] Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Statistics and analysis of shapes*, pages 81–105, 2006.
- [32] 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*, pages 224–236, 2018.
- [33] Chen Jun, Gu Yue, Luo Linbo, Gong Wenping, and Wang Yong. Two-view correspondence learning via complex information extraction. *Multimedia Tools and Applications*, 81(3):3939–3957, 2022.
- [34] Lei Yin, Chong Yu, Yuyi Wang, Bin Zou, and Yuanyan Tang. Ultrarobust support vector registration. *Applied Intelligence*, 51(6):3664–3683, 2021.
- [35] Tangfei Liao, Xiaoqin Zhang, Yuewang Xu, Ziwei Shi, and Guobao Xiao. SGA-Net: A sparse graph attention network for two-view correspondence learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(12):7578–7590, 2023.
- [36] Hua Yang, Yuyang Jiang, Kaiji Huang, and Zhouping Yin. Dynamic attention-based detector and descriptor with effective and derivable loss for image matching. *Journal of Electronic Imaging*, 32(2):023022, 2023.
- [37] Weiyue Zhao, Hao Lu, Zhiguo Cao, and Xin Li. A2B: Anchor to barycentric coordinate for robust correspondence. *International Journal of Computer Vision*, 131:2582–2606, 2023.
- [38] Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Neighbourhood consensus networks. In *Advances in neural information processing systems*, volume 31, 2018.
- [39] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, UK, 2000.
- [40] 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.
- [41] 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*, pages 1625–1632, 2013.