Pairwise Alignment and Compatibility for Arbitrary Irregular and Eroded Image Fragments

Ofir Itzhak Shahar[1], Gur Elkin[1], and Ohad Ben-Shahar[1]

Faculty of Computer and Information Science,
Ben-Gurion University of the Negev, Beer-Sheva, Israel
shofir@post.bgu.ac.il, gurshal@post.bgu.ac.il, ben-shahar@cs.bgu.ac.il

Abstract. Pairwise compatibility calculation is at the core of most archaeological reconstruction algorithms, in particular those designed to solve different types of jigsaw puzzle problems. However, most existing approaches fail, or aren't designed to deal with fragments of realistic geometric properties one encounters in real-life puzzles, where irregular shapes or erosion are abundant. In practice, compatibility is typically needed, formulated, and assessed only for fragments with idealized, uniform, or strongly constrained shapes, limiting real-world applicability. In this paper, we propose an efficient approach for computing optimal alignment configurations for pairs of fragments and scoring their pictorial compatibility without any assumptions about their shapes, sizes, or visual content. We also propose evaluation metrics for pairwise compatibility, and introduce an extensive dataset containing pairs of fragments generated via a novel image fragmentation approach and a formal erosion model that mimics real-world archaeological erosion. We benchmark it with existing and new baselines, demonstrating the superior performance of our proposed approach.

1 Introduction

Although computer vision and image processing have been experiencing significant breakthroughs in the last decade, some real-world problems, even those that could be abstracted and formulated relatively easily, remain unsolved. Such is the problem of robustly reconstructing real-world archaeological artifacts from their fragments, a challenge that is a major motivation for seeking computational solutions to the *jigsaw puzzle problem* (e.g., [23,7,14,30]) and is still largely open despite the vast resources, time, and direct involvement of domain experts recruited to its solution [4,1].

Freeman and Gardner first posed the 2D jigsaw-puzzle problem for unrestricted shapes in 1964 [12], later shown to be NP-complete [6]. Since then most research has focused on square piece puzzles [14,17,28,10] while the original case of fragments having arbitrary realistic shapes remains unsolved. This, of course, limits the applicability. For example, to our best knowledge, most solvers do not successfully reconstruct public real-world archaeological challenges without relying on prior knowledge of the target, such as its image in reconstructed form [30,8].

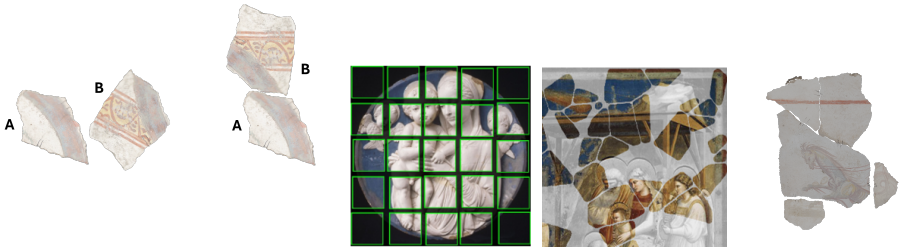


Fig. 1. (A) Successful alignment of two RePAIR fragments [30] using PolEx. Samples from public datasets: **(B)** JPwLEG-5 [27] with square pieces and gaps; **(C)** DAFNE [8] with Voronoi partition and smoothing erosion; **(D)** RePAIR [30] with real-world 3D scanned fragments.

Another critical aspect of the archaeological reconstruction domain, which is often treated poorly or completely ignored by recent works, is the potential *erosion* of the archaeological fragments (see Fig. 1). In realistic scenarios, fragments may be degraded due to poor handling, harsh physical conditions, or simply matter deterioration over time which might drastically alter their physical geometric shape, as is often observed in real pieces [30]. Unfortunately, even when this is taken into consideration, recent works on archaeological reconstruction, puzzle solving, and pairwise alignment tend to simulate it in ways that rarely resemble realistic scenarios [31,32].

With these critical gaps in mind, and to clarify our contributions, in this work we focus on *pairwise compatibility*, the core of any assembly algorithm. We propose: (1) PolEx, a method and algorithm for combining polygonal approximation with extrapolation-based pictorial scoring that handles unrestricted shapes and erosion; (2) adaptive configuration sampling based on fragment geometry rather than uniform grids; (3) pairwise evaluation metrics; and (4) a novel archaeologically-inspired fragmentation approach using Voronoi tessellation, a realistic erosion model based on Perlin noise, and a comprehensive dataset with 1520 fragment pairs from 1000 puzzles generated using these techniques.

All these novelties facilitate robust pairwise compatibility computation for unrestricted, eroded fragments, paving the way for practical archaeological reassembly.

2 Related Work

While compatibility of fragments has been addressed for square pieces (e.g. [10,23,26]) and even polygonal fragments [14,19], pairwise compatibility methods for fragments of unrestricted and possibly eroded shapes are much scarcer. For example, Derech et al. [7] proposed a greedy archaeological puzzle reconstruction algorithm that iteratively picks the best additional fragment while discretizing the (other continuous) configuration space into a finite set of translations and orientations. The compatibility score is based solely on pictorial content, by calculating pixel-wise dissimilarity between extrapolated parts of the pair of fragments.

Although their compatibility function considers erosion, the use of an outdated extrapolation process prohibits useful results on real archaeological data, and the uniform discretization creates a trade-off between runtime and the ability to capture good (and let alone optimal) alignments.

Khoroshiltseva et al. [16] handle irregular shapes via line-continuation of pictorial features while again utilizing restricted orientations, while Cao et al. [3] employ a two-stage global reassembly with region-overlap pruning and a MobileViT classifier. Their method does score polygonal-edge alignments but does not model or handle erosion. Tsesmelis et al. [30] presented two geometric schemes: one replacing pictorial matching with spring-mass optimization [14], the other a genetic global optimizer balancing reconstruction area and overlap, both ignoring pictorial coherence. Recently, Islam et al. [15] proposed ReassembleNet, an end-to-end framework with learnable contour keypoints and diffusion-based pose estimation, achieving improved results on the RePAIR dataset [30] but requiring pretraining on semi-synthetic data.

Unlike the general puzzle-solving literature, the prior art in computational archaeology did put some effort to handle puzzles of unrestricted geometric shapes but at the same time tended to ignore pictorial content (e.g., [1,13,22]). These works typically scan fragments into 3D point clouds, process them into meshes, and extract geometrical features from facets or boundary curves. Differing from these approaches, PolEx handles unrestricted shapes, arbitrary pictorial content, and realistic erosion without predetermined configuration spaces or shape constraints.

3 Archaeologically-inspired Fragmentation

Although many have addressed puzzle-solving challenges, only few archaeological reconstruction datasets are currently available, and none are specifically designed to test pairwise pictorial dissimilarity and alignment. To address this gap and assess our pairwise alignment and compatibility approach, we present a novel dataset of pairs of fragments that better simulate the geometry of broken archaeological artifacts. The pieces are generated using a controlled fragmentation process, inspired by contemporary reconstruction challenges [8,30]. The realism of our synthetic approach is most strongly apparent in the *uneven degree of erosion* along fragment boundaries, thus incorporating another aspect of complexity for reassembly algorithms. The next subsections describe that dataset, which is available in this link.

Image Generation In total, our dataset comprises pairs of fragments from 1000 archaeological puzzles. To generate the base images, we first employed a large language model [2] to create specialized prompts, seeking to describe high-quality fresco images across various artistic styles [11]. These prompts were then fed into Stable Diffusion 3.5 [24] to generate the diverse set of images used in our dataset. Our decision to use synthetically generated images rather than photographs of actual frescos serves multiple purposes: (1) it ensures no publicly available images of real artifacts can be exploited in the compatibility and

alignment process (for example, by granting the solver internet access for reference materials); (2) it avoids potential copyright and cultural heritage considerations while still producing convincing fresco-like images; and (3) to facilitate transparency and reproducibility, all prompts used to generate the images are published alongside the dataset.

Puzzle and Fragment Pairs Generation At the base of our dataset generation pipeline lies a Voronoi tessellation of the image domain, a mathematical process that partitions a plane into disjoint regions based on the distance to predetermined generating sites $\mathcal{V}(s_1, \ldots, s_N) = \{R_1, \ldots, R_N\}$, such that $R_i = \{x \in R^2 \mid \forall j, \|s_i - x\| \leq \|s_j - x\|\}$. Thus, we start by uniformly sampling $s_1, \ldots, s_N$ from the image plane (Fig. 2A) and treat each Voronoi region as a unique piece segment (Fig. 2B).

To simulate the natural erosion patterns observed in archaeological fragments, we implement a controlled degradation process on the Voronoi boundaries. First, we extract the region boundaries using an edge detector (Fig. 2C), creating a binary edge map that precisely localizes the potential erosion regions. We then generate a two-dimensional Perlin noise [21] field matched to the image dimensions (Fig. 2D), which provides smoothly varying values that determine the local erosion intensity. This noise intensity at each edge pixel is then multiplied by the overall *erosion rate* parameter, to determine an "erosion radius" around it. Pixels within this radius are then deleted from the Voronoi segmentation image (Fig. 2E-F). To produce the final puzzle, the eroded partition is applied to an image (Fig. 2A-H). Given the continuity of the noise model, this approach ensures spatial coherence in the erosion pattern while maintaining local variability, as adjacent boundary points with similar noise values undergo similar degrees of erosion. The resulting erosion map is then applied to the original Voronoi regions, discarding the eroded pixels. Finally, we use this eroded segmentation map as a mask to extract the puzzle pieces from the input image, producing a set of realistically degraded fragments that closely mimic the appearance of actual archaeological artifacts.

To extract pairs of adjacent fragments from the generated puzzle (Fig. 2I), we utilized the Delaunay triangulation [5] of the generating sites $s_1, \ldots, s_N$. This graph is known to be dual to the Voronoi diagram over the same set of sites. Hence, fragments that are generated from adjacent sites in the Delaunay triangulation are neighbors in the generated puzzle.

Statistical Properties of the Proposed Dataset The puzzles created by the above Voronoi-based approach exhibit well-known statistical properties [18]. Being generated from randomly distributed points, the expected region size in our Voronoi tessellation is $\frac{1}{N}$ [18] of the total image area. From a topological perspective, the tessellation forms a planar graph whose edges represent interfaces between adjacent fragments. Following Euler's characteristic formula, the expected number of edges per fragment converges to 6 as N increases [18]. Consequently, the expected perimeter is approximately $\frac{4}{\sqrt{N}}$ [18]. As expected, certain properties diverge from theoretical predictions as erosion increases (Fig. 2, bottom row).

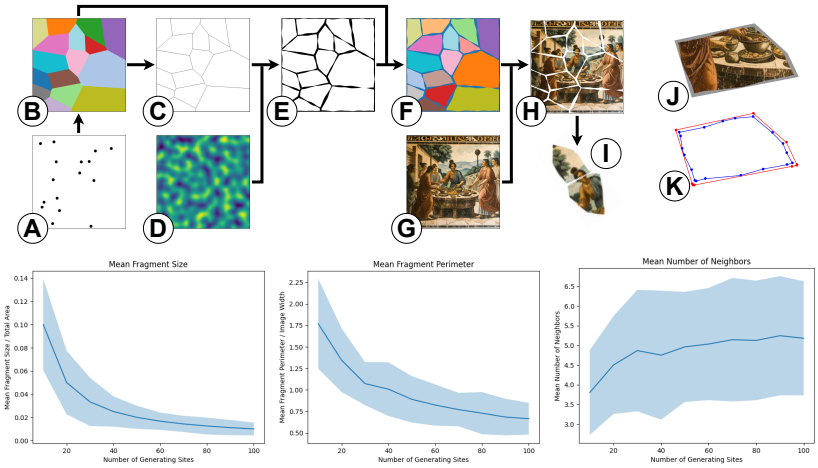


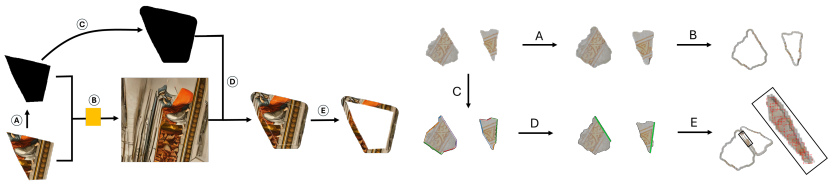
Fig. 2. Puzzle generation and statistical properties. **Top:** Fragmentation process (A-H) from sampling to final puzzle, (I) pair extraction, (J) eroded fragment on original Voronoi region, (K) polygonization showing increased vertices post-erosion. **Bottom:** Empirical properties from non-eroded puzzles with 10-100 fragments: normalized area, perimeter, and neighbor count.

4 Pairwise Alignment and Compatibility for Fragments of Unrestricted Shapes

As briefly described above, the task of pairwise alignment, i.e., computing a valid alignment configuration, out of an infinite space of potential transformations, is challenging. In this context, even the task of just *measuring* compatibility, either geometric or pictorial, in the presence of potential erosion, is already challenging in itself. With both in mind, PolEx is a compatibility measure that is completely free of restrictions regarding the geometric or pictorial features of the inspected fragments, or any information regarding erosion which might have altered them. To do so we take an approach which first reduces the otherwise infinite set of possible alignment configurations by utilizing whatever geometric shape is available of the fragments, and then scores them based on pictorial regions extrapolated beyond their original boundaries. Before we begin, we note that given the task of aligning two fragments, it is common to fix one of them while computing the proper alignment transformation for the other. In the following subsections we denote the latter the *Source*, and the former the *Target*.

4.1 Handling Erosion by Extrapolation

Since fragments are eroded and thus miss regions along their perimeter, we first dilate their geometric shape and *extrapolate* their pictorial content into the dilated area. This is done far enough to exceed their original (pre-eroded) boundary, while optionally utilizing some knowledge on the maximum erosion level. This is done to compensate for the missing near-border pictorial content that is typically used to evaluate the coherence of neighboring fragments.



(A) (B)

Fig. 3. PolEx framework visualization. **(A)** Extrapolation process: (A) fragment mask, (B) diffusion model input, (C) dilated mask M', (D) extrapolated fragment, (E) final extrapolated band. **(B)** Complete compatibility pipeline on RePAIR fragments: (A) fragments with extrapolated bands, (B) polygonal approximations with augmented edges, (C) matched edge pair, (D) computed alignment, (E) pictorial scoring via patch sampling in shared region.

The extrapolation is performed using a frozen, pre-trained Stable Diffusion 1.4 model [24]. Crucially, no fine-tuning is performed on our dataset; thus, PolEx operates as a zero-shot inference method, effectively eliminating the risk of overfitting to the specific distributions of the synthetic benchmark. It should be noted that this pre-trained model was chosen as it was trained for the *inpainting* task, for which extrapolation is a special (albeit extreme and more challenging) case. Hence, with proper adjustments, it could be applied on our task. Moreover, this model can be activated without input text prompts, which greatly simplifies its application in our context.

The extrapolation process extrapolates the fragment image with pictorial content to occupy the entire rectangular region of the image. To focus only on the relevant extrapolation regions, we apply a binary *mask* that confines the newly generated content to a region slightly larger than the fragment's original shape. Slightly more formally, let I be the image of the fragment, and let M be its binary mask (where $M = 1$ indicates a fragment point and $M = 0$ otherwise). Let $M' = \text{dilate}(M, K_{\text{str}})$, be a morphologically dilated mask [25], where K_{str} is a standard structuring element of size $(n_{px} \times n_{px})$ (e.g., $n_{px} = 20$). M is used as a "visual prompt" region for the Stable Diffusion model, so that the generative process augments only the zone surrounding the fragment. The result is cropped back by M', and the original content of the eroded piece is removed, yielding an *extrapolated band* that possesses the extrapolated content around the original eroded fragment. Figure 3A illustrates this process: starting from the fragment's mask (A), both the image and mask are fed into the Latent Diffusion Model (B), while the mask is dilated to M' (C). The model's output is cropped by M' to create an extrapolated fragment (D), from which the original content is removed to create the extrapolated band (E). Figure 3B shows the complete compatibility calculation pipeline on two RePAIR fragments: (A) the two fragments with their extrapolated bands, (B) polygonal approximations with augmented edges (colored), (C) a matched pair of edges (green), (D) the computed alignment transformation, and (E) the pictorial scoring via random patch sampling in the shared region.

4.2 Space of Potential Alignments

The space of potential relative alignments between two fragments is three-dimensional, continuous, and thus infinite. In contrast to earlier proposals, which sample this space uniformly and regularly (i.e., equally spaced samples) to allow reasonable exploration time [7,16], we seek an alternative approach that adapts the sampling to the geometrical shapes of the fragments. Specifically, we first compute explicit edges and augmented edges (see below) from the polygonal approximations of both fragments. Then, we compute the transformation required to align the selected edge of the source fragment to the one of the target, in order to score the pictorial coherency of the two fragments under the corresponding alignment.

Polygonal Approximation and Edge Augmentation: In order to robustly capture the geometric shape of a fragment, we represent it as a binary image $M(i,j) = \begin{cases} 1, & \text{if pixel } (i,j) \text{ belongs to the fragment} \\ 0, & \text{otherwise} \end{cases}$.

To obtain a clean boundary representation of the fragment, we apply a smoothing operation (e.g., Gaussian filtering) to M to reduce noise and irregularities. We then re-binarize the result to extract a refined contour C that is then approximated by a polygon P_C using the Ramer–Douglas–Peucker algorithm [9].

Let p denote the perimeter of P_C and let $\varepsilon = \alpha\, p$, with $\alpha \ll 1$ be a tolerance parameter. A representation simplification step then retains only the vertices essential to the overall structure by removing the middle vertex where three consecutive vertices $\mathbf{p}_i$, $\mathbf{p}_{i+1}$, and $\mathbf{p}_{i+2}$ form an angle $\theta = \arccos\left(\frac{(\mathbf{p}_{i+1}-\mathbf{p}_i)\cdot(\mathbf{p}_{i+2}-\mathbf{p}_{i+1})}{\|\mathbf{p}_{i+1}-\mathbf{p}_i\|\,\|\mathbf{p}_{i+2}-\mathbf{p}_{i+1}\|}\right)$ that is below a pre-defined small threshold $\Delta\alpha$. This process yields a compact polygonal representation of the fragment's shape.

To better capture geometric structure at different scales, and possible differences in polygonization of matching fragments, we augment the initial representation with additional edges. While there are different ways to augment the geometric representation, we chose to consider triples of consecutive *edges* and test whether they can be considered one edge at some larger scale. If the central edge in the triple is sufficiently short and the outer edges are nearly collinear (i.e., their angular difference is below a given threshold), a new edge is synthesized from the first vertex of the first edge to the last vertex of the third edge.

This augmented edge represents a larger-scale, implicit boundary segment that captures the overall geometric structure. In contrast to the edge-merging procedure performed in the prior polygonal simplification step, the augmented edge is *added* to the representation as an extra edge, *without replacing its components*. The augmented representation is no longer a simple polygonization, but a multiscale representation that captures numerous polygonizations concurrently.

Extraction of Candidate Matching Edges: To establish potential correspondences between the two fragments, candidate pairs of matching edges from the two fragments are identified from their polygonal approximations. Recall that each polygon is initially decomposed into base edges defined by consecutive vertices of the polygonization.

Once the augmented representation is available for the source and target fragments, candidate matching between them is established based on edge lengths. Two candidate edges with lengths L_t and L_s are considered compatible if they satisfy $\frac{\min(L_t,L_s)}{\max(L_t,L_s)} \geq \gamma$, where γ is a predefined length ratio. Edge pairs failing to meet this criterion are considered inadmissible and discarded, an operation that filters out most candidate pairs for most practical values of γ (see experimental evaluation). The pairs that survive this test proceed to the next step.

Calculating Potential Alignment Configurations: Given a pair of admissible candidate matching edges, our goal is to compute a rigid transformation that aligns the source edge with the target edge so that they are oppositely oriented and offset properly. Let M_t and M_s be the midpoints of the target and source edges, respectively. To ensure a proper gap between the fragments such that the erosion is accounted for, the target midpoint is offset outward by a distance g along its outward normal $\mathbf{n}_t$, yielding an adjusted midpoint $M_t^{\text{offset}} = M_t + g\,\mathbf{n}_t$. Let the target edge have an orientation θ_t and the source edge an orientation θ_s. The angle $\theta = (\theta_t + \pi) - \theta_s$ thus defines a rotation that aligns the two fragments such that the source edge faces its corresponding target edge, as would be expected if the two fragments indeed match this way. The source edge midpoint is then rotated by θ about the source fragment centroid to obtain M_s^{rot}. The required translation $T = (T_x, T_y)$ is computed as $T = M_t^{\text{offset}} - M_s^{\text{rot}}$ and the overall alignment configuration for this edge pair is thus given by the transformation parameters (θ, T_x, T_y), which are subsequently used to assess the pictorial compatibility of the fragment pair.

4.3 Pictorial Compatibility of Candidate Alignments

Each candidate alignment is finally assessed for its pictorial compatibility by comparing the extrapolated bands of the fragments after the source fragment is transformed by (θ, T_x, T_y). To do so we find the intersection of the two extrapolated bands and sample patches of random sizes in the predetermined range $[P_{\min}, P_{\max}]$ at fixed intervals with stride s. Given corresponding patches from the two bands, we then compute their LAB color difference. Since the latter is sensitive to the size of patch in which it is computed (i.e., indeed, small objects may be washed out in excessively large patches, and large objects may be missed in patches that are smaller), fixed size patches are intrinsically inferior. The random patches thus serve the purpose of better capturing pictorial structure, shapes, and details of unpredictable scales.

More formally, for each patch i, let Ω_i denote its set of pixel indices. For each pixel $k \in \Omega_i$, let $(L_{t,k}, a_{t,k}, b_{t,k})$ and $(L_{s,k}, a_{s,k}, b_{s,k})$ denote the LAB color values in the target and source patches, respectively. The per-patch dissimilarity is then computed as $\Delta E_i = \frac{1}{|\Omega_i|} \sum_{k\in\Omega_i} \Delta E_{i,k}$, where

$$\Delta E_i = \frac{1}{|\Omega_i|} \sum_{k\in\Omega_i} \Delta E_{i,k}, \text{ where } \Delta E_{i,k} = \sqrt{(L_{t,k} - L_{s,k})^2 + (a_{t,k} - a_{s,k})^2 + (b_{t,k} - b_{s,k})^2}\,.$$

Since different pairs of patches yield different differences, all of the latter are aggregated using a p-norm over N patches, namely $S = \left(\frac{1}{N}\sum_{i=1}^{N} \Delta E_i^p\right)^{1/p}$, and the value is finally normalized by the maximum possible LAB distance to yield a unitless measure.

5 Evaluation

As thoroughly described in previous works [20,14,30], evaluating the reconstruction of fragment assemblies is challenging, not only because the evaluation metric must be invariant to rigid transformations, but because erosion introduces uncertainties in the desired outcome. At the same time, the evaluation measures previously proposed emerge from the need to score *assemblies of many fragments*, while in our case we seek to handle just pairs. With this in mind, we propose two pairwise metrics inspired by the classic metrics, adjusted to the global reconstruction problem by Tsesmelis et al. [30], the *Pairwise Relative Position Score* and the *Pairwise Anchored RMSE* (of both the translation T and rotation angle θ), to assess the quality of the predicted configurations relative to the ground truth.

For the **Pairwise Anchored RMSE**, the target fragment (used as the anchor) is first aligned according to its ground truth transformation. The source fragment is then transformed using its predicted configuration relative to the aligned target fragment. Denoting by $\mathbf{c}_p$ and $\mathbf{c}_g$ the centroids of the source fragment in the predicted and ground truth configurations, respectively, the *translation error* is defined as: $\text{RMSE}_T = \sqrt{(c_{p,x} - c_{g,x})^2 + (c_{p,y} - c_{g,y})^2}$. Similarly, let θ_p and θ_g denote the predicted and ground truth rotation angles of the source fragment. The **rotation error** is then defined as $\text{RMSE}_R = \min\big(\,|\theta_p - \theta_g|\,,\; 2\pi - |\theta_p - \theta_g|\,\big)$, which also accounts for periodicity.

The **Pairwise Relative Position Score** quantifies the spatial consistency between the aligned configurations. After aligning the target and source fragments into a common coordinate frame, we extract the shared region (i.e., the intersection of their M shapes as defined in Sec. 4). The score is then defined as the ratio of the area of the shared region to the area of the ground truth fragment: $S_{\text{rel}} = \frac{\text{Area(Shared region)}}{\text{Area(Source fragment)}}$. A higher S_{rel} score indicates better alignment between the fragments, and shorter adjustments are required to align the inspected solution with the ground truth solution. Together, the S_{rel} and the RMSE metrics provide a comprehensive assessment of the geometric and spatial fidelity of the reconstruction.

6 Results

While real-world datasets such as RePAIR and DAFNE provide valuable global reconstruction challenges, they lack the ground-truth pairwise annotations necessary for benchmarking unrestricted fragment alignment quantitatively. Furthermore, current real-world scans do not possess the isolated erosion ground truth

required to measure the specific impact of pictorial extrapolation. Therefore, we utilize RePAIR solely for qualitative validation (Fig. **??**, Fig. 1) to demonstrate domain transferability, while relying on our proposed dataset for rigorous quantitative ablation. We apply *PolEx* to a test set of 1520 pairs of fragments from our dataset (15%), first reporting results for the single highest-ranked candidate per pair (dubbed top-1). This represents the most challenging and practical scenario where only one alignment can be immediately evaluated, crucial for assessing the robustness of the method without relying on post-processing or multiple candidate evaluation. The hyperparameters used were obtained via Optuna optimization (as further described later). The complete code along with full hyper-parameters description will be made public upon acceptance.

Baselines Benchmarking in this domain is severely constrained by the lack of publicly available solvers capable of handling pairwise inputs with unrestricted, eroded shapes. Recent state-of-the-art global solvers (e.g., Sec. 2) are either unavailable for public use or rely on architecture-specific global context that cannot be decoupled for pairwise scoring. Consequently, we compare PolEx against the only viable open-source alternatives: adapted versions of the RePAIR geometric and genetic solvers, and standard feature-matching baselines (SIFT/ORB), which we explicitly adapted to this pairwise setting to provide a fair comparative landscape.

To rigorously isolate the contribution of PolEx's geometric constraints, these feature-matching baselines serve as "pictorial-only" controls. Since standard registration fails on eroded margins, we adapted these pipelines to utilize the same extrapolated content generated by our method. Specifically, alignment is estimated by detecting and matching keypoints via ORB or SIFT, followed by RANSAC. We test two protocols: (i) Ext. Based, which matches features only within the generated extrapolated bands, and (ii) Mixed, which matches the source's extrapolated band against the entire target fragment (including original content). These baselines effectively simulate whether generative extrapolation alone, without our proposed geometric guidance is sufficient for solving the erosion problem.

Main Results Table 1 presents the comparison of PolEx against all baselines for the top-1 candidate (i.e., the single highest-ranked alignment per pair), evaluated on the test test described above. Our method consistently outperforms all baselines across all metrics. Compared to the best baseline (RePAIR Genetic with 0.153 relative position score), PolEx achieves over 3× improvement (0.506), demonstrating the effectiveness of combining adaptive geometric sampling with extrapolation-based pictorial scoring. The registration-based methods (ORB, SIFT) perform poorly, confirming that pictorial reasoning alone is insufficient without geometric guidance.

Ablation Studies Directly ablating only the geometric component of our framework is not feasible, since meaningful pictorial compatibility measurement requires a plausible candidate alignment, hence our registration-based baselines provide this perspective. Here we ablate two key design choices.

Table 1. Pairwise benchmark on the full test set. Metrics: Avg. Relative Position Score (higher is better), RMSE(θ) in degrees, RMSE(T) in pixels (both lower are better). All methods output a single pose per pair (top-1 candidate).

Method	Rel. Pos. ↑	RMSE(θ) ↓	RMSE(T) ↓
RePAIR Greedy [30]	0.097	122.66	208.28
RePAIR Genetic [30]	0.153	122.08	165.15
ORB Ext. Based	0.058	117.31	251.09
SIFT Ext. Based	0.039	119.78	425.34
ORB Mixed	0.048	119.89	350.80
SIFT Mixed	0.040	118.44	494.19
POLEX (ours)	**0.506**	**57.24**	**100.27**

Table 2. Ablation studies on the full test set (top-1 candidate).

Method	Rel. Pos. ↑	RMSE(θ) ↓	RMSE(T) ↓
POLEX (full, g=10)	**0.506**	**57.24**	**100.27**
No gap (g=0)	0.454	74.16	113.61
No extrapolation (g= − 10)	0.405	74.69	118.00

To isolate the effect of the gap parameter under erosion, we compare our standard setting ($g = 10$) to a zero-gap setting where extrapolated bands meet without offset. As shown in Table 2, introducing a positive gap leads to higher alignment accuracy (0.506 vs 0.454 relative position score), demonstrating that properly accounting for erosion is crucial.

To specifically evaluate the impact of extrapolated content, we compare our standard setting to a *negative* gap variant ($g = -10$), in which no extrapolated content is used and only the original fragment content is permitted, with slight overlap allowed. Table 2 shows that the exclusion of extrapolation substantially degrades performance (0.506 vs 0.405 relative position score), confirming extrapolation's necessity for handling erosion.

Analysis: Considering Multiple Candidates While our results, when choosing only the top candidate achieve strong performance, an interesting question arises: how much additional benefit could be gained by considering multiple top-ranked candidates? This scenario is relevant when integrating PolEx into a global optimization framework that can efficiently evaluate shortlists of potential alignments. Figure 4 explores systematically the impact of two principal parameters: candidate-list size (k) and matching edge ratio (γ). Each panel shows best-of-top-k performance, demonstrating that considering multiple candidates yields substantial improvements. When using Top-1 optimized parameters, the relative position score improves from 0.506 ($k = 1$) to 0.646 ($k = 3$) and 0.654 ($k = 10$). More dramatically, when using parameters specifically optimized for Top-10 scenarios, performance reaches 0.804 for the best-of-10 candidates (Table 3). This rapid improvement with small increases in k reveals an important characteristic of the pairwise compatibility problem: while perfect top-1 ranking is intractable due to the NP-complete nature of puzzle-solving [6], the correct alignment is frequently present among the highest-ranked candidates.

Common pitfalls. Ranking errors commonly stem from: (1) *pictorial ambiguities*, such as large monochromatic or low-texture regions (e.g., sky, water,

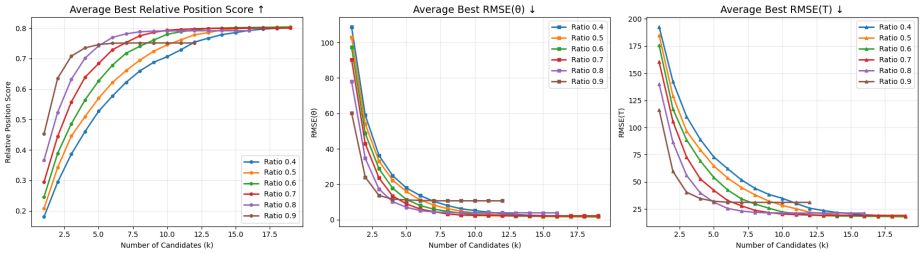


Fig. 4. Impact of candidate-list size (k) and edge ratio (γ) on pairwise performance. Each bar plot shows average best-of-top-k scores for different values of k (horizontal axis) across the three evaluation metrics. Higher γ values (stricter edge matching) reduce the configuration space but may miss correct alignments, while lower γ values consider more candidates at the cost of increased runtime. The rapid improvement from k=1 to k=3 demonstrates the significant value of maintaining candidate shortlists in global reconstruction frameworks.

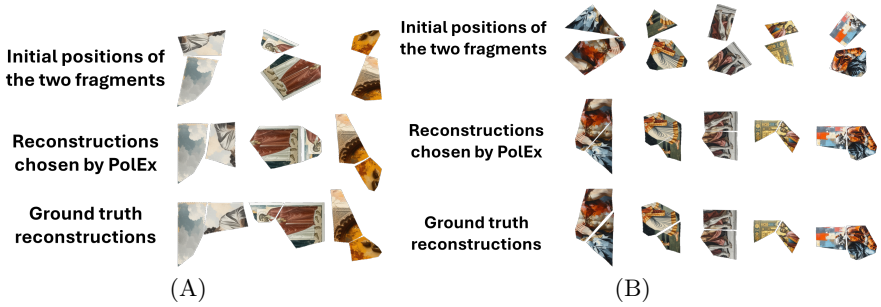


Fig. 5. Qualitative results and common pitfalls. **(A)** Failure modes where monochromatic regions (sky) cause false matches and erosion breaks true neighboring edges. **(B)** Successful alignments demonstrating PolEx's robustness across diverse fragment shapes, erosion patterns, and pictorial content from our test set.

uniform backgrounds) that yield similar dissimilarity scores for incorrect alignments; and (2) *accidental geometric matches* caused by erosion, where fragments' shapes are altered such that non-neighboring edges appear compatible. Moreover, erosion may cause truly neighboring edges to become *broken* into multiple non-collinear segments during polygonization (Fig. 2K), hindering their correct matching. Figure 5 illustrates these pitfalls, along with successful qualitative examples. The top panel shows potential failures, as where monochromatic sky regions lead to false matches. The bottom panel demonstrates representative successful alignments from the test set, showcasing PolEx's ability to handle diverse fragment shapes, erosion patterns, and pictorial content.

Hyperparameter Optimization Given the large hyperparameter space of PolEx (edge length ratios, smoothing parameters, patch sizes, gap distances, etc.), we employed Optuna [29], a hyperparameter optimization framework, to systematically search for optimal parameter configurations. We defined two complementing optimization objectives: **Top-1** optimizes parameters to maximize the quality of the single highest-ranked candidate; and **Top-10** optimizes for

Table 3. Test set performance with different Optuna-optimized parameter sets. Each row shows results using parameters optimized for a specific objective, evaluated at top-k for $k \in \{1, 3, 10\}$. Best-of-top-k means we evaluate the three metrics for the first k ranked candidates and retain the best score.

Optimization	k	Rel. Pos. ↑	RMSE(θ) ↓	RMSE(T) ↓
Top-1 optimized	1	**0.506**	**57.24**	**100.27**
	3	0.646	39.17	60.97
	10	0.654	38.46	58.88
Top-10 optimized	1	0.221	111.89	180.04
	3	0.433	99.24	112.13
	10	**0.804**	**34.05**	**22.02**

the best alignment within the top 10 candidates, which is suitable for global optimization frameworks.

Table 3 shows the test set performance when using parameters optimized for each objective. Notably, Top-1-optimized parameters achieve 0.506 relative position score for the single best candidate, while parameters optimized for Top-3 and Top-10 objectives show different trade-offs when considering multiple candidates. Interestingly, the Top-10-optimized parameters achieve the best performance when considering all 10 candidates (0.804 relative position score), demonstrating that optimization objectives successfully guide the parameter search toward their intended use cases. Please note that the all main results (Table 1) and ablations (Table 2) used Top-1-optimized parameters since it emphasizes single-candidate reliability.

Machine Requirements and Runtime. Experiments were conducted on either a laptop equipped with an NVIDIA RTX 4070 GPU (8GB VRAM) or a workstation with an NVIDIA RTX 4090 GPU (24GB VRAM). The content extrapolation preprocessing step (Stable Diffusion 1.4, executed once per fragment) averaged 27.20 ± 1.28s per piece over 100 samples. PolEx's pairwise alignment process, evaluated on 100 fragment pairs from the test set, averaged 0.076 ± 0.059s per pair when optimized for top-1 candidate, and 0.257 ± 0.078s when optimized for top-10 candidates (suitable for larger reconstruction frameworks). After computing extrapolations once as a pre-processing step, these runtimes, even on limited machines, demonstrate computational efficiency appropriate for employment in broader reconstruction frameworks.

7 Conclusions

In this work, we introduce PolEx as a principled solution for evaluating pairwise compatibility and computing optimal alignments between image fragments of unrestricted geometry, a setting that poses significant challenges for existing methods. Alongside the algorithmic advances, we release a new dataset of fragment pairs, designed to reflect archaeological realism and accompanied by transparent benchmarks against both established and newly developed baselines. Across all quantitative metrics, our approach demonstrates superior performance, underscoring the effectiveness of integrating geometric and pictorial

reasoning in highly degraded and irregular scenarios. Looking forward, extending PolEx within a global optimization framework offers an exciting avenue for future research, potentially enabling robust multi-piece reassembly and further mitigating the complex pitfalls inherent to archaeological reconstruction tasks, as discussed in Sec. 6.

Acknowledgments. This work has been funded in part by the European Union's Horizon 2020 research and innovation programme under grant agreement No 964854 (the RePAIR project).

References

1. Brown, B.J., Toler-Franklin, C., Nehab, D., Burns, M., Dobkin, D., Vlachopoulos, A., Doumas, C., Rusinkiewicz, S., Weyrich, T.: A system for high-volume acquisition and matching of fresco fragments: Reassembling theran wall paintings. ACM transactions on graphics (TOG) **27**(3), 1–9 (2008)
2. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
3. Cao, Y., Fang, Z., Tian, H., Wei, R.: 2d irregular fragment reassembly with deep learning assistance. IEEE Access **12**, 28554–28563 (2024)
4. Caple, C., Garlick, V. (eds.): Studies in Archaeological Conservation. Routledge, London (2020)
5. Delaunay, B.: Sur la sphère vide. a la mémoire de georges voronoï. . (6), 793–800 (1934)
6. Demaine, E.D., Demaine, M.L.: Jigsaw puzzles, edge matching, and polyomino packing: Connections and complexity. Graphs and Combinatorics **23**(Suppl 1), 195–208 (2007)
7. Derech, N., Tal, A., Shimshoni, I.: Solving archaeological puzzles. Pattern Recognition **119**, 108065 (2021)
8. Dondi, P., Lombardi, L., Setti, A.: Dafne: A dataset of fresco fragments for digital anastylosis. Pattern Recognition Letters **138**, 631–637 (2020)
9. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: the international journal for geographic information and geovisualization **10**(2), 112–122 (1973)
10. Elkin, G., Shahar, O.I., Ben-Shahar, O.: Seq2seq models reconstruct visual jigsaw puzzles without seeing them. arXiv preprint arXiv:2511.06315 (2025)
11. Elkin, G., Shahar, O.I., Ohayon, Y., Alali, N., Ben-Shahar, O.: Recognizing artistic style of archaeological image fragments using deep style extrapolation. In: International Conference on Human-Computer Interaction. pp. 115–131. Springer (2025)
12. Freeman, H., Garder, L.: Apictorial jigsaw puzzles: The computer solution of a problem in pattern recognition. IEEE Transactions on Electronic Computers (2), 118–127 (1964)
13. Funkhouser, T., Shin, H., Toler-Franklin, C., Castañeda, A.G., Brown, B., Dobkin, D., Rusinkiewicz, S., Weyrich, T.: Learning how to match fresco fragments. Journal on Computing and Cultural Heritage (JOCCH) **4**(2), 1–13 (2011)
14. Harel, P., Shahar, O.I., Ben-Shahar, O.: Pictorial and apictorial polygonal jigsaw puzzles from arbitrary number of crossing cuts. International Journal of Computer Vision **132**(9), 3428–3462 (2024)

15. Islam, A., Fiorini, S., James, S., Morerio, P., Del Bue, A.: Reassemblenet: Learnable keypoints and diffusion for 2d fresco reconstruction. arXiv preprint arXiv:2505.21117 (2025)
16. Khoroshiltseva, M., Palmieri, L., Aslan, S., Vascon, S., Pelillo, M.: Nash meets wertheimer: Using good continuation in jigsaw puzzles. In: Proceedings of the Asian Conference on Computer Vision. pp. 3223–3238 (2024)
17. Kim, G., Cho, H., Nam, H.: Solving jigsaw puzzles by predicting fragment's coordinate based on vision transformer. Expert Systems with Applications **272**, 126776 (2025)
18. Meijering, J.: Interface area, edge length, and number of vertices in crystal aggregates with random nucleation. Philips Research Reports **8**, 270–290 (1953)
19. Ohayon, Y., Shahar, O.I., Ben-Shahar, O.: Solving convex partition visual jigsaw puzzles. arXiv preprint arXiv:2511.04450 (2025)
20. Paumard, M.M., Picard, D., Tabia, H.: Deepzzle: Solving visual jigsaw puzzles with deep learning and shortest path optimization. IEEE Transactions on Image Processing **29**, 3569–3581 (2020)
21. Perlin, K.: An image synthesizer. ACM Siggraph Computer Graphics **19**(3), 287–296 (1985)
22. Pintus, R., Pal, K., Yang, Y., Weyrich, T., Gobbetti, E., Rushmeier, H.: Geometric analysis in cultural heritage. In: Eurographics Workshops on Graphics and Cultural Heritage. pp. 117–133. The Eurographics Association (2014)
23. Pomeranz, D., Shemesh, M., Ben-Shahar, O.: A fully automated greedy square jigsaw puzzle solver. In: CVPR 2011. pp. 9–16. IEEE (2011)
24. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
25. Serra, J.: Image analysis and mathematical morphology. Academic Press, Inc. (1983)
26. Sholomon, D., David, O., Netanyahu, N.: A generalized genetic algorithm-based solver for very large jigsaw puzzles of complex types. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 28 (2014)
27. Song, X., Jin, J., Yao, C., Wang, S., Ren, J., Bai, R.: Siamese-discriminant deep reinforcement learning for solving jigsaw puzzles with large eroded gaps. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2303–2311 (2023)
28. Talon, D., Del Bue, A., James, S.: Ganzzle++: Generative approaches for jigsaw puzzle solving as local to global assignment in latent spatial representations. Pattern Recognition Letters **187**, 35–41 (2025)
29. Teredesai, A.: Optuna: A next-generation hyperparameter optimization framework. Applied Data Science Track Paper p. 2623 (2019)
30. Tsesmelis, T., Palmieri, L., Khoroshiltseva, M., Islam, A., Elkin, G., Shahar, O.I., Scarpellini, G., Fiorini, S., Ohayon, Y., Alali, N., et al.: Re-assembling the past: The repair dataset and benchmark for real world 2d and 3d puzzle solving. Advances in Neural Information Processing Systems **37**, 30076–30105 (2024)
31. Yang, M., Svirsky, Y., Cheng, Z., Sharf, A.: Self-supervised fragment alignment with gaps. IEEE Transactions on Visualization and Computer Graphics **30**(9), 6235–6246 (2023)
32. Zhou, R., Xia, D., Zhang, Y., Pang, H., Yang, X., Li, C.: Pairingnet: A learning-based pair-searching and-matching network for image fragments. In: European Conference on Computer Vision. pp. 234–251. Springer (2024)