PARTMOSAIC3D: COUPLING PART-AWARE DECOMPOSITION AND JOINT CONSTITUENT SELECTION FOR 3D OBJECT MOSAICS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study 3D object mosaic synthesis from a fixed gallery: reconstructing a textured target mesh by placing whole objects retrieved from a gallery of 3D assets. Any solution makes two coupled decisions—how the target is partitioned into regions, and which object occupies each region—that existing approaches make implicitly or greedily. In this work, we present PARTMOSAIC3D, a formulation that makes both decisions explicit: regions are nodes of a learned semantic part hierarchy, refined where retrieval under a frozen shape representation is poor, and constituents are chosen jointly over the region adjacency graph with fidelity, fit, reuse and interpenetration costs. The fidelity-and-reuse core of this problem is an exactly solvable min-cost flow that contains the baseline's greedy reuse heuristic, and the full problem is solved to certified optimality within candidate pools. A decomposition $\times$ selection factorial with matched-granularity controls tests whether and when each decision matters. On four targets, parts with joint selection raise coverage from 0.527–0.703 to 0.725–0.918 over voxels with greedy selection at matched granularity, and at the baseline's granularity parts reduce the protrusion error from 1.71% to 0.30% and from 1.38% to 0.32% where parts are contiguous. Joint selection helps through its fit and overlap terms, cutting interpenetration volume by 30–54% on three targets, whereas exact optimisation of fidelity and reuse alone changes the mosaics only marginally; retrievability-driven refinement showed no benefit over size-driven subdivision at the thresholds tested, and amodal completion did not help.

1 INTRODUCTION

A photomosaic reconstructs an image from many smaller photographs: from a distance the arrangement reads as the target, and up close each tile is an image in its own right (Silvers & Hawley, 1997; Finkelstein & Range, 1998). We study the three-dimensional analogue, *3D object mosaic synthesis from a fixed gallery*: given a textured target mesh and a gallery of 3D objects, place whole gallery objects, each by a similarity transform with bounded scaling, so that their arrangement approximates the target. Unlike tiles, a region must be matched by a *whole* object, neighbours can interpenetrate or leave surface uncovered, and galleries mix textured and untextured assets.

The closest prior formulation, 3D Collage (Gal et al., 2007), fits database elements to sphere neighbourhoods of sampled surface points and adds them greedily under visibility, overlap and proportion constraints; we adopt its principles of fit, coverage and limited overlap. Its setting leaves open how to handle galleries of thousands of heterogeneous textured objects, whether the target should be partitioned before constituents are chosen, and whether the choices should be made jointly. We make these questions precise by viewing any solution as two coupled decisions: **Decision 1 (region representation)**, how the target is partitioned into regions—by a uniform voxel grid, a semantic part decomposition or an adaptively refined one; and **Decision 2 (constituent assignment)**, which gallery object occupies each region, chosen independently or jointly. A natural modern pipeline, our uniform-voxel baseline, makes both implicitly: it slices the target with a voxel grid, retrieves per cell the best DINOv2 (Oquab et al., 2023) and CLIP (Radford et al., 2021) match, independently apart from a sequential reuse penalty, and aligns it by PCA and ICP (Besl & McKay, 1992).

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

In this work, we present PARTMOSAIC3D (Figure 1), a formulation that makes both decisions explicit. Its novelty lies in three elements. *(i) Retrieval-aware granularity:* the target is decomposed with PartField (Liu et al., 2025) into a hierarchy of semantic parts, optionally completed with HoloPart (Yang et al., 2026), and refined where the gallery cannot represent a node; the rule uses only each node's own candidate pool, so the partition is fixed before selection. To our knowledge, retrieval confidence has not previously been used to select the granularity of a part decomposition. *(ii) Joint selection with certified solutions:* constituents are chosen jointly over the region adjacency graph; the fidelity-and-reuse core of the resulting quadratic semi-assignment problem is a min-cost flow and is exactly the energy that the baseline's reuse penalty optimises greedily (Section 3), so greedy and joint selection are compared on one energy. *(iii) A controlled protocol:* a decomposition $\times$ selection factorial with region-count controls, weights fixed on one target and metrics computed against the original target asks *whether and when* each choice matters (Table 1).

The main contributions of this work are summarized as follows:

1. **Joint constituent selection.** Constituent choice as a labelling problem with fidelity, fit, reuse and interpenetration costs; an exact core that contains the baseline's reuse heuristic as an order-dependent approximation (Propositions 1, 2); and certified optima within candidate pools.
2. **Retrievability-driven part decomposition.** A refinement rule for a semantic part hierarchy, defined independently of selection, that turns retrieval confidence into a granularity criterion.
3. **An end-to-end framework.** A pipeline (Figure 1, Algorithm 1) that builds part regions, retrieves from $16{,}466$ textured objects with frozen features, precomputes placements and interpenetration, and solves selection by min-cost flow or CP-SAT.
4. **A controlled evaluation and its findings.** On four targets, parts beat voxels at matched granularity in coverage and F-score and remove most of the protrusion at the baseline's granularity where parts are contiguous; joint selection helps through its geometric terms, whereas the exact fidelity-and-reuse optimum changes the surface metrics only marginally; retrievability-driven refinement and amodal completion showed no benefit, which we report as findings.

2 RELATED WORK

2D mosaics and collages. Photomosaics select a library image for each grid cell by colour similarity (Silvers & Hawley, 1997; Finkelstein & Range, 1998); decorative mosaics orient tiles along image features (Hausner, 2001; Di Blasi & Gallo, 2005; Battiato et al., 2007), and adaptive mosaics vary tile size and shape with image content (Pavić et al., 2009; Xu et al., 2019). Jigsaw Image Mosaics (Kim & Pellacini, 2002), Digital Tapestry and AutoCollage (Rother et al., 2005; 2006) pose selection as energy minimisation with fidelity, gap and overlap terms, and graph-cut texture synthesis (Kwatra et al., 2003) couples neighbouring patch choices through seam costs. They show that selection should be joint and granularity adaptive, but on the plane: overlap is layering, not interpenetration, and adaptivity follows image edges, not what the library can represent.

3D collage and arrangement. 3D Collage (Gal et al., 2007), our direct predecessor, adds database elements greedily at sphere neighbourhoods of sampled target points, with databases of dozens of elements and no appearance model; its greedy placement gives no measure of distance to the best solution, and its neighbourhood radii do not form a controllable partition. Packing, layout and brick-construction methods (Merrell et al., 2011; Chen et al., 2015; Luo et al., 2015; Cui et al., 2023; Xue et al., 2023b; Debnath et al., 2025) optimise the poses of pieces whose identities are given; in our problem the pose follows from the region and the object, and the free variables are identities chosen from thousands of candidates.

Part decomposition and shape abstraction. Geometric partitions—grids, variational shape approximation (Cohen-Steiner et al., 2004), convex decomposition (Wei et al., 2022) and curvature- or convexity-based segmentation (Shamir, 2008)—are not in general semantic. Semantic segmentation learns category-specific labels (Yi et al., 2016; Mo et al., 2019), zero-shot methods transfer 2D foundation models (Liu et al., 2023b; Abdelreheem et al., 2023; Yang et al., 2024), and PartField (Liu et al., 2025) learns a class-agnostic feature field whose clustering yields hierarchical parts. Shape abstraction fits primitives (Tulsiani et al., 2017; Paschalidou et al., 2019; 2021), adaptively choosing their number by fit (Sun et al., 2019). All define parts intrinsically or by a parametric family,

whereas in mosaic synthesis the useful granularity may depend on what the gallery can represent; the closest mechanisms are 3D Collage's multi-radius fits, which set element size without forming a partition (Gal et al., 2007), and part-retrieval reconstruction steered by the part library (Xu et al., 2023). HoloPart (Yang et al., 2026) completes segmented parts into closed shapes; we use it unchanged as a retrieval query.

Retrieval and part-aware modelling. Multi-view renderings provide strong shape descriptors (Su et al., 2015), and frozen self-supervised (Oquab et al., 2023) and vision–language (Radford et al., 2021) features transfer to retrieval; 3D encoders aligned with language and images (Xue et al., 2023a; Liu et al., 2023a) are an alternative we do not evaluate. Modelling by example and component suggestion (Funkhouser et al., 2004; Chaudhuri et al., 2011; Sung et al., 2017) retrieve parts that become components of a coherent object; part-aware generators synthesise new geometry per part (Chen et al., 2025; Lin et al., 2025; Yang et al., 2025), and reference-augmented generation uses a retrieved shape only as guidance (Wang et al., 2024). In a mosaic, an unrelated whole object is reused unmodified, so fidelity must come from selection and placement.

Discrete optimisation. Unary and pairwise label costs are handled by MAP inference in Markov random fields (Boykov et al., 2001; Kolmogorov, 2006; Kappes et al., 2015); label costs (Delong et al., 2012) favour *fewer* distinct labels, whereas our reuse cost penalises repeated use. One-to-one assignment is solved by the Hungarian method (Kuhn, 1955); with pairwise costs and without the one-to-one constraint it becomes a quadratic semi-assignment problem (Greenberg, 1969), and costs convex in usage counts admit min-cost-flow solutions (Ahuja et al., 1993).

Research gap. To our knowledge, no existing method treats both decisions of 3D object mosaic synthesis as explicit, separately controllable choices: 2D mosaics select jointly but on the plane, 3D Collage selects greedily without a partition, packing optimises poses rather than identities, and part-aware generators synthesise rather than reuse. It is therefore unknown whether semantic parts, gallery-aware granularity or joint selection improve 3D mosaics once the number of regions is controlled; PARTMOSAIC3D addresses this with an explicit formulation, exact solvers and a matched-granularity factorial.

3 PROBLEM FORMULATION

Problem summary. Given a textured target mesh and a fixed gallery, the task is to output placed, unmodified gallery objects that approximate the target with high coverage, low protrusion and little interpenetration. Two decisions are made in sequence: a decomposition operator fixes the regions (Section 3.2), and a selection assigns one object to each region by minimising an energy (Section 3.3). The research problem is which decomposition and which selection rule yield better mosaics once the number of regions is controlled.

3.1 3D OBJECT MOSAIC SYNTHESIS FROM A FIXED GALLERY

Let $\mathcal{M}$ be a textured target mesh with surface $\mathcal{S}$, and let $\mathcal{D} = \{O_1, \dots, O_{|\mathcal{D}|}\}$ be a fixed gallery of 3D objects ($|\mathcal{D}| = 16{,}466$ in our experiments; Section 5). A *3D object mosaic* is a finite set

$$\Phi = \{(m_k, T_k)\}_{k=1}^{n}, \qquad m_k \in \{1, \dots, |\mathcal{D}|\}, \tag{1}$$

where each T_k is a similarity transform with bounded scaling that places the whole object O_{m_k} in the frame of $\mathcal{M}$; an object may appear more than once. Quality is always assessed against the original $\mathcal{M}$. Any procedure that produces Φ decides (1) which regions receive a constituent and (2) which object each region receives; we make both decisions explicit.

3.2 DECISION 1: REGIONS

A *decomposition operator* Π maps the target to a finite region set $\mathcal{R} = \Pi(\mathcal{M}) = \{R_1, \dots, R_n\}$, which is fixed before any constituent is chosen. Each region R_r is a subset of the target geometry, and has a *query geometry* P_r that is used for retrieval. We consider two operators. The uniform-voxel operator Π_{vox} returns the occupied cells of a regular grid over the bounding box of $\mathcal{M}$; the

query geometry of a cell is the target geometry inside it, closed by planar cut faces. The part operator Π_{part} returns a cut of a hierarchy $\mathcal{H}$ of semantic parts, selected by the refinement rule of Section 4; the query geometry of a part is its surface with the new boundary capped, or its amodally completed volume. For Π_{part}, a node r of $\mathcal{H}$ with candidate pool $\mathcal{C}_r$ (defined below) has *retrievability error*

$$\varepsilon_r \;=\; 1 - \max_{m \in \mathcal{C}_r} s_{\text{shape}}(P_r, O_m), \tag{2}$$

and is replaced by its children when $\varepsilon_r > \tau_{\text{err}}$, the node has children, and the children exceed a minimum size. Because ε_r depends only on the node's own candidate pool, $\mathcal{R}$ is fixed before, and independently of, constituent selection.

Pools, placement and adjacency. Frozen shape and appearance representations give similarities $s_{\text{shape}}(P_r, O_m) \in [-1, 1]$ and $s_{\text{app}}(P_r, O_m) \in [0, 1]$, computed from multi-view renderings; the candidate pool $\mathcal{C}_r \subset \mathcal{D}$ holds the K objects most similar in shape, and the assignment pool $\mathcal{C}_r^a \subseteq \mathcal{C}_r$ is a smaller subset (Section 4.2). A fixed alignment procedure $\mathcal{A}$ (PCA, ICP and a bounded scale search, seeded per pair) returns $T_{rm} = \mathcal{A}(R_r, O_m)$, so placement is a deterministic function of (r, m) and every spatial quantity is a function of the chosen identities. The region adjacency graph $\mathcal{G} = (\mathcal{R}, \mathcal{E})$ connects voxel cells that share a face, an edge or a vertex (face adjacency alone misses a fifth of the interpenetrating pairs on baseline mosaics; Appendix D.1) and parts whose surfaces come within a proximity threshold.

3.3 Decision 2: constituent identities

Given $\mathcal{R}$, constituent selection chooses binary variables $x_{rm} \in \{0, 1\}$ for $r \in \mathcal{R}$ and $m \in \mathcal{C}_r^a$, subject to

$$\textstyle\sum_{m \in \mathcal{C}_r^a} x_{rm} = 1 \quad \text{for all } r \in \mathcal{R}, \tag{3}$$

and writes $N_m = \sum_r x_{rm}$ for the number of regions that use object m. The resulting mosaic is $\Phi(x) = \{(m, T_{rm}) : r \in \mathcal{R},\ x_{rm} = 1\}$. We define the joint constituent-selection energy

$$J(x) \;=\; \sum_{r \in \mathcal{R}} \sum_{m} u_{rm}\, x_{rm} \;+\; \lambda_{\text{div}} \sum_{m} N_m^2 \;+\; \sum_{(r,r') \in \mathcal{E}} \sum_{m, m'} w_{rr'}(m, m')\, x_{rm}\, x_{r'm'}, \tag{4}$$

with the following terms.

Unary fidelity and fit. $u_{rm} = -\big[(1-\omega)\, s_{\text{shape}}(P_r, O_m) + \omega\, s_{\text{app}}(P_r, O_m)\big] + \lambda_{\text{fit}}\, \ell_{\text{fit}}(r, m)$, where $\omega = 0.6$ is the appearance weight used by the baseline, and $\ell_{\text{fit}}(r, m) \geq 0$ is the fraction of the region's surface that is not within a distance threshold of O_m placed by T_{rm}.

Reuse. $\lambda_{\text{div}} \sum_m N_m^2$ penalises repeated use, with a marginal cost that grows with each reuse.

Pairwise spatial compatibility. For adjacent regions, $w_{rr'}(m, m') = \lambda_{\text{ov}}\, \text{ov}\big(T_{rm} O_m,\, T_{r'm'} O_{m'}\big)$, where ov is a normalised interpenetration volume of the two placed objects.

Following Gal et al. (2007), fit and overlap are soft costs; since coverage by the union of placements is higher-order in x, ℓ_{fit} is a per-region surrogate. The energy has no scale term, as real-world dimensions are unverified for much of the gallery (Appendix E).

3.4 Structure of the selection problem

By the identity $\sum_m N_m^2 = n + 2\,|\{\{r, r'\} : r \neq r',\, m_r = m_{r'}\}|$, the reuse term is a same-label pairwise cost between *all* pairs of regions that is convex and separable in the usage counts. Minimising Eq. (4) subject to Eq. (3) is therefore a quadratic semi-assignment problem—MAP inference in a pairwise Markov random field with dense same-label costs—which is NP-hard in general, since it contains Potts-model energy minimisation (Boykov et al., 2001); we treat optimality as a property that a solver must certify. Two properties relate it to the baseline.

Proposition 1 (Exactly solvable core)**.** *If $\lambda_{\text{ov}} = 0$, minimising Eq. (4) subject to Eq. (3) is a min-cost flow problem: a unit of flow leaves each region node, traverses an arc of cost u_{rm} to object node m, and reaches the sink through parallel unit-capacity arcs of costs $\lambda_{\text{div}}(2t-1)$, $t = 1, \ldots, n$. Because these costs increase with t, an optimal flow uses them in order, so its cost equals J; integrality of min-cost flow yields an exact binary solution in polynomial time (Ahuja et al., 1993).*

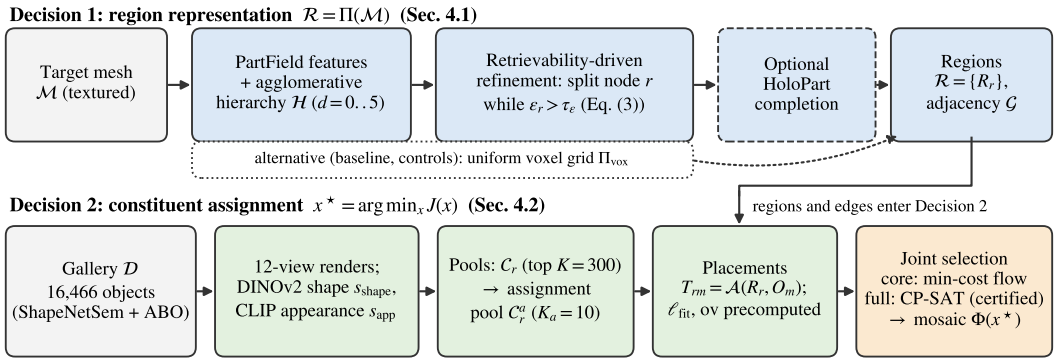

Figure 1: Overview of PARTMOSAIC3D. Decision 1 (top) builds regions from a PartField hierarchy, refined where the best retrieval is poor ($\varepsilon_r > \tau_{\text{err}}$); voxels are the baseline alternative. Decision 2 (bottom) retrieves candidates, precomputes placements and geometric costs, and selects constituents jointly by min-cost flow (core) or CP-SAT (full energy).

Proposition 2 (The baseline's reuse penalty greedily minimises the same core)**.** *The baseline assigns regions in a fixed order, each to the candidate maximising its blended score minus λ_{rep} times the number of earlier regions that chose the same object. Summed over the order, these penalties equal $\lambda_{\text{rep}} \sum_m N_m(N_m - 1)/2 = \frac{\lambda_{\text{rep}}}{2} \sum_m N_m^2 - \frac{\lambda_{\text{rep}}}{2} n$. Its output is therefore a feasible solution of the core energy of Proposition 1 with $\lambda_{\text{div}} = \lambda_{\text{rep}}/2$ and $\lambda_{\text{fit}} = 0$, obtained by a single order-dependent greedy pass; with identical pools, the min-cost flow solution has no greater energy.*

Greedy and joint selection can thus be compared on one energy; whether lower energy yields a better mosaic is RQ4. For the full objective, a solution is called optimal only when an exact solver certifies it.

3.5 CONDITIONS

The factorial crosses the two decisions: A (Π_{vox}, independent greedy), B (Π_{vox}, joint; B-core optimises the core energy and B-full the full energy), C (Π_{part}, greedy) and D (Π_{part}, joint). As region-count controls, A-m and B-m use a coarse voxel grid with $|\mathcal{R}| \approx |\mathcal{R}_{\text{C}}|$, and C-m cuts the part hierarchy into $|\mathcal{R}_{\text{A}}|$ parts. All share every other component; the deployed baseline A_0 (greedy with a top-3 geometric re-rank) is reported outside the controlled comparison (Table 5, Appendix C).

4 PARTMOSAIC3D

PARTMOSAIC3D instantiates the two decisions as a four-stage pipeline (Figure 1; Algorithm 1, Appendix B): it cuts a semantic part hierarchy into regions by retrievability-driven refinement (Section 4.1), forms candidate pools for every region, precomputes placements and geometric costs, and selects all constituents jointly with an exact solver (Section 4.2).

4.1 REGIONS FROM A PART HIERARCHY

Hierarchy. PartField (Liu et al., 2025) predicts a feature for every target face; faces are clustered agglomeratively (Ward linkage) under a face-adjacency constraint that links mesh components through a spanning tree over nearby faces. We keep the complete merge tree; depth d is its cut into 2^d clusters, so nodes at consecutive depths are nested ($d = 0, \ldots, 5$).

Regions and refinement. A node becomes a region as a voxel cell does: its faces keep their materials, untextured faces take the colour of the nearest textured surface, and its new boundary is closed by a fan of triangles to its centroid. Like a voxel cut face, this cap enters the shape embedding as real geometry but not the appearance embedding, whose colour it would falsify. A

node is replaced by its children while $\varepsilon_r > \tau_{\mathrm{err}}$, it has children, and every child covers at least the median area of a production voxel cell, with τ_{err} the median ε of the production voxel cells of the same target.

Amodal completion. HoloPart (Yang et al., 2026) completes each node of a depth-d partition into a closed part conditioned on the whole shape. Completed faces farther than 0.5% of the diagonal from the node's surface replace the cap, take the nearest textured colour and enter both embeddings.

4.2 Joint constituent selection

Scores and pools. Given the regions, the shape embedding concatenates the DINOv2 CLS token with a mask-aware mean of patch tokens that excludes the background; the appearance embedding is the CLIP image embedding with cut faces excluded. s_{shape} and s_{app} are cosine similarities (the latter clipped to $[0, 1]$), as in the baseline, whose selection rule reproduces its saved retrieval lists from our scores. $\mathcal{C}_r$ holds the $K = 300$ objects most similar in shape. Because the fit term requires a placement, the assignment pool is $\mathcal{C}_r^a = \mathrm{top}_{K_a}\{(1-\omega)s_{\mathrm{shape}} + \omega\, s_{\mathrm{app}}\} \cup \{m_r^{\mathrm{greedy}}, m_r^{\mathrm{flow}}\}$, the $K_a = 10$ candidates of $\mathcal{C}_r$ with the highest blended fidelity plus the choice of each compared selection procedure, so that all compared solutions are feasible for the same problem.

Geometric terms. Every pair in the assignment pools is placed once before selection, with the alignment and per-pair seed of the final assembly. $\ell_{\mathrm{fit}}(r, m)$ is the fraction of samples on the region's surface, excluding cut faces, that lie farther than 0.25% of the target diagonal from the placed object. ov is the interior-voxel interpenetration used for evaluation, at pitch $1/800$ of the diagonal, normalised by a unit that scales with region size so that λ_{ov} keeps its meaning across targets and decompositions. For parts, $\mathcal{E}$ connects regions whose surfaces come within 1% of the diagonal. Sampling densities and the normalisation are detailed in Appendix C.1.

Solvers and assembly. The core energy is solved exactly by min-cost flow on the full 300-candidate pools (Proposition 1). The full energy is solved with CP-SAT (Perron & Furnon, 2025): the reuse term uses integer usage counts and their squares, and each positive pairwise cost is carried by a Boolean $y \geq x_{rm} + x_{r'm'} - 1$, which is exact because the cost is non-negative and minimised; the solver is warm-started from the core optimum, and we report its status, bound and gap. Every condition is assembled by the same script, which places exactly the selected object in each region.

Assumptions. (A1) The partition is fixed before selection, so Decision 1 can be varied with Decision 2 held fixed. (A2) A placement depends only on the region–object pair, so all geometric costs are precomputed and selection is purely discrete. (A3) Gallery objects are used whole and unmodified, with bounded scaling and without real-world scale. (A4) All representations are frozen; nothing is trained on the targets or the gallery.

Design rationale. PartField is class-agnostic and hierarchical: not all targets belong to fixed segmentation label sets, and a nested hierarchy makes granularity a single controllable variable. DINOv2 and CLIP are the baseline's frozen representations, so differences between conditions are attributable to the two decisions rather than to features. HoloPart completes a part conditioned on the whole shape, as a query for a partly hidden part requires. CP-SAT returns a certified optimum or a bound, which move-making MRF solvers do not in general, so the effect of the objective is separated from that of the search; the assignment pools keep it tractable (Section 6.4 varies K_a).

5 Experimental Setup

Targets, gallery and shared components. Each research question is answered by a controlled comparison (Table 1). The four textured targets differ in part structure: contiguous parts (Tiger, Dinosaur), two separate objects (Flamingos) and many disconnected panels (Sports Car). The gallery of the deployed baseline has $16{,}466$ objects, $12{,}490$ from ShapeNetSem (Chang et al., 2015; Savva et al., 2015) and $3{,}976$ from ABO (Collins et al., 2022), mixing CAD models and scanned products. All conditions share the renderer (12 views at 224×224 pixels (Matl, 2019)), the frozen representations (DINOv2-base (Oquab et al., 2023), CLS token and mask-aware patch mean, $1{,}536$

Table 1: Research questions, the comparison that answers each (conditions of Section 3.5; E-k: experiment), and the outcome. RQ1, whether decomposition is needed at all, is outside our scope.

RQ	Question	Comparison	Outcome (section)
RQ2	Parts vs. voxels, equal granularity?	C/A-m, D/B-m (E-5)	Parts better: Cov, F, $\mathrm{CD}_{\rightarrow}$ (6.3)
RQ3	Refinement by ε vs. by size?	E-2, C-size (E-6)	No benefit at tested thresholds (6.4)
RQ4	Joint vs. greedy selection?	A, B-core, B-full (E-4)	Via fit and overlap; core marginal (6.1)
RQ5	Decisions complementary?	2×2 factorial (E-5)	Yes; D best in Cov, F, $\mathrm{CD}_{\rightarrow}$ (6.3)
RQ6	Completion, each term needed?	C-holo, term drops (E-6)	Completion hurts; both terms needed (6.4)
RQ7	Computational cost?	runtimes (E-3, E-4)	CP-SAT ≤ 213 s; HoloPart 41–49 s/part

dimensions; CLIP ViT-L/14 (Radford et al., 2021), 768 dimensions), the embeddings and the placement procedure (PCA and ICP (Besl & McKay, 1992) with a bounded scale search).

Uniform voxels (A, B). Each target is divided by a regular grid over its bounding box (for Dinosaur's production grid, its oriented bounding box), and every occupied cell is a region (710, 549, 193 and 1,080 regions). Retrieval uses the deployed baseline's $K = 300$ and $\omega = 0.6$, and greedy selection its reuse penalty $\lambda_{\mathrm{rep}} = 0.15$, i.e. $\lambda_{\mathrm{div}} = 0.075$ (Proposition 2), which all joint conditions share.

Parts (C, D) and controls. C and D use the refinement frontier of Section 4.1 (32, 32, 22 and 30 regions); A-m and B-m use the coarse grid whose occupied-cell count is closest (26, 33, 17, 27 cells), and C-m cuts the PartField tree into as many parts as the production grid has regions. The rule was fixed before these runs with an additional condition—split only if the children are more retrievable—that leaves Dinosaur and Flamingos as a single region; it was removed before any part-based mosaic was assembled.

Weights. $(\lambda_{\mathrm{ov}}, \lambda_{\mathrm{fit}})$ was explored on Tiger only, on $\{0.25, 1, 4\} \times \{0, 0.25, 1\}$ (Appendix D.3); $(4, 0.25)$ was then fixed, with $(4, 0)$ as a secondary setting, for all other targets and conditions; Tiger is therefore not held out. τ_{err} was fixed after the retrievability results of Section 6.2, and the bracketing grids of Section 6.3 and the ablations of Section 6.4 were added after the first part-based results.

Metrics. Metrics are computed on the assembled mosaic against the original target (Appendix C.2): the Chamfer distance per direction, $\mathrm{CD}_{\rightarrow}$ (completeness) and $\mathrm{CD}_{\leftarrow}$ (accuracy; large values indicate protrusion), in % of the target diagonal; coverage and F-score at $\delta = 1\%$; interpenetration, the excess volume of lattice voxels inside two or more constituents, a lower bound validated in Appendix F; and the unique-object ratio. None is a term of the energy, but the overlap cost uses the interpenetration measure at a finer pitch ($1/800$ vs. $1/400$), and ℓ_{fit} is related to coverage.

6 Results

We test joint selection on the baseline's voxels, examine what governs retrievability, cross the two decisions, and ablate components; full tables are in Appendix D.

6.1 Joint selection on uniform voxels (E-4)

Baseline mosaics cover the target (coverage ≥ 0.959) but protrude and interpenetrate in every run (Appendix D.1). Table 2 compares greedy selection (A), the core optimum (B-core) and the full-energy optimum (B-full; certified within the assignment pools). The core optimum changes 61–85% of the constituents yet the surface metrics only marginally (by at most 0.04 percentage points in either Chamfer direction and 0.005 in F-score), and it interpenetrates *more* than greedy selection. The full energy improves coverage, F-score and both Chamfer directions on every target—relative to the baseline A, F-score by 0.009–0.029 and $\mathrm{CD}_{\leftarrow}$ by 0.023–0.110 points—and reduces the interpenetration volume by 30–54% on Tiger, Dinosaur and Sports Car; on Flamingos, whose interpenetration is close to the lattice resolution, it does not (Section 7). The contrast is the main finding: retrieval

Table 2: Selection on the production voxel grids. B-full: full-energy optimum within the assignment pools at $(\lambda_{\text{ov}}, \lambda_{\text{fit}}) = (4, 0.25)$, weights chosen on Tiger (not held out). F at 1%, $\text{CD}_{\leftarrow}$ in % of the diagonal; Int.: interpenetration, % of summed interior volume at pitch $1/400$ (lower bound). Coverage, $\text{CD}_{\rightarrow}$ and pairs: Tables 9, 10.

	Tiger			Dinosaur			Flamingos			Sports Car		
Selection	F	$\text{CD}_{\leftarrow}$	Int.	F	$\text{CD}_{\leftarrow}$	Int.	F	$\text{CD}_{\leftarrow}$	Int.	F	$\text{CD}_{\leftarrow}$	Int.
A: greedy	0.609	1.714	6.3	0.698	1.384	7.6	0.926	0.487	**0.8**	0.942	0.494	2.3
B-core: flow	0.605	1.689	9.7	0.693	1.346	11.1	0.929	0.482	4.6	0.941	0.499	3.1
B-full	**0.626**	**1.636**	**3.5**	**0.727**	**1.274**	**3.5**	**0.935**	**0.452**	2.0	**0.951**	**0.471**	**1.6**

Table 3: Decomposition $\times$ selection at matched granularity. Regions (voxel/part): Tiger 26/32, Dinosaur 33/32, Flamingos 17/22, Sports Car 27/30. Joint: full-energy optimum within the assignment pools at $(4, 0.25)$, certified by CP-SAT. Cov, F at 1%; CD in % of the diagonal.

	Tiger				Dinosaur				Flamingos				Sports Car			
Condition	Cov	F	$\text{CD}_{\rightarrow}$	$\text{CD}_{\leftarrow}$	Cov	F	$\text{CD}_{\rightarrow}$	$\text{CD}_{\leftarrow}$	Cov	F	$\text{CD}_{\rightarrow}$	$\text{CD}_{\leftarrow}$	Cov	F	$\text{CD}_{\rightarrow}$	$\text{CD}_{\leftarrow}$
A-m (voxel, greedy)	0.629	0.481	0.935	1.708	0.703	0.571	0.782	1.425	0.527	0.560	1.462	1.178	0.567	0.574	1.098	1.182
C (parts, greedy)	0.722	0.562	0.815	1.563	0.747	0.619	0.698	1.231	0.805	0.796	0.656	**0.662**	0.699	0.633	0.897	1.258
B-m (voxel, joint)	0.784	0.573	0.667	1.714	0.805	0.665	0.609	1.224	0.621	0.586	1.090	1.348	0.669	0.667	0.886	0.956
D (parts, joint)	**0.798**	**0.657**	**0.648**	**1.199**	**0.845**	**0.714**	**0.568**	**0.992**	**0.918**	**0.817**	**0.413**	0.717	**0.725**	**0.689**	**0.835**	**0.950**

scores are computed per region in isolation and do not see how the placed object fits or collides, so optimising them exactly changes *which* objects are used without improving the mosaic, whereas the geometric terms supply this information.

6.2 Part hierarchy and retrievability (E-1–E-3)

PartField parts are nearly contiguous on Tiger and Dinosaur, span both birds on Flamingos, and are fragmented on Sports Car (1,365 separate panels). Over all hierarchy nodes and five voxel grids (Appendix D.5), retrievability is governed largely by size: ε grows with area for voxel cells on three targets (Spearman $\rho = 0.48$–0.66), and production cells, 5–34 times smaller than depth-5 parts, are the most retrievable regions everywhere. At equal area parts are not more retrievable than voxels (part effect $+0.021$ and $+0.057$ on Tiger and Dinosaur, none on the others), and where refinement by ε splits at all it selects nearly the frontier of a size-only rule on Tiger and Flamingos (ARI 0.74–0.99) but not on Sports Car (0.11–0.54); on Dinosaur the tested thresholds lie above the root's ε, so no refinement occurs. HoloPart makes regions *less* retrievable on Tiger, Flamingos and Sports Car (mean $\Delta\varepsilon = +0.023$ to $+0.048$; Appendix D.6). Any advantage of parts must therefore come from elsewhere than retrievability.

6.3 Decomposition and selection (E-5)

Parts versus voxels (RQ2). At matched granularity (Table 3), parts give higher coverage, higher F-score and lower $\text{CD}_{\rightarrow}$ than voxels on all four targets under both selection rules; $\text{CD}_{\leftarrow}$ favours parts except on Sports Car, whose parts are fragmented. Interpolated to equal region counts between the matched grid and the next grid, parts retain higher coverage and F-score on every target (e.g. Tiger 0.722 vs. 0.668 and 0.562 vs. 0.498), although a 55-cell Tiger grid exceeds the 32 parts in coverage. The coverage gain of parts under greedy selection ranges from 0.044 (Dinosaur) to 0.278 (Flamingos); since parts are not more retrievable than equally sized cells, it is best explained by where boundaries fall.

Joint versus greedy selection (RQ4, RQ5). Joint selection improves coverage and F-score for both decompositions on all four targets, and D is the best cell in coverage, F-score and $\text{CD}_{\rightarrow}$ everywhere; with 17–33 regions interpenetration is negligible (at most 10 pairs), so the gain comes from fit and fidelity. The gains overlap: joint selection adds 0.026–0.113 coverage on parts and 0.094–0.155 on voxels.

Production granularity. With as many parts as the production grid has cells (C-m, Table 13), voxels cover slightly more of Tiger and Dinosaur, but parts remove most of the protrusion: $\mathrm{CD}_{\leftarrow}$ falls from 1.71% and 1.38% to 0.30% and 0.32%, and the F-score rises from 0.61–0.70 to 0.98; Flamingos shows the same trade-off more weakly. On Sports Car, whose fine parts are scattered panel fragments, voxels are better on every surface metric (F-score 0.942 vs. 0.890): parts help where their boundaries follow the object's structure.

6.4 Ablations (E-6)

A size-only frontier matches the refinement frontier on Tiger and Dinosaur, is slightly worse on Flamingos and better on Sports Car: no benefit of retrievability over size. HoloPart-completed regions lower coverage on Tiger, Flamingos and Sports Car. On Tiger, enlarging K_a lowers the energy almost entirely through fit, without the overlap term the surface gains remain but interpenetration exceeds greedy selection, and the overlap term alone lowers it without the surface gains (Appendix D.8).

7 Discussion and Limitations

The results locate the benefit of PartMosaic3D in where region boundaries fall and in the geometric selection terms, not in retrieval scores. Parts are not more retrievable than equally sized cells, yet give better mosaics and remove most protrusion when contiguous, and the fragmented target reverses this—supporting a boundary explanation. Likewise, exact optimisation of fidelity and reuse changes most constituents but not the mosaic, whereas the fit and overlap terms improve it: per-region similarity is a weak proxy for how a whole object fills its region, so selection must see placements. The refinement criterion fails because retrievability is governed by size, so thresholding it mostly reproduces size-driven subdivision. There are limits. Interpenetration is a lower bound, blind to constituents thinner than two voxels (up to 73% of fine parts), and as a single-pitch cost unreliable near the lattice resolution (Flamingos). Each condition ran once on four targets with weights chosen on Tiger; the held-out targets reproduce the Tiger findings (consistency, not significance). RQ1, perceptual quality, scale consistency, a decomposition-free baseline and learned 3D encoders are not evaluated.

8 Conclusion

We cast 3D object mosaic synthesis as two coupled decisions and presented PartMosaic3D, which makes both explicit: regions come from a part hierarchy refined by retrievability, and constituents are chosen jointly under an energy with an exact min-cost-flow core and certified full-energy optima within the candidate pools. In answer to the research questions, parts beat voxels at matched granularity (RQ2) and complement joint selection (RQ5), which helps through its fit and overlap terms rather than exact optimisation of fidelity and reuse (RQ4); refinement by retrievability (RQ3) and completion (RQ6) showed no benefit, both geometric terms are needed (RQ6), and every full-energy problem at $K_a = 10$ was certified optimal in at most 213 s on the held-out targets (RQ7).

Future directions. The findings suggest four concrete directions. (1) Since boundaries, not retrievability, drive the gain of parts, decompositions could be optimised for how whole objects fill regions, e.g. with the fit term as refinement criterion. (2) A multi-pitch or continuous interpenetration measure would remove the lattice-resolution limit on thin constituents. (3) Larger assignment pools with joint optimisation of placements would relax the pool restriction and assumption (A2), and verified real-world dimensions (Appendix E) would allow scale-consistent retrieval. (4) A perceptual study, more targets with repeated runs and a decomposition-free baseline would test how far the conclusions generalise.

Use of Generative AI (required)

An AI assistant was used extensively. It implemented and ran all experiments and analysed and interpreted the results. It contributed to the experimental design, proposing the assignment pools, the geometric-term settings, the controls and the ablations, and selecting the weight grid and the fixed

weights $(4, 0.25)$; the authors chose the refinement threshold and the refinement-rule change from options it proposed. It also drafted and edited the text, checked the formal statements, performed the novelty literature search and audited the reported numbers. The authors take full responsibility for the content of this paper.

Ethics Statement

The gallery combines ShapeNetSem (Savva et al., 2015) (ShapeNet terms of use: non-commercial research and education only) and Amazon Berkeley Objects (Collins et al., 2022) (CC BY 4.0, credit: Amazon.com), both from their official releases; the four target meshes are from Sketchfab and are used under their creators' licences. No personal data are involved. As with other creative tools, outputs could reproduce protected designs present in the gallery.

Reproducibility Statement

Settings, region counts and solver parameters are given in Section 4, Section 5 and Appendix C, the interpenetration measure and its validation in Appendix F, and the pre-registered choices and the one deviation in Section 5. Every experiment stores its configuration, solver status and bounds, runtimes and metrics; sampling and HoloPart use fixed seeds, alignment is seeded per region–object pair, gallery embeddings are identified by a configuration signature, and some HoloPart runtimes were recovered from logs. Code and configurations will be released upon publication. ShapeNetSem must be obtained from ShapeNet (`https://huggingface.co/datasets/ShapeNet/ShapeNetSem-archive`); ABO is public (`https://amazon-berkeley-objects.s3.us-east-1.amazonaws.com/index.html`); the Sketchfab targets will be referenced by link.

References

Ahmed Abdelreheem, Ivan Skorokhodov, Maks Ovsjanikov, and Peter Wonka. SATR: Zero-shot semantic segmentation of 3D shapes. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15120–15133, 2023. doi: 10.1109/ICCV51070.2023.01392.

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

Sebastiano Battiato, Gianpiero Di Blasi, Giovanni Maria Farinella, and Giovanni Gallo. Digital mosaic frameworks – an overview. *Computer Graphics Forum*, 26(4):794–812, 2007. doi: 10.1111/j.1467-8659.2007.01021.x.

Paul J Besl and Neil D McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 14(2):239–256, 1992.

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Siddhartha Chaudhuri, Evangelos Kalogerakis, Leonidas Guibas, and Vladlen Koltun. Probabilistic reasoning for assembly-based 3D modeling. *ACM Transactions on Graphics*, 30(4), 2011.

Minghao Chen, Roman Shapovalov, Iro Laina, Tom Monnier, Jianyuan Wang, David Novotny, and Andrea Vedaldi. PartGen: Part-level 3D generation and reconstruction with multi-view diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5881–5892, 2025. doi: 10.1109/CVPR52734.2025.00552.

Xuelin Chen, Hao Zhang, Jinjie Lin, Ruizhen Hu, Lin Lu, Qi-Xing Huang, Bedrich Benes, Daniel Cohen-Or, and Baoquan Chen. Dapper: Decompose-and-pack for 3D printing. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 34(6):213:1–213:12, 2015. doi: 10.1145/2816795.2818087.

David Cohen-Steiner, Pierre Alliez, and Mathieu Desbrun. Variational shape approximation. *ACM Transactions on Graphics*, 23(3):905–914, 2004.

Jasmine Collins, Shubham Goel, Ken Deng, Angela Luan, Christine Xu, Erhan Zheng, Abhinav Shrivastava, Angjoo Kanazawa, and Jitendra Malik. Abo: Dataset and benchmarks for real-world 3d object understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21126–21136, 2022.

Qiaodong Cui, Victor Rong, Desai Chen, and Wojciech Matusik. Dense, interlocking-free and scalable spectral packing of generic 3D objects. *ACM Transactions on Graphics*, 42(4):1–14, 2023. doi: 10.1145/3592126.

Soumyaratna Debnath, Ashish Tiwari, Kaustubh Sadekar, and Shanmuganathan Raman. RASP: Revisiting 3D anamorphic art for shadow-guided packing of irregular objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5849–5858, 2025. doi: 10.1109/CVPR52734.2025.00549.

Andrew Delong, Anton Osokin, Hossam N. Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 96(1):1–27, 2012.

Gianpiero Di Blasi and Giovanni Gallo. Artificial mosaics. *The Visual Computer*, 21(6):373–383, 2005. doi: 10.1007/s00371-005-0292-4.

Adam Finkelstein and Marisa Range. Image mosaics. In *Electronic Publishing, Artistic Imaging, and Digital Typography (RIDT)*, pp. 11–22. Springer, 1998.

Thomas Funkhouser, Michael Kazhdan, Philip Shilane, Patrick Min, William Kiefer, Ayellet Tal, Szymon Rusinkiewicz, and David Dobkin. Modeling by example. *ACM Transactions on Graphics*, 23(3):652–663, 2004.

Ran Gal, Olga Sorkine, Tiberiu Popa, Alla Sheffer, and Daniel Cohen-Or. 3d collage: expressive non-photorealistic modeling. In *Proceedings of the 5th international symposium on Non-photorealistic animation and rendering (NPAR)*, pp. 7–14, 2007.

Harvey J. Greenberg. A quadratic assignment problem without column constraints. *Naval Research Logistics Quarterly*, 16(3):417–421, 1969.

Alejo Hausner. Simulating decorative mosaics. In *Proceedings of ACM SIGGRAPH*, pp. 573–580, 2001.

Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Thorben Kröger, Jan Lellmann, Nikos Komodakis, Bogdan Savchynskyy, and Carsten Rother. A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision*, 115(2):155–184, 2015.

Junhwan Kim and Fabio Pellacini. Jigsaw image mosaics. *ACM Transactions on Graphics*, 21(3): 657–664, 2002.

Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

Harold W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955. doi: 10.1002/nav.3800020109.

Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick. Graphcut textures: Image and video synthesis using graph cuts. In *ACM SIGGRAPH 2003 Papers*, pp. 277–286, 2003. doi: 10.1145/1201775.882264.

Yuchen Lin, Chenguo Lin, Panwang Pan, Honglei Yan, Yiqiang Feng, Yadong Mu, and Katerina Fragkiadaki. PartCrafter: Structured 3D mesh generation via compositional latent diffusion transformers. *arXiv preprint arXiv:2506.05573*, 2025. doi: 10.48550/arXiv.2506.05573.

Minghua Liu, Ruoxi Shi, Kaiming Kuang, Yinhao Zhu, Xuanlin Li, Shizhong Han, Hong Cai, Fatih Porikli, and Hao Su. OpenShape: Scaling up 3D shape representation towards open-world understanding. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 44860–44879, 2023a.

Minghua Liu, Yinhao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. PartSLIP: Low-shot part segmentation for 3D point clouds via pretrained image-language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21736–21746, 2023b. doi: 10.1109/CVPR52729.2023.02082.

Minghua Liu, Mikaela Angelina Uy, Donglai Xiang, Hao Su, Sanja Fidler, Nicholas Sharp, and Jun Gao. PartField: Learning 3D feature fields for part segmentation and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. arXiv:2504.11451.

Sheng-Jie Luo, Yonghao Yue, Chun-Kai Huang, Yu-Huan Chung, Sei Imai, Tomoyuki Nishita, and Bing-Yu Chen. Legolization: Optimizing LEGO designs. *ACM Transactions on Graphics*, 34(6), 2015.

Matthew Matl. Pyrender: Fast and light offscreen rendering for python. `https://github.com/mmatl/pyrender`, 2019.

Paul Merrell, Eric Schkufza, Zeyang Li, Maneesh Agrawala, and Vladlen Koltun. Interactive furniture layout using interior design guidelines. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 30(4):87:1–87:10, 2011. doi: 10.1145/2010324.1964982.

Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subhashree Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 909–918, 2019.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Despoina Paschalidou, Ali Osman Ulusoy, and Andreas Geiger. Superquadrics revisited: Learning 3D shape parsing beyond cuboids. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10336–10345, 2019. doi: 10.1109/CVPR.2019.01059.

Despoina Paschalidou, Angelos Katharopoulos, Andreas Geiger, and Sanja Fidler. Neural parts: Learning expressive 3D shape abstractions with invertible neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3203–3214, 2021. doi: 10.1109/CVPR46437.2021.00322.

D. Pavić, U. Ceumern, and L. Kobbelt. GIzMOs: Genuine image mosaics with adaptive tiling. *Computer Graphics Forum*, 28(8):2244–2254, 2009. doi: 10.1111/j.1467-8659.2009.01437.x.

Laurent Perron and Vincent Furnon. OR-Tools. `https://developers.google.com/optimization/`, 2025. Version 9.15.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pp. 8748–8763. PMLR, 2021.

Carsten Rother, Sanjiv Kumar, Vladimir Kolmogorov, and Andrew Blake. Digital tapestry. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

Carsten Rother, Lucas Bordeaux, Youssef Hamadi, and Andrew Blake. Autocollage. *ACM Transactions on Graphics*, 25(3):847–852, 2006.

Manolis Savva, Angel X Chang, and Pat Hanrahan. Semantically-enriched 3d models for common-sense knowledge. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 24–31, 2015.

Ariel Shamir. A survey on mesh segmentation techniques. *Computer Graphics Forum*, 27(6):1539–1556, 2008.

Robert Silvers and Michael Hawley. *Photomosaics*. Henry Holt and Company, 1997.

Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 945–953, 2015.

Chun-Yu Sun, Qian-Fang Zou, Xin Tong, and Yang Liu. Learning adaptive hierarchical cuboid abstractions of 3D shape collections. *ACM Transactions on Graphics*, 38(6):1–13, 2019. doi: 10.1145/3355089.3356529.

Minhyuk Sung, Hao Su, Vladimir G. Kim, Siddhartha Chaudhuri, and Leonidas Guibas. ComplementMe: Weakly-supervised component suggestions for 3D modeling. *ACM Transactions on Graphics*, 36(6), 2017.

Shubham Tulsiani, Hao Su, Leonidas J. Guibas, Alexei A. Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1466–1474, 2017. doi: 10.1109/CVPR.2017.160.

Zhenwei Wang, Tengfei Wang, Zexin He, Gerhard Hancke, Ziwei Liu, and Rynson W. H. Lau. Phidias: A generative model for creating 3D content from text, image, and 3D conditions with reference-augmented diffusion. *arXiv preprint arXiv:2409.11406*, 2024. doi: 10.48550/arXiv.2409.11406.

Xinyue Wei, Minghua Liu, Zhan Ling, and Hao Su. Approximate convex decomposition for 3D meshes with collision-aware concavity and tree search. *ACM Transactions on Graphics*, 41(4):1–18, 2022. doi: 10.1145/3528223.3530103.

Pengfei Xu, Jianqiang Ding, Hao Zhang, and Hui Huang. Discernible image mosaic with edge-aware adaptive tiles. *Computational Visual Media*, 5(1):45–58, 2019. doi: 10.1007/s41095-019-0130-7.

Xianghao Xu, Paul Guerrero, Matthew Fisher, Siddhartha Chaudhuri, and Daniel Ritchie. Unsupervised 3D shape reconstruction by part retrieval and assembly. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Le Xue, Mingfei Gao, Chen Xing, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, and Silvio Savarese. ULIP: Learning a unified representation of language, images, and point clouds for 3D understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1179–1189, 2023a. doi: 10.1109/CVPR52729.2023.00120.

Tianyang Xue, Mingdong Wu, Lin Lu, Haoxuan Wang, Hao Dong, and Baoquan Chen. Learning gradient fields for scalable and generalizable irregular packing. In *SIGGRAPH Asia 2023 Conference Papers*, pp. 1–11, 2023b. doi: 10.1145/3610548.3618235.

Yunhan Yang, Yukun Huang, Yuan-Chen Guo, Liangjun Lu, Xiaoyang Wu, Edmund Y. Lam, Yan-Pei Cao, and Xihui Liu. SAMPart3D: Segment any part in 3D objects. *arXiv preprint arXiv:2411.07184*, 2024.

Yunhan Yang, Yufan Zhou, Yuan-Chen Guo, Zi-Xin Zou, Yukun Huang, Ying-Tian Liu, Hao Xu, Ding Liang, Yan-Pei Cao, and Xihui Liu. OmniPart: Part-aware 3D generation with semantic decoupling and structural cohesion. *arXiv preprint arXiv:2507.06165*, 2025. doi: 10.48550/arXiv.2507.06165.

Yunhan Yang, Yuan-Chen Guo, Yukun Huang, Zi-Xin Zou, Zhipeng Yu, Yangguang Li, Yan-Pei Cao, and Xihui Liu. HoloPart: Generative 3D part amodal segmentation. In *International Conference on Learning Representations (ICLR)*, 2026. arXiv:2504.07943.

Li Yi, Vladimir G Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016.

A NOTATION

Table 4 summarises the notation of the main text.

Table 4: Notation.

Symbol	Meaning
$\mathcal{M}, \mathcal{S}$	target mesh and its surface
$\mathcal{D} = \{O_m\}$	gallery of $\lvert\mathcal{D}\rvert = 16{,}466$ objects
$\Pi, \mathcal{R} = \{R_r\}$	decomposition operator and region set ($n = \lvert\mathcal{R}\rvert$)
P_r	query geometry of region r
$s_{\text{shape}}, s_{\text{app}}$	shape and appearance similarities; ω appearance weight
$\mathcal{C}_r, \mathcal{C}_r^a$	candidate pool ($K = 300$) and assignment pool ($K_a = 10$)
$\varepsilon_r, \tau_{\text{err}}$	retrievability error and refinement threshold
$T_{rm} = \mathcal{A}(R_r, O_m)$	placement of object m in region r
x_{rm}, N_m	selection variables and usage count of object m
$\mathcal{G} = (\mathcal{R}, \mathcal{E})$	region adjacency graph
$u_{rm}, \ell_{\text{fit}}$, ov	unary cost, fit loss, pairwise interpenetration
$\lambda_{\text{div}}, \lambda_{\text{fit}}, \lambda_{\text{ov}}, \lambda_{\text{rep}}$	reuse, fit and overlap weights; baseline reuse penalty
J	joint constituent-selection energy, Eq. (4)

B ALGORITHM

Algorithm 1 summarises the complete pipeline of Section 4; Figure 1 shows the same stages schematically.

Algorithm 1 PARTMOSAIC3D: region construction, joint selection and assembly

Require: target $\mathcal{M}$, gallery $\mathcal{D}$, weights $(\lambda_{\text{div}}, \lambda_{\text{fit}}, \lambda_{\text{ov}})$, pool sizes K, K_a, threshold τ_{err}
1: $\mathcal{H} \leftarrow$ PartField hierarchy of $\mathcal{M}$ (depths $0, \ldots, 5$); compute ε_r for every node (Eq. (2))
2: $\mathcal{R} \leftarrow \{\text{root}\}$; **while** some $r \in \mathcal{R}$ has $\varepsilon_r > \tau_{\text{err}}$ and children of sufficient area **do** replace r by its children
3: close each region's new boundary by a cap, or (optional) complete the region with HoloPart
4: **for** $r \in \mathcal{R}$ **do**
5: render P_r (12 views); $\mathcal{C}_r \leftarrow$ top-K by s_{shape}; $\mathcal{C}_r^a \leftarrow$ top-K_a by blended fidelity $\cup \{m_r^{\text{greedy}}, m_r^{\text{flow}}\}$
6: **for** $m \in \mathcal{C}_r^a$ **do** $T_{rm} \leftarrow \mathcal{A}(R_r, O_m)$ with a per-pair seed; compute u_{rm} and $\ell_{\text{fit}}(r, m)$
7: **end for**
8: **end for**
9: compute $\text{ov}(T_{rm}O_m, T_{r'm'}O_{m'})$ for all $(r, r') \in \mathcal{E}$, $m \in \mathcal{C}_r^a$, $m' \in \mathcal{C}_{r'}^a$
10: $x^{\text{core}} \leftarrow$ min-cost flow on the core energy (Proposition 1)
11: $x^\star \leftarrow$ CP-SAT on J, warm-started by ICM from x^{core}; record status, bound and gap
12: **return** mosaic $\Phi(x^\star)$, assembling each selected O_m at T_{rm}

C ADDITIONAL SETUP

Table 5 lists the evaluated conditions of Section 3.5.

Table 5: Evaluated conditions. The factorial A–D varies only the decomposition operator and the selection rule. A_0 is reported outside the controlled comparison.

Condition	Decision 1 (regions)	Decision 2 (identities)	Role
A_0	Π_{vox}	greedy + top-3 geometric re-rank	deployed baseline
A	Π_{vox}	independent greedy	factorial
B	Π_{vox}	joint	factorial
C	Π_{part}	independent greedy	factorial
D	Π_{part}	joint	factorial
A-m, B-m	coarse Π_{vox}, $\lvert\mathcal{R}\rvert \approx \lvert\mathcal{R}_{\text{C}}\rvert$	as A, B	region-count controls
C-m	Π_{part} cut, $\lvert\mathcal{R}\rvert = \lvert\mathcal{R}_{\text{A}}\rvert$	as C	region-count control

Table 6 lists the production uniform-voxel grid of each target.

Table 6: Uniform-voxel regions per target. Embeddings: the cached retrieval run whose chunk and gallery embeddings all conditions reuse.

Target	Grid	$\lvert\mathcal{R}\rvert$	Gallery
Tiger	$12 \times 12 \times 12$	710	full (16,466)
Dinosaur	$10 \times 14 \times 21$	549	full
Flamingos	$12 \times 12 \times 12$	193	full
Sports Car	$12 \times 12 \times 12$	1,080	full

C.1 Implementation details

Geometric terms. Since $T_{rm} = \mathcal{A}(R_r, O_m)$ depends only on the pair, every pair in the assignment pools is placed once before selection, with the alignment and per-pair seed of the final assembly. $\ell_{\mathrm{fit}}(r, m)$ is the fraction of 2,000 samples on the region's surface (excluding cut faces and, on Sports Car, the untextured original faces that the chunker assigns the same material) farther than 0.25% of the target diagonal from 20,000 samples on the placed object. ov is the interpenetration of two placed objects under the interior-voxel measure used for evaluation, at pitch $1/800$ of the diagonal, divided by $V_{\mathrm{bbox}}\rho/|\mathcal{R}|$ (ρ the occupied fraction of the production grid), which scales with region size and equals the production cell volume to within 0.2% (except on Dinosaur's OBB-aligned grid, $1.18\times$; runs on that grid use the cell volume), so that λ_{ov} keeps its meaning across targets and decompositions. For parts, $\mathcal{E}$ connects regions whose surfaces come within 1% of the target diagonal.

Solvers and assembly. The core energy is solved exactly by min-cost flow on the full 300-candidate pools (Proposition 1). The full energy is solved with CP-SAT (Perron & Furnon, 2025): one-hot constraints are exact, the reuse term uses integer usage counts and their squares, and each positive pairwise cost is carried by a Boolean $y \geq x_{rm} + x_{r'm'} - 1$, which is exact because the cost is non-negative and minimised. Costs are scaled to integers at resolution 10^{-5} and energies are recomputed in floating point; the solver is warm-started with iterated conditional modes from the core optimum, and we report its status, bound and gap. Every condition is assembled by the same script with a single candidate per region, so the placed object is the selected one, and alignment is seeded per pair, so a placement does not depend on processing order.

C.2 Metrics

All metrics are computed on the assembled mosaic against the original target and are not terms of the energy, but two are related to optimised terms: the overlap cost uses the interpenetration measure below at a finer pitch ($1/800$ instead of $1/400$), and ℓ_{fit} is a per-region target-surface distance related to coverage. With $d_{\rightarrow}$ the distance from a point on $\mathcal{S}$ to the mosaic surface and $d_{\leftarrow}$ the reverse (2×10^5 area-weighted samples, as fractions of the target's bounding-box diagonal), we report the Chamfer distance per direction, $\mathrm{CD}_{\rightarrow} = \mathbb{E}[d_{\rightarrow}]$ (completeness) and $\mathrm{CD}_{\leftarrow} = \mathbb{E}[d_{\leftarrow}]$ (accuracy; large values indicate protrusion beyond the target); coverage at δ, the fraction of target samples with $d_{\rightarrow} \leq \delta$; and the F-score at δ, combining coverage with the fraction of mosaic samples within δ of the target ($\delta = 1\%$). Interpenetration is the excess volume $\sum_v (k_v - 1)$ over lattice voxels v that lie inside $k_v \geq 2$ constituents, where a voxel counts as inside a constituent only if it is filled and not crossed by that constituent's own surface, so touching constituents contribute nothing; it is a lower bound, blind to constituents thinner than about two voxels, and is validated in Appendix F. Diversity is the unique-object ratio $\mathrm{UOR} = |\{m_r\}|/|\mathcal{R}|$.

C.3 Decomposition-free placement (SPG, not evaluated)

The surface-point greedy baseline (SPG) has no operator Π. It samples points on $\mathcal{S}$ and, around each, sphere neighbourhoods of several radii (Gal et al., 2007); each neighbourhood p receives a candidate pool $\mathcal{C}_p$ computed with the same representations as above. The candidates are pairs (p, m) with indicator variables $y_{pm} \in \{0, 1\}$ and *no* one-per-region constraint: which neighbourhoods

are used at all is part of the decision. SPG adds pairs greedily, subject to overlap and visibility constraints, until a target surface-coverage level is reached. The corresponding joint problem is not a labelling problem over a fixed set of variables; if the requirement is to cover all surface samples at minimum cost, it contains weighted set cover as a special case. This structural difference is the formulation-level contrast between SPG and all Π-based conditions. We evaluate only the greedy form.

D ADDITIONAL RESULTS

This appendix gives the full tables and per-target details that are summarised in Section 6.

D.1 DIAGNOSTICS OF EXISTING UNIFORM-VOXEL MOSAICS (E-0)

Before running the controlled conditions, we measured the hypotheses of Section 1 on nine mosaics that the uniform-voxel baseline had already produced for the four targets. These runs were not generated under one configuration: retrieval settings and assembly versions differ between them and are not recoverable for two of the Dinosaur runs, and every Sports Car run that keeps constituents as separate objects was assembled with a containment step that confines each constituent to its region's oriented bounding box. They are therefore descriptive of the existing baseline outputs and are not part of the controlled comparison; conditions A_0 and A are re-run under a single configuration (Section 6.1). Table 7 lists the runs.

Table 7: Diagnostics of existing uniform-voxel mosaics. rep: sequential reuse penalty λ_{rep}; w: appearance weight; "?": not recoverable; C: assembled with containment. UOR: unique-object ratio; max: most regions filled by one object. Cov and F at $\delta = 1\%$ of the target diagonal. CD in % of the diagonal. Int.: interpenetration volume in % of the summed interior volume of the constituents, at lattice pitch $1/400$ of the diagonal (a lower bound; Appendix F); inv.: % of constituents too thin to have an interior voxel at that pitch, which the measure cannot see.

Target	Retrieval (w, rep)	$\lvert\mathcal{R}\rvert$	UOR	max	Cov	F	$CD_{\rightarrow}$	$CD_{\leftarrow}$	Int.	inv.
Tiger	0.6, 0.15	710	0.63	6	0.999	0.60	0.30	1.73	3.7	42
Dinosaur	?, ?	549	0.66	7	0.966	0.68	0.37	1.37	5.8	54
Dinosaur	?, ?	616	0.17	161	0.999	0.73	0.25	1.16	7.3	29
Dinosaur	?, ?	616	0.16	134	0.999	0.75	0.25	1.11	5.9	40
Flamingos	1.0, 0.3	193	0.75	7	0.988	0.94	0.26	0.44	1.7	56
Flamingos	1.0, 0	193	0.32	44	0.997	0.92	0.23	0.50	11.1	42
Flamingos	0.35, 0	193	0.31	56	0.998	0.96	0.22	0.38	0.8	41
Sports Car	0.35, 0.3 (C)	1074	0.54	10	0.959	0.95	0.41	0.44	0.8	68
Sports Car	0.35, 0 (C)	1074	0.07	246	0.967	0.95	0.39	0.43	0.3	68

Reuse. Constituent repetition is large whenever the reuse penalty is off. With $\lambda_{\text{rep}} = 0$, the unique-object ratio is between 0.07 and 0.32, and a single gallery object fills up to 246 of the $1{,}074$ Sports Car regions. With $\lambda_{\text{rep}} \in \{0.15, 0.3\}$ it is between 0.54 and 0.75, and no object fills more than 10 regions. For Flamingos (appearance weight 1.0) and Sports Car, the paired runs share the chunks, candidate pool and appearance weight and differ only in λ_{rep}; the other differences in the table are confounded with target and configuration.

Coverage. Uncovered target surface is small in every run: at least 95.9% of the target surface lies within 1% of the diagonal of the mosaic, and at least 99.8% within 2%. Gaps, one of the hypothesised failure modes, are not observed at this scale in these outputs.

Protrusion. For Tiger and Dinosaur, $CD_{\leftarrow}$ is 3.7–5.8 times $CD_{\rightarrow}$ (1.73% vs. 0.30% for Tiger), and the F-score at 1% is 0.60–0.75 although coverage is at least 0.966: the mosaic contains surface that lies outside the target. The ratio is 1.7–2.2 for Flamingos and 1.1 for Sports Car. This asymmetry is consistent with constituents extending beyond their regions, but it does not measure interpenetration between constituents.

Interpenetration. With the interior-voxel measure of Appendix F, interpenetration is non-zero in every run, including both Sports Car runs assembled with containment (36 and 15 interpenetrating pairs). As a volume it is between 0.3% and 11.1% of the constituents' summed interior volume, and it involves up to 45% of the placed constituents in a run. Between 29% and 68% of constituents are too thin for the measure at this pitch, so these values are lower bounds and are not comparable across targets. Containment to each region's oriented bounding box therefore does not exclude interpenetration, since the boxes of neighbouring regions can intersect.

Which pairs interpenetrate. Across the nine runs, 820 pairs of constituents interpenetrate. Of these, 644 (78.5%) occupy face-adjacent cells and 808 (98.5%) occupy cells that share at least a vertex. A pairwise overlap cost defined on face adjacency alone would leave about a fifth of the observed interpenetrating pairs uncharged; we therefore define voxel adjacency by the 26-neighbourhood (Section 3), which contains all but 12 of the 820 pairs.

Fragmentation. In every run the occupied cells form a single face-connected component. Whether voxel boundaries cross semantic part boundaries requires part annotations and is examined with the part hierarchy (Section 6.2).

D.2 Joint and sequential selection on the core energy (E-4, Tiger)

We compare the baseline's sequential selection with the exact solution of the fidelity-and-reuse core (Proposition 1) on Tiger's 710 uniform-voxel regions. Candidate pools and scores are rebuilt from the embeddings cached by the baseline run; applying the baseline's own selection rule to them reproduces its saved top-5 lists for all 710 regions, with scores equal to within 5×10^{-7}. Sequential greedy selection uses $\lambda_{\text{rep}} = 2\lambda_{\text{div}}$ (Proposition 2); the min-cost flow uses the same 300-candidate pools. As an independent check, both the flow and CP-SAT were solved on pools restricted to 20 candidates per region; CP-SAT certified optimality and its energy equalled the flow's up to integer rounding at every λ_{div}.

Table 8: Core energy J (Eq. (4) with $\lambda_{\text{ov}} = \lambda_{\text{fit}} = 0$) on Tiger, 710 regions, identical 300-candidate pools. $\sum$fid: summed blended fidelity; unique: distinct objects selected; changed: regions whose selected object differs.

λ_{div} (λ_{rep})	J greedy	J flow	$\sum$fid greedy	$\sum$fid flow	unique (g / f)	changed
0.025 (0.05)	-574.48	-578.79	592.43	597.19	706 / 697	602
0.075 (0.15)	-538.93	-543.18	592.18	596.43	710 / 710	603
0.15 (0.30)	-485.68	-489.93	592.18	596.43	710 / 710	603
0.25 (0.50)	-414.68	-418.93	592.18	596.43	710 / 710	603

Table 8 lists the results. At every λ_{div} the flow energy is lower than the greedy energy, by 4.2–4.3. Per region this is small (mean blended fidelity 0.834 for greedy and 0.840 for the flow), yet the two selections differ in 603 of the 710 regions. For $\lambda_{\text{div}} \geq 0.075$ both use every object at most once, and the core problem reduces to a maximum-weight bipartite matching that the sequential rule approximates in a single confidence-ordered pass. Both solvers run in under 1.5 s. Whether the lower energy corresponds to a different mosaic is examined in Section D.3.

Selection in the deployed baseline. With $\lambda_{\text{rep}} = 0.15$, the baseline's retrieval selects 710 distinct objects for Tiger, yet its deployed mosaic uses 444 (Table 7). The deployed assembly re-ranks the top three retrieved candidates by geometric fit, and the object it places is the retrieval choice in only 254 of the 710 regions. Repetition in the deployed mosaic is therefore introduced after retrieval, which is why the controlled conditions place exactly the selected object.

D.3 Joint selection with geometric terms (E-4, Tiger)

For the full energy (Eq. (4)) each region's assignment pool contains its 10 best candidates by blended score and, so that both earlier selections remain feasible, the greedy and flow choices ($8{,}122$ region–candidate pairs). Every pair was placed with the same alignment procedure as the final assembly,

with a fixed seed per pair. ℓ_{fit} was computed at $\delta = 0.25\%$ of the target diagonal, and pairwise interpenetration on all $6{,}587$ 26-neighbour edges with the interior-voxel measure at pitch $1/800$ of the diagonal, in units of one grid-cell volume. The reuse weight was fixed at $\lambda_{\mathrm{div}} = 0.075$. CP-SAT, warm-started from iterated conditional modes, proved optimality within the assignment pools at all nine points of the grid $\lambda_{\mathrm{ov}} \in \{0.25, 1, 4\} \times \lambda_{\mathrm{fit}} \in \{0, 0.25, 1\}$ in 7–142 s; iterated conditional modes alone ended 1.5–8.9 energy units above the optimum.

The precomputed pairwise terms reproduce the assembled mosaics. At pitch $1/400$ they predict 182, 253, 170, 112 and 125 interpenetrating 26-neighbour pairs for the five assembled selections of Table 9, against 182, 250, 171, 112 and 123 measured.

Table 9: Assembled Tiger mosaics, 710 regions, each constituent placed exactly as selected. B-full: optimal selection of the full energy at the given $(\lambda_{\mathrm{ov}}, \lambda_{\mathrm{fit}})$. Int. and pairs at pitch $1/400$ (lower bounds; Appendix F); pairs_{200} at pitch $1/200$. CD in % of the diagonal.

Selection	UOR	Cov	F@1%	$\mathrm{CD}_{\rightarrow}$	$\mathrm{CD}_{\leftarrow}$	Int. (%)	pairs	pairs_{200}
A: greedy	1.000	0.9988	0.609	0.311	1.714	6.3	191	57
B-core: flow	1.000	0.9983	0.605	0.306	1.689	9.7	264	101
B-full $(1, 0.25)$	0.939	0.9992	0.628	0.273	1.630	5.2	188	59
B-full $(4, 0)$	0.986	0.9983	0.609	0.310	1.678	3.5	120	38
B-full $(4, 0.25)$	0.930	0.9994	0.626	0.274	1.636	3.5	128	47

Table 9 compares the assembled mosaics. The fidelity-and-reuse optimum (B-core) changes the surface metrics only marginally relative to greedy selection, and it interpenetrates more (264 against 191 pairs). With the overlap term at $\lambda_{\mathrm{ov}} = 4$, the number of interpenetrating pairs falls by 37% ($\lambda_{\mathrm{fit}} = 0$) and 33% ($\lambda_{\mathrm{fit}} = 0.25$), and the excess volume by 45% and 44%, with the same direction at pitch $1/200$. Adding the fit term improves the surface metrics well beyond both A and B-core: at $(4, 0.25)$ the F-score at 1% rises from 0.609 to 0.626 and $\mathrm{CD}_{\rightarrow}$ falls by 12%, at the cost of a unique-object ratio of 0.930. The two geometric terms compete: at $\lambda_{\mathrm{ov}} = 0.25$, raising λ_{fit} from 0 to 1 raises the modelled number of interpenetrating pairs from 248 to 379, because constituents that cover their region more completely extend into neighbouring cells. These results come from one target and one run per condition, and the weights were chosen on the same target; Section D.4 applies them without re-tuning to the other targets.

D.4 Fixed weights on the remaining targets

Conditions A, B-core and B-full were run on Dinosaur, Flamingos and Sports Car with every setting fixed from Tiger (Section 5). The core-energy comparison transfers: with identical 300-candidate pools, the min-cost flow reaches a lower core energy than sequential greedy selection on all three targets (Dinosaur -425.04 vs. -421.17; Flamingos -145.67 vs. -144.73; Sports Car -843.55 vs. -835.51), and the two selections differ in $441/549$, $117/193$ and $904/1{,}080$ regions.

Table 10: Assembled mosaics on the held-out targets, weights fixed from Tiger. Columns as in Table 9.

Target	Selection	UOR	Cov	F@1%	$\mathrm{CD}_{\rightarrow}$	$\mathrm{CD}_{\leftarrow}$	Int. (%)	pairs
Dinosaur	A: greedy	1.000	0.9963	0.698	0.302	1.384	7.6	133
	B-core: flow	1.000	0.9954	0.693	0.297	1.346	11.1	195
	B-full $(4, 0.25)$	0.953	0.9983	0.727	0.259	1.274	3.5	110
	B-full $(4, 0)$	1.000	0.9952	0.699	0.299	1.339	5.4	81
Flamingos	A: greedy	1.000	0.9902	0.926	0.269	0.487	0.8	13
	B-core: flow	1.000	0.9925	0.929	0.259	0.482	4.6	31
	B-full $(4, 0.25)$	0.995	0.9973	0.935	0.222	0.452	2.0	26
	B-full $(4, 0)$	1.000	0.9914	0.929	0.264	0.482	2.7	23
Sports Car	A: greedy	1.000	0.9822	0.942	0.372	0.494	2.3	91
	B-core: flow	1.000	0.9847	0.941	0.368	0.499	3.1	136
	B-full $(4, 0.25)$	0.952	0.9932	0.951	0.337	0.471	1.6	107
	B-full $(4, 0)$	1.000	0.9850	0.942	0.369	0.497	2.3	72

CP-SAT certified optimality within the assignment pools for every held-out target and weight setting, in 0.5–213 s.

Flamingos. Flamingos interpenetrates little under any selection: the modelled overlap of greedy selection is 0.031 cell volumes, against 8.35 on Tiger. As on Tiger, the core optimum interpenetrates more than greedy selection (31 against 13 pairs), and the full objective with the fit term improves the surface metrics: coverage at 1% rises from 0.9902 to 0.9973 and $\mathrm{CD}_{\rightarrow}$ falls by 17%. Unlike on Tiger, the full objective does not reduce interpenetration below greedy selection (23–26 against 13 pairs at pitch $1/400$). The precomputed terms reproduce these counts at pitch $1/400$ (13, 31, 25 and 23 predicted), so the discrepancy is not caused by assembly. It comes from the lattice: the overlap term is evaluated at pitch $1/800$, where the full-objective selection has one-tenth of the overlap of greedy selection (0.003 against 0.031 cell volumes), whereas at pitch $1/400$ the order is reversed (0.098 against 0.017). For interpenetrations close to the lattice resolution, the interior-voxel measure at one pitch is therefore not a reliable surrogate for the measure at another. We keep the pre-specified single-pitch term for this comparison and discuss the limitation in Section 7.

Dinosaur. Dinosaur reproduces the Tiger result without re-tuning. The core optimum interpenetrates more than greedy selection (195 against 133 pairs; excess volume 11.1% against 7.6%). At $(4, 0.25)$ the full objective reduces the excess volume by 54% and the number of interpenetrating pairs by 17%, raises the F-score at 1% from 0.698 to 0.727, and lowers $\mathrm{CD}_{\rightarrow}$ by 14% and $\mathrm{CD}_{\leftarrow}$ by 8%, with a unique-object ratio of 0.953; at $(4, 0)$ it reduces the number of pairs by 39%. CP-SAT certified both solutions optimal (in 39 and 213 s).

Sports Car. On Sports Car the core optimum again interpenetrates more than greedy selection (136 against 91 pairs). The full objective at $(4, 0.25)$ improves every surface metric (coverage at 1% from 0.9822 to 0.9932, F-score from 0.942 to 0.951, $\mathrm{CD}_{\rightarrow}$ by 9%) and reduces the excess volume by 30%, but it has more interpenetrating pairs than greedy selection (107 against 91); at $(4, 0)$ it reduces the number of pairs by 21%. Half of the Sports Car constituents (479–534 of 1,080) are too thin for the interpenetration measure at pitch $1/400$, the largest fraction among the targets, so its interpenetration values are the least informative.

Summary. With the weights fixed on Tiger, three findings transfer to every held-out target: the core optimum interpenetrates more than independent greedy selection; the full objective with the fit term improves coverage, F-score and both Chamfer directions; and the geometric terms, not joint selection by itself, are responsible for these changes. The reduction of interpenetration transfers to Dinosaur, is partial on Sports Car and absent on Flamingos, where interpenetration is close to the lattice resolution (Section 7).

D.5 Part hierarchy and retrievability (E-1, E-2)

PartField features were extracted for every target face and clustered agglomeratively with PartField's own connectivity-constrained procedure; we keep the full merge tree. Depth d denotes the cut into 2^d clusters, so that nodes at consecutive depths are nested. The parts follow semantic structure (head, torso, limbs and tail; wings and legs; wheels, roof and body panels). They are nearly contiguous on Tiger and Dinosaur (92–99% of each part's area lies in one connected piece at $K \in \{4, 8, 16, 32\}$), span both birds on Flamingos (72–82%), and are strongly fragmented on Sports Car (40–46%), whose mesh consists of 1,365 separate panels that PartField groups by type.

D.6 Amodal completion (E-3)

HoloPart was applied to every depth-d partition ($d = 1, \ldots, 5$) of every target at its default settings, and each node was rebuilt with the generated surface in place of the cap (Section 4.1). A visual check confirms that the generated surface is genuine completion—for example, a leg continued into the body, or a closing surface on a cut side—rather than drift from the original surface. Table 11 reports the change in retrievability, $\Delta\varepsilon = \varepsilon_{\mathrm{completed}} - \varepsilon_{\mathrm{capped}}$, and the cost.

Table 11: Effect of HoloPart completion on retrievability ($\Delta\varepsilon < 0$: completion helps), over all hierarchy nodes of depth 1–5; 95% bootstrap interval; ρ: Spearman correlation of $\Delta\varepsilon$ with the generated area relative to the original part area. Cost on one shared NVIDIA L40S.

Target	nodes	mean $\Delta\varepsilon$	improved	ρ	s/part	peak GiB
Tiger	51	$+0.048$ [$+0.034$, $+0.061$]	18%	$+0.57$	40.7	7.0
Dinosaur	51	$+0.005$ [-0.001, $+0.012$]	43%	$+0.21$	48.5	7.0
Flamingos	51	$+0.028$ [$+0.017$, $+0.039$]	31%	$+0.13$	40.6	15.2
Sports Car	47	$+0.023$ [$+0.016$, $+0.030$]	17%	$+0.36$	48.1	7.0

Completion makes part regions *less* retrievable on Tiger, Flamingos and Sports Car and has no detectable effect on Dinosaur; the loss grows with the amount of generated surface. Because the shape representation already sees the cap (it is real geometry that bounds the region), the comparison isolates the geometry of the completion: marking the generated surface as a cut face instead changes ε by at most 6×10^{-4}. Completing all 62 nodes of one target's hierarchy costs 42–50 GPU-minutes.

Table 12: Retrievability error ε of PartField hierarchy nodes and uniform-voxel cells. Median ε at depths 0, 3 and 5, and for the production and 3^3 voxel grids. β_{part}: difference in ε between parts and voxel cells of equal surface area (regression of ε on log area and a part indicator over the common size range, 95% bootstrap interval). ch.<par.: fraction of nodes more retrievable than their parent. ARI: agreement between the refinement frontier at the median ε threshold and a size-only frontier with the same number of regions; ARI $= 1$ for Dinosaur and Flamingos is trivial, because at that threshold neither target is refined (single-region frontier).

Target	d=0	d=3	d=5	voxel	3^3	β_{part}	ch.<par.	ARI
Tiger	0.517	0.448	0.321	0.191	0.354	$+0.021$ [0.001, 0.040]	0.92	0.74
Dinosaur	0.234	0.338	0.310	0.175	0.260	$+0.057$ [0.041, 0.072]	0.65	1.00
Flamingos	0.320	0.363	0.311	0.254	0.371	$+0.010$ [-0.009, 0.030]	0.69	1.00
Sports Car	0.252	0.229	0.236	0.204	0.243	$+0.012$ [-0.000, 0.025]	0.40	0.54

D.7 Production granularity (E-5)

Table 13 compares the production voxel grid (A) with a cut of the PartField tree into the same number of parts (C-m), both with greedy selection; Section 6.3 discusses it.

Table 13: Production granularity (equal region counts, greedy). Int. at pitch $1/400$; inv.: constituents too thin for it.

Target	Regions	$\lvert\mathcal{R}\rvert$	Cov	F@1%	CD$_\rightarrow$	CD$_\leftarrow$	Int. (%)	inv.
Tiger	A: voxels	710	**0.999**	0.609	0.311	1.714	6.3	325
	C-m: parts	710	0.991	**0.981**	**0.246**	**0.303**	0.0	515
Dinosaur	A: voxels	549	**0.996**	0.698	0.302	1.384	7.6	279
	C-m: parts	549	0.987	**0.977**	**0.267**	**0.324**	0.5	392
Flamingos	A: voxels	193	**0.990**	0.926	**0.269**	0.487	0.8	95
	C-m: parts	193	0.946	**0.951**	0.345	**0.343**	0.2	121
Sports Car	A: voxels	1,080	**0.982**	**0.942**	**0.372**	**0.494**	2.3	534
	C-m: parts	1,080	0.917	0.890	0.492	0.841	1.9	610

D.8 Ablations (E-6)

A size-only frontier coincides with the refinement frontier on Tiger and Dinosaur, where every node exceeds τ_{err}, and is slightly worse on Flamingos and better on Sports Car (coverage 0.788 vs. 0.805; 0.725 vs. 0.699): the retrievability criterion shows no benefit over size. HoloPart-completed regions lower coverage on Tiger, Flamingos and Sports Car (0.584, 0.700, 0.645 vs. 0.722, 0.805, 0.699). On Tiger, $K_a = 5, 10, 20$ gives optimal energies -494.9, -498.2 and -504.7, almost entirely through fit; at $K_a = 20$ CP-SAT stops at its time limit with a gap of 0.08–0.11. Without the

overlap term, the surface gains remain (F-score 0.621 vs. 0.626) but interpenetration exceeds greedy selection (283 vs. 191 pairs); the overlap-only optimum lowers it (120 pairs) without the surface gains.

E Reliability of Gallery Scale Metadata (E-meta)

A scale-consistency term in the selection energy would need the real-world size of every gallery object. We audited the size metadata of both gallery sources against the loaded geometry (Table 14).

ShapeNetSem. We sampled $4{,}904$ models (up to 40 per category for categories with at least 10 models, and 100 from the remaining categories). Raw mesh extents match the metadata dimensions `aligned.dims` within $\pm 10\%$ for only 2.3% of models, because meshes are stored in mixed units. For the $4{,}674$ sampled models with a `unit` value, mesh extent $\times$ `unit` $\times\, 100$ matches `aligned.dims` within $\pm 10\%$ for 99.6% (median ratio 1.000). This agreement is by construction: `aligned.dims` is the mesh extent converted with `unit`, so it confirms internal consistency, not real-world size. Across the full metadata file, `unit` is missing for $2{,}771$ of $12{,}288$ models (22.6%). In every sampled category whose models lack it, the median ratio of raw extent to `aligned.dims` is $0.394 \approx 1/2.54$, i.e. inches were assumed. A further 202 of the $12{,}490$ ShapeNetSem meshes in our gallery have no metadata row. Some category medians are physically implausible (e.g. Thumbtack 1.30 m and Key 1.18 m, both categories without a `unit` value); a per-category plausibility test was not carried out.

Amazon Berkeley Objects. Of 350 sampled ABO models, 341 have catalogue dimensions. The ratio of mesh extent to catalogue dimension has median 1.000 and interquartile range 0.996–1.011; 86.2% of models are within $\pm 10\%$ and 90.9% within $\pm 20\%$.

Table 14: Gallery scale metadata against loaded geometry (largest extent vs. largest listed dimension). dims: ShapeNetSem `aligned.dims`.

Comparison	n	Median	within $\pm 10\%$	within $\pm 20\%$
ShapeNetSem: raw extent / dims	4,904	0.427	2.3%	4.9%
ShapeNetSem: extent·`unit`·100 / dims	4,674	1.000	99.6%	99.8%
ABO: extent / catalogue dimension	341	1.000	86.2%	90.9%

Consequence. ABO sizes are usable, but real-world sizes are unverified for a large part of the ShapeNetSem portion of the gallery, and the scale applied during placement is expressed in each asset's own file units. The selection energy therefore contains no scale term (Section 3.3).

F Validation of the Interpenetration Measure

Definition. Each placed constituent is voxelised on a lattice of pitch h that is shared by all constituents (voxel centres at integer multiples of h). A voxel is *interior* to a constituent if it lies inside the constituent's filled voxelisation and is not crossed by the constituent's own surface. The interpenetration volume is h^3 times the number of voxels that are interior to two or more constituents. Constituents that only touch—for example adjacent regions sharing a cut face—have no common interior voxel and contribute zero. Penetration shallower than about one voxel is not detected, so the measure is a lower bound; we report it at $h \in \{1/100, 1/200, 1/400\}$ of the target diagonal.

A retired measure. A simpler measure counts voxels occupied (surface or filled) by more than one constituent. Applied to the target's own voxel chunks, which have disjoint interiors, it reports 24–47% of the occupied volume as shared at $h = 1/200$, as much as on the mosaics themselves, because it counts contact at shared faces. We do not use it.

Synthetic control. Two unit cubes are offset along one axis so that their true intersection volume is d. Table 15 lists the measured interpenetration.

Table 15: Interpenetration measured on two unit cubes with true intersection volume d, at three pitches h (fractions of the pair's bounding-box diagonal).

d	$h = 1/100$	$h = 1/200$	$h = 1/400$
0 (touching)	0	0	0
0.05	0.023	0.048	0.042
0.10	0.067	0.080	0.087
0.30	0.285	0.278	0.287

Touching cubes give exactly zero, and every positive overlap is detected and underestimated. The estimate does not increase monotonically as h decreases, because the offset of the intersection relative to the lattice changes with h.

Controls on real geometry. As a negative control, we apply the measure to the voxel chunks of each target, whose interiors are disjoint by construction. For all five chunk sets (Tiger, two Dinosaur grids, Flamingos, Sports Car) and all three pitches, the measured interpenetration is exactly zero, with no overlapping pair. As a positive control, we add exact copies of a random 10% (71) of the Tiger chunks. Each copy shares its whole interior with its original, so the expected interpenetration equals the interior volume of the copied chunks; the measured value equals it exactly at every pitch.

Sensitivity. Constituents thinner than about two voxels have no interior voxel and are invisible to the measure. Table 16 lists the fraction of target chunks without an interior voxel. At $h = 1/400$ it is 13–15% for Tiger and Dinosaur, 41% for Flamingos and 66% for Sports Car, whose chunks are thin shell fragments. Reported interpenetration is therefore a lower bound whose tightness differs between targets, and we report the corresponding fraction for the placed constituents with every result.

Table 16: Fraction of target voxel chunks with no interior voxel, by lattice pitch h (fraction of the target diagonal).

Chunk set	chunks	$h = 1/100$	$h = 1/200$	$h = 1/400$
Tiger (12^3)	710	82%	26%	13%
Dinosaur (16^3)	616	61%	34%	15%
Dinosaur ($10 \times 14 \times 21$)	549	62%	30%	15%
Flamingos (12^3)	193	91%	64%	41%
Sports Car (12^3)	1,080	95%	83%	66%