Beyond Complete Shapes: A Benchmark for Quantitative Evaluation of 3D Shape Surface Matching Algorithms

V. Ehm[†,1,3] ⓘ, N. El Amrani[†,2,4] ⓘ, Y. Xie[1] ⓘ, L. Bastian[1,3] ⓘ, M. Gao[1,3] ⓘ, W. Wang[2,4] ⓘ,
L. Sang[1,3] ⓘ, D. Cao[2,4] ⓘ, T. Weißberg[2,4] ⓘ, Z. Lähner[2,4] ⓘ, D. Cremers[1,3] ⓘ, and F. Bernard[2,4] ⓘ

[1]Technical University of Munich, Germany,   [2]University of Bonn, Germany
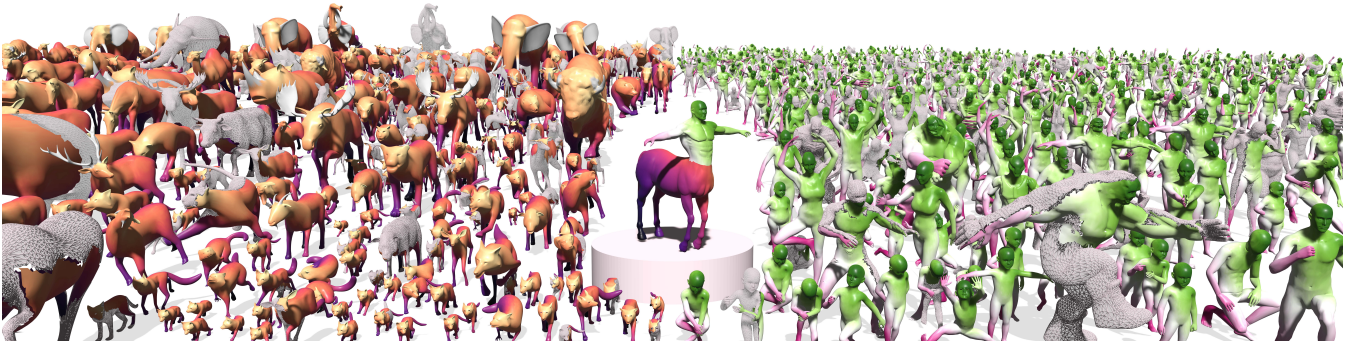[3]Munich Center for Machine Learning, Germany   [4]Lamarr Institute, Germany

Figure 1: *Our benchmark* `BeCoS` *provides **correspondences between thousands of shapes** (visualised via colour transfer). As a special focus, we consider the generation of partial shapes, where we can generate partial shape matching instances across diverse four-legged animals and humanoid categories (grey parts indicate the missing parts of partial shapes).*

Abstract
*Finding correspondences between 3D deformable shapes is an important and long-standing problem in geometry processing, computer vision, graphics, and beyond. While various shape matching datasets exist, they are mostly static or limited in size, restricting their adaptation to different problem settings, including both full and partial shape matching. In particular the existing partial shape matching datasets are small (fewer than 100 shapes) and thus unsuitable for data-hungry machine learning approaches. Moreover, the type of partiality present in existing datasets is often artificial and far from realistic. To address these limitations, we introduce a generic and flexible framework for the procedural generation of challenging full and partial shape matching datasets. Our framework allows the propagation of custom annotations across shapes, making it useful for various applications. By utilising our framework and manually creating cross-dataset correspondences between seven existing (complete geometry) shape matching datasets, we propose a new large benchmark **BeCoS** with a total of 2543 shapes. Based on this, we offer several challenging benchmark settings, covering both full and partial matching, for which we evaluate respective state-of-the-art methods as baselines. Visualisations and code of our benchmark can be found at:* `https://nafieamrani.github.io/BeCoS/.`

1. Introduction

Finding correspondences between deformable shapes is a fundamental problem in geometry processing, with various applications such as texture and deformation transfer [DYT05; OBS*12], or statistical shape analysis [HSS*09; BRLB14; ECB24]. The problem has been widely studied in the literature [OBS*12; ELC20; AEOW23; WWO24; CRB23], and can vary significantly in complexity. For example, current methods still struggle when estimating severe *non-isometric* deformations for shapes of different categories [DLR*20; PKO22; ELC20; BXNL24; CEE*24], or when they are only partially observed [RSCB22; EGR*24]. Current non-isometric shape

datasets are often limited in size and flexibility [ZKJB17; MMR*19; DLR*20; MRSO22; LTT*21], restricting their ability to support dense correspondence propagation or custom annotations, which reduces their applicability to broader tasks.

In real-world scenarios, it is particularly common to encounter shapes with missing parts due to occlusions, limited sensor coverage, or acquisition errors during 3D scanning. The setting in which at least part of one shape does not exist on the other is called partial shape matching [RCB*17; APO21; RSCB22; ERE*23; EGR*24]. Despite its practical relevance, partial shape matching has received limited attention in the context of machine learning-based approaches, partly because of the increased difficulty, but largely due to the lack of suitable and realistic large-scale datasets, a requirement for data-hungry learning algorithms. The full-to-full matching setting is not only simpler but also benefits from wider data availability which has supported the development of data-driven methods that already display convincing performance, even without supervision [HLR*19; RSO19; CRB23]. In contrast, existing datasets for partial shape correspondence often suffer from a small number of shapes, limited realism of generated partiality (e.g., cuts or holes) [CRB*16], and a lack of diversity in shape categories [BDK23; APO21]. This limited data availability hampers the progress of data-driven methods for partial shape matching; recent approaches have thus required pre-training with large full-geometry datasets [CRB23], or rely on ground-truth correspondences [APO21] which are challenging to obtain in practice.

To address the absence of large datasets and foster research on (partial) shape correspondence, we introduce a procedural data generation framework to generate diverse shape pairs from a shape network. By simulating more realistic partiality patterns using ray casting, we bridge the gap between existing synthetic datasets and real-world partial shape correspondence problems. Furthermore, we propose BeCoS, a carefully curated benchmark based on existing datasets comprising diverse shape categories, thus enabling to assess the performance of existing and foster the development of novel methods. Through extensive experiments of current state-of-the-art partial and full shape matching methods, we demonstrate the limitations of partial methods and highlight the need for further research in this area.

Contributions. We summarise our main contributions as follows:

- **Shape Network for Correspondence Propagation:** We unify multiple existing mesh-based datasets through manual cross-category annotations, constructing a large and diverse shape network that indicates between which shapes correspondences exist. This shape network facilitates the propagation of dense correspondences and custom annotations across shapes, enhancing its applicability to various shape analysis tasks.
- **Scalable Framework for Partial Shape Correspondence:** We introduce a flexible and scalable framework that enables the generation of training examples for partial shape correspondence tasks that is applicable to new datasets.
- **Benchmark and Performance Analysis:** As a concrete realization of our framework, we present a challenging non-rigid shape dataset, serving as a standardised benchmark for shape matching

methods. We further conduct an extensive evaluation of state-of-the-art techniques on this benchmark, offering insights into their strengths, limitations, and key challenges. Our results indicate that partial shape matching remains an open problem, motivating future research in the field.

2. Related Work

2.1. Datasets

Although several 3D shape datasets exist, we align with established deformable shape matching literature by focusing on humanoid and four-legged animal shapes. These categories support continuous deformations and offer well-defined dense correspondences, making them particularly suitable for our study. In the following, we provide an overview of existing datasets for full-to-full (F2F), partial-to-full (P2F), and partial-to-partial (P2P) shape matching datasets. Table 1 summarises existing deformable shape matching datasets.

Full-to-Full Datasets. Full-to-full shape correspondence datasets, such as FAUST [BRLB14] and SCAPE [ASK*05], have been essential in methodological progress on shape correspondence. They encompass humans in various poses, initially emphasising point-wise correspondence recovery under isometric deformations. SHREC'19 [MMR*19] was introduced to evaluate the generalisation capabilities of methods to different mesh connectivity on human scans. Datasets like TOSCA [BBK08], SMAL [ZKJB17], SHREC'20 [DLR*20], and DeformingThings4D [MRSO22; LTT*21] have shifted the emphasis towards non-isometric correspondence between various animal categories and humanoid figurines. Although deep learning methods have achieved state-of-the-art performance in shape correspondence, non-isometric datasets remain relatively small and static, limiting their applicability in training and evaluating deep learning-based correspondence methods. While full-to-full shape matching has received considerable attention, the partial setting has been explored less (see Table 1).

Partial-to-Full Datasets. In this setting, a partial shape is deformably matched to a shape template with complete geometry. SHREC'16 [CRB*16], based on the TOSCA [BBK08] dataset, was the first benchmark for this problem setting. PFAUST [BDK23] and PFARM [APO21] were later introduced using a similar approach based on the FAUST [BRLB14] and FARM [KMP*21] datasets, respectively. SHREC'20 [DLR*20] also contains a subset of full-to-partial correspondence pairs.

Partial-to-Partial Datasets. Partial-to-partial shape matching involves matching two partial shapes with unknown overlap, making it considerably more challenging than partial-to-full matching. In [APO21] the authors proposed the first benchmark for this setting with CP2P [APO21] based on the SHREC'16 CUTS [CRB*16] benchmark. Recently, PSMAL [EGR*24] was introduced, extending the SMAL [ZKJB17] dataset to include partial-to-partial correspondence pairs.

As can be seen in Table 1, most existing datasets for partial shape matching (P2F and P2P) focus only on near-isometric problem settings. Furthermore, they contain only a limited number of shapes, impeding the capabilities of learning-based approaches, particularly when considering unsupervised settings for partial non-isometric correspondence settings.

Setting	Dataset	# Unique Shapes	Shape Type	Non Isometry
F2F	FAUST [BRLB14]	100	⚲	✗
	SCAPE [ASK*05]	71	⚲	✗
	SMAL [ZKJB17]	49	🐾	✓
	DT4D Human [MRSO22; LTT*21]	248	⚲	✓
	DT4D Animal [MRSO22; LTT*21]	1950	🐾	✗
	TOSCA Human [BBK08]	45	⚲	✗
	TOSCA Four-legged [BBK08]	37	🐾	✗
	KIDS [RRW*14]	32	⚲	✗
	SHREC'20 [DLR*20]	11	🐾	✓
	SHREC'19 [MMR*19]	44	⚲	✓
P2F	SHREC'16 [BBK08; CRB*16]	76	⚲/🐾	✗
	PFAUST [BDK23; BRLB14]	10	⚲	✗
	FARM [KMP*21]	5	⚲	✗
	PFARM [APO21; KMP*21]	28	⚲	✗
	SHREC'20 [DLR*20]	14	🐾	✓
P2P	CP2P [APO21; CRB*16]	76	⚲/🐾	✗
	PSMAL [EGR*24; ZKJB17]	43	🐾	✓
Both	**BeCoS** (ours)	**2543**	⚲/🐾	✓

Table 1: *An Overview of 3D Deformable Shape Datasets, categorised as full-to-full (F2F), partial-to-full (P2F) and partial-to-partial (P2P), including the number of available shapes, shape type (Humanoid ⚲ or Four-legged 🐾), and deformation class (Near-Isometric (✗) and Non-Isometric (✓)). Due to limited correspondence pairs, FARM and PFARM are typically only used for evaluation. Note that in previous datasets, not all unique shapes have annotated correspondences, limiting the number of possible correspondence pairs, e.g., DT4D Animal (maximum 62 452 correspondence pairs) compared to our benchmark (maximum 1 414 179 correspondence pairs). As SHREC20 [DLR*20] contains three partial shapes, it can be used as a partial-to-full or full-to-full dataset.*

2.2. Methods for Shape Correspondence

We proceed by discussing a selection of shape matching methods and refer to recent surveys [Sah20; DYDZ22] for a more thorough treatment. While point-cloud matching methods also exist [LH22; BLZ*20; HGU*21; YQH*23], we focus on mesh-based methods, which are better aligned with the surface representations emphasised in our dataset. Methods addressing shape correspondence can be categorised into axiomatic and learning-based. Among the axiomatic approaches are discrete optimization methods [BST20] and spectral methods utilising the powerful framework of functional maps [OBS*12; RCB*17; MRR*19]. However, the performance of the latter class typically degrades as assumptions of isometry are violated, such as in partial correspondence settings.

Full-to-Full Shape Matching. Spectral methods based on the eigenfunctions of the Laplace-Beltrami Operator (LBO) have been successful in full-to-full matching settings [OBS*12; LRR*17; DSO20; BXNL24; CRB23], yielding a compact solution. In contrast to spectral approaches, discrete optimization formulations have also been introduced for full-to-full matching [BST20; RSCB22; RAC*24; RB24]. These algorithms often come with guaranteed bounds on optimality, sometimes incorporating geometric consistency constraints.

Partial-to-Full Shape Matching. Rodola et al. [RCB*17] demonstrate that for partial-to-full matching, the functional map has a slanted diagonal structure, which has, for example, been used in [LRB*16] to solve multi-part partial-to-partial correspondence; however, under the assumption that a full template is given. Several methods also operate in the spectral domain in learning-based settings [APO21; CB22; CRB23]. Other works do not use functional map layer [BDK23; BDK24] and explicitly tackle the partial-to-full case.

Partial-to-Partial Shape Matching. Few works thus far explicitly address the partial-to-partial matching setting, which is particularly challenging as the overlapping region between shapes is unknown. As such, the overlapping region must simultaneously be identified and matched under an unknown deformation. DPFM [APO21] demonstrates that spectral methods are applicable in this domain using an additional attention mechanism, albeit using ground-truth supervision for feature learning. In [WSSC11] shape matching is formulated as an integer linear program (ILP) over the space of orientation-preserving diffeomorphisms. SM-COMB [RSCB22] demonstrated that the latter constitutes a strong prior to guide partial-to-partial matching. However, holes must be synthetically filled to obtain a closed manifold. Recently, GC-PPSM [EGR*24] formulated a geometrically consistent partial-to-partial shape matching non-linear integer program. While both SM-COMB and GC-PPSM are compatible with various surface descriptors, they achieve SOTA results only with high-quality pre-trained input features (e.g. from ULRSSM [CRB23]), which can be prohibitive when these are not available [RSCB22; EGR*24].

2.3. Feature Descriptors for Shape Correspondence

Many axiomatic and learning-based methods for shape matching rely on surface feature descriptors as guidance for correspondence computation. Feature descriptors, like SHOT [TSD10], can be hand-crafted based on extrinsic quantities like XYZ-coordinates or surface normals. Many surface descriptors are sensitive to surface degradation, such as holes, or are not invariant under rotations, as is the case for XYZ coordinates. The common intrinsic heat kernel (HKS) [SOG09] and wave kernel signatures (WKS) [ASC11] are invariant under isometric deformations; however, they are not effective under partiality as the isometry assumption can be heavily violated. On the other hand, methods like DiffusionNet [SACO22] aim to learn robust surface descriptors from data while still using some features like XYZ coordinates or WKS descriptors as input. Considering that 3D shapes of many datasets are typically oriented in an upright manner [BRLB14; ASK*05; BBK08; RRW*14; DLR*20], or generally positioned in a canonical pose [ENK*21] this can lead to biased predictions. We discuss these biases in detail in Sec. 5.4. Moreover, learned features require large training datasets and a high-quality (self-)supervision signals which is challenging to achieve under severe partiality. Since obtaining per-vertex feature descriptors directly on 3D shapes is challenging, [WHC*16] addresses this by establishing correspondences between 3D shapes through correspondences between their 2D rendered images. Recent methods [DMM24; DLH23; KMS22; FLN*24] leverage 2D image foundation model features [ODM*23; RBL*22] from multiple rendered views of a 3D shape to aggregate per-vertex features. However, such

an approach can suffer from significant bias due to missing semantic information in the rendered views.

3. Our Cross-Dataset Shape Matching Network

We propose a novel framework enabling the procedural generation of shape matching instances with ground-truth correspondences based on a given collection of 3D shapes. Our framework supports shapes from multiple categories to generate shape matching instances of non-isometric shape pairs. Furthermore, we manually link correspondences between multiple datasets to vastly increase the diversity and representativeness. Our framework provides researchers with the possibility to propagate dense correspondences or custom annotations (e.g. left/right annotations, which we later consider as demonstrative example) through the entire collection of 3D shapes. Our pipeline supports re-meshing while preserving shape correspondences within a collection and thus fosters discretisation-agnostic method development. To overcome the limitations in realism and sample size of existing partial shape benchmarks, we specifically focus on partial shape generation. We use ray casting to generate more realistic partial shapes than existing benchmarks based on synthetic planar cuts and holes. We summarise the main ideas in Fig. 2. A detailed description of our data generation can be found in the supp. mat. Sec. 9. We apply our framework to construct a new large benchmark combining different shape correspondence datasets (Section 4) and show open challenges in the field through an evaluation of existing methods on this benchmark, as detailed in Sec. 5.

3.1. Cross-Dataset/Category Correspondences

Current datasets commonly lack correspondences between different categories, limiting data availability and diversity [BBK08; LTT*21]. We address this shortcoming by manually establishing correspondences across categories *and* across different datasets. To this end, we treat each category for which a cross-category correspondence is not available due to missing annotation as a separate shape group. Then, we select a template shape for each group and manually choose a set of corresponding landmarks on pairs of templates using FaceForm [Fac23] (see supp. mat. Sec. 9). In the case of non-isometry, more manual landmarks are needed to guide the deformation, see Fig. 11 in supp. mat. for an example. We then morph one template into another to establish dense point-to-surface correspondences (represented in barycentric coordinates to account for different discretisations). Using the tightly aligned shapes, dense correspondences are obtained by projecting points onto triangles of the other surface. After the templates are in dense correspondence, we propagate them to the entire shape category using the existing intra-category correspondences. Overall, we build a shape network of pairwise dense correspondences that cover all possible shape templates (cross-category and cross-dataset). The detailed shape network is presented on our project page. A colourmap visualization of the correspondences can be seen in Fig. 2.

3.2. Scaling of Shapes

Our framework integrates multiple mesh-based datasets, each with varying shape scales. To harmonise these datasets, we manually define scaling factors for each dataset, ensuring shapes are scaled according to semantic criteria (e.g., the same species across datasets have consistent sizes). Consequently, our framework offers diverse non-isometrically deformed shapes with varying scaling factors (e.g., a cat is smaller than an elephant; see Fig. 1 for examples).

3.3. Alignment of Shapes

Many correspondence datasets, specifically partial ones, such as CP2P [APO21] and SHREC'16 [CRB*16], keep shapes in the same canonical pose during training and testing. This bias can prevent methods from generalising to shapes under random rigid transformations, see also [ALO23]. As this assumption is often unrealistic in real-world settings, we randomly rotate the shapes around the z-axis (with respect to their canonical pose as present in the original dataset) and thus generate shapes in random upright positions. We show an ablation study about rotation in Sec. 5.4.

3.4. Custom Annotation Propagation

Shape annotations like movable parts labels used in robotics [QYW*25] or part segmentation labels [KHS10; CCZZ24] are often indispensable for data-driven training or evaluation. Our framework allows for the propagation of any type of information defined on the vertices from a few annotated examples to all shapes in the dataset. By leveraging the shape network, we can find a path to the closest annotated example from each shape, establishing correspondences to a shape that offers ground-truth information. To demonstrate the effectiveness of our method, we choose to propagate left/right annotations from only 12 manually annotated shapes to the whole dataset (see Figure 3). While in theory, even fewer annotations would suffice, we selected 12 shapes to reduce propagation error (see the examples in the supplementary material in Figure 9).

3.5. Partiality Generation

Partiality via Ray Casting. Given a shape $\mathcal{X}$, we introduce partiality by simulating a simple unidirectional 3D scanning procedure based on ray casting, as shown in Fig. 2 right. Doing so leads to a variety of different partial shape boundaries and overlapping behaviour (see Fig. 1 for examples). To perform ray casting, we first centre and scale each full shape $\mathcal{X}$ to be within a fixed unit bounding box and then choose random camera poses on the unit sphere. Partiality is obtained by casting R rays from the sampled camera pose onto the full shape $\mathcal{X}$. Rays stop upon hitting a triangle. All hit triangles that form the largest connected component (by area) are kept to be part of the partial shape $\hat{\mathcal{X}}$. Finally, each partial shape $\hat{\mathcal{X}}$ is scaled and translated back to its original size and location. To generate a shape pair for partial-to-partial shape matching we get a partial shape $\hat{\mathcal{Y}}$ from a second full shape $\mathcal{Y}$.

Random camera poses could lead to non-existent or prohibitively small overlap in the partial-to-partial matching setting. Therefore, we sample two camera poses with the constraint that their maximum angular disparity in both azimuth and elevation does not exceed α degrees. Through this, we attain an overlap of a certain size to ensure meaningful matching instances. For the overlap computation, we rigidly align the full shapes $\mathcal{X}$ and $\mathcal{Y}$ with Procrustes analysis using
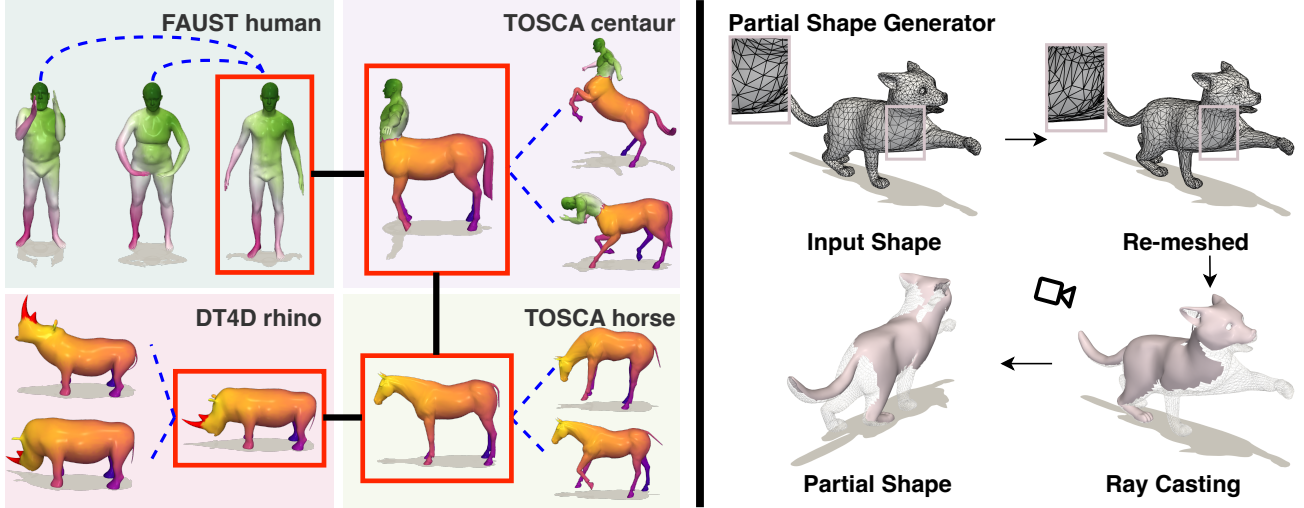
Figure 2: *Left: We use a **collection of multiple complete geometry datasets** to generate novel shape pairs. Using FaceForm (see supplementary material), we manually annotate cross-category and cross-dataset correspondences (black lines, see Sec. 3.1) between (manually selected) templates (shapes in red boxes). These are then combined with existing intra-dataset correspondences (blue dashed lines) to construct a shape network, through which dense correspondences can be propagated. **Right**: Given an input shape with full geometry from our shape collection, we first **re-mesh** the shape, then **apply ray casting** to simulate a unidirectional 3D scanning procedure to introduce partiality (see Sec. 3.5), and finally apply a deterministic random **rotation**.*
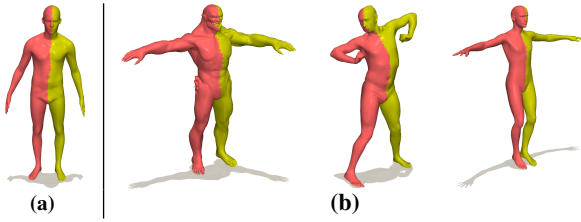

Figure 3: *Our framework allows for the **propagation of custom annotations** throughout all shapes. We show the case of left/right annotations. **(a):** Left/Right annotations are manually annotated on the shape considered as template. **(b):** Left/Right annotations are propagated from the closest template in the shape network (see Sec. 8.2 in supp. mat. for more examples).*

the given correspondences to ensure that both shapes are in a similar position i.e. both shapes are up-right and facing the same direction. Then, we perform the previously described partiality generation on both shapes to generate $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$. Subsequently, we quantify the overlap as the percentage of vertices from $\hat{\mathcal{X}}$ that are mapped to $\hat{\mathcal{Y}}$ and vice-versa. If neither of these quantities falls within a predefined range (see Sec. 4 for details), the process proceeds with new random camera poses until the overlap meets the specified criteria, or until a maximum of m iterations is reached. In the latter case, the partial shapes whose overlap is the closest to the desired range are selected (see Sec. 9.3 in supp. mat. for more details).

Flexible Generation of Shape Matching Instances. Existing partial shape datasets provide a precomputed and fixed number of holes/cuts [CRB*16; BDK23; KMP*21; APO21; DLR*20;

EGR*24], creating partial shape pairs from identical meshes in varying poses. Instead, our proposed framework can generate partial shapes from a larger and more diverse pool of shapes, allowing for pairing compatible shapes from different categories, potentially originating from different datasets. This creates a more challenging scenario encompassing simultaneous non-isometric deformations and partiality, representing an important step towards more realistic settings. Furthermore, our framework introduces more variety by using different camera poses for ray casting and re-meshing shapes to different resolutions.

3.6. Semantic Missing Parts

By combining multiple categories and datasets, we encounter cases where dense correspondences cannot be reliably established between certain parts of shapes without introducing inaccuracies (e.g., matching a deer's horn to a horse's head). While such ambiguities can be managed when considering semantic correspondences [AEOW23], they remain problematic for dense correspondences. To address this, for the first time we create large cross-dataset correspondences with semantic missing parts annotations, as shown in Fig. 4. To this end, we modify our pipeline described in Sec. 3.1 by manually removing the semantic missing parts from the shapes and match the rest of the shapes to get the dense correspondences. We show an example in Fig. 11 in supp. mat. This creates a novel objective for shape matching methods that warrants future research.

4. The `BeCoS` Benchmark

In addition to the procedural data generation framework, we propose the `BeCoS` benchmark, which is a specific instantiation of our general framework introduced in Sec. 3. `BeCoS` contains realistic full
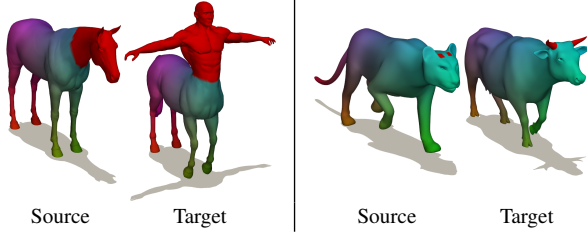
Figure 4: *For the first time we define **semantic missing parts in full shapes**. The horse and puma do not possess a human upper body or horns, making these features unmatchable in the centaur and cow, respectively.*

and partial shapes from established datasets with cross-category and cross-dataset correspondences. The purpose of `BeCoS` is to challenge current state-of-the-art shape matching methods, specifically in the partial case, and facilitate progress in the field through a standardised evaluation. To create our benchmark, we apply the pipeline introduced in Sec. 3 on a combination of several existing datasets. We train and evaluate various existing shape correspondence methods on `BeCoS` (see Sec. 5.2 for an overview), demonstrating the shortcomings of existing methods, particularly for partial-to-partial matching.

Datasets. `BeCoS` is composed of seven different datasets with four-legged creatures and humanoid shapes encompassing challenging and diverse shape matching scenarios for both isometric and non-isometric deformations: TOSCA [BBK08], FAUST [BRLB14], SCAPE [ASK*05], KIDS [RRW*14], DT4D [MRSO22], SMAL [ZKJB17] and SHREC'20 [DLR*20]. See Table 1 for more information about the used datasets. From these datasets, we use a total of 2543 shapes (496 are humanoid shapes and 2047 are four-legged shapes) to generate pairwise matching problems. We partition our dataset into predefined training/validation/test splits, ensuring evaluation is performed on unseen categories. To combat dataset imbalance, we oversample the smaller datasets in our training set. The number of shape matching instances in the train/validation/test splits is 10185/137/142. We limit the test set size to easily compare to axiomatic methods, which are often computationally heavy.

Predefined Benchmark Parameters. We choose the following parameters in the data generation pipeline to create a challenging but reasonable benchmark. For re-meshing, we use quadric decimation [GH97] and we fix the range of vertices to a random number between $9000 - 10,000$ to facilitate training of current supervised/unsupervised methods and limit memory consumption. Note that re-meshing the same shape multiple times results in different re-meshed versions of the shape due to the random selection of the number of vertices. Our main setting introduces a random rotation around the z-axis, assuming each shape has a natural 'upright' direction. The maximum camera angle for the partial-to-partial setting is set to $\alpha = \frac{\pi}{4}$. Similar to [APO21] and [EGR*24], these parameters are chosen to maintain an overlap of $10\% - 90\%$, using $m = 10$ maximum iterations (see supp. mat. Sec. 9.3 for distribution of overlap). All default parameters to generate the BeCoS benchmark

are explicitly stated and explained in details in our implementation repository.

The parameter configuration of `BeCoS` is chosen to significantly challenge state-of-the-art methods while still being attainable. In the future, if the benchmark is saturated due to the development of more accurate shape matching methods, the customisable nature of our framework enables a straightforward construction of more challenging `BeCoS` variants (e.g. random rotation around all axes).

Methods	Type	F2F	P2F	P2P	Input features
GeomFMaps [DSO20]	*Sup*	✓	✓	✗	wks, xyz
DPFM [APO21]	*Sup/Unsup*	✓	✓	✓	wks, xyz
ULRSSM [CRB23]	*Unsup*	✓	✓	✗	wks, xyz
Smooth Shells [ELC20]	*Axio*	✓	✗	✗	shot
PFM [RCB*17]	*Axio*	✗	✓	✗	shot
SM-COMB [RSCB22]	*Axio*	✗	✗	✓	xyz
GC-PPSM [EGR*24]	*Axio*	✗	✗	✓	xyz

Table 2: *Overview of baselines with their methodology type (**Sup**ervised, **Unsup**ervised or **Axio**matic), their possible problem statements (F2F: Full-to-Full, P2F: Partial-to-Full and P2P: Partial-to-Partial) and their corresponding input features. All default values are used for all baselines to produce the results shown in this paper.*

5. Evaluation

We evaluate different shape matching methods, specifically showcasing the difficulty posed by partiality. We additionally present numerous ablation studies and further experiments with custom annotations (left/right annotations), showcasing the challenges and possibilities of `BeCoS`.

5.1. Evaluation Metrics

For quantitative evaluation, we consider well-established quantitative measures: we use the **Geodesic Error** to quantify the correspondence accuracy, as well as the **Intersection over Union** (IoU) and the **F1 Score** to quantify the overlap. For the ablation study on custom annotation with left/right prediction, we use a **Left/Right Accuracy** metric.

Geodesic Error. We evaluate the geodesic error of correspondences using the Princeton Protocol [KLF11]. This evaluation entails normalising the geodesic error between the ground truth and the computed correspondences by the shape diameter, defined as the square root of the area of the target shape. For partial shapes, we use the square root of the area of their corresponding full shape. To handle the partial-to-partial matching paradigm, we follow [EGR*24] and set an infinite geodesic error for unmatched vertices that should be matched and for vertices that are matched but should not be. In contrast to the partial-to-full (P2F) and full-to-full (F2F) settings, where no overlapping region is determined, the overlapping region prediction affects the geodesic error curve for the partial-to-partial setting.

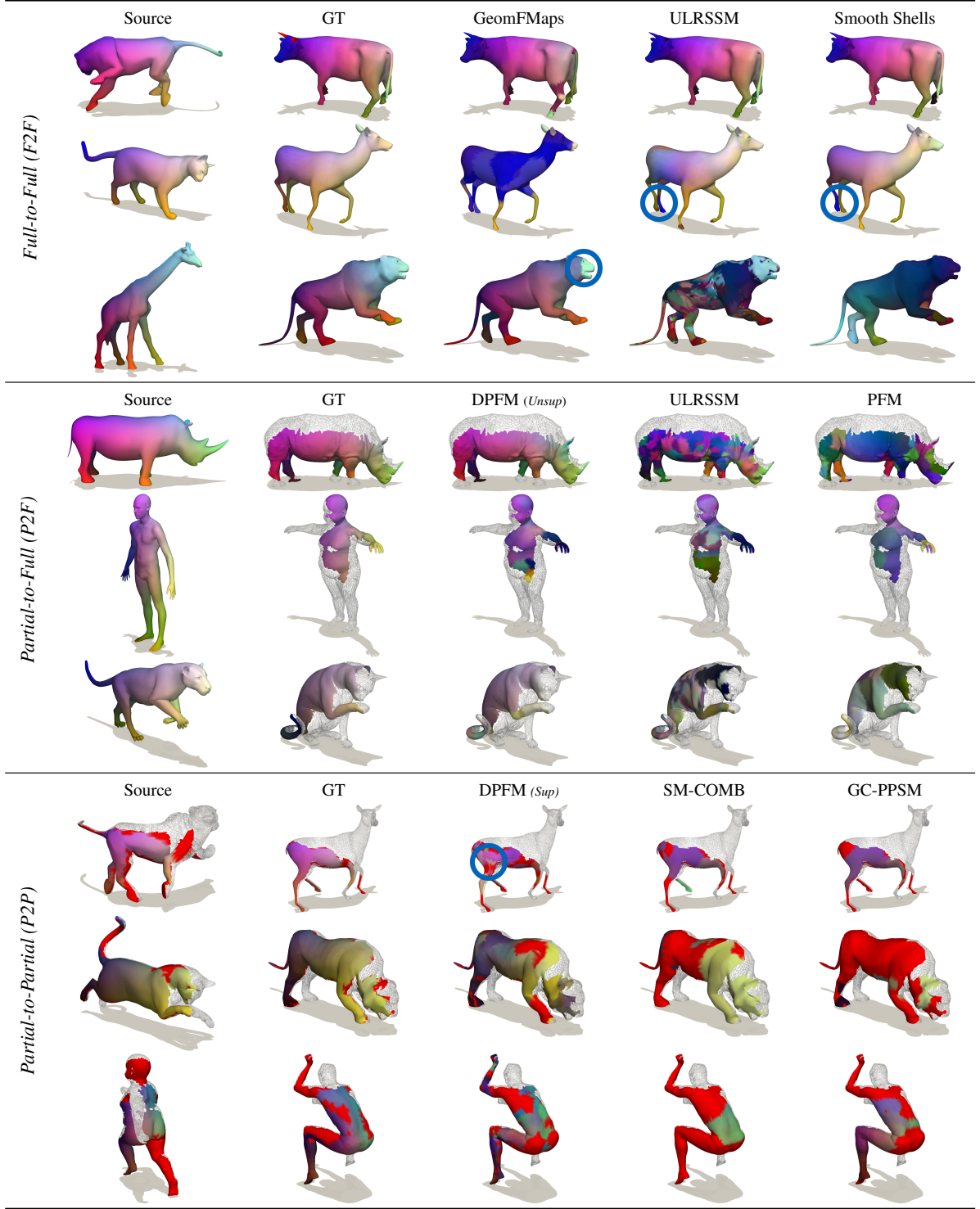**Intersection over Union.** To evaluate the quality of the overlapping

Figure 5: *Cross-dataset and cross-category shape matching results visualised via colour transfer (red denotes missing parts). All baselines show sub-optimal performance on all three settings (some examples are highlighted with blue circles). The green square shows the semantic missing parts between the source and target mesh. As GC-PPSM and SM-COMB are solved on very low resolution when upsampling, we do nearest neighbour search in the features space of the matched patches (see [EGR*24] for more information), ensuring geometric consistency in low resolution. Both methods show patch-wise inaccuracies due to this low resolution solution.*
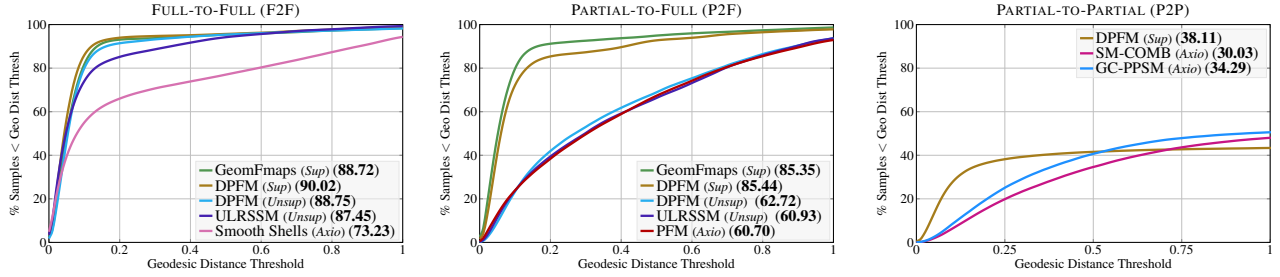
Figure 6: *We show **geodesic errors (AUC x100)** for the different methods on different settings: F2F (left), P2F (middle), P2P (right). The partial-to-partial setting is the most difficult and more realistic case, in which errors are computed for* all *vertices. Unmatched vertices are set to an infinite geodesic error; matchings that predict wrong overlaps lead to curves that do not reach 100%.*

areas induced by a matching, we consider the Intersection over Union (IoU). To this end, we define $P \in \{0,1\}^{(|V| \times 1)}$ and $G \in \{0,1\}^{(|V| \times 1)}$ as vectors (predicted and ground truth, respectively), indicating which vertices are part of the matching. We quantitatively evaluate the overlapping region in the partial-to-partial setting using the intersection over union (IoU), calculated as $\text{IoU} = \frac{|P \cap G|}{|P \cup G|}$.

F1 Score. To give further insight into the overlap prediction, we use the F1 score. It integrates precision ($\frac{\sum P \cap G}{\sum P}$) and recall ($\frac{\sum P \cap G}{\sum G}$) through the harmonic mean: F1 score = $\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$. Its possible value spans $[0,1]$ with 1 indicating a perfect precision and recall.

Left/Right Accuracy. For a shape $\mathcal{X}$ with N vertices, the left/right accuracy is defined as

$$acc_\chi = \frac{1}{N} \sum_{i \in N} \mathbb{1}(\chi_i = \chi_i^{gt}) \tag{1}$$

where χ_i is the left/right prediction for the i-th vertex in $\mathcal{X}$, $\mathbb{1}$ is the indicator function, χ_i^{gt} is the ground truth of the left/right annotation of the i-th vertex in $\mathcal{X}$. We use a similar approach for the partial-to-partial case as when using the geodesic error. For unmatched vertices that should be matched and vertices that are matched but should not be, we set an accuracy of 0.

5.2. Baselines

We choose a mix of learning-based and axiomatic methods for each of the three settings (full-to-full, partial-to-full and partial-to-partial), including SOTA methods in the supervised, unsupervised and axiomatic case. We train all learning-based methods on `BeCoS`. We show an overview of the methods in Table 2. For the partial-to-partial problem setting, an unsupervised approach has not yet been proposed in the literature. Therefore, we include two axiomatic methods instead. One reason that learning-based approaches are not extensively used for partial-to-partial shape matching is the lack of sufficient training data, which we explicitly address in this work. The severe difficulty of partial-to-partial shape matching is manifested by a significant performance degradation compared to other matching settings, see Fig. 6.

5.3. Results

We show the results on full-to-full, partial-to-full and partial-to-partial settings in Fig. 6. We additionally show the quality of the overlapping region of the partial-to-partial case in Table 3. Qualitative results can be found in Fig. 5.

Discussion. While the learning-based methods yield mostly promising results on the full geometries (see Fig. 5 F2F), existing methods struggle with the challenging partial scenarios (P2F and P2P) presented in our benchmark, demonstrating a need for further methodological advances: In the partial-to-partial case DPFM returns noisy overlapping region predictions (see Fig. 5 P2P, third column). Both SM-COMB and GC-PPSM perform poorly, probably because they rely on informative and reliable feature descriptors for computing the matching energy. Since these are not available for `BeCoS`, we resort to XYZ coordinates for computing respective energies (see supp. mat. Sec. 8.1 for a justification and analysis of this choice). In the partial-to-full case, both axiomatic and unsupervised methods return bad correspondences, both qualitatively (see Fig. 5 P2F, third and fifth column) and quantitatively (see Fig. 6 middle). We discuss further in Sec. 5.4, why our benchmark is so challenging for these settings. We believe that our work will contribute to the development of sophisticated and reliable machine learning approaches for these partial shape matching settings.

	DPFM	SM-COMB	GC-PPSM
mIoU ($\uparrow$)	46.52	46.91	49.24
F1 Score($\uparrow$)	60.28	60.61	62.42

Table 3: *Overlapping Region Prediction: We show the mean IoU (x100) and F1 Score (x100) of the partial-to-partial shape matching methods on* `BeCoS`*.*

5.4. Ablation Studies

Multiple challenges can be simultaneously present in the shape matching problem: partiality, varying scale, non-isometry, different orientations/poses, and non-alignment of pairs of shapes. Previous datasets, especially partial shape matching datasets, simplify these challenges (e.g. by aligning the shapes). In contrast, we include all

the above mentioned challenges in our benchmark, which makes it significantly more challenging.

In order to separate the effects of partiality, scaling and rotation on existing methods we conduct two ablation studies. First, we present the effects of the scaling challenge. Then, we demonstrate how random rotations adversely affect the performance of existing SOTA methods for partial-to-full shape matching.

	Methods	Default (↑)	Normalised (↑)
F2F	GeomFmaps (*Sup*)	88.72	88.96 (-0.24)
	DPFM (*Sup*)	90.20	91.47 (-1.27)
	DPFM (*Unsup*)	88.75	89.40 (-0.65)
	ULRSSM (*Unsup*)	87.45	**94.06** (-6.61)
	Smooth Shells (*Axio*)	73.23	73.93 (-0.7)
P2F	GeomFmaps (*Sup*)	85.35	89.92 (-4.57)
	DPFM (*Sup*)	85.44	**90.02** (-4.58)
	DPFM (*Unsup*)	62.72	61.85 (+0.87)
	ULRSSM (*Unsup*)	60.93	61.00 (-0.07)
	PFM (*Axio*)	60.70	62.83 (-2.03)
P2P	DPFM (*Sup*)	38.11	**39.58** (-1.47)
	SM-COMB (*Axio*)	30.03	25.09 (+4.94)
	GC-PPSM (*Axio*)	34.29	32.81 (+1.48)

Table 4: *Scale Analysis: Correspondence prediction accuracy measured by Area-under-the-curve (**AUC x100**) for baselines on all three shape matching settings. Unsupervised ULRSSM in the full-to-full setting, supervised GeomFmaps in partial-to-full and DPFM in all settings exhibit a notable performance drop (underlined) in the default compared to the normalised scale setting. The best performing method in each setting is in **bold**.*

Scaling of Shapes. To illustrate the effects of scaling, we compare the performance of multiple shape matching methods using the *default* setting of BeCoS (as discussed in Sec. 3.2) and a *normalised* setting, where the surface area of each (full) shape in the benchmark is normalised to 1. The results in Table 4 show that the performance drops when using the *default* setting for most methods. While full-to-full methods can easily handle the full-to-full variant of BeCoS under the *default* setting by normalising the surface area to 1, resulting in similarly scaled shapes. This is non-trivial for the partial-to-full and partial-to-partial settings, as each partial shape represents an unknown percentage of its full counterpart, making the normalisation non-trivial. Using the normalised surface area under these two partial settings thus increases performance, especially for the supervised learning-based methods DPFM and GeomFmaps. On the other hand, the axiomatic and unsupervised methods generally perform poorly on BeCoS, making it hard to draw a conclusion about the methods' performance in the *normalised* and *default* settings.

Alignment of Shapes. To demonstrate the impact of shape alignment on the performance of shape matching methods, we compare two different settings to probe existing methods for robustness to shape rotation on the partial-to-full setting:

1. *With rotation*: Our default realistic benchmark setting *with rotation*, where all shapes in the train and test sets have a random

rotation around the *z*-axis (i.e. all pairs of shapes are in different poses).
2. *Without rotation*: The train and test are not rotated, and shapes are kept as present in their respective datasets (most shapes are in a canonical pose).

We select the partial-to-full methods for this ablation, as the partial-to-partial methods perform prohibitively poorly, resulting in a mostly random prediction accuracy. Many full-to-full methods use wave kernel signature features [ASC11] as input, making them independent of rotations.

We show the results of our ablation in Fig. 7. A noticeable decline in performance is observed when rotation is introduced compared to when the shapes are kept as present in their datasets. This indicates that current methods benefit from biased datasets. The bias errantly favours XYZ features as input features when shapes are all aligned in a canonical pose. While reliable intrinsic (and therefore rotation-invariant) features such as wave kernel signature (wks) exist, these are not robust to partiality. This presents a challenge for future research in the domain of partial shape matching involving rotated partial shapes. One important question is how to define rotation-invariant feature descriptors that are robust to partiality. We believe our benchmark may serve as an important contribution to tackling this challenge.
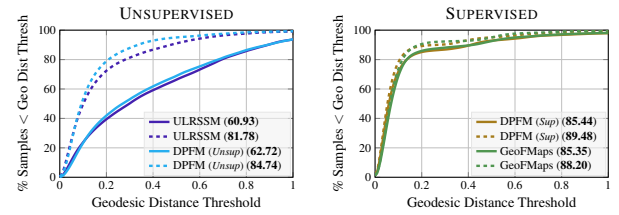

Figure 7: *Rotation evaluation for the test set of BeCoS presented in Sec. 5.4 with (solid) and without (dashed) rotation. Unsupervised methods struggle the most with rotation, whereas supervised methods can better accommodate rotational changes in the data due to their supervision signals.*

5.5. Custom Annotations: Left/Right Predictions

As a proof of concept, we show our framework's ability to propagate custom annotations throughout the dataset. To this end, we provide a quantitative evaluation of the performance of shape matching methods in predicting left/right parts on our benchmark. In this experiment, we use the **left/right accuracy** as evaluation metric (see Sec. 5.1 for more details) and compare all baselines presented in Sec. 5.2. We present the results on our benchmark in Table 5. Similar to the dense correspondence prediction, the supervised methods perform best in this task. While unsupervised methods perform well in full-to-full shape matching, in the partial-to-full case, they perform poorly due to the XYZ coordinates as input (shapes are by default randomly rotated around the *x*-axis in our benchmark). Similarly, the partial-to-partial methods perform unconvincingly.

	Methods	Left/Right Accuracy (↑)
F2F	GeomFmaps (*Sup*)	88.09
	DPFM (*Sup*)	**89.05**
	DPFM (*Unsup*)	86.92
	ULRSSM (*Unsup*)	82.13
	Smooth Shells (*Axio*)	72.88
P2F	GeomFmaps (*Sup*)	78.96
	DPFM (*Sup*)	**80.84**
	DPFM (*Unsup*)	51.24
	ULRSSM (*Unsup*)	53.39
	PFM (*Axio*)	52.85
P2P	DPFM (*Sup*)	**29.17**
	SM-COMB (*Axio*)	14.23
	GC-PPSM (*Axio*)	16.82

Table 5: *Chirality benchmark: While methods in the full-to-full setting and supervised methods in the partial-to-full setting perform well, unsupervised methods in the partial-to-full setting and all methods in the partial-to-partial setting perform poorly.*

6. Discussion and Future Work

The cross-category and cross-dataset correspondences of `BeCoS` are generated pairwise for the selected template shapes. Although they are of good quality, for cross-category shapes they may give rise to correspondence ambiguity (e.g. hooves to paws). Ambiguities may potentially accumulate if many pairwise ground truth correspondences are chained (i.e. for shapes that are distant in our shape network of shapes), which deteriorates the composed ground truth quality (see left example in Figure 9 in the supplementary material). This could potentially be improved by incorporating cycle-consistency or synchronisation techniques [HG13; PKS13; BCT21] in the process of generating ground truth correspondences.

7. Conclusion

While several shape matching datasets exist, they are typically static and limited in size, which constrains their adaptability to diverse problem settings and custom use cases. To address this we connect multiple existing mesh-based datasets through manual cross-category annotations, enabling dense correspondence and annotation propagation across a diverse set of shapes. Building on this foundation, we introduced a scalable framework for generating unlimited training instances for partial shape correspondence, simulating realistic occlusions and semantic variability. As a concrete instantiation, we release a challenging non-rigid shape surface matching benchmark and provided a comprehensive evaluation of current state-of-the-art methods. Our findings highlight persistent challenges, particularly in partial shape matching, and underscore the need for more robust and generalizable solutions. We hope our benchmark and framework will foster future advancements in shape matching research.

Acknowledgments

References

[AEOW23] ABDELREHEEM, AHMED, ELDESOKEY, ABDELRAHMAN, OVSJANIKOV, MAKS, and WONKA, PETER. "Zero-shot 3d shape correspondence". *SIGGRAPH Asia 2023 Conference Papers*. 2023, 1–11 1, 5.

[ALO23] ATTAIKI, SOUHAIB, LI, LEI, and OVSJANIKOV, MAKS. "Generalizable Local Feature Pre-training for Deformable Shape Analysis". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023 4.

[APO21] ATTAIKI, SOUHAIB, PAI, GAUTAM, and OVSJANIKOV, MAKS. "Dpfm: Deep partial functional maps". *2021 International Conference on 3D Vision (3DV)*. IEEE. 2021, 175–185 2–6, 17.

[ASC11] AUBRY, MATHIEU, SCHLICKEWEI, ULRICH, and CREMERS, DANIEL. "The wave kernel signature: A quantum mechanical approach to shape analysis". *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. Nov. 2011, 1626–1633. DOI: 10.1109/ICCVW.2011.6130444. (Visited on 10/20/2023) 3, 9.

[ASK*05] ANGUELOV, DRAGOMIR, SRINIVASAN, PRAVEEN, KOLLER, DAPHNE, et al. "Scape: shape completion and animation of people". *ACM SIGGRAPH 2005 Papers* (2005), 408–416 2, 3, 6.

[BBK08] BRONSTEIN, ALEXANDER, BRONSTEIN, MICHAEL, and KIMMEL, RON. *Numerical Geometry of Non-Rigid Shapes*. 1st ed. Springer Publishing Company, Incorporated, 2008. ISBN: 0387733000 2–4, 6.

[BCT21] BERNARD, FLORIAN, CREMERS, DANIEL, and THUNBERG, JOHAN. "Sparse quadratic optimisation over the Stiefel manifold with application to permutation synchronisation". *Advances in Neural Information Processing Systems* 34 (2021), 25256–25266 10.

[BDK23] BRACHA, AMIT, DAGÈS, THOMAS, and KIMMEL, RON. "On Partial Shape Correspondence and Functional Maps". *arXiv preprint arXiv:2310.14692* (2023) 2, 3, 5.

[BDK24] BRACHA, AMIT, DAGÈS, THOMAS, and KIMMEL, RON. "Wormhole Loss for Partial Shape Matching". *arXiv preprint arXiv:2410.22899* (2024) 3.

[BLZ*20] BAI, XUYANG, LUO, ZIXIN, ZHOU, LEI, et al. "D3feat: Joint learning of dense detection and description of 3d local features". *CVPR*. 2020 3.

[BRLB14] BOGO, FEDERICA, ROMERO, JAVIER, LOPER, MATTHEW, and BLACK, MICHAEL J. "FAUST: Dataset and evaluation for 3D mesh registration". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, 3794–3801 1–3, 6.

[BST20] BERNARD, FLORIAN, SURI, ZEESHAN KHAN, and THEOBALT, CHRISTIAN. "MINA: Convex Mixed Integer Programming for Non-Rigid Shape Alignment". *CVPR* (2020) 3.

[BXNL24] BASTIAN, LENNART, XIE, YIZHENG, NAVAB, NASSIR, and LÄHNER, ZORAH. "Hybrid Functional Maps for Crease-Aware Non-Isometric Shape Matching". *CVPR* (2024) 1, 3.

[CB22] CAO, DONGLIANG and BERNARD, FLORIAN. "Unsupervised deep multi-shape matching". *European Conference on Computer Vision*. Springer. 2022, 55–71 3.

[CCZZ24] CHEN, ZHIQIN, CHEN, QIMIN, ZHOU, HANG, and ZHANG, HAO. "DAE-Net: Deforming Auto-Encoder for fine-grained shape co-segmentation". *ACM SIGGRAPH 2024 Conference Papers*. SIGGRAPH '24. Denver, CO, USA: Association for Computing Machinery, 2024. ISBN: 9798400705250. DOI: 10.1145/3641519.3657528. URL: https://doi.org/10.1145/3641519.3657528 4.

[CEE*24] CAO, DONGLIANG, EISENBERGER, MARVIN, EL AMRANI, NAFIE, et al. "Spectral Meets Spatial: Harmonising 3D Shape Matching and Interpolation". *CVPR*. 2024 1.

[CRB*16] COSMO, LUCA, RODOLA, EMANUELE, BRONSTEIN, MICHAEL M, et al. "SHREC'16: Partial matching of deformable shapes". *Proc. 3DOR* 2.9 (2016), 12 2–5.

[CRB23] CAO, DONGLIANG, ROETZER, PAUL, and BERNARD, FLORIAN. "Unsupervised Learning of Robust Spectral Shape Matching". *ACM Transactions on Graphics (TOG)* (2023). DOI: 10.1145/3592107 1–3, 6, 13, 17.

[DLH23] DECATUR, DALE, LANG, ITAI, and HANOCKA, RANA. "3d highlighter: Localizing regions on 3d shapes via text descriptions". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, 20930–20939 3.

[DLR*20] DYKE, ROBERTO M, LAI, YU-KUN, ROSIN, PAUL L, et al. "SHREC'20: Shape correspondence with non-isometric deformations". *Computers & Graphics* 92 (2020), 28–43 1–3, 5, 6.

[DMM24] DUTT, NILADRI SHEKHAR, MURALIKRISHNAN, SANJEEV, and MITRA, NILOY J. "Diffusion 3D Features (Diff3F): Decorating Untextured Shapes with Distilled Semantic Features". *CVPR*. 2024 3.

[DSO20] DONATI, NICOLAS, SHARMA, ABHISHEK, and OVSJANIKOV, MAKS. "Deep Geometric Maps: Robust Feature Learning for Shape Correspondence". *CVPR* (2020) 3, 6, 17.

[DYDZ22] DENG, BAILIN, YAO, YUXIN, DYKE, ROBERTO M., and ZHANG, JUYONG. "A Survey of Non-Rigid 3D Registration". en. *Computer Graphics Forum* 41.2 (2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14502, 559–589. ISSN: 1467-8659. DOI: 10.1111/cgf.14502. (Visited on 10/20/2023) 3.

[DYT05] DINH, HUONG QUYNH, YEZZI, ANTHONY, and TURK, GREG. "Texture transfer during shape transformation". *ACM Transactions on Graphics (ToG)* 24.2 (2005), 289–310 1.

[ECB24] EL AMRANI, NAFIE, CAO, DONGLIANG, and BERNARD, FLORIAN. "A Universal and Flexible Framework for Unsupervised Statistical Shape Model Learning". *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Oct. 2024 1.

[EGR*24] EHM, VIKTORIA, GAO, MAOLIN, ROETZER, PAUL, et al. "Partial-to-Partial Shape Matching with Geometric Consistency". *CVPR* (2024) 1–3, 5–7, 13, 17.

[ELC20] EISENBERGER, MARVIN, LÄHNER, ZORAH, and CREMERS, DANIEL. "Smooth Shells: Multi-Scale Shape Registration With Functional Maps". *CVPR*. 2020 1, 6, 17.

[ENK*21] EISENBERGER, MARVIN, NOVOTNY, DAVID, KERCHENBAUM, GAEL, et al. "Neuromorph: Unsupervised shape interpolation and correspondence in one go". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 7473–7483 3.

[ERE*23] EHM, VIKTORIA, ROETZER, PAUL, EISENBERGER, MARVIN, et al. "Geometrically Consistent Partial Shape Matching". *International Conference on 3D Vision (3DV)* (2023) 2.

[Fac23] FACEFORM LLC. *FaceForm*. https://faceform.com/. Version 2023.11.4. Version: 2023.11.4. 2023. URL: https://faceform.com/ 4, 14.

[FLN*24] FISCHER, MICHAEL, LI, ZHENGQIN, NGUYEN-PHUOC, THU, et al. "NeRF Analogies: Example-Based Visual Attribute Transfer for NeRFs". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 4640–4650 3.

[GH97] GARLAND, MICHAEL and HECKBERT, PAUL S. "Surface simplification using quadric error metrics". *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. 1997, 209–216 6.

[HG13] HUANG, QI-XING and GUIBAS, LEONIDAS. "Consistent shape maps via semidefinite programming". *Computer graphics forum*. Vol. 32 (5). 2013, 177–186 10.

[HGU*21] HUANG, SHENGYU, GOJCIC, ZAN, USVYATSOV, MIKHAIL, et al. "Predator: Registration of 3d point clouds with low overlap". *CVPR*. 2021 3.

[HLR*19] HALIMI, OSHRI, LITANY, OR, RODOLA, EMANUELE, et al. "Unsupervised learning of dense shape correspondence". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, 4370–4379 2.

[HSS*09] HASLER, NILS, STOLL, CARSTEN, SUNKEL, MARTIN, et al. "A statistical model of human pose and body shape". *Computer graphics forum*. Vol. 28 (2). Wiley Online Library. 2009, 337–346 1.

[KHS10] KALOGERAKIS, EVANGELOS, HERTZMANN, AARON, and SINGH, KARAN. "Learning 3D mesh segmentation and labeling". *ACM Trans. Graph.* 29.4 (July 2010). ISSN: 0730-0301. DOI: 10.1145/1778765.1778839. URL: https://doi.org/10.1145/1778765.1778839 4.

[KLF11] KIM, VLADIMIR G, LIPMAN, YARON, and FUNKHOUSER, THOMAS. "Blended intrinsic maps". *ACM transactions on graphics (TOG)* 30.4 (2011), 1–12 6.

[KMP*21] KIRGO, MAXIME, MELZI, SIMONE, PATANE, GIUSEPPE, et al. "Wavelet-based Heat Kernel Derivatives: Towards Informative Localized Shape Analysis". *Computer Graphics Forum*. Vol. 40 (1). Wiley Online Library. 2021, 165–179 2, 3, 5.

[KMS22] KOBAYASHI, SOSUKE, MATSUMOTO, EIICHI, and SITZMANN, VINCENT. "Decomposing nerf for editing via feature field distillation". *Advances in Neural Information Processing Systems* 35 (2022), 23311–23330 3.

[LH22] LI, YANG and HARADA, TATSUYA. "Lepard: Learning partial point cloud matching in rigid and deformable scenes". *CVPR*. 2022 3.

[LRB*16] LITANY, OR, RODOLÀ, EMANUELE, BRONSTEIN, ALEXANDER M, et al. "Non-rigid puzzles". *Computer Graphics Forum*. Vol. 35 (5). Wiley Online Library. 2016, 135–143 3.

[LRR*17] LITANY, OR, REMEZ, TAL, RODOLA, EMANUELE, et al. "Deep functional maps: Structured prediction for dense shape correspondence". *Proceedings of the IEEE international conference on computer vision*. 2017, 5659–5667 3.

[LTT*21] LI, YANG, TAKEHARA, HIKARI, TAKETOMI, TAKAFUMI, et al. "4dcomplete: Non-rigid motion estimation beyond the observable surface". *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, 12706–12716 2–4, 14.

[MMR*19] MELZI, S., MARIN, R., RODOLÀ, E., et al. "Matching Humans with Different Connectivity". *Eurographics Workshop on 3D Object Retrieval*. Ed. by BIASOTTI, SILVIA, LAVOUÉ, GUILLAUME, and VELTKAMP, REMCO. 2019 2, 3.

[MRR*19] MELZI, SIMONE, REN, JING, RODOLÀ, EMANUELE, et al. "ZoomOut: spectral upsampling for efficient shape correspondence". *ACM Trans. Graph.* 38.6 (Nov. 2019). ISSN: 0730-0301. DOI: 10.1145/3355089.3356524 3.

[MRSO22] MAGNET, ROBIN, REN, JING, SORKINE-HORNUNG, OLGA, and OVSJANIKOV, MAKS. "Smooth non-rigid shape matching via effective dirichlet energy optimization". *2022 International Conference on 3D Vision (3DV)*. IEEE. 2022, 495–504 2, 3, 6, 14.

[OBS*12] OVSJANIKOV, MAKS, BEN-CHEN, MIRELA, SOLOMON, JUSTIN, et al. "Functional maps: a flexible representation of maps between shapes". *ACM Transactions on Graphics (TOG)* 31.4 (2012), 30:1–30:11 1, 3.

[ODM*23] OQUAB, MAXIME, DARCET, TIMOTHÉE, MOUTAKANNI, THÉO, et al. "Dinov2: Learning robust visual features without supervision". *arXiv preprint arXiv:2304.07193* (2023) 3.

[PKO22] PANINE, MIKHAIL, KIRGO, MAXIME, and OVSJANIKOV, MAKS. "Non-Isometric Shape Matching via Functional Maps on Landmark-Adapted Bases". *Computer graphics forum*. Vol. 41 (6). Wiley Online Library. 2022, 394–417 1.

[PKS13] PACHAURI, DEEPTI, KONDOR, RISI, and SINGH, VIKAS. "Solving the multi-way matching problem by permutation synchronization". *Advances in neural information processing systems* 26 (2013) 10.

[QYW*25] QIU, XIAOWEN, YANG, JINCHENG, WANG, YIAN, et al. "Articulate AnyMesh: Open-vocabulary 3D Articulated Objects Modeling". *arXiv preprint arXiv:2502.02590* (2025) 4.

[RAC*24] ROETZER, PAUL, ABBAS, AHMED, CAO, DONGLIANG, et al. "DiscoMatch: Fast Discrete Optimisation for Geometrically Consistent 3D Shape Matching". *In Proceedings of the European conference on computer vision (ECCV)*. 2024 3.

[RB24] ROETZER, PAUL and BERNARD, FLORIAN. "SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 14543–14553 3.

[RBL*22] ROMBACH, ROBIN, BLATTMANN, ANDREAS, LORENZ, DOMINIK, et al. "High-resolution image synthesis with latent diffusion models". *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, 10684–10695 3.

[RCB*17] RODOLÀ, EMANUELE, COSMO, LUCA, BRONSTEIN, MICHAEL M, et al. "Partial functional correspondence". *Computer graphics forum*. Vol. 36 (1). Wiley Online Library. 2017, 222–236 2, 3, 6, 17.

[RRW*14] RODOLA, EMANUELE, ROTA BULO, SAMUEL, WINDHEUSER, THOMAS, et al. "Dense non-rigid shape correspondence using random forests". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, 4177–4184 3, 6.

[RSCB22] ROETZER, PAUL, SWOBODA, PAUL, CREMERS, DANIEL, and BERNARD, FLORIAN. "A scalable combinatorial solver for elastic geometrically consistent 3d shape matching". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, 428–438 1–3, 6, 13, 17.

[RSO19] ROUFOSSE, JEAN-MICHEL, SHARMA, ABHISHEK, and OVSJANIKOV, MAKS. "Unsupervised deep learning for structured shape matching". *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, 1617–1627 2.

[SACO22] SHARP, NICHOLAS, ATTAIKI, SOUHAIB, CRANE, KEENAN, and OVSJANIKOV, MAKS. "Diffusionnet: Discretization agnostic learning on surfaces". *ACM Transactions on Graphics (TOG)* 41.3 (2022), 1–16 3.

[Sah20] SAHILLIOĞLU, YUSUF. "Recent advances in shape correspondence". en. *The Visual Computer* 36.8 (Aug. 2020), 1705–1721. ISSN: 1432-2315. DOI: 10.1007/s00371-019-01760-0. (Visited on 04/11/2022) 3.

[SOG09] SUN, JIAN, OVSJANIKOV, MAKS, and GUIBAS, LEONIDAS. "A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion". en. *Computer Graphics Forum* 28.5 (2009), 1383–1392. ISSN: 1467-8659. (Visited on 10/20/2023) 3.

[TSD10] TOMBARI, FEDERICO, SALTI, SAMUELE, and DI STEFANO, LUIGI. "Unique signatures of histograms for local surface description". *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11*. Springer. 2010, 356–369 3.

[WHC*16] WEI, LINGYU, HUANG, QIXING, CEYLAN, DUYGU, et al. "Dense human body correspondences using convolutional networks". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 1544–1553 3.

[WSSC11] WINDHEUSER, THOMAS, SCHLICKEWEI, ULRICH, SCHMIDT, FRANK R, and CREMERS, DANIEL. "Geometrically consistent elastic matching of 3d shapes: A linear programming solution". *2011 International Conference on Computer Vision*. IEEE. 2011, 2134–2141 3.

[WWO24] WIMMER, THOMAS, WONKA, PETER, and OVSJANIKOV, MAKS. "Back to 3D: Few-Shot 3D Keypoint Detection with Back-Projected 2D Features". *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024 1.

[YQH*23] YU, HAO, QIN, ZHENG, HOU, JI, et al. "Rotation-invariant transformer for point cloud matching". *CVPR*. 2023 3.

[ZKJB17] ZUFFI, SILVIA, KANAZAWA, ANGJOO, JACOBS, DAVID, and BLACK, MICHAEL J. "3D Menagerie: Modeling the 3D Shape and Pose of Animals". *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. July 2017 2, 3, 6.

Beyond Complete Shapes: A Benchmark for Quantitative Evaluation of 3D Shape Surface Matching Algorithms

Supplementary Material

The supplementary material provides more information about our framework and the `BeCoS` benchmark. In Sec. 8, we show additional results on the partial-to-partial (P2P) setting and qualitative results of custom annotation propagation. In Sec. 9, we provide details on the procedure followed to create our framework. Additionally, we present details about the `BeCoS` benchmark in Sec. 10. We discuss the licenses of our framework and the assets used to create the `BeCoS` benchmark and our long-term preservation plan in Sec. 11. In Sec. 12, we provide information about the computation resources used for the baseline methods.

8. The `BeCoS` Benchmark: Additional Results

8.1. Alternative Energies on the Partial-to-Partial Setting

In the main paper, we use XYZ coordinates for the axiomatic methods SM-COMB [RSCB22] and GC-PPSM [EGR*24] for the partial-to-partial setting. While these methods work best with learned features (e.g. unsupervised features from ULRSSM [CRB23]), the latter features are not reliable for our challenging benchmark dataset (see Fig. 6 in the main paper for the performance of ULRSSM in the partial-to-full setting). In our case, we found that XYZ coordinates as feature descriptors are most suited for our benchmark. We compare the XYZ features to the original energy formulation used for SM-COMB (a combination of bending and membrane energy) and the rotation-invariant SHOT features. We show the results in Fig. 8 for the geodesic distance comparison and Table 6 for the overlapping region prediction. Both show that XYZ as input features work best in the given setting. This is the case because in `BeCoS` the shapes are rotated only around the z-axis, which leaves the z-coordinates as reliable information. While this is not always a realistic setting, as it would not work for completely random rotations, we choose this setting to be able to evaluate the axiomatic methods in a (reasonably) meaningful manner. As these methods will benefit from better features, we hope to encourage further research in this direction.

Method		Metrics	
		mIoU (↑)	F1 Score (↑)
SM-COMB	XYZ	**46.91**	**60.61**
	Original	23.81	33.78
	Shot	24.47	34.29
GC-PPSM	XYZ	**49.24**	**62.42**
	Original	43.69	53.72
	Shot	39.89	51.20

Table 6: *Overlapping Region Prediction: We show the mean IoU (x100) and F1 Score (x100) of the partial-to-partial shape matching methods. Out of the three options, XYZ features are the input features that suit both methods the best.*
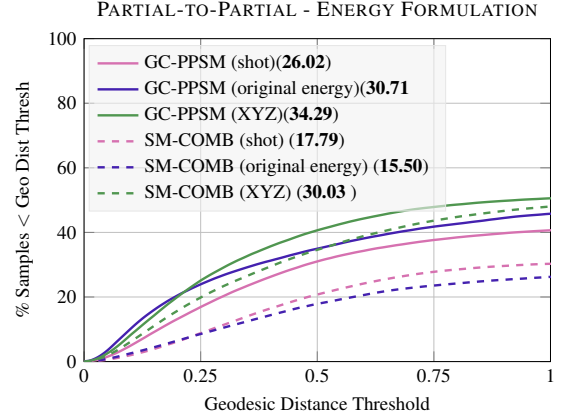

Figure 8: *We compare **different energy formulations** as input for SM-COMB and GC-PPSM. For both methods, XYZ coordinates work best as input features.*

8.2. Custom Annotation Propagation

We show more qualitative results of Left/Right annotation propagation using our framework in Fig. 9. We also include failure cases of the annotation propagation. This propagation relies on the manually created dense correspondence between templates in our framework; small errors caused by the manual annotation get amplified throughout the propagation and cause unwanted artefacts, especially in regions like the tails of animals.
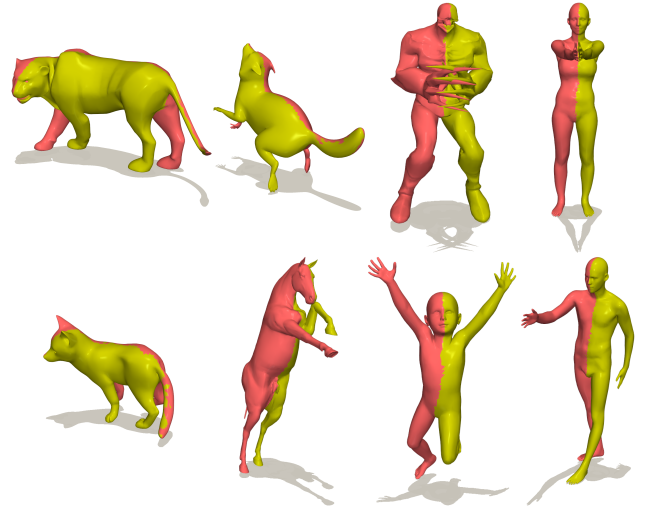

Figure 9: *Our framework allows for the **propagation of custom annotations** throughout all shapes. We show the case of left/right annotations. The Left/Right annotations exhibit errors in thin parts of the shapes (e.g. tails of animals).*

9. Details on the Procedural Data Generation

9.1. Correspondence Generation

This procedure is similar to the one used by [MRSO22], and we use the FaceForm software with an academic license [Fac23]. The steps of our procedure can be summarised as follows:

1. We select a template for each dataset with cross-category correspondence and a template for each category in each dataset that does not have cross-category correspondences. The template should ideally be in a neutral A-or T-pose for humanoid shapes and in an "upright" standing position for four-legged animals.
2. We build a shape network connecting all templates in such a way that the semantic differences between two connected shapes are small. This makes the morphing step easier. For example, we try to connect all cats from all datasets together (see the interactive graph in our project page).
3. For each edge in the shape network, we load the two connected templates *A* and *B* in FaceForm using the "LoadGeom" module (see Fig. 11 **(c)**).
4. Some template shapes do not completely match the other template as some parts are missing (see Fig. 12). For these shapes, we first cut the shapes so that we get a matchable and a semantic missing part. For these template shapes, we only load the matchable part in FaceForm.
5. We connect the two modules to a "SelectPointPairs" module that allows manually selecting corresponding landmarks on the templates *A* and *B* (see Fig. 11 **(a)**). Depending on the semantic difference between the two template shapes, we select between 15 and 95 landmarks.
6. Once the landmarks are placed manually on both shapes, we use the "Wrapping" module on FaceForm to morph template *A* to template *B*. This process took around 25 hours of manual work.
7. We then save the morphed template $\hat{A}$ (see Fig. 11 **(b)**).
8. To get the correspondences between the two shapes, we find for every vertex $\hat{A}$ the closest triangle of template shape *B*. We then project every vertex of $\hat{A}$ on the closest triangle and return the barycentric coordinates. The correspondences of semantic missing parts are set to -1.
9. Once steps [3-7] are done for all edges in the tree, we are able to generate the correspondences between any two input shapes.

9.2. Correspondence Computation between a Shape Pair

Given any two shapes in the dataset, we compute the ground-truth correspondences between them using the following steps: First, we extract the path between those two shapes over the given template shapes. In the interactive shape network on our project page, we show the chosen templates and their inter-connectivity. We then extract the correspondences between any two shapes in the shape network either through the given datasets or by using our manually annotated correspondences. We then propagate the correspondences along the path.

9.3. Overlap Region for the P2P Setting

For the overlap computation, we rigidly align the shapes $\mathcal{X}$ and $\mathcal{Y}$ using the given correspondences to ensure that both partial shapes

are semantically similar and perform the in the main paper described partiality generation on both shapes. Subsequently, we quantify the overlap as the percentage of vertices from $\hat{\mathcal{X}}$ that are mapped to $\hat{\mathcal{Y}}$ and vice-versa ($\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ are the partial shape generated using ray casting from shapes $\mathcal{X}$ and $\mathcal{Y}$, respectively). If both overlaps do not fall within a predefined range, the process iterates with new random camera poses until the overlap meets the specified criteria, or until a maximum of *m* iterations is reached. In the latter case, the partial shapes whose overlap is the closest to the desired range are selected.

In the `BeCoS` default configuration for the partial-to-partial setting, shapes have an overlap between 10% and 90%. We show a histogram of the distribution of the overlapping region of each shape pair going from shape $\mathcal{X}$ to shape $\mathcal{Y}$ and vice-versa (see Fig. 10).
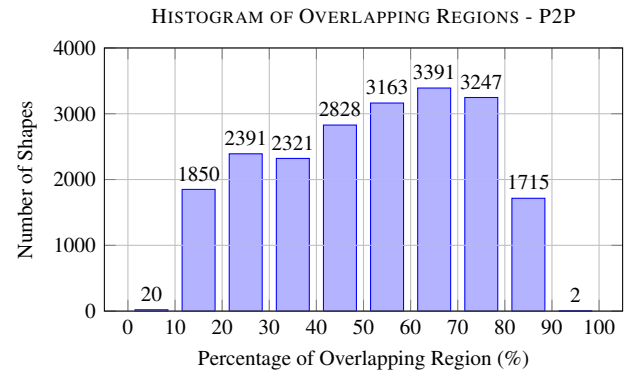

Figure 10: *Partial-to-Partial **histogram of overlapping region** of `BeCoS`. The pairs that are below 10% and the ones above 90% are the ones that reached the maximum number of iterations m = 10. Our framework produces partial-to-partial shapes, creating this histogram when the predefined parameters discussed in Sec. 4 of the main paper are used.*

10. Details about the `BeCoS` Benchmark and the Framework

As mentioned in the main paper, we use seven different full datasets in our benchmark. We use each dataset entirely with two exceptions: For TOSCA we exclude the gorilla as it is excluded from most shape matching analyses due to its difficulty. The original dataset [LTT*21] of DT4D provides animated meshes for each category with a lot of categories. Some categories in the original dataset are difficult to semantically match to other categories, e.g. dragons. Therefore, we use the same categories that were selected from DT4D-Matching [MRSO22]. This dataset provides its own cross-dataset correspondences but only for a version of the dataset where each mesh is around 8000 vertices. Since the shapes are only re-meshed and not in the original resolution, we use the original animated files but extract the same categories and frames from the animation as DT4D-Matching. This ensures that we have higher resolution input shapes and less error propagation along the correspondence paths, as the shapes in the same category have the same triangulation. We end up with 39 categories for four-legged animals and eight categories of humanoid shapes.
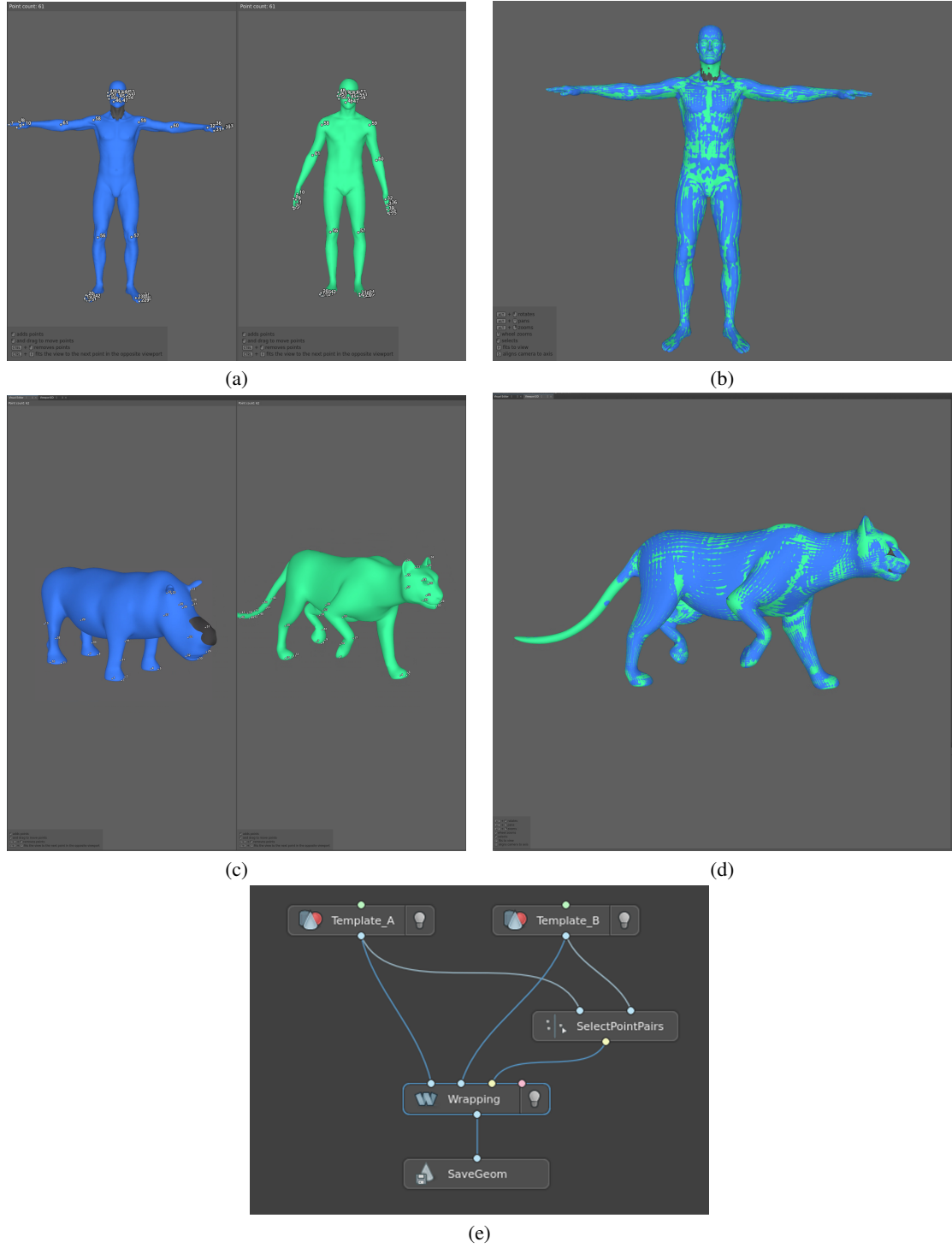
(a)

(b)

(c)

(d)

(e)

Figure 11: *Usage of FaceForm to generate the cross-category and cross-dataset ground-truth correspondences. **(a)** Given two templates, we manually place corresponding landmarks on both templates. **(b)** Overlap of the **template A** (blue) and the **morphed template B** (green) on top of each other after running the "Wrapping" module. **(c)** and **(d)** We show a non-isometric deformation between a puma and rhino shape. Note that the rhino is missing its horn which is considered a missing semantic part and we only match the rhino shape without its horn to the puma. **(e)** The modules used to perform the morphing of **template B** to **template A**.*
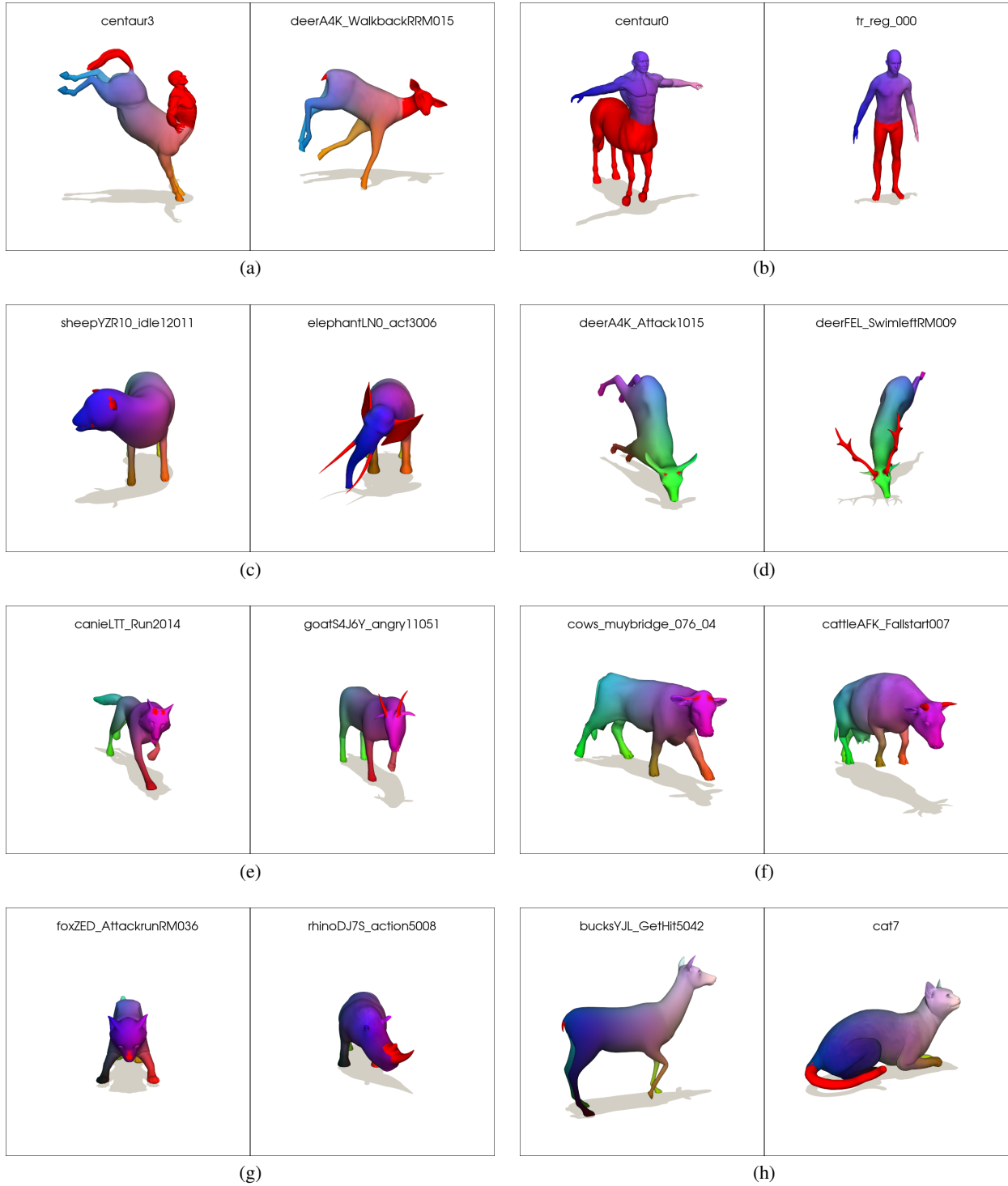
Figure 12: *Correspondence between pairs with semantic missing parts. The colour is transferred from the left shape to the right shape. The red parts are the parts that cannot be semantically matched between the two shapes.* **(a)** *and* **(b)** *The centaur shape is a special case that can be matched to both humanoid and four-legged animals.* **(c)** *The elephant's tusks and ears cannot be matched to a sheep.* **(d), (e)** *and* **(f)** *The horns are not always matched between shapes.* **(g)** *The rhino's horn is not semantically matched to any other animal in* BeCoS. **(h)** *The cat's long tail is not matched to the buck's small fluffy tail.*

11. Hosting, Licences, Maintenance Plan & Long-Term Preservation of `BeCoS`

Our pipeline is based on existing 3D shape datasets. The original datasets have to be downloaded and can only be used with the corresponding licences from the original datasets. We ensure this correct usage by letting the user download the data directly from the corresponding websites. Our code for generating the (partial) shape pairs will be available on GitHub with a CC BY-NC-SA 4.0 (open source for non-commercial use) licence.

Our main product is a publicly accessible framework, which we intend to permanently host on our GitHub repository. Since the `BeCoS` benchmark is a specific instantiation of this universal framework, its accessibility depends on the availability of certain 3D shape datasets. The biggest risk regarding accessibility is therefore related to the `BeCoS` benchmark in case some of the involved 3D shape datasets become unavailable. However, given that the 3D shape datasets used in our framework are well-established, we anticipate their continued public availability. In the event that a dataset is taken down by the originators, our framework's functionality will remain intact for the remaining datasets.

12. Compute and Resources

We use an internal cluster to compute the given results with the different models. For SM-COMB and GC-PPSM, we use an Intel Xeon E5-2697 with 16 cores and up to 36GB of RAM per run. For the learning based methods, we use an Intel Xeon E5-2697 with 16 cores and one Nvidia Titan GPU with 12G VRAM. As PFM's precompiled code runs only on Windows, we use an Intel i7-14700KF with 20 cores and up to 32GB of RAM. We present the compute time per method used in our paper in Table. 7.

Method	Training time	Evaluation time (per shape pair)
DPFM [APO21]	8-20h	1.1s
SM-COMB [RSCB22]	N/A	0.08h
GC-PPSM [EGR*24]	N/A	2.5h
PFM [RCB*17]	N/A	0.2h
ULRSSM [CRB23]	16-86h	20s
Smooth Shells [ELC20]	N/A	0.2h
GeomFMaps [DSO20]	3-7h	1.1s

Table 7: *A breakdown of the **compute time** per method presented as a baseline in our work. Axiomatic methods do not require any training time. We show the training time for the entire training data and the evaluation time per shape pair.*

13. All Framework Options

In Table 8, we provide a full list of the general options available in our framework. We invite the reader to check our GitHub repository for further information.

V. Ehm, N. El Amrani et al. / BeCoS: Beyond Complete Shapes

Table 8: ***All options in the main configuration file.*** *The default values of these options are stored on our framework code and allow the generation of the* `BeCoS` *benchmark.*

Option	Description
`data_dir`	Path to data directory.
`datasets`	True if the dataset is used.
`combinations`	Type of shapes to generate, possible values: human, four-legged, human_centaur, four-legged_centaur, or all.
`setting`	The setting to generate, possible values: full_full/partial_full/partial_partial.
`remesh`	True if the shapes should be re-meshed after loading.
`cam_pos_regime`	Specifies how far the pair of camera poses are apart, possible values: high/medium/low.
`store_vis`	If True, store visualization of each pair of shapes.
`show_output`	If True, an open3d visualization is shown during generation.
`original_settings`	If True, multiple assertions are activated to ensure that `BeCoS` is generated.
`use_precompute_remeshing`	If True, use the cached re-meshed shapes if available.
`update_precomputed_remeshed`	If True, update the caching directory for re-meshed meshes.
`use_precomputed_partial_raycasting`	If True, use the cached triangle IDs of each partial shape if available.
`update_precomputed_raycasting`	If True, update the caching directory for triangle IDs.
`one_axis_rotation`	If True, perform a random rotation around the z-axis after generating the pair of shapes.
`n_cam_pos`	The number of camera poses to sample to get within the desired range of overlap.
`min_overlap`	Minimum overalp between partial shapes desired.
`max_overlap`	Maximum overalp between partial shapes desired.