# Surface Networks via General Covers

Niv Haim *           Nimrod Segol *        Heli Ben-Hamu        Haggai Maron
Yaron Lipman

Weizmann Institute of Science
Rehovot, Israel

## Abstract

*Developing deep learning techniques for geometric data is an active and fruitful research area. This paper tackles the problem of sphere-type surface learning by developing a novel surface-to-image representation. Using this representation we are able to quickly adapt successful CNN models to the surface setting.*

*The surface-image representation is based on a covering map from the image domain to the surface. Namely, the map wraps around the surface several times, making sure that every part of the surface is well represented in the image. Differently from previous surface-to-image representations, we provide a low distortion coverage of all surface parts in a single image. Specifically, for the use case of learning spherical signals, our representation provides a low distortion alternative to several popular spherical parameterizations used in deep learning.*

*We have used the surface-to-image representation to apply standard CNN architectures to 3D models including spherical signals. We show that our method achieves state of the art or comparable results on the tasks of shape retrieval, shape classification and semantic shape segmentation.*

## 1. Introduction

Adapting deep learning methods to geometric data (*e.g.*, shapes) is a vibrant research area that has already produced state of the art algorithms for several geometric learning tasks (*e.g.*, [32, 33, 38]).

Two prominent approaches are: (i) mapping the geometric data to tensors (*e.g.*, images) and using off-the-shelf convolutional neural network (CNN) architectures and optimization techniques [38, 45, 36, 25]; and (ii) developing novel architectures and optimization techniques that are tailored to the geometric data [26, 32, 33]. An important benefit of (i) is in reducing the geometric learning task to an image learning one, allowing to harness the *huge* algorithmic

---
*Equal contribution

progress of neural networks for images directly to geometric data.

Some previous attempts, following (i), to perform learning tasks on geometric data use projections to 2D planes, *e.g.*, by rendering the shapes [38]. Such projections are not injective and suffer from occlusions, thus often require a collection of projections for a single shape. Other methods embed the shape in an encapsulating 3D grid [45, 27]; these methods require dealing with higher dimensional tensors and are usually less robust to deformations. Other methods [36, 25] try to find low distortion 2D mappings to an image domain. In this case the intrinsic dimensionality of the data is preserved, however, these maps suffer from high distortion and/or ignore the difference in the topologies of the surface (no boundary) and the image (with boundary).

In this paper, we advocate a novel 2D mapping method for representing sphere-type (genus zero, *e.g.*, the human model in Figure 1a, left) surfaces as images. The challenge in using an image to represent a surface has two aspects: geometrical and topological. Geometrically, a general curved surface cannot be mapped to a flat domain (*i.e.*, the image) without introducing a significant distortion. Topologically, an image has a boundary while sphere-type surfaces do not; hence, any mapping between the two will introduce cuts and discontinuities. Furthermore, a naive application of 2D convolution to the image would be ambiguous on the surface (see Figure 2 and Subsection 3.1).

To address these challenges we think of the image as a periodic domain (*i.e.*, a torus) and relax the notion of a one-to-one mapping to that of a *covering map* from the image domain onto the surface. That is, we construct a mapping from the image domain to the surface that covers the surface several times. For example, Figure 1a visualizes a degree-5 covering map. Meaning, the surface appears 5 times in the image; note how each part of the surface appears with low distortion at-least once in the image. The image generated by our covering map is periodic, namely its left and right boundaries as well as its bottom and top boundaries correspond, making the image boundaryless. Importantly,

since image convolution is well defined on a torus, it will translate to a continuous convolution-like operator on the surface [25].

Applying our method to surface learning is easy: use a covering map to transfer functions of interest over the input surfaces (*e.g.*, the coordinate functions) to images and apply one's favorite CNN with periodic padding.

We tested our method in two scenarios: spherical signal learning [9, 7], and surface collection learning. For spherical signal learning, our approach provided state of the art results among all spherical methods on a shape retrieval dataset (SHREC17 [35]) and a shape classification dataset (ModelNet40 [45]). For surface collection learning, our method produced state of the art results on a surface segmentation dataset (Humans [25]).

Our contributions are:

- We introduce a broad family of low distortion surface-to-toric image representations. The toric image representation allows applying off-the-shelf CNNs to general genus-zero surfaces.

- In particular, we provide a framework for learning spherical signals using CNNs.

- We introduce a practical algorithm for computing toric covers of genus zero surfaces.

## 2. Previous work

Applying deep learning techniques to geometric data has proved to be a huge success in the last few years. A wide variety of methods were suggested, where the most popular approaches are: volumetric based methods (*e.g.*, [45, 27]), rendering based methods (*e.g.*, [38, 44, 47]), spectral based methods (*e.g.*, [5, 10]) and methods that operate directly on the surface itself (*e.g.*, [26]). A popular related problem is the problem of learning on point clouds which received a lot of attention lately (see *e.g.*, [32, 2, 23]).

Here, we restrict our attention to intrinsic or parameterization-based surface methods and refer the reader to the above mentioned works and a recent survey [4] for further information.

**Local parameterization.** Such methods (*e.g.*, [26, 3, 28]) extract local surface patches and use them in order to learn point representations. In [26] the authors use local polar coordinates as the patch operator. In a follow-up work, [3] use projections on oriented anisotropic diffusion kernels, where [28] learn the patch operator using a Gaussian mixture model. In contrast to these works, we employ a global parameterization which represents the shape using a single image.

**Global parameterization.** Other methods use global parametrization of the surface to a canonical domain. [36, 37] use an area-preserving parameterization and map surfaces to a planar domain (going through a sphere); the global area-preserving parameterization cannot cover the
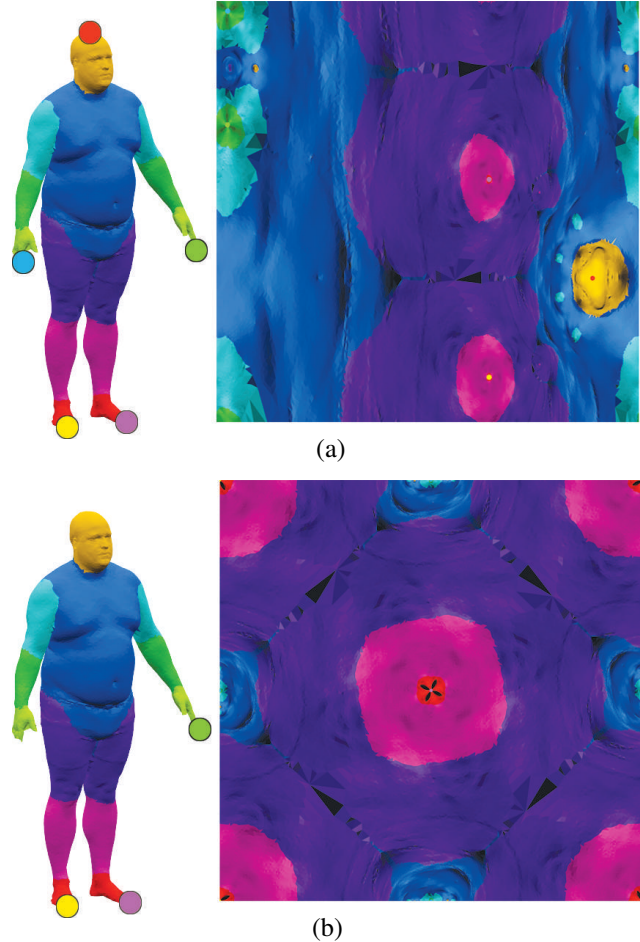


(a)



(b)

Figure 1. (a) A periodic general cover with $k = 5$ branch points (colored dots) and degree $d = 5$ of a human model, computed using our method. (b) The orbifold cover of [1] with $k = 3$ branch points and degree $d = 4$. Note that the cover in (b) has four (rotated) repetitions of the same mapping; is missing the head and right arm; and the torso (blue) suffers from a considerable downscale.

surface with low distortion everywhere and depends on the specific cut made on the surface.

The most similar method to ours is [25] that proposes gluing four copies of the surface into a torus and map it conformally (*i.e.*, preserving angles) to a flat torus, where the convolution is well defined. Their map is defined by a choice of three points on the surface, and suffers from significant angle and scale distortion, see Figure 1b (*e.g.*, the head, right arm and torso). In order to cover each point on the surface reasonably well, the authors sample multiple triplets of points from each surface where each triplet focuses on a different part of the surface. In a follow up work, [14] use the same parameterization as a surface representation for Generative Adversarial Networks (GANs) [13]. In order to deal with the high distortion of each single parameterization, the authors devise a multi-chart structure and rely on given sparse correspondences between the surfaces.

**Convolutions on tangent planes.** [42] define convolutions on surfaces by working on the tangent planes. [29] also define the convolutions on tangent planes and relate convolutions on nearby points using parallel transport. [30] define convolutions on surfaces by extending the notion of a signal on a surface into a directional signal and build layers that are equivariant to the choice of reference directions. [16] utilizes 4-rotational symmetric field to define a domain for convolution on a surface.

**Convolutions of spherical signals.** Our work targets learning of general genus zero surfaces. In particular, it can facilitate learning of spherical signals, a task that has received growing interest in the last few years. [39, 9, 48] note that an equirectangular projection of a spherical signal suffers from large distortions and suggest network architectures that try to compensate for these distortions. [6] perform 2D convolution on spherical strips extracted from the spherical signal. [17] suggest to define the convolution of a spherical signal as a linear combination of differential operators with learned weights. In a different line of work, [7, 12, 21] propose networks that are invariant to the natural action of $SO(3)$ on spherical signals. [8] advocate the notion of gauge equivariance as the correct equivariance notion on manifolds, and construct gauge equivariant networks on spheres.

**Other methods.** [46] tackle the shape segmentation problem by a novel architecture that operates on local features (such as normals) and global features (such as distances) and then fuses them together. [22] propose an improved graph neural network model based on the Dirac operator.

# 3. Preliminaries

In this section we discuss our choice of periodic images (*i.e.* images with toric topology) and introduce branched covering maps, the main mathematical tool used in our approach.

## 3.1. Convolutions on flattened spheres

A standard way to apply CNNs to a signal on a sphere-type surface is to represent it as an image and apply standard 2D convolution. Since representing a sphere as an image requires cutting and duplicating the cuts, different boundary segments in the image represent the same segment on the sphere.

In the case where the transformation in the image domain between the two duplicated boundary segments is a pure translation then the result of applying 2D convolution at any two matching points on these segments will result in exactly the same value. In other cases, such as equirectangular spherical projection [39] or octahedron spherical projection [31, 36], 2D convolution on two matching points result in two different values. Figure 2 shows an example where duplicated image boundary segments are marked with the



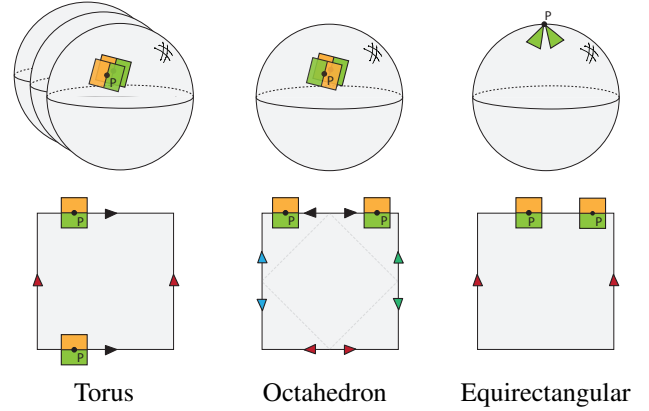| Torus | Octahedron | Equirectangular |

Figure 2. Standard 2D convolution applied to image representations (bottom row) of spherical topology (top row). In each example, the points indicated by P represent the same point on the sphere. Only for the toric topology the convolution in the image domain defines a consistent, continuous operator everywhere on the sphere.

same color arrows; a pair of matching points (marked P) are shown in each example along with an illustration of a convolution kernel. Note that only in the toric topology the kernel is consistent at the duplicated points. A similar point of view for toric images was suggested in [25]. We extend it to a more general family for toric images of sphere-type surfaces.

## 3.2. Branched covering maps

This section provides a brief introduction to branched covering maps (for more details see [15]). We start with a formal definition:

**Definition 1.** *Let $X$ and $Y$ be topological spaces. A map $E : X \rightarrow Y$ is a **branched covering map** if every point $y \in Y$ except for a finite set of points $\{b_1, \ldots, b_k\}$ has a neighborhood $U \subseteq Y$, such that $E^{-1}(U)$ is a disjoint union of homeomorphic [1] copies of $U$.*

*The set of points $\{b_1, \ldots, b_k\}$ are called **branch points**.*

A simple example for a branched covering map is $E(z) = z^d$, for $X = Y = \mathbb{C}$, and for some integer $d$. The function $E$ has one branch point at $b_1 = 0$. Every point $y \in Y \setminus \{0\}$, has $d$ *distinct* pre-images $E^{-1}(y) = \{x_1, \ldots, x_d\} \subset X$. However, the point $y = 0$ has a single pre-image $E^{-1}(y) = \{0\}$. We say that the point $y = 0$ has $d$ pre-images located at $0$, or that $0$ is a pre-image with multiplicity $d$. The **ramification index** of $x$ over $E(x)$ is the multiplicity at $x$, namely $1$ for all $x \in E^{-1}(Y \setminus \{0\})$ and $d$ for $x = E^{-1}(b_1) = 0$. We denote it as $r(x|E(x))$. Figure 3b shows this example for $d = 4$. In fact, this example captures all the local behaviors of cov-

---

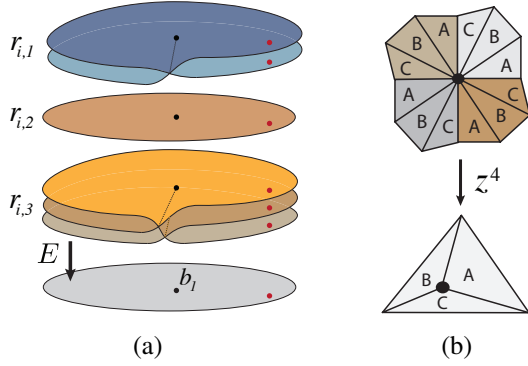[1] A homeomorphism is continuous map with a continuous inverse.

Figure 3. (a) A branch point $b_1$ (in black) with ramification structure $\rho_1 = [2, 1, 3]$. Note that all other points, such as the marked red point, have six distinct pre-images. (b) Shows an example of a branched covering map with degree 4 on a triangular mesh; The central point has a single pre-image (above it, also in black). The map looks like the map $z \mapsto z^4$. Each color represents a different copy of the neighborhood of the branch point. A triangle and its pre-images are marked with the same letter.

ering maps: around a point $x \in X$ with $r(x|E(x)) = r$ the map $E$ looks like the map $z \mapsto z^r$.

Let us give another example: Consider the function $E(z) = z^2(z - 7)$. It has a branch point at $y = 0$ with *two* distinct pre-images. Namely, $E^{-1}(0) = \{0, 7\}$. Here, the ramification index of 0 over $E(0)$ is 2 and the ramification index of 7 over $E(7)$ is 1. We say that the *ramification structure* of 0 is $[2, 1]$, formally:

**Definition 2.** *Let $E : X \rightarrow Y$ be a branched covering map, $b_i$ a branch point and $l_i = |E^{-1}(b_i)|$, the number of pre-images of $b_i$. The **ramification structure** of $b_i$ is the multi-set of ramification indices of its pre-images, denoted by $\rho_i = [r_{i,1}, \ldots r_{i,l_i}]$. The **ramification type** of $E$ is the collection of its ramification structures, $\rho = [\rho_1, \ldots, \rho_k]$.*

Figure 3a depicts a branch point $b_1$ with three distinct pre-images, $l_1 = 3$, and ramification structure $\rho_1 = [2, 1, 3]$. Note that the ramification structure of a non-branch point is a trivial multi-set of ones: $[1 \ldots 1]$, see *e.g.*, the red dot in Figure 3a.

The sum of the ramification indices of any point in $X$ is independent of the choice of the point (see [11], page 44 Proposition 7), namely

$$\sum_{j=1}^{l_i} r_{i,j} = d. \tag{1}$$

Lastly, $d$ is called the **degree of the covering**. Intuitively, the degree of the covering counts how many times $X$ covers $Y$, or alternatively how many copies of $Y$ can be found in $X$.

### 3.3. Riemann-Hurwitz formula

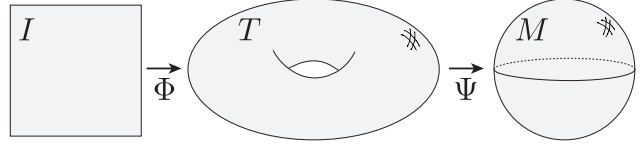A key fact about ramification types of branched covering maps between (boundaryless) surfaces is the **Riemann-**



Figure 4. Construction of the covering map $E : I \rightarrow M$.

**Hurwitz formula** (RH), which connects the genus (*i.e.*, number of handles) of the surfaces with the ramification type. In our case, we map a torus to a sphere-type surface and get the corresponding RH formula:

$$\sum_{i=1}^{k} \sum_{j=1}^{l_i} (r_{i,j} - 1) = 2d \tag{2}$$

A quick derivation of this formula is given in the Supplementary

Therefore, the RH formula sets a necessary condition on the possible ramification types $\rho$ of such branched covering maps. For example, the ramification type $[[2], [2], [2], [2]]$ satisfies the RH equations but the ramification type $[[2], [2]]$ does not (in this case $d = 2$, $k = 2$, $l_i = 1$), implying that there is no covering map with this ramification type. We note that Equations (1) and (2) are necessary but not sufficient conditions.

## 4. Approach

Our goal is transferring signals (*i.e.*, functions) from a sphere-type surface $M$ to the image domain $I$ (*i.e.*, the flat torus: unit square $[0, 1]^2$ with opposite ends identified). This is done by constructing a branched covering map

$$E : I \rightarrow M \tag{3}$$

and pulling back the signals to the image using $E$. That is, given a signal $f : M \rightarrow \mathbb{R}^n$ that we want to transfer, the value of a pixel $p \in I$ is set to $f(E(p))$. We represent the surface $M$ using a triangular mesh.

We build the covering map $E : I \rightarrow M$ in two steps, as a composition of two functions:

$$E = \Psi \circ \Phi$$
$$\Phi : I \rightarrow T$$
$$\Psi : T \rightarrow M.$$

where $T$ is a torus-type surface built out of $d$ copies of $M$, $\Psi$ is a branched covering map, and $\Phi$ is a homeomorphism between the two tori $I$ and $T$ (see Figure 4 for illustration).

### 4.1. Computing the branched covering map $\Psi$

In this section we describe how we construct the mesh $T$ out of the mesh $M$ and the branched covering map $\Psi : T \rightarrow$

$M$. The idea is to cut and glue together several copies of the input surface $M$ in a way that generates a toric covering space corresponding to a specific choice of $\rho$.

First, we choose $k$ branch points $b_1, \ldots, b_k$ from the set of vertices of $M$ (using farthest point sampling), a degree $d$ and a valid ramification type $\rho$ satisfying Equations (1)-(2). Our algorithm then consists of the following steps:

**Step (i):** We cut the mesh along $k$ disjoint paths, all emanating from the same (arbitrary) vertex $v_0$ in $M$ and ending at the branch points $b_i$ for $i \in [k]$. Figure 5 shows this for $k = d = 5$. Topologically, $M_{disk}$ is a disk, with all branch points at its boundary.

**Step (ii):** $M_{disk}$ is then duplicated $d$ times, to form copies $M_{disk}^{(1)}, \ldots, M_{disk}^{(d)}$. Figure 5 shows the 5 copies with $v_0$ as a white dot and the branch points as colored dots.
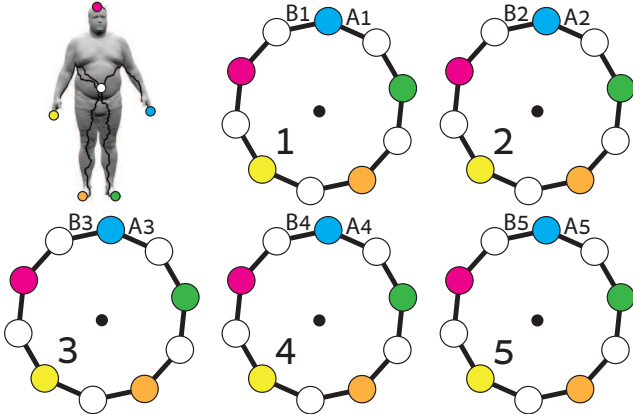


Figure 5. Illustration of the mesh cutting (top-left), and the gluing algorithm. Branch points are visualized as colored dots.

**Step (iii):** We glue the $d$ copies of $M_{disk}$ to create the surface $T$ as follows. Consider a branch point $b_i$; it has $d$ copies located in each of the copies of $M_{disk}$, see *e.g.*, the blue dots in Figure 5. Denote by $B_j$ and $A_j$ the two boundary edges emanating from the $j$-th copy of $b_i$. Note that on the original surface $A_j$ is glued to $B_j$; since every $B_{j'}$ is a duplicate of $B_j$, $A_j$ can be glued to any $B_{j'}$, $j' \in [d]$. Therefore, to describe the gluing of the edges emanating from $b_i$ we use a permutation $\sigma_i \in S_d$ (a permutation is a bijection $[d] \to [d]$): $A_j$ is glued to $B_{\sigma_i(j)}$. The collection of all permutations (one permutation per branch point)

$$\Sigma = \{\sigma_1, \ldots, \sigma_k\} \qquad (4)$$

is called the **gluing instructions**. Given gluing instructions $\Sigma$ we use it to stitch the boundary of the $d$ copies of $M_{disk}$ to construct the toric surface $T$ (*i.e.*, genus one). The mapping $\Psi : T \to M$ is then defined by: map $v \in T$ to its original version in $M$, and extend linearly in each triangle (*i.e.*, face) of $T$. $\Psi$ is a well defined branched covering map. The gluing procedure is summarized in Algorithm 1. In Subsection

4.1.1 we describe the algorithm for computing the gluing instructions given the desired ramification type $\rho$.

---

**Data:** cut mesh copies $M_{disk}^{(j)}$, $j \in [d]$;
       gluing instructions $\Sigma = \{\sigma_1, \ldots, \sigma_k\}$
**Result:** The torus $T$ and a branched covering map
       $\Psi : T \to M$ with ramification type $\rho$
**for** *every branch point $b_i$, $i \in [k]$* **do**
     **for** *every copy $j \in [d]$* **do**
         stitch $A_j^{(i)}$ and $B_{\sigma_i(j)}^{(i)}$
     **end**
**end**
$\psi : T \to M$ is defined by mapping $v$ to the unique
    vertex in $M$ that originated $v$.

**Algorithm 1:** Gluing algorithm.

### 4.1.1 Computing the gluing instructions

In this paper we limit our attention to ramification types of the form

$$\rho = \left[[1^{d-r}, r]^k\right], \qquad (5)$$

where $d$ is the cover degree, $k$ is the number of branch points, and $r$ is the maximal multiplicity of the branch points' pre-images. The motivation in choosing these ramification types is two-fold: First, we want all branch points to be treated equally by the cover. Second, applying higher ramification order improves area distortion of protruding parts (see *e.g.*, [19]); See Figure 1 and Subsection 4.3 for an example.

First, let us compute necessary conditions for $\rho$ defined in (5) to be a feasible ramification type. Equation (1) is automatically satisfied since $d - r + r = d$. Plugging $\rho$ in (2) we get

$$k(r - 1) = 2d. \qquad (6)$$

This sets a trade-off between $r$ and $d$: higher values of $r$, while reducing distortion of protruding parts would force higher degree $d$ of the cover, which will produce more copies of $M$ in the image. Practically, we found that $k = 5, r = 5, d = 10$ and $k = 6, r = 2, d = 3$ are both good options that strike a good balance between $r$ and $d$.

To compute gluing instructions $\Sigma$ we start with $k, r, d$ satisfying (6). The next theorem (proved in the Supplementary) provides a necessary and sufficient condition for the gluing instructions $\Sigma$ to furnish a cover with ramification type $\rho$:

**Theorem 1.** *A set of gluing instructions $\Sigma = \{\sigma_1, \ldots, \sigma_k\}$ yields a branched covering map with ramification type $\rho$ if and only if the following conditions hold:*

*(i) The cycle structure of $\sigma_i$ equals the ramification structure of $b_i$, i.e., $\rho_i = [r_{i,1}, \ldots, r_{i,l_i}]$.*

*(ii) $\Sigma$ is a product one tuple. That is, $\sigma_1 \cdot \sigma_2 \cdots \sigma_k = I_d$.*

*(iii)* *The group $G$ generated by $\Sigma$ is a transitive subgroup of $S_d$. Namely, for each $i, j \in [d]$ there exists $\sigma \in G$ so that $j = \sigma(i)$.*

Theorem 1 indicates that we should search for permutations $\sigma_i$ with prescribed cycle structures. That is, the permutations $\sigma_i$, if exist, are in some prescribed conjugacy classes of the permutation group. Algorithm 2 (in the Supplementary) performs such a search, more or less exhaustively, using conditions (ii) and (iii) to prune options that cannot lead to a solution $\Sigma$.

Since theoretically not all $k, r, d$ satisfying Equation (6) have a corresponding covering map, Algorithm 2 can terminate without finding gluing instructions. In this case, according to Theorem 1 we know that there is no covering map with ramification type $\rho$. Nevertheless, it is rare to find such examples in practice and indeed we did not encounter such a case in our experiments. In the Supplementary, we present a table containing the results of Algorithm 2 for any permissible $k, r, d$ with $k \leq 6, d \leq 10$ so that they can be used as input to Algorithm 1.

### 4.2. Flattening the toric surface

The last part of our covering map computation is the computation of the map $\Phi : I \rightarrow T$. Equivalently, we compute $\Phi^{-1}$. To that end we use a version of the Orbifold-Tutte embedding [1]. We first cut $T$ along the two generating loops of the torus (using [18, Algorithm 5]) to get a disk-type surface $T_{disk}$. Second, we compute a bijective piecewise affine map $\Phi^{-1} : T \rightarrow I$ by solving a sparse linear system of equations $Ax = b$, where $A \in \mathbb{R}^{m \times m}$ and $x, b \in \mathbb{R}^{m \times 2}$, and $m$ is the number of vertices in the disk-like mesh $T_{disk}$. This system is a discrete version of the Poisson equation [24], see Supplementary for details on how to construct $A, b$. We use $x$ to map the vertices of $T$ to $I$ and extend linearly to get the piecewise affine map $\Phi^{-1}$.

The resulting map is discrete harmonic [24], approximately conformal up to a linear transformation, and as proven in [1], a bijection.

### 4.3. Example

Figure 1a depicts the case $k = 5$, $r = 3$, $d = 5$. Thus $\rho = \left[[1, 1, 3]^5\right]$; every branch point $b_i$ has three distinct pre-images, where two have ramification one, and one with order-3 ramification. The gluing instructions in this case, computed using Algorithm 2 in the Supplementary, are:

$$\Sigma = \{(1)(2)(345), (1)(4)(235), (3)(4)(152),$$
$$(3)(4)(125), (1)(5)(243)\}.$$

Note that each of these permutations has a cycle structure $[1, 1, 3]$ as required in Theorem 1 *(i)*; conditions *(ii)-(iii)* can be checked as well. These gluing instructions were used to glue the 5 copies of $M_{disk}$ (as shown in Figure 5 and
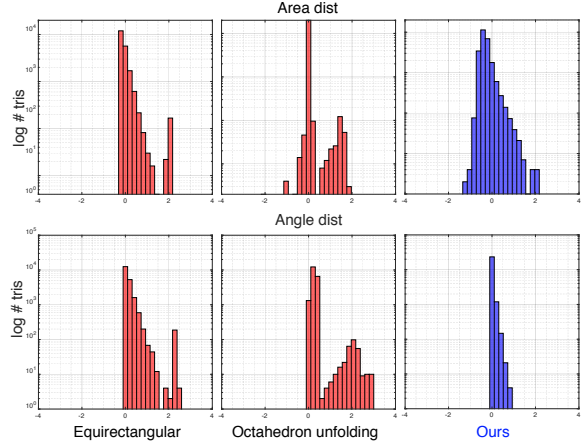


Figure 6. Area and scale distortion (log scale) of our method (right, in blue), equirectangular parametrization [39] and parametrization by octahedron unfolding [31].

described in Algorithm 1) to generate the representation $E : I \rightarrow M$ shown in Figure 1a.

## 5. Experiments

To evaluate the efficacy of our method we tested it in two main scenarios: learning signals on the sphere, and learning sphere-type surface data.

### 5.1. Evaluation

In this section we compare the geometric properties of our representation to standard or existing techniques. Figure 6 shows the area and scale distortion of our method (right, in blue) and two other popular methods for sphere flattening: Equirectangular projection (see *e.g.*, [39]) and octahedron unfolding projection, see [31]. Area distortion is computed as the determinant of the differential of the cover map $E$ (treated as affine over each triangle of $M$), and angle distortion is the condition number of the differential. Since our image representation contains several copies of each triangle of $M$ we use the least distorted one for the histogram, as we want each part of the surface to appear in the image at-least once with low distortion. As can be seen in Figure 6, our projection has better angle preservation with only a mild sacrifice to area distortion.

In Figure 7 we repeat this experiment with a sphere-type model of a human and compare the area and angle distortion of five different types of image representations. While the method of [36] (leftmost, in red) preserves area better, it suffers from significant angle distortion. The orbifold covering of [25] (second to the left, in red) is angle-preserving, but suffers from notable area shrinking. Our covering maps (green and blue) strike a balance between angle and area preservation. The covering of type $\left[[1^5, 5]^5\right]$ (middle, in blue) has the least area distortion and we chose it for the segmentation task (below).

Table 1. Comparison of our method and the top results in each category of the SHREC17 shape retrieval task.

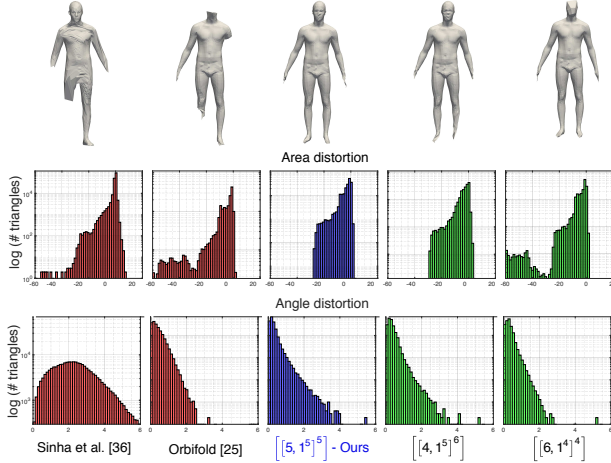| Method | P@N | R@N | F1@N | mAP | NDCG |
|---|---|---|---|---|---|
| FURUYA_DLAN | **0.814** | 0.683 | 0.706 | 0.656 | 0.754 |
| Tatsuma_ReVGG | 0.705 | **0.769** | 0.719 | 0.696 | 0.783 |
| SHREC16-Bai_GIFT | 0.678 | 0.667 | 0.661 | 0.607 | 0.735 |
| Deng_CM-VGG-6DB | 0.412 | 0.706 | 0.472 | 0.524 | 0.642 |
| Spherical CNN [7] | 0.701 | 0.711 | 0.699 | 0.676 | 0.756 |
| SO(3) Equivariant CNNs [12] | 0.717 | 0.737 | - | 0.685 | - |
| Ours | 0.749 ($2^{nd}$) | 0.741 ($2^{nd}$) | **0.734** | **0.709** | **0.794** |



Figure 7. Top row: meshes reproduced from image representations of several methods (left to right): area preserving method of [36], orbifold covering of [25] and three different covering maps produced by our method, of ramification types $[[1^5, 5]^5]$, $[[1^5, 4]^6]$ and $[[1^4, 6]^4]$. Middle and bottom rows: area and angle distortion (log scale). The histograms are taken from ten randomly chosen human models.

The top row of Figure 7 compares the different image representations by reconstructing the original model. Specifically, for each vertex of the mesh we sampled its $x, y, z$ coordinates directly from the image at the vertex location (we used $512 \times 512$ images here). In our representation, we take the coordinates from the vertex copy with the least area distortion. Note that the image representations of [36] and [25] do not represent well significant parts of the surface (*e.g.*, the right leg and the head).

## 5.2. 3D shape retrieval

The first application of our method is 3D shape retrieval. We use the SHREC2017 benchmark [35] that contains 51162 3D models from 55 different categories. There are two separate challenges: (i) the shapes are consistently aligned (ii) the shapes are randomly rotated. We tackle the (harder) second challenge.

Since the shapes are not of genus zero we follow the protocol of [7] that project the meshes on a bounding sphere using ray casting, and record six functions on this sphere:
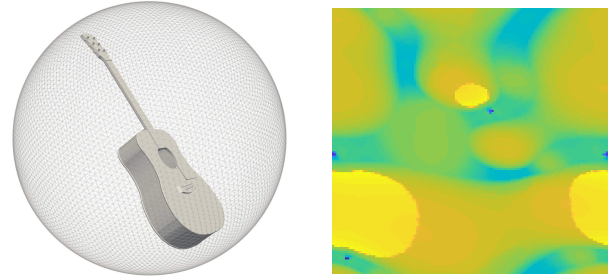


Figure 8. Left: A guitar mesh inside an encapsulating sphere. Right: The image representation of a spherical signal for this mesh. The signal is generated as follows: for each point on the sphere we cast a ray towards the origin and record the ray length at the intersection with the mesh.

distance to the model, $cos/sin$ of the model angles (this is done for both the model and its convex-hull). We then use our method to transfer these six spherical signals to periodic images (flat torus). See Figure 8 for an example of such shape representation.

We compare our method to the top methods in each category, the Spherical CNN method [7], and the recent SO-3 equivariant networks suggested in [12]. The results are summarized in Table 1; note that in the F1 measure we score first among all methods.

For this application we use a slight modification of the inception v3 architecture [41]. We train the network with ADAM optimizer [20] for 100 epochs with learning rate 0.05, batch size of 32, and learning rate decay of 0.995. Training took 15 minutes per epoch on a Tesla V100 Nvidia GPU. In evaluation time we average the output of the network on 5 randomly rotated copies of the query model.

## 5.3. Surface classification

We apply our method to the ModelNet40 surface classification benchmark [45] that contains 12311 3D models from 40 different categories. As in the shape retrieval task, we follow the protocol of [7] to generate input signals on a sphere. We then use our method to represent the spherical signals as periodic images and apply the same inception v3 model as in the shape retrieval task. We present peak performance results (following [17]) for two scenarios that are popular in the literature: (i) the shapes are rotated randomly about

Table 2. Results on ModelNet40 dataset.

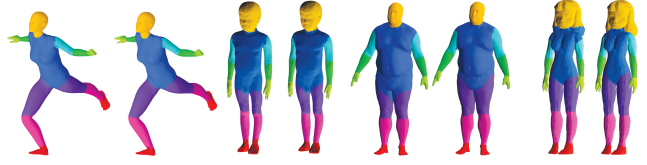| Method | Inputs | Accuracy |
|---|---|---|
| Learning Gims [36] | mesh | 83.9% |
| 3DShapeNets [45] | voxels | 84.7% |
| VoxNet [27] | voxels | 85.9% |
| Pointnet[32] | points | 89.2% |
| Pointnet++ [33] | points | 91.9% |
| Dynamic graph CNN [43] | points | 92.2% |
| PCNN [2] | points | 92.3% |
| Spherical CNN [7] | spherical | 85.0% |
| SO(3) Equivariant CNNs [12] | spherical | 88.9% |
| Spherical on unstructured grid [17] | spherical | 90.5% |
| Octahedron unfolding (rot $z$) | spherical | 90.2% |
| Equirectangular projection (rot $z$) | spherical | 90.1% |
| **Ours** | **spherical** | **91.6%** |
| **Ours (rot $z$)** | **spherical** | **91.0%** |



Figure 9. Segmentation results of our method. For each pair: left is our result and right is ground truth.

Table 3. Results on the human segmentation dataset.

| Method | Inputs | Accuracy |
|---|---|---|
| Toric CNN [25] | WKS,AGD,curv | 88.00% |
| Geodesic Conv [26] | 3D coords | 76.49% |
| Pointnet++ [33] | 3D coords | 90.77% |
| Dynamic graph CNN [43] | 3D coords | 89.72% |
| Multi-directional Conv [30] | 3D coords | 88.61% |
| Learning Gims [36] | 3D coords | 84.53% |
| **Ours** | **3D coords** | **91.31%** |

the $z$ axis; and (ii) the shapes are learned in their original orientation. We train the network with ADAM optimizer [20] for 100 epochs for scenario (1) and 300 for scenario (ii) with learning rate 0.0005, batch size 16, and learning rate decay 0.995. Training took 19 minutes per epoch for the first scenario (that contains 10 rotation augmentations) and 3 minutes per epoch for the second scenario on a Tesla V100 Nvidia GPU.

Table 2 compares our results with several recent methods including the baselines of equirectangular projection (*e.g.*, [39]) and octahedron unfolding projection [31]. Our results are the best among all spherical learning methods.

### 5.4. Surface segmentation

While our first two application targeted spherical signals, our last applications learns signals defined on general sphere-type human models. In particular, we perform human model semantic segmentation. We use the benchmark from [25] that consists of 373 train models from multiple sources and 18 test models. 5% randomly sampled train models were used as a validation set (18 models). All models are given as triangular meshes. For each model, each face is labeled according to a predefined partition of the human body (*e.g.*, head, torso, hands, total of 8 labels). The task is to label the triangles of a new unseen human model with these labels. For each model we generate an augmented set of 120 images per mesh, by permuting the order of the branch points, multiplying the vertices by a random orthogonal matrix and a uniform scale sampled from [0.85, 1.15] as suggested in [30], and small periodic image translations of $\pm 15$ pixels. In evaluation, as the toric image contains $d$ values for each triangle on the original mesh, we use the label of the triangle with the largest area. Furthermore, we use 10 random augmentations of test images and label each mesh face using a majority vote. Table 3 sum-

marizes the results of this experiment, where our method outperformed previous methods; Figure 9 shows typical segmentation results.

For this application we used the U-net architecture [34] with 16 layers (see Table 2 for details). We used a weighted loss with equal probability labels, and trained the network using stochastic gradient descent with momentum [40] for 50 epochs with learning rate 0.2, batch size 2, and learning rate decay of 0.995. Training takes $\sim 3$ hours per epoch on a Tesla V100 Nvidia GPU.

### 6. Conclusions

In this paper, we introduce a new method for representing sphere-type surfaces as toric images that can be used in standard Convolutional Neural Network frameworks for shape learning tasks. The method allows faithful representation of *all parts of the surface in a single image*, thus alleviating the need to generate multiple maps to cover each surface. Our method is general and can target both spherical signal learning tasks as well as more general learning tasks that involve signals on different genus zero surfaces. Practically, we showed that off-the-shelf CNN models applied to images generated with our method lead to state of the art performance in the tasks of shape retrieval, shape classification and surface segmentation.

The main limitation of this work is its restriction to genus-zero surfaces. This kind of models are abundant, but certainly do not exhaust all 3D models. We would like to seek a generalization of this method to point clouds, depth images and more general topological types.

### 7. Acknowledgements

# References

[1] Noam Aigerman and Yaron Lipman. Orbifold tutte embeddings. *ACM Trans. Graph.*, 34(6):190–1, 2015.

[2] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point convolutional neural networks by extension operators. *ACM Trans. Graph.*, 37(4):71:1–71:12, July 2018.

[3] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 3189–3197, 2016.

[4] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[5] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[6] Zhangjie Cao, Qixing Huang, and Ramani Karthik. 3d object classification via spherical projections. In *2017 International Conference on 3D Vision (3DV)*, pages 566–574. IEEE, 2017.

[7] Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *arXiv preprint arXiv:1801.10130*, 2018.

[8] Taco S Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral cnn. *arXiv preprint arXiv:1902.04615*, 2019.

[9] Benjamin Coors, Alexandru Paul Condurache, and Andreas Geiger. Spherenet: Learning spherical representations for detection and classification in omnidirectional images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 518–533, 2018.

[10] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pages 3844–3852, 2016.

[11] Simon Donaldson. *Riemann surfaces*. Oxford University Press, 2011.

[12] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 52–68, 2018.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[14] Heli Ben Hamu, Haggai Maron, Itay Kezurer, Gal Avineri, and Yaron Lipman. Multi-chart generative surface modeling. *arXiv preprint arXiv:1806.02143*, 2018.

[15] Allen Hatcher. *Algebraic topology*. Cambridge Univ. Press, Cambridge, 2000.

[16] Jingwei Huang, Haotian Zhang, Li Yi, Thomas Funkhouser, Matthias Nießner, and Leonidas J Guibas. Texturenet: Consistent local parametrizations for learning from high-resolution signals on meshes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4440–4449, 2019.

[17] Chiyu Jiang, Jingwei Huang, Karthik Kashinath, Philip Marcus, Matthias Niessner, et al. Spherical cnns on unstructured grids. *arXiv preprint arXiv:1901.02039*, 2019.

[18] Miao Jin, Xianfeng Gu, Ying He, and Yalin Wang. *Conformal Geometry: Computational Algorithms and Engineering Applications*. Springer, 2018.

[19] Liliya Kharevych, Boris Springborn, and Peter Schröder. Discrete conformal mappings via circle patterns. *ACM Transactions on Graphics (TOG)*, 25(2):412–438, 2006.

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

[21] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–gordan nets: a fully fourier space spherical convolutional neural network. In *Advances in Neural Information Processing Systems*, pages 10138–10147, 2018.

[22] Ilya Kostrikov, Zhongshi Jiang, Daniele Panozzo, Denis Zorin, and Joan Bruna. Surface networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)(June 2018)*, volume 3, 2017.

[23] Yangyan Li, Rui Bu, Mingchao Sun, and Baoquan Chen. Pointcnn. *arXiv preprint arXiv:1801.07791*, 2018.

[24] László Lovász. Discrete analytic functions: an exposition. *Surveys in differential geometry*, 9(1):241–273, 2004.

[25] Haggai Maron, Meirav Galun, Noam Aigerman, Miri Trope, Nadav Dym, Ersin Yumer, Vladimir G Kim, and Yaron Lipman. Convolutional neural networks on surfaces via seamless toric covers. *ACM Trans. Graph*, 36(4):71, 2017.

[26] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015.

[27] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 922–928. IEEE, 2015.

[28] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proc. CVPR*, volume 1, page 3, 2017.

[29] Hao Pan, Shilin Liu, Yang Liu, and Xin Tong. Convolutional neural networks on 3d surfaces using parallel frames. *arXiv preprint arXiv:1808.04952*, 2018.

[30] Adrien Poulenard and Maks Ovsjanikov. Multi-directional geodesic neural networks via equivariant convolution. *arXiv preprint arXiv:1810.02303*, 2018.

[31] Emil Praun and Hugues Hoppe. Spherical parametrization and remeshing. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 340–349. ACM, 2003.

[32] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification

and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[33] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5099–5108, 2017.

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[35] Manolis Savva, Fisher Yu, Hao Su, Asako Kanezaki, Takahiko Furuya, Ryutarou Ohbuchi, Zhichao Zhou, Rui Yu, Song Bai, Xiang Bai, et al. Shrec17 track large-scale 3d shape retrieval from shapenet core55.

[36] Ayan Sinha, Jing Bai, and Karthik Ramani. Deep learning 3d shape surfaces using geometry images. In *European Conference on Computer Vision*, pages 223–240. Springer, 2016.

[37] Ayan Sinha, Asim Unmesh, Qixing Huang, and Karthik Ramani. Surfnet: Generating 3d shape surfaces using deep residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6040–6049, 2017.

[38] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[39] Yu-Chuan Su and Kristen Grauman. Learning spherical convolution for fast features from 360 imagery. In *Advances in Neural Information Processing Systems*, pages 529–539, 2017.

[40] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.

[41] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[42] Maxim Tatarchenko, Jaesik Park, Vladlen Koltun, and Qian-Yi Zhou. Tangent convolutions for dense prediction in 3d. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3887–3896, 2018.

[43] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *arXiv preprint arXiv:1801.07829*, 2018.

[44] Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1544–1553, 2016.

[45] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes.

In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[46] Haotian Xu, Ming Dong, and Zichun Zhong. Directionally convolutional networks for 3d shape segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2698–2707, 2017.

[47] Tan Yu, Jingjing Meng, and Junsong Yuan. Multi-view harmonized bilinear network for 3d object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 186–194, 2018.

[48] Qiang Zhao, Chen Zhu, Feng Dai, Yike Ma, Guoqing Jin, and Yongdong Zhang. Distortion-aware cnns for spherical images. In *IJCAI*, pages 1198–1204, 2018.