



ZerNet: Convolutional Neural Networks on Arbitrary Surfaces Via Zernike Local Tangent Space Estimation

Zhiyu Sun,¹  Ethan Rooke,² Jerome Charton,¹ Yusen He,¹ Jia Lu³ and Stephen Baek^{1,2} 

¹Department of Industrial and Systems Engineering, The University of Iowa, USA
zhiyu-sun@uiowa.edu, charton@lab327.net, {yusen-he, stephen-baek}@uiowa.edu

²Applied Mathematical and Computational Sciences, The University of Iowa, USA
ethan-rooke@uiowa.edu

³Department of Mechanical Engineering, The University of Iowa, USA
jia-lu@uiowa.edu

Abstract

In this paper, we propose a novel formulation extending convolutional neural networks (CNN) to arbitrary two-dimensional manifolds using orthogonal basis functions called Zernike polynomials. In many areas, geometric features play a key role in understanding scientific trends and phenomena, where accurate numerical quantification of geometric features is critical. Recently, CNNs have demonstrated a substantial improvement in extracting and codifying geometric features. However, the progress is mostly centred around computer vision and its applications where an inherent grid-like data representation is naturally present. In contrast, many geometry processing problems deal with curved surfaces and the application of CNNs is not trivial due to the lack of canonical grid-like representation, the absence of globally consistent orientation and the incompatible local discretizations. In this paper, we show that the Zernike polynomials allow rigorous yet practical mathematical generalization of CNNs to arbitrary surfaces. We prove that the convolution of two functions can be represented as a simple dot product between Zernike coefficients and the rotation of a convolution kernel is essentially a set of 2×2 rotation matrices applied to the coefficients. The key contribution of this work is in such a computationally efficient but rigorous generalization of the major CNN building blocks.

Keywords: 3D shape matching, modeling, databases of geometric models/shape retrieval, computer vision - shape recognition, methods and applications

ACM CCS: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling — Curve; surface; solid and object representations

1. Introduction

Many areas of scientific research deal with geometric data. The shapes of objects often encompass critical information for understanding scientific trends and phenomena. Such problems often boil down to identifying latent geometric patterns behind a diversity of shapes and correlating these geometric characteristics with certain physical phenomena.

In this regard, convolutional neural networks (CNNs) have demonstrated a substantially enhanced capacity in capturing and recognizing geometric features from images or signals, conceiving the unprecedented advances in computer vision and artificial intelligence over the last several years. However, the success was concen-

trated mostly on computer vision applications where one can enjoy the canonical grid-structured representation of data and, thus, the basic operations of CNNs (i.e. convolutions and poolings) could be seamlessly defined.

On the contrary, the geometry processing community defines many visual recognition problems on discretely sampled arbitrary surfaces. For instance, the segmentation of an LiDAR scan [AML18, MPS09, DUK*11] can be understood as a point-wise classification problem defined over point-cloud-approximated surfaces. Feature detection [SX14, HSB17, NWB15], correspondence matching [LRR*17, ZBVH09, SHG*17, RRBW*14] and shape classification [ESKBC17] are other typical examples of manifold-based visual recognition problems.

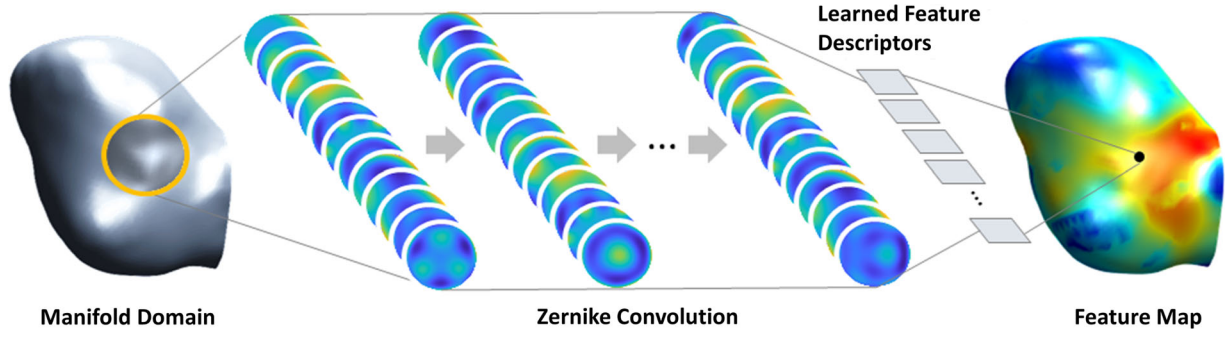


Figure 1: A schematic overview of a Zernike convolutional neural network.

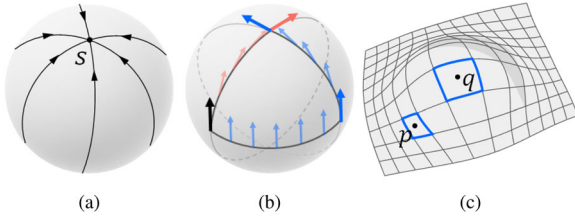


Figure 2: Problems associated with manifolds when expanding the notion of convolution. (a) a topological sphere cannot have a grid defined on it without a singularity (S); (b) the end-up direction of a vector parallel-transported along a loop on a manifold is path-dependency; (c) irregularity of a grid causes varying-size receptive fields.

By their nature, many of these visual recognition problems defined on arbitrary surfaces are analogous to those in the Euclidean setting. In practice, however, extending CNNs to manifolds is not trivial due to the lack of fundamental operations which exist in Euclidean space. As Bronstein *et al.* [BBL*17] point out, ‘(CNNs) have been most successful on data with an underlying Euclidean or grid-like structure, and in cases where the invariants of these structures are built into networks used to model them’. For instance, the hairy ball theorem [EG79] shows that it is impossible to install a grid on a domain homeomorphic to a sphere without creating a singular point where we lose the notion of orientation (Figure 2(a)). And even though there was a grid structure somehow defined without a singularity, when we transport a vector in parallel across the manifold, the direction of the vector can end up differently according to the path we take, which is so-called holonomy (Figure 2(b)). Therefore, the sliding convolution kernels along a loop path on a surface can cause a change of orientation when it comes back to the starting point. Moreover, the grid may be irregular, causing varying sizes of the effective receptive fields across the manifold (Figure 2(c)).

In this work, we propose a novel mathematical formulation of CNNs on arbitrary surfaces as illustrated in Figure 1. The key innovation of the proposed approach resides in the local piece-wise parameterization of a tensor field using the Zernike polynomial bases [vF34], as well as the discovery of the following analytic properties that are useful for computation: the Zernike orthogonal

basis functions allow a rigorous but straightforward generalization of convolution that ultimately result in a simple dot product between Zernike polynomial coefficients (Section 3.3). Furthermore, rotations of convolution kernels on the tangent spaces, which are inevitable in relevant manifold CNN literature, can be greatly simplified to 2×2 matrix rotation under the Zernike setting (Section 3.2). This enables concise definition of the angular pooling, as well as introducing new possibilities for direction preserving convolutions (Section 4.4). Therefore, the main contribution of this paper is to generalize the representation of CNN operators on surfaces while maintaining their simplicity. Finally, building upon these new theoretical grounds, we propose a new approach called ZerNet (Zernike CNN) and demonstrate its performance in comparison with the other state-of-the-art methods in both classification and regression tasks.

2. Related Works

Although there has been vigorous research activity surrounding CNNs, only a few have focused on non-Euclidean CNNs, despite the numerous applications and benefits mentioned above. Existing work on non-Euclidean CNNs falls into one of the two following categories: *spectral methods* and *spatial methods*.

Spectral methods. Pioneering works [BZSL13, HBL15, DBV16, YSGG17] defined the convolution operation on manifolds by employing a *spectral graph processing* approach. The main theoretical foundation for this is the convolution theorem, which states that the convolution $f * g$ of two functions f and g is equivalent to a simple element-wise product in the Fourier (spectral) domain: $\mathcal{F}(f * g) = \mathcal{F}(f) \odot \mathcal{F}(g)$, where $\mathcal{F}$ denotes the Fourier transform and $\odot$ is the element-wise Hadamard product. The convolution theorem generalizes to manifolds quite effortlessly as when we let U be the linear Fourier operator on a manifold, the convolution theorem gives:

$$f * g = U^\top \{ (Uf) \odot (Ug) \}. \quad (1)$$

The manifold Fourier operator U is essentially the eigenfunctions of the Laplace–Beltrami operator Δ_g defined on $(\mathcal{M}, g)$. The Laplace–Beltrami operator Δ_g is a generalization of the second-order derivative, Laplacian, on Riemannian manifolds and the

discretization of the Laplace–Beltrami operator is well understood in literature, for example, as in [MDSB03].

This elegant generalization suffers from three main flaws in practice. First, the computation of U is prohibitively expensive when there is a large set of vertices, as it requires the eigendecomposition of the linear operator Δ_g . More recent works [DBV16, KW16, BMM*15] have introduced circumventing measures to limit the computation of eigenfunctions to local regions by using computational techniques such as *windowed* Fourier transformations [SRV16].

Aside from these speed concerns, a more fundamental issue lies in the numerical behaviour of the eigenfunctions. It is well known to the geometry processing community that the eigenfunctions can flip signs and change orderings [LZ10, RCG08]. In the context of geometric data analysis, this can be critical since the eigenbases across different geometric models may vary and may not be compatible, making it impossible to represent convolution kernels consistently. Finally, spectral kernels are rotationally symmetric as the spectral bases are isotropic [BZSL13, BMM*15], which limits the expressiveness substantially [CWKW19].

Spatial methods. Spatial approaches follow a more straightforward and explicit generalization of CNNs, as in [DMI*15, NAK16, HCQ17]. In contrast to the spectral formulations where the convolution is implicitly reparameterized using spectral bases, in spatial formulations, convolution is defined more explicitly and intuitively on tangent spaces of the manifold. For example, Masci *et al.* [MBBV15] applied convolution kernels on local geodesic disks to approximate the tangent spaces. This idea quite naturally generalizes the explicit notion of convolution onto manifold domains. In a similar work, Boscaini *et al.* 2016 used anisotropic heat kernels to enable higher expressivity. Monti *et al.* 2017 further improved the local geodesic disk convolution idea by introducing the notion of trainable local parameterization such that the network learns coordinate values of the neighbouring points on geodesic disks from data. More recently, Honocka *et al.* [HHF*19] similarly discretized tangent spaces on one-ring neighbourhoods of edges on a triangular mesh.

One problem is a majority of these works [MBBV15, MRB*16, MBM*17], via the use of *angular pooling* resolve the directional ambiguity caused by holonomy, end up suppressing the features directionality. Recent works such as [HHF*19] instead introduce the notion of invariant convolutions. Unfortunately, these methods achieve directional invariance at the cost of feature directionality giving them, in essence, the same issues. Verma *et al.* 2018, instead, proposed a data-driven scheme to attain directional correspondences among patches, by designing the network to learn the correspondence between the kernels. Some very recent works [PO18, SDL18, CWKW19] contain a more mathematically rigorous solution to these directional ambiguities by introducing a vector field indicating the orientation of the local coordinate charts.

In summary, as opposed to the spectral methods, the spatial approaches do not assume global function bases, so that kernels are made explicit and compatible across different domains. Further, compare to spectral approaches, spatial methods tend to be computationally more efficient as they do not require an eigenvector

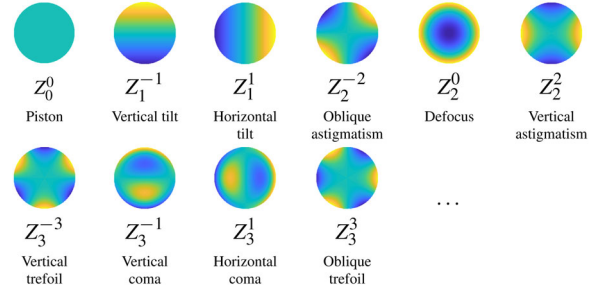


Figure 3: Examples of Zernike bases.

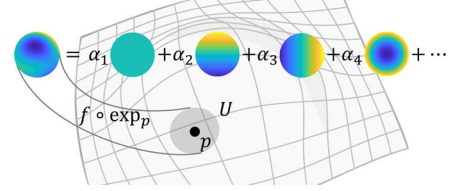


Figure 4: Zernike decomposition of a function.

computation. However, at the same time, mathematical rigour is often lost during the spatial discretization process. Furthermore, local neighbourhood topology can vary across different locations on the manifold, rendering another obstacle for convolving kernels consistently.

The proposed method combines the advantages of spectral and spatial approaches. In our formulation, the convolution kernels are applied on local tensor fields spatially extracted from the surface. We parametrize the local tensor field via a set of corresponding coefficients of Zernike polynomials, which preserves the mathematical rigour similarly to spectral approaches.

3. Zernike CNNs

In this section, we first introduce Zernike polynomials with their formal definition and examine their analytic properties. We employ them to describe the local geometry of a surface on the local tangent spaces. We will then define manifold convolution by using the notion of tangent spaces under Zernike formulation. Finally, we discretize the continuous formulation of the Zernike convolution to define ZerNet.

3.1. Zernike polynomials

Zernike polynomials Z_i are an orthogonal polynomial basis for functions defined over the unit disk $\Omega \in \mathbb{R}^2$ such that $\langle Z_i, Z_j \rangle = \int_{\Omega} Z_i(t) Z_j(t) dt = 0$ for all $i \neq j$, where $\langle \cdot, \cdot \rangle$ denotes the inner product. The formal definition of the Zernike polynomials is separated into even and odd sequences denoted in $Z_n^m(r, \theta)$ and $Z_n^{-m}(r, \theta)$, respectively:

$$\begin{aligned} Z_n^m(r, \theta) &= R_n^m(r) \cos(m\theta), \\ Z_n^{-m}(r, \theta) &= R_n^m(r) \sin(m\theta), \end{aligned} \quad (2)$$

where m and n are non-negative integer indices with $m \leq n$; $r \in [0, 1]$ is the radial distance and θ is the azimuthal angle on the disk. Here, $R_n^m(r)$ is called the Zernike radial polynomial, defined as:

$$R_n^m(r) = \begin{cases} \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! (\frac{n+m}{2}-k)! (\frac{n-m}{2}-k)!} r^{n-2k}, & \text{if } n-m \text{ even} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Figure 3 illustrates the first few Zernike basis functions. Due to simple analytic properties, Zernike polynomials are widely used in optometry and ophthalmology to describe aberrations of the cornea or to represent lens distortions in optics. As we will see in later sections, we take advantage of such simple analytic properties of Zernike polynomial bases for extending CNNs on manifolds.

In practice, Zernike polynomials can be normalized with a normalization factor $\sqrt{\frac{2-\delta(m,0)}{\pi}}$ such that the integral over the unit disk becomes unity:

$$\hat{Z}_n^m(r, \theta) = Z_n^m(r, \theta) \sqrt{\frac{2-\delta(m,0)}{\pi}}, \quad (4)$$

where $\hat{Z}_n^m$ denotes the normalized Zernike polynomial and δ is the Kronecker delta function. Due to their orthonormality over the unit disk, the normalized Zernike polynomials can serve as bases for decomposing a complex function as a weighted sum, such that any function $f(r, \theta)$ defined on the domain $[0, 1] \times [0, 2\pi)$ can be expressed as:

$$f(r, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \alpha_{nm} \hat{Z}_n^m(r, \theta). \quad (5)$$

For the sake of brevity, in the rest of the paper we use the term *Zernike bases* standing for the normalized Zernike polynomials, denoted as $Z_i(r, \theta)$ with index i corresponding to a certain pair of (n, m) in (5) following the ordering suggested by Figure 3. Thus (5) can be expressed as:

$$f(r, \theta) = \sum_{i=1}^{\infty} \alpha_i Z_i(r, \theta), \quad (6)$$

and we use the term *Zernike coefficients* to represent the set of weights α_i corresponding to *Zernike bases* Z_i , which can be interpreted as the coordinate vector of f in the Zernike base space.

3.2. Rotation

We want to be able rotate our functions in later sections, i.e., be able to compute $f(r, \theta + \phi)$. To facilitate this, we establish some rotational properties of Zernike polynomials. Using the sum of angle formulae for trigonometric functions, we obtain

$$\begin{aligned} Z_n^m(r, \theta + \phi) &= Z_n^m(r, \theta) \cos(m\phi) - Z_n^{-m}(r, \theta) \sin(m\phi), \\ Z_n^{-m}(r, \theta + \phi) &= Z_n^{-m}(r, \theta) \cos(m\phi) + Z_n^m(r, \theta) \sin(m\phi). \end{aligned} \quad (7)$$

Therefore, for a function f decomposed as in (5), we then derive the following representation of its rotation with an angle offset ϕ as:

$$\begin{aligned} f(r, \theta + \phi) &= \sum_{\text{even}} \alpha_n^m Z_n^m(r, \theta + \phi) \\ &\quad + \sum_{\text{odd}} \alpha_n^{-m} Z_n^{-m}(r, \theta + \phi) \\ &= \sum_{\text{even}} \tilde{\alpha}_n^m Z_n^m(r, \theta) \\ &\quad + \sum_{\text{odd}} \tilde{\alpha}_n^{-m} Z_n^{-m}(r, \theta), \end{aligned} \quad (8)$$

where $\tilde{\alpha}_i$ is computed via the rotational transform:

$$\begin{bmatrix} \tilde{\alpha}_n^m \\ \tilde{\alpha}_n^{-m} \end{bmatrix} = \begin{bmatrix} \cos(m\phi) & \sin(m\phi) \\ -\sin(m\phi) & \cos(m\phi) \end{bmatrix} \begin{bmatrix} \alpha_n^m \\ \alpha_n^{-m} \end{bmatrix}, \quad (9)$$

To lessen the notational burden, for a function decomposed as in (6), we will represent its rotation with an angle offset ϕ as:

$$(\text{Rot}(\phi))(f) = f(r, \theta + \phi) = \sum_{i=1}^{\infty} \tilde{\alpha}_i(\phi) Z_i(r, \theta), \quad (10)$$

where $\tilde{\alpha}_i(\phi)$ corresponds to the rotated coefficient from (9).

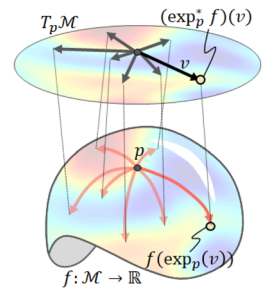
3.3. Zernike convolution

We generally view convolution on two-dimensional Cartesian space as the ‘sliding’ of a filter over the image and measuring how much the image matches our filter. Using this visual, we seek to define convolution on the manifold similarly. Our filter will remain tangent to the manifold as we slide it around, then we locally parameterize our space using the tangent space and do the usual convolution. Thus, we let $g : T_p \mathcal{M} \rightarrow \mathbb{R}$ be our ‘filter’ and $f : \mathcal{M} \rightarrow \mathbb{R}$ be our ‘image’. The aim here is to show that if we represent f and g locally at p using Zernike polynomials as in (6), the convolution on a manifold is nothing but a simple vector dot product of Zernike coefficient vectors on a tangent space and, thus, the convolution operation on a manifold is greatly simplified. We will also show that the rotation of a kernel also becomes a set of simple 2×2 matrix rotations under this formulation. Hence, without compromising mathematical rigour, ZerNet can be defined on arbitrary surfaces efficiently while remaining compatible with existing tensor-based Euclidean CNN software packages.

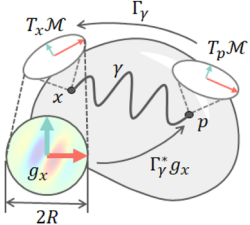
3.3.1. Convolution on manifold

We would like to define the convolution on manifolds by using the notion of tangent spaces. First, for a given function $f : \mathcal{M} \rightarrow \mathbb{R}$ defined on a manifold $\mathcal{M}$, we can locally parameterize f in terms of tangent vectors attached at $p \in \mathcal{M}$ by pulling back f to the tangent space $T_p \mathcal{M}$ by the exponential map:

$$f_p(v) \stackrel{\text{def}}{=} (\exp_p^* f)(v) = (f \circ \exp_p)(v), \quad (11)$$



for all $v \in T_p\mathcal{M}$ and $\|v\| < R$, where R defines the radius of the local neighbourhood around p . The radius R later will be weaved into the ZerNet implementation as a user-defined parameter concerning the size of the convolution kernel. Intuitively, we can think of (11) as projecting v onto the manifold $\mathcal{M}$, sampling the function value of f at the tip of the arrow, and bringing it back to $T_p\mathcal{M}$, for each and every $\|v\| < R$.



Now, let us consider a convolution kernel $g : \mathcal{M} \rightarrow \mathbb{R}$ defined around some ‘origin’ $x \in \mathcal{M}$. Note that the origin x here is a temporary reference we introduce to facilitate the discussion and will eventually disappear in the formulation.

Hence, any arbitrary point on $\mathcal{M}$ can be chosen to serve as our x . As in (11), we can pullback the kernel g to the tangent space $T_x\mathcal{M}$ by $g_x = g \circ \exp_x$. Under this setting, we can parallel transport g_x from the tangent space at x to the tangent space at p along some path $\gamma \subset \mathcal{M}$ connecting x and p . From the perspective of p , this is a process of pulling back the function g_x from $T_x\mathcal{M}$ to $T_p\mathcal{M}$ (i.e. coordinate transform from $T_x\mathcal{M}$ to $T_p\mathcal{M}$). Therefore, when the mapping $\Gamma_\gamma : T_p\mathcal{M} \rightarrow T_x\mathcal{M}$ is the isomorphism (i.e. an invertible mapping) induced by the parallel transport along γ , the transported kernel $g_p : T_p\mathcal{M} \rightarrow \mathbb{R}^c$ is simply the pullback Γ_γ^* of the function g_x , or formally, $g_p = \Gamma_\gamma^* g_x = g_x \circ \Gamma_\gamma$, where the asterisk (*) is used to denote the pullback. Under this setting, the convolution $f * g$ at p is defined as:

$$(f * g)(p) = \langle f_p, g_p \rangle = \int_{T_p\mathcal{M}} f_p(v) g_p(v) dA. \quad (12)$$

Here, by representing v in polar coordinates $v = (r, \theta)$ and by (6), we achieve

$$\begin{aligned} (f * g)(p) &= \iint_{r,\theta} \sum_i \alpha_f^i Z_i(r, \theta) \sum_j \alpha_g^j Z_j(r, \theta) r dr d\theta \\ &= \iint_{r,\theta} \sum_{i,j} \alpha_f^i \alpha_g^j Z_i(r, \theta) Z_j(r, \theta) r dr d\theta. \end{aligned} \quad (13)$$

Note, the area integral is about r and θ , but not the Zernike coefficients α^i . Therefore, by the orthonormality $\int Z_i Z_j dA = \delta_i^j$, where δ_i^j is the Dirac delta function, the equation simplifies to:

$$(f * g)(p) = \sum_i \alpha_f^i \alpha_g^i, \quad (14)$$

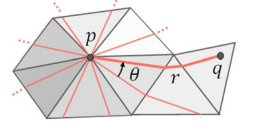
which is merely a vector dot product between the Zernike coefficient vectors α_f and α_g at p .

Furthermore, if we assume the isomorphism Γ_γ is an isometry between $T_p\mathcal{M}$ and $T_x\mathcal{M}$, g_x and g_p are the same up to rotation. Therefore, in practice, we do not need to actually do anything to pullback g_x to $T_p\mathcal{M}$ but, instead, just reuse the same r, θ representation of g_x for g_p with an angle offset ϕ , such that $g_p = (\text{Rot}(\phi))(g_x)$. Meanwhile, from the proof given in Section 3.2, the rotational transformation $\text{Rot}(\phi)$ of g_x can be achieved simply as a set of 2×2 matrix rotations of the Zernike coefficient vectors, presenting an analytic, vector-space formula for the parallel transport.

3.3.2. Discretization

With the above continuous theory, discretization is rather straightforward. Here, with a remark that the above continuous theory can be discretized to virtually any type of discretization, be it point clouds, polygonal meshes and parametric surfaces, we demonstrate here only the triangular mesh case for brevity.

Exponential map. Computation of a discrete exponential map has been widely studied in computer graphics and relevant areas, for a variety of purposes such as interactive drawing on surfaces [SGW06] and local remeshing [MR12]. These methods aim to find the shortest geodesic paths from a reference point to its neighbours using variants of the Dijkstra’s algorithm. In this work, we employ the method presented in Melvaer *et al.* [MR12] to compute the exponential map at each point. Furthermore, by measuring the angle from some reference direction to the geodesic curve, we can also determine the azimuth θ . Here, the reference direction typically is arbitrarily picked, most commonly based on the local ordering (indexing) of the neighbouring points. To this end, we again use the method of Melvaer *et al.* [MR12] for the results presented in this paper. However, there are methods such as Crane *et al.* [CWW17], which can be extended to a much broader scope including point clouds, subdivision surfaces, noisy/partial meshes and spline surfaces.



Zernike decomposition. From (6) and the *geodesic polar coordinates* (r_j, θ_j) computed via the discrete exponential map, the relationship $f_p(r_j, \theta_j) = \sum_i \alpha_f^i Z_i(r_j, \theta_j)$ holds true for all (r_j, θ_j) in the neighbourhood (Figure 4). This in fact is a linear system of equations solved locally at p . In implementation, the summation is approximated with a finite number of terms k instead of the infinite sum. The number of points in the neighbourhood may be lesser than k . In this case, we simply sample more points by linearly blending the existing points and their function values. Practically, we sample a greater number of points in the neighbourhood than k and solve the linear system in a least square manner.

Convolution. With the above discretization, the local parameterization f_p can be represented as a $k \times d$ tensor, where k is the number of Zernike polynomials used in decomposition and d is the dimensionality of the input, analogous to the number of image channels in image-based CNNs. If a manifold is discretized with N points, the stack $F = [f_p]$ of parameterized functions is an $N \times k \times d$ tensor. Meanwhile, the convolution kernel g can also be similarly discretized as $k \times d$. However, due to the rotation of the kernel induced by parallel transport, the tensor G corresponding to g requires an additional axis for the rotation parameter, such that G becomes $k \times d \times s$, where s is the *angular resolution*. Finally, from the fact that Zernike convolution is a simple dot product between the coefficient vectors (i.e. (14)), the implementation of the Zernike convolution layer becomes a simple tensor dot product between F and G along k and d axes, producing an $N \times s$ response.

Angular pooling. With the response produced via convolution, the activation maps defined across different direction configurations

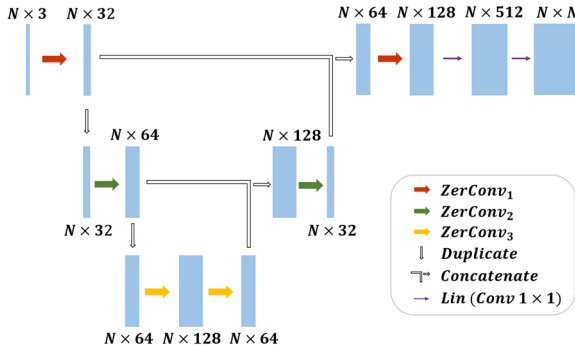


Figure 5: Multi-scale ZerNet architecture.

along the angular axis can be down-sampled via angular-pooling operation. Similar to the (spatial) max-pooling layers in conventional CNNs, angular-pooling selects the maximum value among the activations produced by convolution kernels in different orientations, producing N response.

4. Experiments

We validate the proposed approach against two supervised learning tasks, namely, classification and regression. For classification, we test our algorithm against two popular problems in geometry processing: point-wise correspondence matching and semantic segmentation, which are both commonly formalized as per-vertex classification problems. For regression, we introduce a new dataset in which the goal is to predict a scalar field defined on a surface.

4.1. 3D shape correspondence

The goal of 3D shape correspondence problem is to find semantically meaningful one-to-one matching between points on a query surface and points on a reference surface.

Dataset. We employ the FAUST human data set [BRLB14] with the similar experimental setup as in other state-of-the-art methods [MBBV15, BMR*16, MBM*17, VBV18]. The dataset consists of 100 watertight meshes of 10 different subjects with 10 different poses for each. Each mesh contains 6890 vertices and the semantic correspondence among the vertices are already established in the dataset. We utilize this as a ground truth for the training, but pretend no such information is provided in testing.

Input processing. We normalize the 100 mesh models in FAUST data set to have the same surface area, $15,000 \text{ cm}^2$ with an approximate shape diameter of 200 cm. For initial input, we take the canonical XYZ-coordinates of mesh vertices as the input to the network.

Network architecture and parameter setting. For fair comparison with the other state-of-the-art methods [MBBV15, BMR*16, MBM*17, VBV18], we first test $\text{Conv}64 \rightarrow \text{Conv}128 \rightarrow \text{Conv}256 \rightarrow \text{Lin}512 \rightarrow \text{Lin}6890 + \text{softmax}$, a single-scale architecture that generalizes the other methods. The numbers in the layer names indicate the dimension of the output channel. In each of the

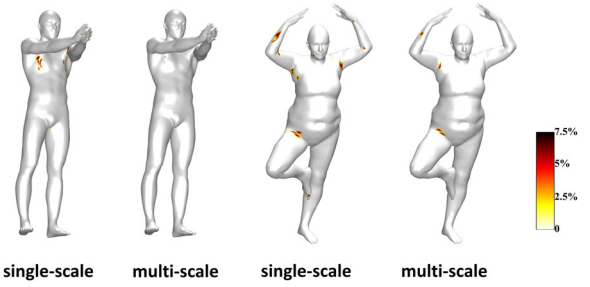


Figure 6: Geodesic errors (in % of the shape diameter) on two test shapes using single-scale and multi-scale ZerNet architectures.

Table 1: Correspondence accuracy on the FAUST human dataset of our method and recent state-of-the-art manifold convolution approaches. Accuracies for the compared methods [MBBV15, BMR*16, MBM*17, VBV18] are directly taken from the corresponding papers.

Method	Input	Accuracy
ACNN [BMR*16] w/o refinement	SHOT	60.6%
ACNN [BMR*16] w/refinement [OBCS*12]	SHOT	62.4%
GCNN [MBBV15] w/o refinement	SHOT	65.4%
GCNN [MBBV15] w/refinement [OBCS*12]	SHOT	42.3%
MoNet [MBM*17] w/o refinement	SHOT	73.8%
MoNet [MBM*17] w/refinement [VLR*17]	SHOT	88.2%
FeaStNet [VBV18] w/o refinement	XYZ	88.1%
FeaStNet [VBV18] w/refinement [VLR*17]	XYZ	92.2%
ZerNet (Ours) w/o refinement	XYZ	94.7%
FeaStNet [VBV18] multi-scale	XYZ	98.7%
ZerNet (Ours) multi-scale	XYZ	96.9%

Conv block, we set the kernel size $r_0 = 5.5 \text{ cm}$ for computing the local exponential map. For the discretization of local exponential map, we first uniformly sample 12,000 surface points across the entire surface via the Poisson disk sampling [PPA13], and collect 50 sampled points in the neighbourhood of every mesh vertex within the radius r_0 . The first 21 *Zernike bases* ($k = 21$) are used for *Zernike* decomposition.

Meanwhile, a multi-scale architecture to incorporate higher contextual information is proposed in [VBV18], and we also compare our method in the multi-scale setting. To this end, we set the kernel sizes (r_0) as 4.5, 5.5 and 6.75 cm, respectively, for three scales of *Conv* blocks (*ZerConv*₁, *ZerConv*₂ and *ZerConv*₃ as in Figure 5). For discretization, we uniformly sample 18,000, 12,000 and 8000 surface points over the mesh surface, respectively, for the three scales of *Conv* blocks. Note that the total number of sample points over the surface is inversely proportional to the square of r_0 . Thus, for different scales, the number of discretized samples surrounding the mesh vertex within its local exponential map remains approximately the same. We collect 50 discretized samples in the neighbourhood and use the first 21 *Zernike bases* ($k = 21$) for *Zernike* decomposition.

The two models above were trained using the Adam optimizer [KB14] with the sparse categorical cross-entropy loss as the objective. The training of our model at a single-scale took on average

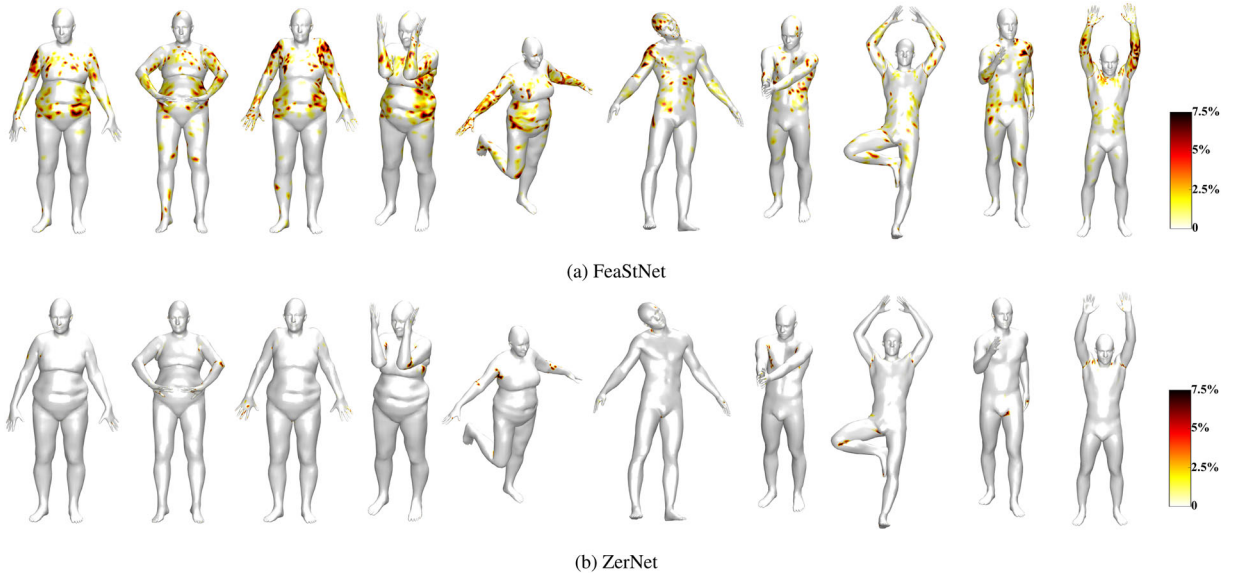


Figure 7: Point-wise geodesic error (in % of the shape diameter) of ZerNet versus FeaStNet on the FAUST human dataset [BRLB14]. Results are generated based on raw performance of both approaches shared a similar single-scale architecture.

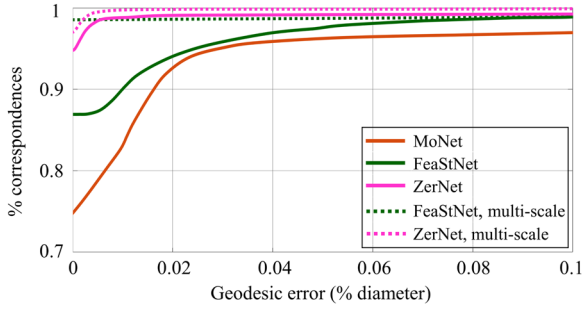


Figure 8: Shape correspondence quality obtained by different methods on the FAUST human dataset. Results are generated based on the raw performance of each method without additional post-process refinements.

174 s per epoch on a machine with a single NVIDIA GeForce GTX 1080 Ti graphics card. The multi-scale model in Figure 5 took 413 s per epoch on average.

Results. To evaluate the performance, we first compare the classification accuracy, defined as:

$$\text{Accuracy} = \frac{\text{true classification}}{\text{total number of surface points}}. \quad (15)$$

The test result is reported in Table 1. As shown in the table, with single-scale architecture, ZerNet outperformed all other state-of-the-art methods with the accuracy of 94.7%, even without additional refinement post-processes [OBCS*12, VLR*17]. In multi-scale setting, FeaStNet [VBV18] outperformed ZerNet with less than 2% margin.

We further evaluated the quality of correspondence using the Princeton Benchmark [KLF11]. The quality of correspondence is measured by the percentage of correctly predicted matches within a geodesic disk around the ground-truth. The results were plotted in Figure 8 with varying radii of the geodesic disk from 0% to 10% of shape diameter. In single scale, our method demonstrates a significantly better quality of correspondence than all the other benchmark methods. In multi-scale, FeaStNet shows the top performance at the zero radius (smallest error tolerance). With larger radii, however, ZerNet begins to outperform FeaStNet.

Figure 6 visualizes the point-wise geodesic correspondence errors on two representative shapes of the two test subjects using our single-scale and multi-scale ZerNet architectures. Figure 7 shows the point-wise geodesic correspondence error of our method in comparison with the most recent state-of-the-art method, FeaStNet [VBV18], based on the similar single-scale architecture.

4.2. Semantic segmentation

In addition, we validate ZerNet on semantic segmentation problem. We compared ZerNet against other state-of-the-art methods, including Toric-cover CNN [MGA*17], PointNet++ [QYSG17], Dynamic graph CNN [WSL*18] and MDGCNN [PO18].

Dataset. We use the human segmentation benchmark in [MGA*17] for comparison. The training set consists of 370 models collected from SCAPE, FAUST, MIT and Adobe Fuse [Ado16]. All models are manually segmented into eight labels, one for the head, one for the torso, three for the arms and three for the legs. The test set is the 18 models collected from the SHREC07 dataset in ‘human’ category.

Table 2: Segmentation accuracy on the human body dataset introduced in [MGA*17] of our method and several state-of-the-art methods. Accuracies for the compared methods [MGA*17, QYSG17, WSL*18, PO18] are directly taken from [PO18].

Method	Input	Accuracy
Toric cover [MGA*17]	WKS,AGD,curv.	88%
Pointnet++ [QYSG17]	XYZ	90.8%
DynGraphCNN [WSL*18]	XYZ	89.7%
MDGCNN [PO18]	XYZ	88.6%
ZerNet (Ours)	XYZ	88.7%



Figure 9: Semantic segmentation on human body shapes.

Experiment setting. We normalize all human models to have the same surface area, 15,000 cm², and take XYZ-coordinate of mesh vertices as the input to ZerNet. For the network architecture and parameters setting, we follow the same setting as the single-scale architecture used for the shape correspondence experiment (Section 4.1). Compared to the dense correspondence task, as the classification classes required for segmentation is significantly reduced (6,890 → 8), a similar architecture but with less output chan-

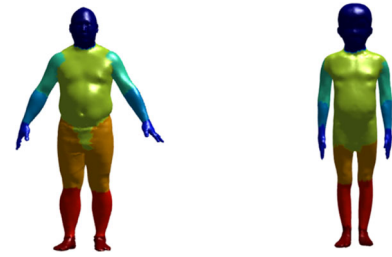


Figure 10: Inconsistent segment labelling in human-labelled ground truth.

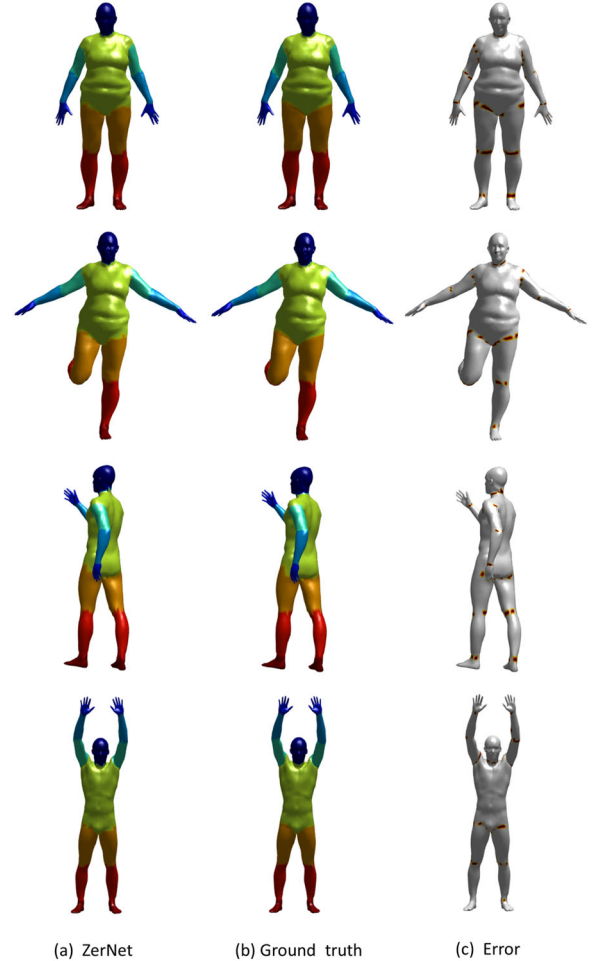


Figure 11: Semantic segmentation on the FAUST human dataset.

nel dimensions of each layer is adopted: *Conv32* → *Conv64* → *Conv128* → *Lin256* → *Lin8* + *softmax*. We trained this architecture on 370 models in the data set, which took 16 min 38 s per each epoch on average.

Results. As shown in Table 2, ZerNet performs the segmentation task in a high accuracy, comparable to the other state-of-the-art methods. As shown in Figure 9 (a), the overall segmentation quality of ZerNet was satisfactory, except near the boundaries between

Table 3: Point-wise regression result of ZerNet vs FeaStNet over five validation cases. Performance was measured with the following criteria: mean absolute percentage error (MAPE), relative root mean square error (RRMSE), Pearson correlation coefficient (PCC) and hit-rate (HR). Subscripts under HR represents the tolerance threshold (10%, 20% and 30%) and HR was calculated as the percentage ratio of the number of the vertices that have the scalar values accurately predicted over the total number of mesh vertices.

Model ID	MAPE	RRMSE	ZerNet			MAPE	RRMSE	PCC	FeaStNet		
			PCC	HR ₁₀	HR ₂₀				HR ₁₀	HR ₂₀	HR ₃₀
TP1a105I	10.11%	13.04%	0.91	65.84%	90.81%	16.09%	19.98%	0.80	48.66%	74.37%	86.36%
TP1a166I	9.57%	12.97%	0.87	68.92%	90.75%	13.38%	15.86%	0.79	45.72%	79.79%	92.89%
TP1a182I	14.32%	18.70%	0.87	45.63%	77.60%	19.99%	24.60%	0.76	35.71%	64.03%	79.27%
TP1a32I	13.80%	16.32%	0.88	50.48%	82.53%	23.89%	25.73%	0.65	28.37%	54.48%	73.38%
TP1a33I	9.60%	13.98%	0.90	65.30%	90.70%	13.78%	17.71%	0.82	52.25%	78.92%	89.92%

Table 4: Point-wise regression result of ZerNet under direction preserving setting over five validation cases.

Model ID	MAPE	ZerNet w/direction preserving			
		RRMSE	PCC	HR ₁₀	HR ₂₀
TP1a105I	9.97%	13.15%	0.91	68.64%	91.31%
TP1a166I	9.90%	12.87%	0.87	65.33%	91.18%
TP1a182I	13.07%	17.91%	0.88	54.05%	82.49%
TP1a32I	13.56%	16.02%	0.88	55.76%	84.56%
TP1a33I	9.37%	12.81%	0.91	67.13%	90.51%

Table 5: Computational time for ZerNet with different angular resolution.

Angular resolution	$s = 1$	$s = 4$	$s = 8$	$s = 16$
Training time	2.08 s	7.55 s	14.58 s	28.26 s
Testing time	1.55 s	4.55 s	8.26 s	15.52 s

segments. For these, we noticed inconsistencies in human-labelled ground truth across the dataset, as in the example illustrated in Figure 10, where the boundaries between the thigh and pelvis did not agree between ground truth labels. In fact, such an error prevailed in the data set from visual inspection. Hence, the minor differences in accuracy within 2% margin in Table 2 could only be interpreted as a sanity check.

Additional experiment on FAUST dataset. Complementary to the previous experiment, here we provide another result of analysis conducted only on the FAUST dataset this time. We found that the segmentation labels in the FAUST dataset are more consistent and reliable since the ground truth labels have been obtained by registering a pre-labelled mesh (details on the registration can be found in the FAUST paper [BRLB14]). The FAUST dataset is comprised of dynamic 3D scans of 10 individuals in 10 different postures each. We used 80 FAUST models corresponding to eight randomly selected individuals in different poses; the remaining 20 models corresponding to the two left-out subjects were used for testing. We adopted the same setting as in the previous experiment for the network architecture and the hyper-parameters.

In this additional experiment, we achieved a segmentation accuracy of 96.6% on average. Figure 11 visualizes some of the test re-

sults. This result proves the good segmentation accuracy of ZerNet, complementary to the result in Table 2.

4.3. Aneurysm wall stress estimation

Lastly, we validate our method on a scalar field regression task. Specifically, the problem is to estimate the mechanical stress distributed over the surface of cerebral aneurysm.

Data set. To this end, we introduce a new benchmark data set comprised of 3D surface meshes of 26 cerebral aneurysm cases. According to relevant literature [LZR07, LZR08, LL16, LFB18], the magnitude of the mechanical stress distribution on aneurysm is known to be correlated with the local surface geometry. The goal here, therefore, is to utilize CNNs to predict the stress distribution on aneurysm based on surface geometry. This is essentially a scalar-field regression problem defined on a surface. The ground-truth values are computed from finite-element (FE) simulations. The aneurysm models are different in mesh topology such that the number of vertices and how the vertices are connected are inconsistent across different models. The total number of mesh vertices across the aneurysm models varies from 1,135 to 8,197 and the surface area is in a range between 27.72 and 169.23.

Input processing. We first normalize the 26 aneurysm models to have the same surface area of 100. colored Based on the fact that the wall stress is proportional to the square root of surface area as the governing physics equation states, the wall stress value is also scaled accordingly. We then uniformly sample 8,000 random points on each mesh surface alongside the stress value, with an assumption that the stress distribution is piece-wise linear colored on each triangle.

Network architecture and parameter setting. We again use the similar architecture as in the single-scale shape correspondence experiment: $Conv128 \rightarrow Conv256 \rightarrow Conv512 \rightarrow Lin800 \rightarrow Lin1$. We set $r_0 = 0.6$ for computing the local exponential map. We use the first 21 *Zernike bases* ($k = 21$) for Zernike decomposition. The Adam optimizer was used for training, with the mean squared error (MSE) loss between ground truth and stress prediction on all sampled points of each aneurysm as the objective. It took 58s on average for one training epoch by using a single GTX 1080 graphics card.

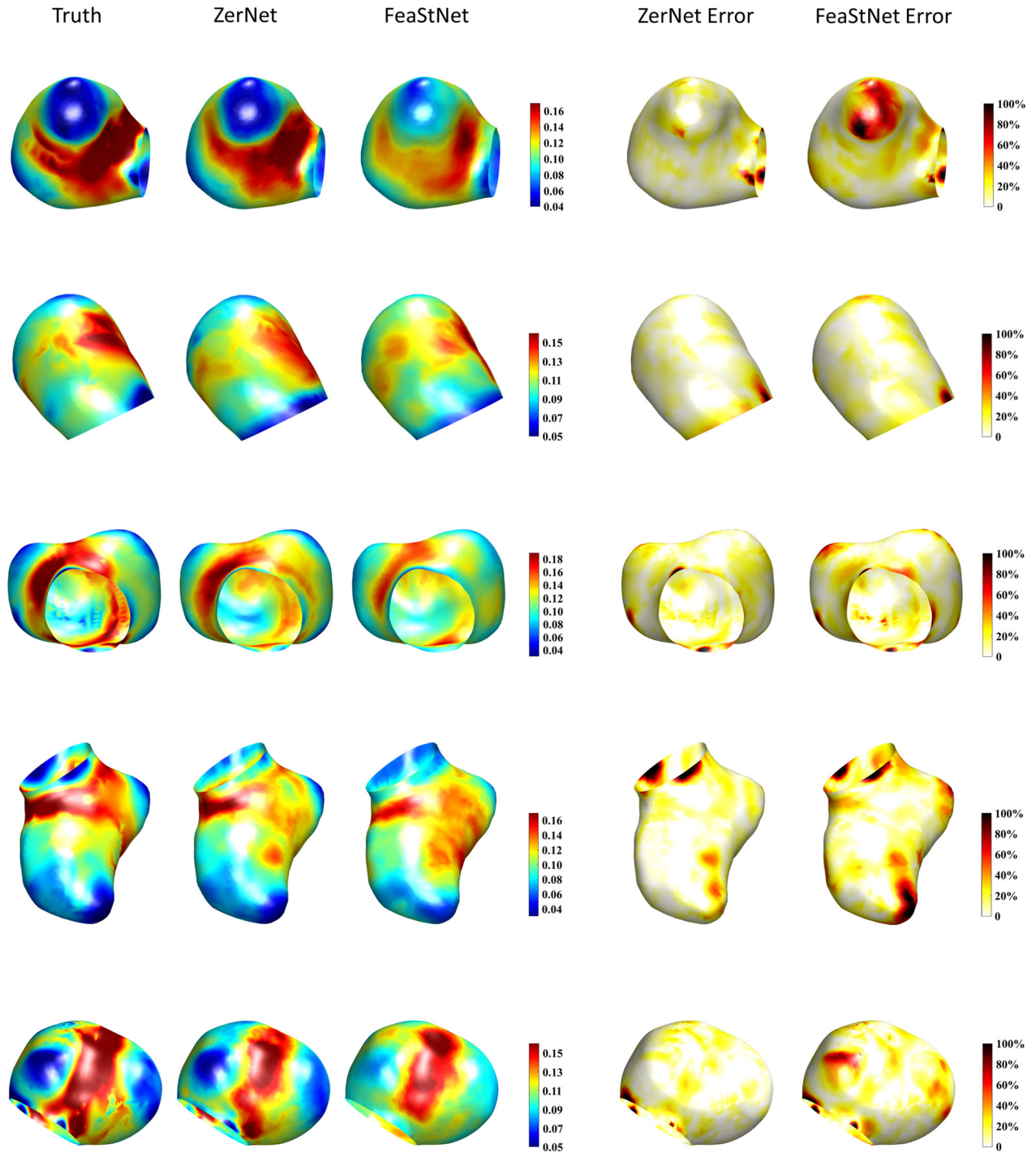


Figure 12: Visualization of the estimated wall stress distribution and the distributed absolute percentage errors on five validation cases. Each row corresponds to the row of the same order in Table 3.

Results. We cross-validated our method along with the current state-of-the-art FeaStNet [VBV18]. Among the 26 aneurysm surfaces, we randomly selected five cases for performing leave-one-out cross validations. For each of the cases among the five, we left it out as the validation set and used the rest 25 mod-

els in our data set as the training set to train the ZerNet or FeaStNet.

The result is reported in Table 3. In addition, Figure 12 visualizes the stress estimation results on the five validation cases and

their point-wise error (in percentage of the true stress value on each vertex). We can notice that ZerNet outperformed FeaStNet for all metrics as presented in Table 3, and achieved a satisfactory quality of stress estimation significantly better than FeaStNet.

It should be noted that neither ZerNet nor FeaStNet is successful in predicting correct values near the boundary. This may be due to incorrect boundary conditions ('zero-padding' is enforced currently). Hence, it would be worthwhile to investigate practical ways to assign boundary conditions in future research.

4.4. Directional convolution

As discussed in Section 2, a critical bottleneck in many geometric CNNs is the suppression of feature directionalities, resulting from the path dependency of the kernel rotation during parallel transport. In the above experiments, we applied angular pooling after each convolution. However, directional features would have been neglected under this setting as the angular pooling applied after each convolution layer suppresses the development of directionality in convolved features. Similar to the idea proposed in [PO18], we can define a *directional function* $f_p(r, \theta; \phi)$ on each tangent space $T_p\mathcal{M}$, as a function $f_p(r, \theta)$ parameterized by an additional direction parameter ϕ . Then, the activation maps in CNNs are modelled as directional functions to resolve ambiguities in feature directions. The advantage of ZerNet, that is, the fact that a rotation of a function is simply a 2×2 matrix rotation, allows us to write an analytic formula for directional functions, as opposed to discrete angular bin based heuristics used in [PO18].

With the idea of directional function, the single-scale correspondence matching experiment in Section 4.1 and the wall stress estimation experiment in Section 4.3 have been redone. For both experiments, we set the angular resolution $s = 4$ and preserve the direction axis in each activation map between *Conv* blocks. For the correspondence matching experiment, we observe a noticeable increase in accuracy from 94.7% to 96.1%. For the wall stress estimation experiment, we also observed meaningful improvement in prediction accuracy as in Table 4. We also have tested in higher angular resolutions, that is, $s = 8, 16$, but the improvement of the performance was limited despite the increase of the computational time (Table 5). Hence, this awaits further investigation and optimization in the future.

5. Conclusion

In this paper, we introduced a new concept of Zernike convolution as a way to generalize convolution to curved surfaces. We showed that Zernike convolution seamlessly generalizes convolution operations to an arbitrary surface in a mathematically rigorous and concise manner. In particular, we proved that manifold convolution can be formalized through decomposition of the local geometry defined on the tangent spaces and that convolution operations became simple dot products of Zernike polynomial coefficients. In addition, we showed that rotations of convolution kernels, which could be critical in manifold settings, could also be rigorously represented as simple 2×2 rotational transforms. Building upon this, we further demonstrated a promising vision of equipping our formulation as a theoretical foundation for direction-preserving convolution, which

will bring more mathematical rigour to the generalization of geometric CNNs in fundamental. As a scalable algorithm developed upon our formulation, ZerNet also illustrated our competitiveness against other state-of-the-art methods on both classification and regression tasks.

For the future work, it would be worth exploring ways to further equip ZerNets with the other essential building blocks of CNNs such as spatial pooling/unpooling, transposed convolution, boundary padding and so forth. Particularly, encoder-decoder type networks on manifolds would be an interesting direction of research, as it can benefit potentially a large amount of computational geometry applications that requires parametrization (i.e. latent space embedding) of geometric shapes (e.g. [BL12, kSFMG12, FB12, BWS*13, LMR*15, BL16, PWH*17]). Furthermore, the injectivity over the tangent plane is not always guaranteed in the current manifold CNN approaches including the present work. For example, thin cylindrical shapes like fingers may have the exponential map defined on a non-injective geodesic disk, as the geodesic disk may self-overlap. Hence, how to properly account such features would also be a valuable direction for research investigation in the future.

References

- [Ado16] Adobe: Adobe fuse 3d characters.
- [AML18] ATZMON M., MARON H., LIPMAN Y.: Point convolutional neural networks by extension operators. arXiv preprint arXiv:1803.10091 (2018).
- [BBL*17] BRONSTEIN M. M., BRUNA J., LECUN Y., SZLAM A., VANDERGHEYNST P.: Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.
- [BL12] BAEK S.-Y., LEE K.: Parametric human body shape modeling framework for human-centered product design. *Computer-Aided Design* 44, 1 (Jan. 2012), 56–67.
- [BL16] BAEK S.-Y., LEE K.: Statistical foot-shape analysis for mass-customisation of footwear. *International Journal of Computer Aided Engineering and Technology* 8, 1/2 (Jan. 2016), 80–98.
- [BMM*15] BOSCAINI D., MASCI J., MELZI S., BRONSTEIN M. M., CASTELLANI U., VANDERGHEYNST P.: Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. *Computer Graphics Forum* 34 (2015), 13–23.
- [BMR*16] BOSCAINI D., MASCI J., RODOLÀ E., BRONSTEIN M. M., CREMERS D.: Anisotropic diffusion descriptors. *Computer Graphics Forum* 35 (2016), 431–441.
- [BMRB16] BOSCAINI D., MASCI J., RODOLÀ E., BRONSTEIN M.: Learning shape correspondence with anisotropic convolutional neural networks. *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)* (2016), 3197–3205.

- [BRLB14] BOGO F., ROMERO J., LOPER M., BLACK M. J.: FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Piscataway, NJ, USA, June 2014), IEEE.
- [BWS*13] BAEK S.-Y., WANG J. H., SONG I., LEE K., LEE J., KOO S.: Automated bone landmarks prediction on the femur using anatomical deformation technique. *Computer-Aided Design* 45, 2 (Feb. 2013), 505–510.
- [BZSL13] BRUNA J., ZAREMBA W., SZLAM A., LECUN Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013).
- [CWKW19] COHEN T. S., WEILER M., KICANAOGLU B., WELLING M.: Gauge equivariant convolutional networks and the icosahedral CNN. In *Proceedings of the 36th International Conference on Machine Learning* (2019), 1321–1330.
- [CWW17] CRANE K., WEISCHEDEL C., WARDETZKY M.: The heat method for distance computation. *Communications of the ACM* 60, 11 (2017), 90–99.
- [DBV16] DEFFERRARD M., BRESSON X., VANDERGHEYNST P.: Convolutional neural networks on graphs with fast localized spectral filtering. *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)* (2016), 3844–3852.
- [DMI*15] DUVENAUD D. K., MACLAURIN D., IPARRAGUIRRE J., BOMBARELL R., HIRZEL T., ASPURU-GUZZI A., ADAMS R. P.: Convolutional networks on graphs for learning molecular fingerprints. *Proceedings of the 28th International Conference on Neural Information Processing Systems* (2015), 2224–2232.
- [DUK*11] DOUILLARD B., UNDERWOOD J., KUNTZ N., VLASKINE V., QUADROS A., MORTON P., FRENKEL A.: On the segmentation of 3D LIDAR point clouds. In *Proceedings of 2011 IEEE International Conference on Robotics and Automation (ICRA)*, (2011), pp. 2798–2805.
- [EG79] EISENBERG M., GUY R.: A proof of the hairy ball theorem. *The American Mathematical Monthly* 86, 7 (1979), 571–574.
- [ESKBC17] EZUZ D., SOLOMON J., KIM V. G., BEN-CHEN M.: Gwcnn: A metric alignment layer for deep shape analysis. *Computer Graphics Forum* 36 (2017), 49–57.
- [FB12] FREIFELD O., BLACK M. J.: Lie bodies: A manifold representation of 3D human shape. In *European Conference on Computer Vision (ECCV)* (Oct. 2012), Part I, LNCS 7572, Springer-Verlag, pp. 1–14.
- [HBL15] HENAFF M., BRUNA J., LECUN Y.: Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163 (2015).
- [HCQ17] HECHTLINGER Y., CHAKRAVARTI P., QIN J.: A generalization of convolutional neural networks to graph-structured data. arXiv preprint arXiv:1704.08165 (2017).
- [HHF*19] HANOCCA R., HERTZ A., FISH N., GIRYES R., FLEISHMAN S., COHEN-OR D.: MeshCNN: A network with an edge. In *ACM SIGGRAPH* (2019).
- [HSB17] HARIK R., SHI Y., BAEK S.: Shape terra: mechanical feature recognition based on a persistent heat signature. *Computer-Aided Design and Applications* 14, 2 (2017), 206–218.
- [KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014).
- [KLF11] KIM V. G., LIPMAN Y., FUNKHOUSER T.: Blended intrinsic maps. *ACM Transactions on Graphics* 30 (TOG) (2011), 79.
- [kSFMG12] KAI SHEN K., FRIPP J., MÉRIAudeau F., et al.: Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models. *NeuroImage* 59, 3 (2012), 2155–2166.
- [KW16] KIPF T. N., WELLING M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016).
- [LFBL18] LUO Y., FAN Z., BAEK S., LU J.: Machine learning-aided exploration of relationship between strength and elastic properties in ascending thoracic aneurysm. *International Journal for Numerical Methods in Biomedical Engineering* 34, 6 (2018), e2977.
- [LL16] LU J., LUO Y.: Solving membrane stress on deformed configuration using inverse elastostatic and forward penalty methods. *Computer Methods in Applied Mechanics and Engineering* 308 (2016), 134–150.
- [LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Transactions in Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.
- [LRR*17] LITANY O., REMEZ T., RODOLÀ E., BRONSTEIN A., BRONSTEIN M.: Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 5659–5667.
- [LZ10] LÉVY B., ZHANG H. R.: Spectral mesh processing. In *ACM SIGGRAPH 2010 Courses* (New York, NY, USA, 2010), SIGGRAPH '10, ACM, pp. 8:1–8:312.
- [LZR07] LU J., ZHOU X., RAGHAVAN M. L.: Inverse elastostatic stress analysis in pre-deformed biological structures: Demonstration using abdominal aortic aneurysms. *Journal of Biomechanics* 40, 3 (2007), 693–696.
- [LZR08] LU J., ZHOU X., RAGHAVAN M. L.: Inverse method of stress analysis for cerebral aneurysms. *Biomechanics and Modeling in Mechanobiology* 7, 6 (2008), 477–486.
- [MBBV15] MASCI J., BOSCAINI D., BRONSTEIN M., VANDERGHEYNST P.: Geodesic convolutional neural networks on

- Riemannian manifolds. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (2015), pp. 37–45.
- [MBM*17] MONTI F., BOSCAINI D., MASCI J., RODOLA E., SVOBODA J., BRONSTEIN M. M.: Geometric deep learning on graphs and manifolds using mixture model CNNs. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017), vol. 1, p. 3.
- [MDSB03] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics III*. Springer, 2003, pp. 35–57.
- [MGA*17] MARON H., GALUN M., AIGERMAN N., TROPE M., DYM N., YUMER E., KIM V. G., LIPMAN Y.: Convolutional neural networks on surfaces via seamless toric covers. *ACM Transactions on Graphics* 36, 4 (2017), 71–1.
- [MPS09] MOOSMANN F., PINK O., STILLER C.: Segmentation of 3D LIDAR data in non-flat urban environments using a local convexity criterion. In *2009 IEEE Intelligent Vehicles Symposium* (2009), IEEE, pp. 215–220.
- [MR12] MELVÆR E. L., REIMERS M.: Geodesic polar coordinates on polygonal meshes. *Computer Graphics Forum* 31, 8 (2012), 2423–2435.
- [MRB*16] MASCI J., RODOLÀ E., BOSCAINI D., BRONSTEIN M. M., LI H.: Geometric deep learning. In *SIGGRAPH ASIA 2016 Courses* (2016), ACM, p. 1.
- [NAK16] NIEPERT M., AHMED M., KUTZKOV K.: Learning convolutional neural networks for graphs. In *International Conference on Machine Learning* (2016), pp. 2014–2023.
- [NWB15] NURUNNABI A., WEST G., BELTON D.: Outlier detection and robust normal-curvature estimation in mobile laser scanning 3d point cloud data. *Pattern Recognition* 48, 4 (2015), 1404–1419.
- [OBCS*12] OVSJANIKOV M., BEN-CHEN M., SOLOMON J., BUTSCHER A., GUIBAS L.: Functional maps: A flexible representation of maps between shapes. *ACM Transactions on Graphics* 31, 4 (July 2012), 30:1–30:11.
- [PO18] POULENARD A., OVSJANIKOV M.: Multi-directional geodesic neural networks via equivariant convolution. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), Article No. 236.
- [PPA13] PEYROT J.-L., PAYAN F., ANTONINI M.: Feature-preserving direct blue noise sampling for surface meshes. In *Eurographics (Short Papers)* (2013), pp. 9–12.
- [PWH*17] PISHCHULIN L., WUHRER S., HELTEN T., THEOBALT C., SCHIELE B.: Building statistical shape spaces for 3d human modeling. *Pattern Recognition* 67 (2017), 276–286.
- [QYSG17] QI C. R., YI L., SU H., GUIBAS L. J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems* (2017), pp. 5099–5108.
- [RCG08] RONG G., CAO Y., GUO X.: Spectral mesh deformation. *The Visual Computer* 24, 7-9 (2008), 787–796.
- [RRBW*14] RODOLÀ E., ROTA BULO S., WINDHEUSER T., VESTNER M., CREMERS D.: Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 4177–4184.
- [SDL18] SCHONSHECK S. C., DONG B., LAI R.: Parallel transport convolution: A new tool for convolutional neural networks on manifolds. arXiv preprint arXiv:1805.07857 (2018).
- [SGW06] SCHMIDT R., GRIMM C., WYVILL B.: Interactive decal compositing with discrete exponential maps. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 605–613.
- [SHG*17] SUN Z., HE Y., GRITSENKO A., LENDASSE A., BAEK S.: Deep spectral descriptors: Learning the point-wise correspondence metric via Siamese deep neural networks. arXiv Preprint: arXiv:1710.06368 (Oct. 2017).
- [SRV16] SHUMAN D. I., RICAUD B., VANDERGHEYNST P.: Vertex-frequency analysis on graphs. *Applied and Computational Harmonic Analysis* 40, 2 (2016), 260–291.
- [SX14] SONG S., XIAO J.: Sliding shapes for 3d object detection in depth images. In *European Conference on Computer Vision* (2014), Springer, pp. 634–651.
- [VBV18] VERMA N., BOYER E., VERBEEK J.: FeastNet: Feature-steered graph convolutions for 3D shape analysis. In *IEEE Conference on Computer Vision and Pattern Recognition* (2018).
- [vF34] VON F. Z.: Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica I*, 7-12 (1934), 689–704.
- [VLR*17] VESTNER M., LITMAN R., RODOLÀ E., BRONSTEIN A. M., CREMERS D.: Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 6681–6690.
- [WSL*18] WANG Y., SUN Y., LIU Z., SARMA S. E., BRONSTEIN M. M., SOLOMON J. M.: Dynamic graph CNN for learning on point clouds. arXiv preprint arXiv:1801.07829 (2018).
- [YSGG17] YI L., SU H., GUO X., GUIBAS L. J.: SyncSpecCNN: Synchronized spectral CNN for 3D shape segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 2282–2290.
- [ZBVH09] ZAHARESCU A., BOYER E., VARANASI K., HORAUD R.: Surface feature detection and description with applications to mesh matching. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), pp. 373–380.