Learning Deep Features for Shape Correspondence with Domain Invariance

Praful Agrawal, Ross T. Whitaker, Shireen Y. Elhabian

arXiv:2102.10493v1 [cs.CV] 21 Feb 2021

Abstract—Correspondence-based shape models are key to various medical imaging applications that rely on statistical analysis of populations of anatomical shapes. Such shape models are expected to represent consistent anatomical features across the population for meaningful, population-specific shape statistics. Early approaches for correspondence placement rely on nearest neighbor search for simpler anatomies. Coordinate transformations for shape correspondence hold promise to address the increasing anatomical complexities. Nonetheless, due to the inherent shape-level geometric complexity and population-level shape variation, the coordinate-wise correspondence often does not translate to the anatomical correspondence. An alternative, group-wise approach for dense correspondence placement is to explicitly model the trade-off between the geometric description of each shape and the statistical compactness of the population. However, due to primitive assumptions, these models achieve limited success in resolving nonlinear shape correspondence. Recent works have addressed this limitation by adopting an application-specific notion of correspondence through lifting positional data to a higher dimensional feature space (e.g., sulcal depth, brain connectivity, and geodesic distance to anatomical landmarks). However, they heavily rely on the manual expertise to create domain-specific features and the consistent placement of landmarks. This paper proposes an automated feature learning approach, using *deep* convolutional neural networks, to extract correspondence-friendly features from shape ensembles. Further, an unsupervised domain adaptation scheme is introduced to augment the pretrained geometric features with new anatomies. Results on anatomical datasets of human scapula, femur, and pelvis bones demonstrate that features learned in supervised fashion show improved performance for correspondence estimation compared to the manual features. Further, the unsupervised learning is demonstrated to learn features for a complex anatomy using the supervised domain adaptation from features learned on a simpler anatomy.

Index Terms—shape correspondence, feature learning, Siamese network, domain adversarial training.

I. INTRODUCTION

Statistical shape models have important applications in various medical imaging tasks, such as image segmentation (e.g., [1]), hypothesis testing (e.g., [2]), anatomical reconstruction (e.g., [3]), and pathology detection (e.g., [4]). The utility

of such models is influenced by a shape representation that facilitates statistical analysis. In this regard, *landmarks* are a popular choice as a lightweight but effective representation. Compared to embedding shapes in the image intensity values at voxels, landmarks-based representation is more intuitive and promotes visual communication of the results [5]. To perform shape statistics, landmarks should be defined consistently within a given population to refer to the same anatomical position on every shape instance, a concept known as *correspondence*. Such correspondences are often created manually, but this is time-/labor-intensive, requiring qualified specialists (e.g., radiologists), and cost-prohibitive for 3D images and large collections of imaging data.

Advances in shape modeling have achieved some success in automatic placement of correspondences on anatomical structures [6]–[9], and have been effective in a range of biomedical applications (e.g., [10]–[19]). However, this success is limited to anatomies that conform to the modeling assumptions in existing methods. Typically, the assumptions include mapping to primitive shapes (such as a sphere) or predefined initial atlas, linear shape variations, and pairwise comparisons with no consideration of population-based shape variations. Such assumptions often do not generalize well for complex anatomies [20]. The lunate surface of the acetabulum in a pelvis, for example, has a horseshoe-like shape that does not adhere to a standard primitive. Pairwise comparisons do not observe the entire population, leading to biased and suboptimal models [21]–[23], and in many cases fail to find surface-to-surface correspondences across populations. Anatomical variability can be far more complex than linear approximations, in which case nonlinear variations normally exist, e.g., bending fingers, soft tissue deformations, and vertebrae with different types (lumbar, thoracic, and cervical). Furthermore, the complexity of a shape anatomy can be quantified as a comparison between the statistical shape variations and the local shape features (such as Feret's diameter [24]). The complexity of the shape increases as the shape features get smaller compared to the observed variations among different samples. Due to this smaller feature size, the task of establishing correspondence becomes more challenging and thus limits the success of existing methods.

Previous works in the context of brain imaging have included sulcal depth [21], brain connectivity [25], and anatomical landmarks [26] as additional domain knowledge to guide the correspondence optimization for complex shapes. Although promising, these approaches are specific to a particular anatomy and thus not generalizable to other anatomies. Further, their success heavily relies on the domain-specific

expertise (and time) to find the geometric features or landmarks consistently on shape samples across the population. Furthermore, such domain knowledge cannot be transferred to new anatomies.

In [27], we presented an automated feature-learning to aide the generation of point-wise shape correspondence models. The idea is motivated by recent advances in computer vision and computer graphics that use deep convolutional neural networks (CNNs) to learn shape features to establish pairwise correspondences between shapes (e.g., [28]–[30]). In [27], shape-specific local geometric features are learned via a surrogate task of correspondence vs noncorrespondence classification for a pair of points sampled from the shape surface(s). At the specific point location, a snapshot of local surface geometry is used as input to the convolutional neural network to facilitate the feature learning. The results indicate that the learned features represent the inherent shape anatomy and are able to successfully guide the optimization of point-wise correspondences. This paper provides an extended description of methods and experimental results presented in [27]. Further, we propose a methodology to generate geometric features for a new complex anatomy via unsupervised domain adaptation of pretrained features. We use orthopedic data with femur and pelvis shapes to showcase successful feature adaptation. The contributions of this work can be summarized as follows:

- supervised deep feature learning using *direct surface geometry* rather than relying on predefined surface descriptors;
- incorporating the *learned features* in the optimization of dense correspondences for complex shape ensembles.
- unsupervised adaptation of pretrained features to new, complex anatomies; and
- comprehensive experiments showcasing supervised feature learning and unsupervised feature adaptation.

II. RELATED WORK

The relevant work falls into three categories: existing methods used to generate shape correspondence models, deep learning methods used to establish shape correspondence, and domain adaptation schemes, with a particular focus on feature-level adaptation.

A. Correspondence Estimation

Correspondence-based shape models represent each surface as a point cloud, where the notion of correspondence is established by the index position of each point. In an optimal correspondence model, a particular index should represent the same anatomical location across all the point clouds. The problem of solving point-wise correspondence is challenging due to the variations observed across different shapes. Existing shape modeling approaches to estimating correspondences fall into two broad families: a *pairwise* approach that relies on pairwise shape comparisons and a *groupwise* approach that observe the entire population.

Pairwise methods that rely on finding nearest points between surfaces (e.g. [31] and their variants) work well for limited anatomies with large sized shape features but are prone to

mismatches with complex geometries having high curvature regions and small sized shape features. Alternatively, other pairwise methods try to match predefined parameterizations of surfaces while maintaining regularity. For example, the spherical harmonics point distribution model (SPHARM-PDM) [9] relies on a smooth one-to-one mapping from each shape instance to the unit sphere. The mapping to a predefined surface topology and the a priori assumptions on smoothness, rather than population-specific shape features, limit the class of anatomies and populations for shape modeling. Correspondence estimation based on pairwise coordinate transformations (e.g., [8], [32], [33]) holds promise, but they fail to adapt the deformation metric to the inherent modes of variability in the population, and in many cases fail to find surface-to-surface correspondences across entire populations.

Groupwise approaches use the statistics of the population itself to drive the matching of shape features. For instance, minimum description length (MDL) [7] optimizes point correspondences using an information content objective, but it relies on intermediate spherical surface parameterizations, which places limitations on the types of shapes and the optimization process. Forgoing the parameterization, the particle-based shape modeling (PSM) approach has offered a flexible nonparametric and general framework to establish dense point correspondences across shape ensembles without constraining surface topology or parameterization [6], [22], [34]. The PSM approach explicitly models the inherent trade-off between the geometric description of each shape and the statistical compactness of the population via an entropy optimization scheme. Recently, it has been show that the PSM approach outperforms other pairwise [9] and groupwise [8] methods in consistently estimating anatomical measurements and capturing clinically relevant population-level shape variations [20]. In this paper, we thus focus on PSM and its variants to demonstrate the efficacy of learned shape features for correspondence optimization.

Current PSM implementations rely on a Gaussian model in the *shape space*, which is the vector space formed by the spatial coordinates of surface correspondences (modulo a similarity transform). However, the distribution of anatomical structures can be far more complex than the Gaussian assumes, and surface locations are not always indicative of their correspondence. To address these shortcomings, Datar *et al.* [26] proposed to use geodesic distances to user-identified landmarks on individual shapes. These point-wise distance values were used to guide the entropy-based optimization of shape correspondence. Further, Oguz *et al.* (in the context of brain imaging) considered sulcal depth [21] and brain connectivity [25] as additional features in the PSM optimization. While promising, such approaches are tailored to a particular dataset, anatomy and application. In [27], we proposed to automatically learn the shape-specific anatomical features in supervised fashion. Here, we propose to adapt the pretrained features for new anatomies.

B. Deep Learning for Shape Correspondence

Shape matching methods in computer vision and computer graphics represent shapes as 3D meshes, and use point-wise

matching to establish correspondence between a pair of shapes. Here, we focus on scenarios where point-wise matching is simultaneously optimized with population-level statistics.

Recent works have trained deep convolutional neural networks (CNNs) to estimate the point-wise matching between an input 3D mesh and a template mesh [35]. Boscaini *et al.* [29] proposed a supervised learning of pointwise shape correspondence using anisotropic CNNs. The network was trained using a predefined correspondence model. Thus, this method requires retraining for a new set of correspondences. Groueix *et al.* [36] estimated shape correspondence with respect to a predefined template using feature descriptors. The choice of template shape impacts the resulting statistical variations, and thereby does not lead to a generic population-based statistical model. Kleiman and Ovsjanikov [37] adopted a graph matching approach to establish region-wise correspondence across shapes. The method works with any given shape representation and relies on inherent symmetric nature of shapes. This symmetric property limits the success for shapes with significantly varying topologies. Sun *et al.* [38] trained a Siamese network for a correspondence vs noncorrespondence classification of an input pair. The training relies on predefined spectral shape descriptors to enable shape matching. Although predefined descriptors may not generalize to complex shapes, this method does not require selection of a template. Hence, we adopt this idea of a Siamese network for a correspondence vs noncorrespondence classification. As input, we have chosen to use a direct snapshot of local surface geometry in place of spectral descriptors, enforcing the network to automatically learn the shape features.

C. Domain Adaptation

Here, our goal is to learn a set of geometric features that may generalize to unknown classes of shapes. The supervised learning of such generic features requires a large pool of exhaustive training data. Further, for complex shapes, the labeled training pairs are often expensive and require manual expertise. Thus, to alleviate the need for labeled data, we refine the pretrained features on simple anatomies to model new anatomies in an unsupervised way borrowing ideas from unsupervised domain adaptation techniques.

A recent survey on domain adaptation [39] illustrates the different techniques to learn semantically meaningful and domain invariant features. Broadly, the domain adaptation techniques follow either adversarial training paradigms or a feature space matching approach. In the feature space matching, the source and target distributions are made to align in the output feature space. For instance, Sun *et al.* [40] formulated the problem as minimizing the distance between covariance matrices of source and target domain feature spaces, thus resulting in aligning the source domain feature space with that of the target domain. Fernando *et al.* [41] used subspace matching to map the input data from the source domain onto the target domain. Although simple, Fernando *et al.* 's approach is dependent on the correct subspace estimation, which typically requires a large number of samples. In the case of a limited number of samples in either or both domains, these methods may lead to overfitting.

Adversarial training is a more generic approach where a discriminator function is trained on the output features, with an adverse objective function. The goal of this adversarial objective is to promote confusion for the discriminator function, to distinguish between source and target domains. There are two types of models for adversarial training; 1) based on a generative model, which tries to generate impostor samples for the discriminator network [42]–[44]; and 2) based on learning discriminative features by directly using them as inputs to the discriminator network. The goal is to ensure that the output features perform poorly at discriminating between samples from the two domains, while being good at the original task [45]–[49].

In this paper, we adopt the domain adversarial training proposed by Ganin and Lempitsky [45], where the output features are used to solve a parallel domain classification problem. The sign-reversed gradients from the augmented discriminator network are backpropagated to encourage domain confusion.

III. METHODOLOGY

In this paper, we propose to train a deep neural network to learn features that capture the geometry of any given anatomy. Specifically, the features should highlight the shape properties of the anatomy, such as convexity/concavity, saddle structures, and flat regions. Such a feature bank can help correlate anatomical regions across different shape samples, and thus guide the correspondence optimization. We devise a supervised feature learning scheme via a paired-classification task. As input to the deep neural network, we extract local surface geometrical characteristics using circular geodesic patches in the neighborhood of given point locations on a shape's surface (see Figure 1). Once trained, the network should encode similar anatomical regions with similar feature vectors. We further use the unsupervised domain adversarial training [45] to adapt the trained network for new anatomies. The trained feature maps, supervised and/or unsupervised, are then used to aide shape correspondence optimization in the context of particle-based shape modeling (PSM) [6], [34].

In this section, we discuss the details of input patch extraction, followed by the deep network configurations used for supervised feature learning and unsupervised feature adaptation to new anatomies. We conclude the section with details to incorporate the extracted features into an existing method for correspondence optimization and evaluation criterion used for the resulting shape-correspondence models.

A. PSM Background

Particle-based shape modeling (PSM) is a method for constructing compact statistical point-based models of shape ensembles while not relying on any specific surface parameterization [22], [34], [50], [51]. PSM uses a set of dynamic particle systems, one for each shape, in which particles interact with one another with mutually repelling forces to cover optimally, and therefore describe the surface geometry. An image segmentation process is generally used to extract the anatomy of interest in which the surface geometry is described implicitly as the interface between foreground and background

regions. PSM makes use of the signed distance transform (DT) of this binary representation for finite numerical calculation of surface geometry. Here we give a brief review of the PSM method.

PSM formulation: Consider a cohort of shapes $\mathcal{S} = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N\}$ of N surfaces, each with its own set of M corresponding particles $\mathbf{z}_n = [\mathbf{z}_n^1, \mathbf{z}_n^2, ..., \mathbf{z}_n^M] \in \mathbb{R}^{dM}$ where each particle $\mathbf{z}_n^m \in \mathbb{R}^d$ whose ordering implies correspondence among shapes (typically $d = 3$ for anatomical structures). This representation involves *two types of random variables*: a shape space variable $\mathbf{Z} \in \mathbb{R}^{dM}$ and a particle position variable $\mathbf{X}_n \in \mathbb{R}^d$ that encodes particles distribution on the n−th shape (configuration space). For groupwise modeling, shapes in $\mathcal{S}$ should share the same world coordinate system. Hence, generalized Procrustes alignment [52] is used to estimate a rigid transformation matrix $\mathbf{T}_n$ that can transform the particles in the n−th shape local coordinate $\mathbf{x}_n^m$ to the world common coordinate $\mathbf{z}_n^m$ such that $\mathbf{z}_n^m = \mathbf{T}_n \mathbf{x}_n^m$. Correspondences are established by minimizing a combined shape correspondence and surface sampling cost function $Q = H(\mathbf{Z}) - \sum_{n=1}^{N} H(\mathbf{X}_n)$, where H is an entropy estimation assuming Gaussian shape distribution with covariance $\mathbf{\Sigma}$ in the shape space and Euclidean particle-to-particle repulsion in the configuration space [34]. In particular, $H(\mathbf{Z}) \approx \frac{1}{2}\log|\mathbf{\Sigma}| = \frac{1}{2}\sum_{j=1}^{N}\log\lambda_j$ where λ_j are the eigenvalues of $\mathbf{\Sigma}$. Since $N \ll dM$, covariance is estimated in the dual space of dimension N which defines the hyperplane in $\mathbb{R}^{dM}$ where all the N−samples inhabit. This formulation favors a compact ensemble representation in shape space (first term) against a uniform distribution of particles on each surface for accurate shape representation (second term). The optimization process is defined via gradient descent, by moving individual points on the surface to minimize Q, as described in [34].

B. Geodesic Patch Extraction

The success of deep convolutional neural networks (CNNs) has been demonstrated in analyzing functions defined on Euclidean grid-like domains such as images. Nonetheless, non-Euclidean data, particularly surface geometry, do not directly benefit from these deep models in which operations such as convolution and pooling are not readily defined. However, using a 3D patch from a volumetric shape representation (e.g., label maps or signed distance maps) to represent the local surface geometry may not yield optimal results. Furthermore,

using 3D convolutions will result in a very large number of network parameters with a limited field of view.

Recently, local charting methods (e.g., [28]) have been proposed as a generalization of convolution to non-Euclidean domains where a *patch operator* is defined to extract the local surface patches that are subsequently associated with predefined *shape descriptors*. Here, we rely directly on surface geometry to compute such local shape descriptors. Specifically, a spatial function of surface distance to the tangent space can encode local geometrical properties in a local neighborhood. Hence, we propose to use signed normal distance to the tangent space sampled in the geodesic neighborhood of a surface point $\mathbf{x}$ as a snapshot of the local surface geometry. We use the principal curvature directions $(\vec{u}, \vec{v})$ to define the local intrinsic coordinate system at the surface point $\mathbf{x}$. As illustrated in Figure 1, neighboring points that lie in the geodesic ball $\mathcal{B}(\mathbf{x}) = \{\mathbf{x}' : d_X(\mathbf{x}, \mathbf{x}') \leq \rho_o\}$ with a radius $\rho_o > 0$ are sampled on a circular geodesic patch. Each ring in the patch corresponds to a particular geodesic distance $d_X(\mathbf{x}, \mathbf{x}') = \rho$ with $\rho \in [0, \rho_o]$. Total 64 concentric rings are sampled with radii increasing by a fixed geodesic distance upto a maximum of rho_o. Every ray originating from $\mathbf{x}$, perpendicular to the geodesic rings, is inclined at an angle $\theta \in [0, 2\pi)$ to the first principal curvature direction at $\mathbf{x}$. Total 64 rays are uniformly sampled to cover the 2π angle. Thus, resulting in a rectangular patch of size 64×64. The parameter ρ_o is set to 5% of the maximum shape diameter in the ensemble.

Principal curvature directions are estimated using least squares fitting of a quadratic [53]. We enforce the right-hand rule on local coordinates with the normal direction representing the z-axis. However, the principal curvature directions are accurate only up to the sign of the vector. To address this ambiguity, we extract two patches per surface point (see Figure 1c), one with bases defined by the positive sign of principal directions and the other by negative sign, and use them as a two-channel image for input to the neural network. The other two combinations with opposite signs of the two directions are invalid as they violate the right hand rule.

There still remains ambiguity about order of the two patches in the two-channel input image. We perform the supervised training of paired-classification such that the network is invariant to this ambiguity. Thus we use all four possible combinations of an input pair as independent training samples to help the network become invariant to this sign ambiguity.

C. Supervised Feature Learning

We use deep CNNs to learn local shape features for establishing shape correspondence. The goal is to make some of the hidden layers of the network respond similarly at corresponding locations across the given shape ensemble, and dissimilarly at noncorresponding ones. Therefore, we design the feature learning problem as a paired-classification task (corresponding vs noncorresponding).

Given a pair of inputs, the task is to identify if they belong to a similar anatomical location. To accomplish this task, we use a Siamese network configuration, which consists of two identical copies of the same deep CNN. We train the
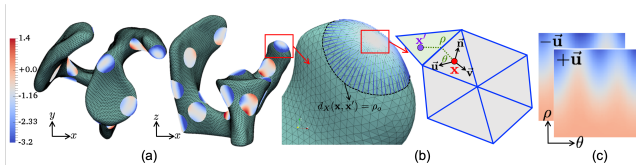

Fig. 1. Geodesic patch extraction at a surface point $\mathbf{x}$: (a) Sample patches on a human scapula mesh. (b) Finding a point ($\mathbf{x}'$ at a geodesic distance ρ in direction that makes angle θ with first principal curvature direction ($\mathbf{u}$) in the tangent space of $\mathbf{x}$. (c) Two channels of an input to the CNN, representing geodesic patches where every pixel corresponds to the signed normal distance of point $\mathbf{x}'$ in the patch to the tangent plane at $\mathbf{x}$.
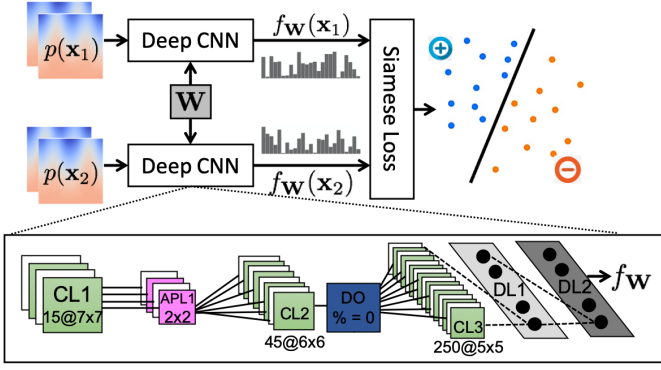
Fig. 2. The Siamese network architecture used to learn features for correspondence.

Fig. 3. The combined network configuration used for adversarial training of the Siamese network.

Siamese CNN to learn the binary classification problem, where the positive class indicates correspondence between a pair of input patches, and the negative class suggests that patches possess different shape characteristics. Contrastive loss [30] is optimized to facilitate the Siamese training. After training, one copy of the two identical deep CNNs in the Siamese network becomes a nonlinear filter bank (Figure 2) that is computed on all mesh vertices to produce feature maps on the surface. We rely on a given point-cloud based correspondence model to generate the supervised training samples, i.e., the corresponding and noncorresponding pairs. The training data consists of a small set of quality controlled point clouds generated using position-based [50] and geodesics-based [26] PSM methods. The geodesic patches extracted at the point locations are fed as inputs to the Siamese network.

For the CNN configuration, we modify the network configuration of Chopra *et al.* [30] with added regularization using batch normalization and dropout layers; see Figure 2. Beyond the regularization layers, the adapted network configuration includes following updates:

- In our experiments, we found that using dropout layer does not impact the classification performance; therefore, we set the probability of dropout to zero.
- Batch normalization is used at the end of convolutional layers.
- It is critical for the correspondence optimization methods that the output features exhibit smooth transitions from one shape feature to another. Therefore, we replace the max pooling layers with the average pooling.
- We use the softplus activation function in place of ReLU. Softplus activation ensures smooth output features, while maintaining faster convergence behavior similar to ReLU.
- All the weights are initialized using a zero mean normal distribution with 0.05 standard deviation, and bias is initialized with zero.
- We subtract the mean of the training data from all training patches (and the same mean is subtracted from testing patches for feature extraction) such that the input data can align with the distribution of initial network weights (spanning the hypersphere).
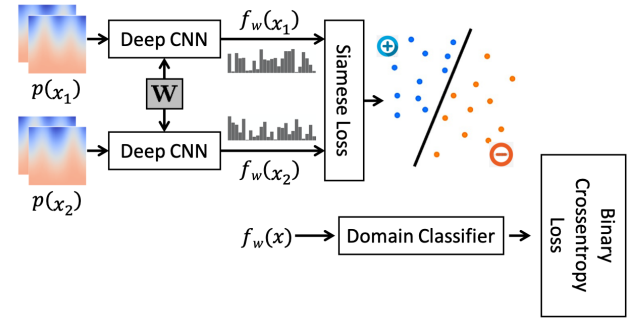
D. Unsupervised Feature Adaptation

The supervised feature learning scheme relies on a given correspondence model to provide training correspondences for the Siamese network. Positional PSM or any other shape modeling approach can provide such training samples. Nonetheless, for highly variable anatomical populations, optimal correspondences are not expected, hence the purpose of learning deep features. In this regard and to alleviate the need for manual intervention through the selection of anatomical landmarks for network training, we make use of recent advances in machine learning for domain adaptation, in which correspondence models optimized for simple varying anatomies (or even synthetic shape ensembles with analytical correspondences), a.k.a. source domain, will leverage populations of unknown correspondences, a.k.a. target domain, with deep learned features that are invariant w.r.t. the shift in shape distribution between the two domains.

Given a pretrained feature (Siamese) network for a particular anatomy S, the objective is to generalize the output features to a new anatomy, without providing the labeled corresponding/noncorresponding samples for this new anatomy T. To accomplish this objective, we adopt the idea of adversarial training, proposed by Ganin and Lempitsky [45]. In [45], feature-level domain adaptation is achieved by augmenting a domain classifier onto the original learning network (Siamese network in our case, see Figure 3). The domain classifier is tasked to identify the domain of an incoming sample, represented by the feature vector $f_w(\mathbf{x})$. The gradients from this domain classifier network are backpropagated into the Siamese network after a sign reversal, typically performed by a gradient reversal layer. This sign reversal facilitates adversarial training of the domain classifier network, and makes it difficult to identify the domain of the input sample. As a result of this confusion, the CNN in the Siamese network is forced to update the parameters so as to combine the topology of the new anatomy T with the shape characteristics of S that is being learned via supervision.

Similar to the configuration used by Ganin and Lempitsky [45], we use a two-layered network for the domain classification. The two layers contain 500 and 1000 convolution filters of size 1×1. The binary crossentropy loss function is used for the domain classifier ($Loss_{domain}$) with a weight λ. Combined with the contrastive loss function [30] as $Loss_{siamese}$,

the total loss for the joint network becomes, $Loss_{total} = Loss_{siamese} + \lambda Loss_{domain}$. The value of $\lambda = 0.01$ achieved stable training of the combined network. The Siamese part of the network is initialized using the weights from pretrained network. Glorot initialization [54] is used for the domain classifier. The initialized network is trained to minimize the total loss, using the samples from S and T. The samples from S have corresponding/noncorresponding labels as well as the domain identifying label (i.e., source domain), whereas samples from T possess only domain identification (i.e., target domain). At the end of training, the CNN is expected to produce features that capture shape characteristics from both S and T anatomies.

E. Deep Feature PSM

We incorporate our learned features into the framework of entropy-based PSM [6], [34]. The method uses a set of dynamic particle systems, one for each shape, in which particles interact with one another using mutually repelling forces to sample (and therefore describe) the surface geometry. Consider a cohort of shapes $\mathcal{S} = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N\}$ containing N surfaces, each with its own set of M corresponding particles. Each shape vector $\mathbf{z}_n = [\mathbf{z}_n^1, \mathbf{z}_n^2, ..., \mathbf{z}_n^M] \in \mathbb{R}^{dM}$ such that $\mathbf{z}_n^m \in \mathbb{R}^d$ is the position vector of $m-$th particle on $n-$th shape, and ordering implies correspondence across shapes. Correspondences are established by minimizing an entropy-based cost function that consists of shape correspondence and surface sampling cost $Q = H(\mathbf{Z}) - \sum_{n=1}^{N} H(\mathbf{X}_n)$. Here, $H(\mathbf{Z})$ measures the differential entropy, assuming multivariate normal distribution over samples in the dM-dimensional shape space. Minimizing this entropy enables correspondence among shape vectors. Further, $H(\mathbf{X}_n)$ is a nonparametric density estimator in the (3D) configuration space of the $n-$th shape. Maximizing this entropy ensures uniform sampling of the $n-$th shape surface. For groupwise modeling, shapes in $\mathcal{S}$ should share the same world coordinate system. Generalized Procrustes alignment [52] is used to estimate a rigid transformation matrix $\mathbf{T}_n$ per shape instance, such that $\mathbf{z}_n^m = \mathbf{T}_n\mathbf{x}_n^m$. In particular, correspondence entropy relies solely on particle positions to achieve compact ensemble representation in shape space (first term) against a uniform distribution of particles on each surface for accurate shape representation (second term).

In this paper, we propose to modify the correspondence term $H(\mathbf{Z})$ by incorporating the features extracted from a trained CNN. Optimizing the updated objective function results in a compact statistical shape model for complex anatomies. The extracted features encode only local geometric properties and require position information, unlike geodesic distances used in [26], which can encode global shape characteristics. Therefore, we also keep the particle positions to guide optimization for locating corresponding points. The updated shape vector becomes $\mathbf{z}_n = [\mathbf{z}_n^1, ..., \mathbf{z}_n^M]^T \in \mathbb{R}^{(d+L)M}$, where $\mathbf{z}_n^m = [f^1(\mathbf{x}_n^m), ..., f^L(\mathbf{x}_n^m), (\mathbf{T}_n\mathbf{x}_n^m)^T]$, and $f^l(\mathbf{x}_n^m)$ is the $l-$th deep feature extracted at the $m-$th particle on the $n-$th shape.

F. Evaluation Criterion

We assess the optimized correspondence models using qualitative inspection and quantitative metrics. The qualitative assessment of shape correspondence models entails examining the anatomical correctness and integrity of the mean shape. Further, modes of variation, computed using principal component analysis (PCA), should reveal population-specific variations. For a quantitative assessment, the resulting shape generative model is evaluated. To construct the generative model, a low-dimensional PCA subspace is learned from the high-dimensional shape space. This generative model is evaluated using the following quantitative metrics:

- *Variance plot:* We plot the percentage of variance with respect to PCA (principal component analysis) modes to demonstrate the compactness of sample distribution in the resulting subspace (higher is better).
- *Generalization:* Proposed by Davies *et al.* [55], this quantifies the ability of the generative model to represent the unknown samples. The reconstruction error of the correspondence model for an unseen shape instance, using the PCA model built from training samples, is evaluated in a leave-one-out fashion. We plot the mean of the point-wise physical distance between test samples and their reconstruction with respect to the number of PCA modes (lower is better).
- *Specificity:* Proposed by Davies *et al.* [55], this measures the ability of the generative model to produce plausible shapes. It is quantified as the Euclidean distance between a sampled shape (from PCA model built using all training samples) and the closest training sample (lower is better). The average over 1000 samples is used.

IV. RESULTS AND DISCUSSION

We showcase the results of supervised feature learning and unsupervised adaptation of pretrained features using synthetic and clinical datasets. First, we demonstrate the success of supervised-learned features in establishing a good shape-correspondence model for complex anatomies. We then showcase the adaptation of features learned from a simpler anatomy to a complex anatomy.

A. Shape Correspondence Models Using Supervised Features
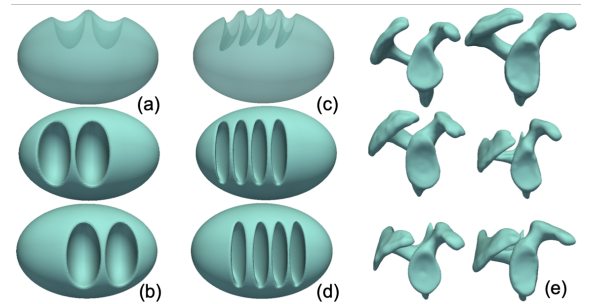

Fig. 4. Sample shapes from synthetic and clinical datasets: (a) side view and (b) top view of Bean2, (c) side view and (d) top view of Bean4, (e) scapula. Left column – controls, right column – patients.
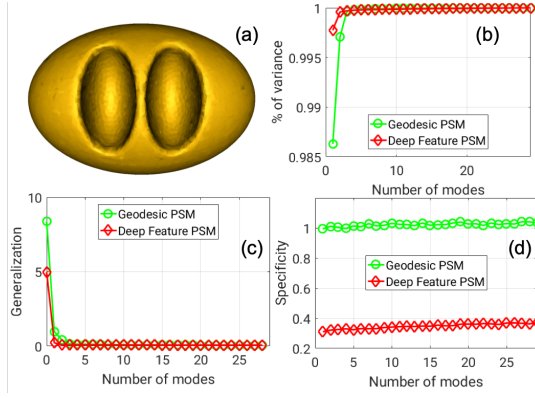
Fig. 5. Bean2 results: (a) mean shape from deep feature PSM, (b) variance plot, (c) generalization (mm), (d) specificity (mm), voxel size=1mm.
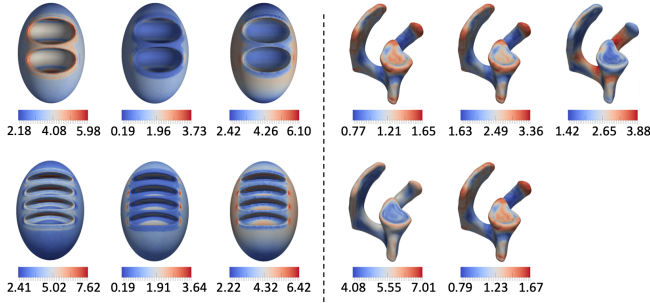

Fig. 6. Sample feature maps from networks trained using Bean2 and Scapula datasets.

Two synthetic datasets and a clincal dataset are used to demonstrate the ability of deep feature PSM to generate improved statistical models for complex shapes compared to relying on positional or handcrafted features. The proposed method (deep feature PSM) is compared with positional PSM [6], [34] and geodesic PSM [26].
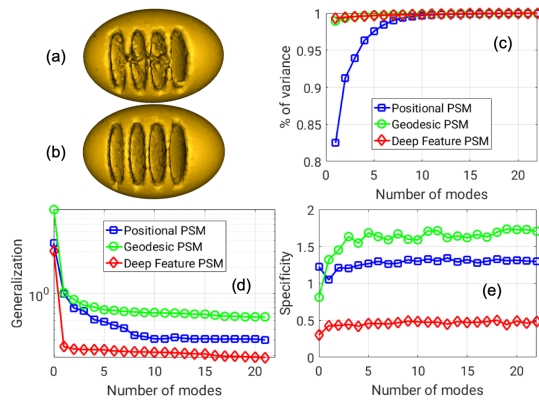

Fig. 7. Bean4 results: (a) mean shape from positional PSM, (b) mean shape from deep feature PSM, (c) variance plot, (d) generalization (mm), (e) specificity (mm), voxel size = 1 mm.

1) Proof-of-Concept: The first synthetic dataset (Bean2) contains 30 shapes that represent a coffee bean with a spatially varying structure. The second dataset (Bean4) is comprised of 23 samples of a more complex coffee bean shape, with a group of closely located thin structures collectively varying in

position (see Figure 4). To generate training samples for the Siamese network, we use an optimized statistical shape model with 3072 points for Bean2, obtained using the geodesic PSM [26]. Patches extracted using correspondence points from 6 randomly selected shapes are used to generate the training data. The CNN configuration with $L = 10$ output features yields an optimal Siamese classification performance of 0.92 AUC (area under the curve) of the ROC (receiver operating characteristics) curve, resulting in a 90% true positive ratio (TPR) at the expense of a 20% false positive ratio (FPR). Figure 6 shows sample features. Given that there are multiple regions with similar shape characteristics, which may lead to a higher FPR, 20% is a relatively small penalty. Moreover, using position information in deep feature PSM will help reduce the impact of false positives. All 10 features are used in PSM to generate a shape model with 4096 correspondence points. Figure 5 presents the mean shape and quantitative evaluation of the correspondence model. Results indicate compactness of the statistical shape model and better generalization and specificity performance over the geodesic PSM.

Bean4 shapes have similar characteristics to those of Bean2, and therefore, we use the same trained Bean2 network to extract features for Bean4 shapes (i.e., without feature adaptation). Figure 6 presents samples of the same feature on the two datasets. Comparative results on Bean4 data, presented in Figure 7, highlight the better performance of the proposed method over positional PSM in generating a more compact statistical shape model. The generative model from the proposed method also outperforms in its ability to generalize over unseen data and to produce plausible shape samples.

2) Clinical Data: A clinical dataset of 20 scapula shapes (10 controls and 10 patients with osseous Hill-Sachs lesions) is used for experiments. Samples are rigidly aligned with respect to the glenoidal plane, and a set of 16 anatomical landmarks are defined manually. Reconstructed scapula meshes are then clipped to the glenoid, acromion, and coracoid to model the areas of high geometric curvature related to constraint of the humeral head; Figure 4 illustrates sample shapes. The shape model with 2432 points and 6 randomly selected shapes are optimized using the geodesic PSM to generate the training data. Using $L = 5$ output features gives optimal classification
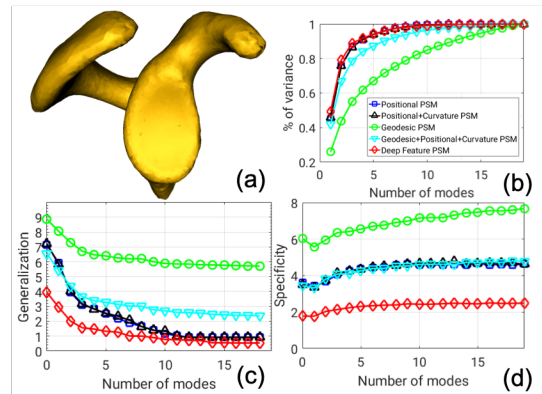

Fig. 8. Scapula results: (a) mean shape from deep feature PSM, (b) variance plot, (c) generalization (mm), (d) specificity (mm), voxel size = 0.5 mm.

performance of AUC= 0.80. Figure 6 shows sample feature maps.

For additional comparison, we augment geodesic and positional PSM with local surface curvature. Figure 8 showcases the mean shape from correspondence model generated using the proposed deep features, and shows the improved quantitative measures in comparison to handcrafted features. It is important to note that the proposed method is able to achieve generalization and specificity of about 2 voxels using dominant modes (5 modes) in contrast to a minimum of 3 voxels from other methods.

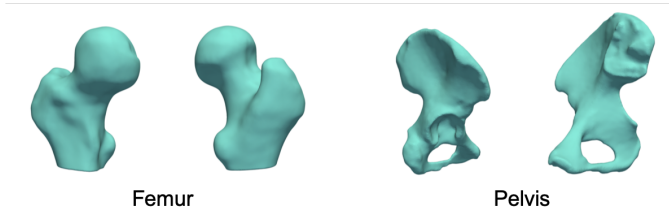### B. Shape Correspondence Models Using Adapted Features

Fig. 9. Anatomical complexity in sample femur and pelvis shapes.

Using the coffee bean and human scapula datasets, we demonstrated the improved performance of statistical shape models using the proposed deep learned features in the PSM optimization. However, for highly complex medical anatomies, labeled training data are often not available. Therefore, the goal is to capitalize on supervised learning of relatively simpler anatomies and adapt the learned features to complex shapes, without supervision.

Here, we use the femur and pelvis anatomies to demonstrate the feature adaptation via domain adversarial learning. Figure 9 shows sample shapes of the two anatomies that have inherently different topological features. Of the two, femur shapes can be successfully modeled using the position-based shape models [10], and hence is used as the source anatomy or domain. Pelvis shape, on the other hand, has complex small-sized shape features and hugely varying curvature. Therefore, it is challenging for the position-based models to generate a good shape-correspondence model. In this work, we showcase automated learning of geometric features for the pelvis shape. Unlike Section IV-A, the goal here is to learn these features without requiring a correspondence model for the pelvis. We first train a CNN in the Siamese framework, using the labeled samples from a femur correspondence model. The trained network is then refined/adapted using the unsupervised domain adversarial training, to adapt to the pelvis anatomy (as the target domain). As a result, the pretrained features are evolved into a combined feature space of femur and pelvis anatomies.

We use an optimized statistical shape model with 2045 correspondence points on a dataset of 8 femur shapes to train the Siamese network. The CNN configuration with 10 output features performed best for corresponding/noncorresponding classification, yielding a performance of 0.90 AUC on training samples (80% of the total samples) and 0.87 AUC on the validation set (10% of the total samples). Figure 10 showcases sample features from the network. The learned features highlight key anatomical aspects of the femur shape, similar to the coffee bean and scapula shapes in Figure 6.
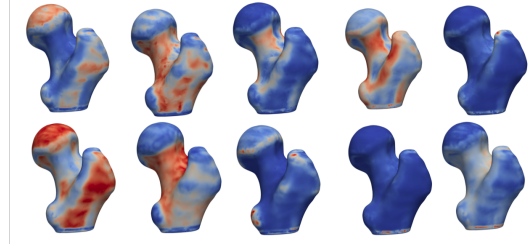

Fig. 10. Sample features for the femur shape, learned using the Siamese network.

To validate the pelvis features, we independently train a Siamese network using a quality-controlled model of 8 pelvis shapes and 2048 correspondence points. Figure 11 shows the sample features thus obtained. To highlight the need for feature adaptation, we use the CNN trained on femur shapes to extract features for the pelvis shape without adaptation; Figure 12 shows the output feature maps. In comparison, the feature maps extracted using the femur trained model (Figure 12) do not generalize for the pelvis anatomy.
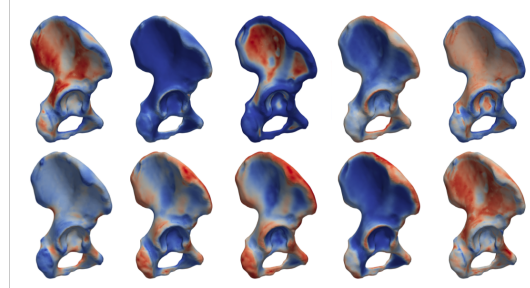

Fig. 11. Sample pelvis features obtained via supervised training using the quality-controlled pelvis correspondence model.
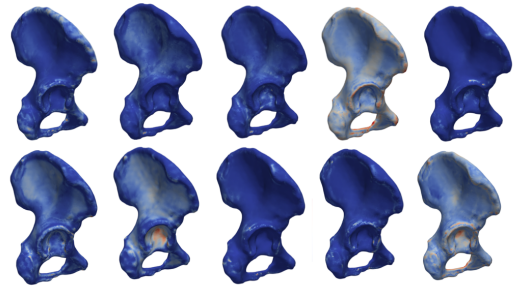

Fig. 12. Sample features extracted on the pelvis shape using the CNN trained on the femur shape.

The femur trained model is then adapted using the domain adversarial training. Figure 13 shows the adapted features extracted on the pelvis shape. As a result of the feature adaptation, the new features are able to model the high curvature regions and the flat anatomical regions in the iliac
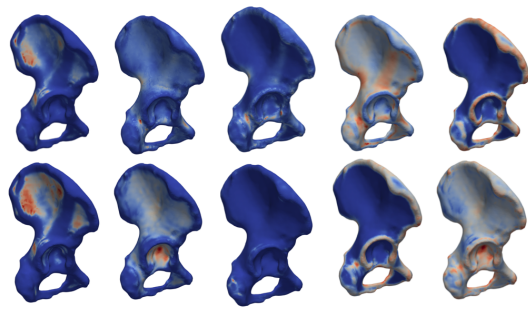
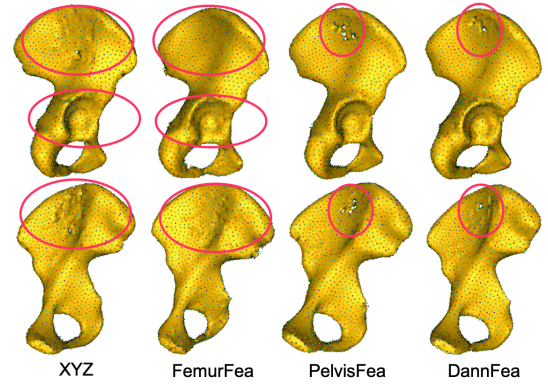Fig. 13. Sample adapted features extracted for the pelvis shape.

Fig. 14. Comparing mean shapes from positional and the three deep feature PSM models. The highlighted regions contain incorrect correspondence points.
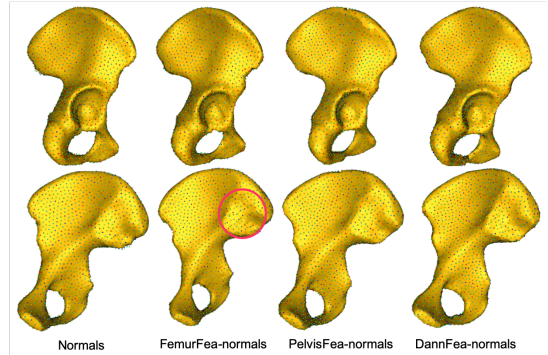

Fig. 15. Comparing mean shapes after including normals to positional and deep feature PSM models. The highlighted regions contain incorrect correspondence points.

wing of pelvis. We use the deep learning features to optimize for the shape correspondence model for a dataset of 23 pelvis shapes. The proposed deep learning features are trained using the geodesic patch, which captures the local surface geometry and is invariant to the location and orientation of the shapes. Such invariance is useful for the generalizability of trained networks onto new datasets. As discussed earlier, we use position information with the deep learning features to resolve the correspondence in similar anatomical regions across the shape. For the pelvis anatomy, orientation information plays a vital role in resolving complex regions, such as the very thin iliac wing. Therefore, we use the position and orientation information along with the deep features in PSM optimization. Specifically, we use the following combinations of position, orientation (surface normals), and deep learning features to compare the performance:

- XYZ: positional PSM [6], [34];
- Normals: position and surface normals in the PSM framework;
- FemurFea: position and femur-trained features (Figure 12) in the PSM framework;
- PelvisFea: position and pelvis-trained features (Figure 11) in the PSM framework;
- DannFea: position and domain-adapted features (Figure 13) in the PSM framework;
- FemurFea-normals: position, surface normals, and femur-trained features (Figure 12) in the PSM framework;
- PelvisFea-normals: position, surface normals, and pelvis-trained features (Figure 11) in the PSM framework; and
- DannFea-normals: position, surface normals, and domain-adapted features (Figure 13) in the PSM framework.

The pelvis is a complex shape with very thin structures, and therefore, both position and orientation provide crucial support to other geometric features. First, we showcase the mean shapes generated using models involving position and deep learning features; see Figure 14. As expected, the pelvis-trained and domain-adapted features are able to significantly reduce the number of incorrect correspondences compared to the positional PSM (XYZ) and the FemurFea models. The few unresolved correspondences in the PelvisFea and DannFea models belong to the highly thin region (up to a voxel thickness of 0.5 mm) in the iliac wing (highlighted by the red circle in Figure 14). Adding surface normals to the PSM further resolves all the incorrect correspondences; see Figure 15. Figure 16 presents the quantitative comparison of all PSM variants. Figure 16(a) compares the variance plots for the different models. PelvisFea-normals produce the most compact model, followed by the PelvisFea, DannFea-normals, normals, DannFea, XYZ, FemurFea-normals, FemurFea, respectively. This trend highlights that the adapted features provide an improvement over traditional PSM models (positional and position with surface normals), and are upper-bounded by the performance of deep features learned with supervision (PelvisFea and PelvisFea-normals). The quantitative performance plots in Figure 16(b-c) closely follow a similar trend, thus highlighting the need for domain-specific geometric features to generate improved shape correspondence models.

V. CONCLUSION

This paper proposed deep learning methods to resolve the challenges in shape correspondence estimation for complex anatomies. Supervised and unsupervised learning approaches are proposed to facilitate geometric feature learning. The unsupervised technique is especially useful when a training correspondence model is difficult to obtain for complex anatomies. Thus far, such models have been generated with the help of manual expertise, which may limit the throughput of studies with a large number of sample shapes. Quantitative
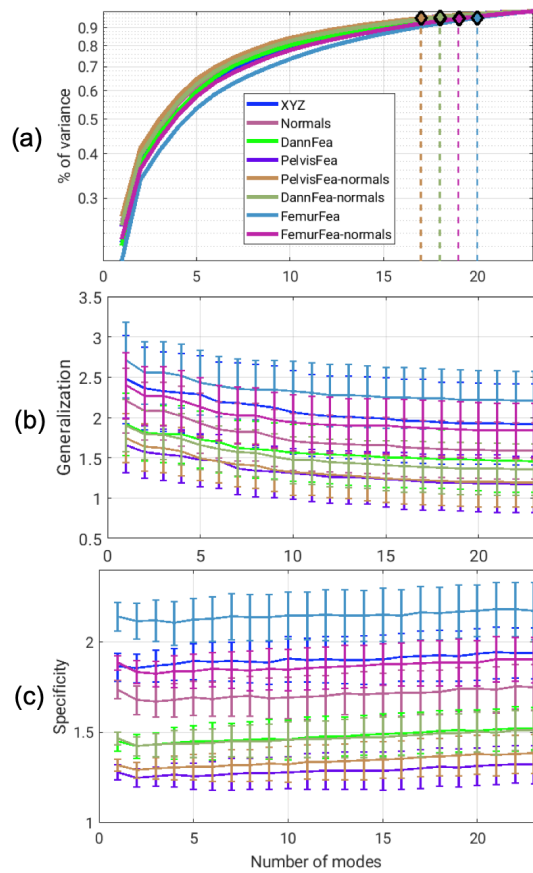
Fig. 16. Pelvis results: (a) variance plot, (b) generalization (mm), (c) specificity (mm), voxel size=0.5mm.

results obtained using synthetic and clinical datasets showcase the success of the proposed method to improve shape correspondence models. The results showed that position and orientation still play a crucial role in the correspondence optimization. Next steps should involve learning geometric features that also encapsulate this global information and thus reduce the high computation cost of Hessian, required for any normals-based PSM. Furthermore, image-based feature learning can help reduce the preprocessing overhead of shape segmentation.

REFERENCES

[1] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *MedIA*, vol. 13, no. 4, pp. 543–563, August 2009.

[2] T. L. Bredbenner, T. D. Eliason, R. S. Potter, R. L. Mason, L. M. Havill, and D. P. Nicolella, "Statistical shape modeling describes variation in tibia and femur surface geometry between control and incidence groups from the osteoarthritis initiative database," *J. Biomech.*, vol. 43, no. 9, pp. 1780–1786, June 2010.

[3] S. Balestra, S. Schumann, J. Heverhagen, L. Nolte, and G. Zheng, "Articulated statistical shape model-based 2D-3D reconstruction of a hip joint," in *IPCAI*. Springer, 2014, pp. 128–137.

[4] K.-K. Shen, J. Fripp, F. Mériaudeau, G. Chételat, O. Salvado, and P. Bourgeat, "Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models," *Neuroimage*, vol. 59, no. 3, pp. 2155–2166, February 2012.

[5] N. Sarkalkan, H. Weinans, and A. A. Zadpoor, "Statistical shape and appearance models of bones," *Bone*, vol. 60, pp. 129–140, March 2014.

[6] J. Cates, S. Elhabian, and R. Whitaker, "Shapeworks: particle-based shape correspondence and visualization software," in *Statistical Shape and Deformation Analysis*, G. Zheng, S. Li, and G. Székely, Eds. Academic Press, 2017, ch. 10, pp. 257–298.

[7] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling," *IEEE TMI*, vol. 21, no. 5, pp. 525–537, May 2002.

[8] S. Durrleman, M. Prastawa, N. Charon, J. R. Korenberg, S. Joshi, G. Gerig, and A. Trouvé, "Morphometry of anatomical shape complexes with dense deformations and sparse parameters," *NeuroImage*, vol. 101, pp. 35–49, 2014.

[9] M. Styner, I. Oguz, S. Xu, C. Brechbühler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig, "Framework for the statistical shape analysis of brain structures using SPHARM-PDM," *Insight J.*, no. 1071, pp. 242–250, 2006.

[10] M. D. Harris, M. Datar, R. T. Whitaker, E. R. Jurrus, C. L. Peters, and A. E. Anderson, "Statistical shape modeling of cam femoroacetabular impingement," *J. Orthop. Res.*, vol. 31, no. 10, pp. 1620–1626, July 2013.

[11] P. R. Atkins, S. K. Aoki, R. T. Whitaker, J. A. Weiss, C. L. Peters, and A. E. Anderson, "Does removal of subchondral cortical bone provide sufficient resection depth for treatment of cam femoroacetabular impingement?" *Clin. Orthop. Relat. Res.*, vol. 475, no. 8, pp. 1977–1986, August 2017.

[12] P. R. Atkins, S. Y. Elhabian, P. Agrawal, M. D. Harris, R. T. Whitaker, J. A. Weiss, C. L. Peters, and A. E. Anderson, "Quantitative comparison of cortical bone thickness using correspondence-based shape modeling in patients with cam femoroacetabular impingement," *J. Orthop. Res.*, vol. 35, no. 8, pp. 1743–1753, August 2017.

[13] P. R. Atkins, Y. Shin, P. Agrawal, S. Y. Elhabian, R. T. Whitaker, J. A. Weiss, S. K. Aoki, C. L. Peters, and A. E. Anderson, "Which two-dimensional radiographic measurements of cam femoroacetabular impingement best describe the three-dimensional shape of the proximal femur?" *Clin. Orthop. Relat. Res.*, vol. 477, no. 1, pp. 242–253, January 2019.

[14] R. Bhalodia, L. A. Dvoracek, A. M. Ayyash, L. Kavan, R. Whitaker, and J. A. Goldstein, "Quantifying the severity of metopic craniosynostosis: A pilot study in application of advanced machine learning in craniofacial surgery," *J. Craniofac. Surg.*, January 2020.

[15] N. Kozic, S. Weber, P. Büchler, C. Lutz, N. Reimers, M. Á. G. Ballester, and M. Reyes, "Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets," *Medical image analysis*, vol. 14, no. 3, pp. 265–275, 2010.

[16] F. Galloway, M. Kahnt, H. Ramm, P. Worsley, S. Zachow, P. Nair, and M. Taylor, "A large scale finite element study of a cementless osseointegrated tibial tray," *Journal of biomechanics*, vol. 46, no. 11, pp. 1900–1906, 2013.

[17] R. Bryan, P. B. Nair, and M. Taylor, "Use of a statistical model of the whole femur in a large scale, multi-model study of femoral neck fracture risk," *Journal of biomechanics*, vol. 42, no. 13, pp. 2171–2176, 2009.

[18] Z. Zhao, W. D. Taylor, M. Styner, D. C. Steffens, K. R. R. Krishnan, and J. R. MacFall, "Hippocampus shape analysis and late-life depression," *PLoS One*, vol. 3, no. 3, 2008.

[19] D. Wang, L. Shi, J. F. Griffith, L. Qin, D. T. Yew, and C. M. Riggs, "Comprehensive surface-based morphometry reveals the association of fracture risk and bone geometry," *Journal of Orthopaedic Research*, vol. 30, no. 8, pp. 1277–1284, 2012.

[20] A. Goparaju, I. Csecs, A. Morris, E. Kholmovski, N. Marrouche, R. Whitaker, and S. Elhabian, "On the evaluation and validation of off-the-shelf statistical shape modeling tools: a clinical application," in *ShapeMI*. Springer, 2018, pp. 14–27.

[21] I. Oguz, J. Cates, T. Fletcher, R. Whitaker, D. Cool, S. Aylward, and M. Styner, "Cortical correspondence using entropy-based particle systems and local features," in *2008 5th IEEE International Symposium on Biomedical Imaging: From nano to macro*. IEEE, 2008, pp. 1637–1640.

[22] I. Oguz, J. Cates, M. Datar, B. Paniagua, T. Fletcher, C. Vachet, M. Styner, and R. Whitaker, "Entropy-based particle correspondence for shape populations," *International journal of computer assisted radiology and surgery*, vol. 11, no. 7, pp. 1221–1232, 2016.

[23] S. K. Balci, P. Golland, M. E. Shenton, and W. M. Wells, "Free-form b-spline deformation model for groupwise registration," *Med Image Comput Comput Assist Interv*, 2007.

[24] S. Dravzic, N. Sladoje, and J. Lindblad, "Estimation of feret's diameter from pixel coverage representation of a shape," *Pattern Recognition Letters*, vol. 80, pp. 37–45, 2016.

[25] I. Oguz, M. Niethammer, J. Cates, R. Whitaker, T. Fletcher, C. Vachet, and M. Styner, "Cortical correspondence with probabilistic fiber connectivity," in *International Conference on Information Processing in Medical Imaging*. Springer, 2009, pp. 651–663.

[26] M. Datar, I. Lyu, S. Kim, J. Cates, M. A. Styner, and R. Whitaker, "Geodesic distances to landmarks for dense correspondence on ensembles of complex shapes," in *MICCAI*, 2013, pp. 19–26.

[27] P. Agrawal, R. T. Whitaker, and S. Y. Elhabian, "Learning deep features for automated placement of correspondence points on ensembles of complex shapes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 185–193.

[28] D. Boscaini, J. Masci, E. Rodolà, and M. Bronstein, "Learning shape correspondence with anisotropic convolutional neural networks," in *NIPS*, 2016, pp. 3189–3197.

[29] D. Boscaini, J. Masci, E. Rodolà, M. M. Bronstein, and D. Cremers, "Anisotropic diffusion descriptors," *Computer Graphics Forum*, vol. 35, no. 2, pp. 431–441, May 2016.

[30] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR*, vol. 1, 2005, pp. 539–546.

[31] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE TPAMI*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[32] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, pp. S151–S160, 2004.

[33] S. C. Joshi and M. I. Miller, "Landmark matching via large deformation diffeomorphisms," *IEEE transactions on image processing*, vol. 9, no. 8, pp. 1357–1370, 2000.

[34] J. Cates, P. T. Fletcher, M. Styner, M. Shenton, and R. Whitaker, "Shape modeling and analysis with entropy-based particle systems," in *IPMI*, 2007, pp. 333–345.

[35] O. Van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Computer Graphics Forum*, vol. 30, no. 6, pp. 1681–1707, July 2011.

[36] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "3D-CODED: 3D Correspondences by Deep Deformation," in *ICCV*, 2018, pp. 230–246.

[37] Y. Kleiman and M. Ovsjanikov, "Robust structure-based shape correspondence," *Computer Graphics Forum*, vol. 38, no. 1, pp. 7–20, Apr 2019.

[38] Z. Sun, Y. He, A. Gritsenko, A. Lendasse, and S. Baek, "Embedded spectral descriptors: Learning the point-wise correspondence metric via siamese neural networks," 2017.

[39] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, October 2018.

[40] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *AAAI*, 2016, pp. 2058–2065.

[41] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *CVPR*, 2013, pp. 2960–2967.

[42] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *CVPR*, 2017, pp. 3722–3731.

[43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[44] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *NIPS*, 2016, pp. 469–477.

[45] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," September 2014.

[46] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *JMLR*, vol. 17, no. 1, pp. 2096–2030, January 2016.

[47] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *CVPR*, 2017, pp. 7167–7176.

[48] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *ECCV*, 2015, pp. 4068–4076.

[49] E. Tzeng, C. Devin, J. Hoffman, C. Finn, P. Abbeel, S. Levine, K. Saenko, and T. Darrell, "Adapting deep visuomotor representations with weak pairwise constraints," November 2015.

[50] J. Cates, M. Meyer, T. Fletcher, and R. Whitaker, "Entropy-based particle systems for shape correspondence," in *MFCA*, 2006, pp. 90–99.

[51] J. Cates, P. T. Fletcher, M. Styner, H. C. Hazlett, and R. Whitaker, "Particle-based shape analysis of multi-object complexes," in *MICCAI*, 2008, pp. 477–485.

[52] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[53] S. Rusinkiewicz, "Estimating curvatures and their derivatives on triangle meshes," in *3DPVT*, 2004, pp. 486–493.

[54] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *AISTATS*, 2010, pp. 249–256.

[55] R. Davies, C. Twining, and C. Taylor, *Statistical Models of Shape: Optimisation and Evaluation*. London, UK: Springer Science & Business Media, 2008.