Geo-SIC: Learning Deformable Geometric Shapes in Deep Image Classifiers

Anonymous Author(s)
Affiliation
Address
email

Abstract

Deformable shapes provide important and complex geometric features of objects presented in images. However, such information is oftentimes missing or under-utilized as implicit knowledge in many image analysis tasks. This paper presents Geo-SIC, the first deep learning model to learn deformable shapes in a deformation space for an improved performance of image classification. We introduce a newly designed framework that (i) simultaneously derives features from both image and latent shape spaces with large intra-class variations; and (ii) gains increased model interpretability by allowing direct access to the underlying geometric features of image data. In particular, we develop a boosted classification network, equipped with an unsupervised learning of geometric shape representations characterized by diffeomorphic transformations within each class. In contrast to previous approaches using pre-extracted shapes, our model provides a more fundamental approach by naturally learning the most relevant shape features jointly with an image classifier. We demonstrate the effectiveness of our method on both simulated 2D images and real 3D brain magnetic resonance (MR) images. Experimental results show that our model substantially improves the image classification accuracy with an additional benefit of increased model interpretability. All code and data produced in this research will be made publicly available online.

1 Introduction

Deformable shapes have been identified to aid image classification for decades, as they capture geometric features that describe changes and variability of objects with complex structures from images (5; 18; 26). Bountiful literature demonstrates that the robustness of shapes to variations in image intensity and texture (e.g., noisy or corrupted data) makes it a reliable cue for image analysis tasks (29; 26; 41). For example, abnormal shape changes of anatomical structures are strong predictors of serious diseases and health problems, e.g., brain shrinkage caused by neurodegenerative disorders (16; 47), irregular heart motions caused by cardiac arrhythmia (3; 8), and pulmonary edema caused by infectious lung diseases (40; 42). Existing methods have studied various representations of geometric shapes, including landmarks (7; 9; 14), point clouds (1), binary segmentations (10; 36), and medial axes (30). A very recent research area in geometric deep learning (11; 31) has investigated mathematical representations of shapes in the form of analytic graphs or points and then uses them to synthesize shapes. These aforementioned techniques often ignore objects' interior structures; hence do not capture the intricacies of complex objects in images. In contrast, deformation-based shape representations (based on elastic deformations or fluid flows) focus on highly detailed shape information from images (13; 33). With the underlying assumption that objects in many generic classes can be described as deformed versions of an ideal template, descriptors in this class arise naturally by matching the template to an input image. This procedure is also known as *atlas building* (22; 38; 46). The resulting transformation is then considered as a shape that reflects

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

geometric changes. In this paper, we will feature deformation-based shape representations that offer more flexibility in describing shape changes and variability of complex structures. However, our developed framework can be easily adapted to other types of representations, including those characterized by landmarks, binary segmentations, curves, and surfaces.

Inspired by the advantages of incorporating shape information in image analysis tasks, current deep learning-based classification networks have been mostly successful in using pre-extracted shapes from images (43; 27; 7). However, these methods require preprocessed shape data and oftentimes achieve a suboptimal solution in identifying shape features that are most representative to differentiate different classes of images. An explicit learning of deformable shapes in deep image classifiers has been missing. This limits the power of classification models where quantifying and analyzing geometric shapes is critical.

In this paper, we introduce a novel deep learning image classification model, named as Geo-SIC, that jointly learns deformable shapes in a multi-template deformation space. More specifically, Geo-SIC provides an unsupervised learning of deformation-based shape representations via a newly designed sub-network of atlas building. Different from previous deep learning based atlas building approaches (15; 20), we employ an efficient parameterization of deformations in a low-dimensional Fourier space (44) to speed up the training inference. The major contribution of Geo-SIC is three folds:

(i) In contrast to previous approaches treating shapes as preprocessed objects from images for classification tasks, Geo-SIC provides a more fundamental approach by merging shape features naturally in the learning process of classification. To the best of our knowledge, we are the first to unify these related tasks in a joint learning framework.

(ii) Geo-SIC performs a simultaneous feature extraction from both image and learned shape spaces. With these integrated features, Geo-SIC achieves an improved accuracy of image classification. An additional benefit of Geo-SIC has an increased model interpretability because of the direct access to underlying geometric features of image data.

(iii) Geo-SIC provides an efficient sub-network of atlas building in a low dimensional shape space. This reduces the computational complexity of model training, especially for high-dimensional image data (i.e., 3D brain MR images of large size).

We demonstrate the effectiveness of Geo-SIC on both synthetic 2D images and real 3D brain MR images. Experimental results show that our model substantially improves the classification accuracy compared to a wide variety of models without jointly learned geometric features. We then visualize the class activation maps by using gradient-weighted class activation mapping (Grad-CAM) (34). The highlighted regions show that Geo-SIC attracts more attention to geometric shape features that positively contribute to the accuracy of classifiers.

2 Background: Deformation-based Shape Representations via Atlas Building

This section briefly reviews the concept of atlas building (22), which is commonly used to derive deformation-based shape representations from images. With the underlying assumption that the geometric information in the deformations conveys a shape, descriptors in this class arise naturally by matching a template to an input image with smoothness constraints on the deformation field.

Given a number of N images $\{I_1, \cdots, I_N\}$, the problem of atlas building is to find a mean or template image I and deformation fields $\phi_1, \cdots, \phi_N$ that minimize the energy function

$$E(I, \phi_n) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \text{Dist}[I \circ \phi_n^{-1}, I_n] + \text{Reg}(\phi_n), \tag{1}$$

where σ^2 is a noise variance and $\circ$ denotes an interpolation operator that deforms image I with an estimated transformation ϕ_n. The $\text{Dist}(\cdot, \cdot)$ is a distance function that measures the dissimilarity between images, i.e., sum-of-squared differences (6), normalized cross correlation (4), and mutual information (39). The $\text{Reg}(\cdot)$ is a regularization that guarantees the smoothness of transformations.

2.1 Shape Representations In The Tangent Space Of Diffeomorphisms

In many applications, it is natural to require the deformation field to be a diffeomorphism, i.e., a differentiable, bijective mapping with a differentiable inverse. Shape representations in the space of diffeomorphic transformations highlight a set of desirable features: (i) they capture large geometric variations within image groups; (ii) the topology of objects in the image remains intact; and (iii) no non-differentiable artifacts, such as creases or sharp corners, are introduced. Moreover, a theoretical framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) defines a metric in the space of diffeomorphic transformations that in turn induces a distance metric on the shape space (6).

Given an open and bounded d-dimensional domain $\Omega \subset \mathbb{R}^d$, we use $\mathrm{Diff}(\Omega)$ to denote a space of diffeomorphisms (i.e., a one-to-one smooth and invertible smooth transformation) and its tangent space $V = T\mathrm{Diff}(\Omega)$. The LDDMM algorithm (6) provides a distance metric in the deformation-based shape space, which is used as a regularization of atlas building in Eq. 1. Such a distance metric is formulated as an integral of the Sobolev norm of the time-dependent velocity field $v_n(t) \in V(t \in [0, 1])$ in the tangent space, i.e.,

$$\mathrm{Reg}(\phi_n) = \int_0^1 (\mathcal{L}v_n(t), v_n(t))\, dt, \quad \text{with} \quad \frac{d\phi_n^{-1}(t)}{dt} = -D\phi_n^{-1}(t) \cdot v_n(t), \tag{2}$$

where $\mathcal{L} : V \to V^*$ is a symmetric, positive-definite differential operator that maps a tangent vector $v(t) \in V$ into its dual space as a momentum vector $m(t) \in V^*$. We typically write $m(t) = \mathcal{L}v(t)$, or $v(t) = \mathcal{K}m(t)$, with $\mathcal{K}$ being an inverse operator of $\mathcal{L}$. In this paper, we adopt a commonly used Laplacian operator $\mathcal{L} = (-\alpha\Delta + \mathrm{Id})^3$, where α is a weighting parameter that controls the smoothness of transformation fields and Id is an identity matrix. The $(\cdot, \cdot)$ is a dual paring, which is similar to an inner product between vectors. The operator D denotes a Jacobian matrix and $\cdot$ represents an element-wise matrix multiplication.

According to a well known geodesic shooting algorithm (37), the minimum of Eq. (2) is uniquely determined by solving a Euler-Poincaré differential equation (EPDiff) (2; 25) with a given initial condition. This nicely proves that the deformation-based shape descriptor ϕ_n can be fully characterized by an initial velocity field $v_n(0)$, which lies in the tangent space of diffeomorphisms. A recent research further identifies an equivalent but more efficient way to reparameterize these initial velocity fields in a low-dimensional bandlimited space (44; 45).

Let $\widetilde{\mathrm{Diff}}(\Omega)$ and $\tilde{V}$ denote the bandlimited space of diffeomorphisms and velocity fields respectively. The EPDiff is reformulated in a complex-valued Fourier space with much less dimensions, i.e.,

$$\frac{\partial \tilde{v}_n(t)}{\partial t} = -\tilde{\mathcal{K}} \left[(\tilde{\mathcal{D}}\tilde{v}_n(t))^T \star \tilde{\mathcal{L}}\tilde{v}_n(t) + \tilde{\nabla} \cdot (\tilde{\mathcal{L}}\tilde{v}_n(t) \otimes \tilde{v}_n(t)) \right], \tag{3}$$

where $\star$ is the truncated matrix-vector field auto-correlation. Here $\tilde{\mathcal{K}}$ is an inverse operator of $\tilde{\mathcal{L}}$, which is the Fourier transform of a Laplacian operator $\mathcal{L}$. The operator $\tilde{\mathcal{D}}$ represents the Fourier frequencies of a Jacobian matrix with central difference approximation, and $*$ is a circular convolution with zero padding to avoid aliasing [1]. The operator $\tilde{\nabla}\cdot$ is the discrete divergence of a vector field, and $\otimes$ represents a tensor product between Fourier frequencies.

We are now ready to optimize the problem of atlas building (Eq. (1)) with reduced computational complexity in a low-dimensional bandlimited space as

$$E(I, \phi_n) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \mathrm{Dist}[I \circ \phi_n^{-1}, I_n] + (\mathcal{L}\tilde{v}_n(0), \tilde{v}_n(0)), \quad \text{s.t. Eq. (2)\& (3).} \tag{4}$$

The deformation ϕ_n^{-1} corresponds to $\tilde{\phi}_n^{-1}$ in Fourier space via the Fourier transform $\mathcal{F}(\phi_n^{-1}) = \tilde{\phi}_n^{-1}$, or its inverse $\phi_n^{-1} = \mathcal{F}^{-1}(\tilde{\phi}_n^{-1})$. Note that we will drop the time index t, i.e., $\tilde{v}_n(0) \triangleq \tilde{v}_n$, for simplified notations in next sections.

[1] To prevent the domain from growing infinity, we truncate the output of the convolution in each dimension to a suitable finite set.

3 Our Method: Geo-SIC

In this section, we present a novel deep image classifier (Geo-SIC) that explicitly learns geometric shape representations for an improved performance of accuracy, as well as increased model interpretability. Geo-SIC consists of two modules: an unsupervised learning of geometric shapes via an atlas building network, and a boosted classification network that integrates features from both image and learned shape spaces. Details of our network architecture are introduced as follows.

Geometric shape learning based on an atlas building network. Let (θ_g^E, θ_g^D) be the parameters of an encoder-decoder in our geometric learning network. Consider a number of J image classes, there exists a number of $N_j, j \in \{1, \ldots, J\}$ images in each class. Our atlas building network will learn the shape representations, also known as initial velocity fields $\tilde{v}_n(\theta_g^E, \theta_g^D), n \in \{1, \cdots, N_j\}$, with an updated atlas I_j. We adopt the architecture of UNet (32) in this work, however, other network structures such as UNet++ (48) and TransUNet (12) can be easily applied.

Boosted image classification network. Let θ_c^E be the parameters of an encoder that extracts features from image spaces. We develop a feature fusion module that integrates geometric shape and image features in a latent space parameterized by $\theta_{gc}(\theta_g^E, \theta_c^E)$. This boosted classification network will predict a class label $y_{nj}(\theta_{gc})$ for each input image.

Network loss. The loss function of Geo-SIC includes the loss from both geometric learning and classification network. Given a set of image class labels $\hat{y}_{nj}$, we define $\Theta = (\theta_{\mathrm{g}}^{\mathrm{E}}, \theta_{\mathrm{g}}^{\mathrm{D}}, \theta_{\mathrm{c}}^{\mathrm{E}}, \theta_{\mathrm{gc}})$ for all network parameters and formulate the total loss of Geo-SIC as

$$l(\Theta) = \sum_{n=1}^{N_j} \sum_{j=1}^{J} [\frac{1}{\sigma_j^2} \| I_j \circ \phi_{nj}^{-1} \left(\tilde{v}_{nj}(\theta_g^E, \theta_g^D) \right) - I_{nj} \|^2 + (\tilde{\mathcal{L}}_j \tilde{v}_{nj}(\theta_g^E, \theta_g^D), \tilde{v}_{nj}(\theta_g^E, \theta_g^D)) \\ - \lambda \hat{y}_{nj} \cdot \log y_{nj}(\theta_{gc})] + \mathrm{reg}(\Theta), \quad \text{s.t.} \quad \text{Eq. (2)\& (3)}, \tag{5}$$

where $\mathrm{reg}(\cdot)$ is a regularity term on the network parameters.

An overview of our proposed Geo-SIC network is shown in Fig. 1.

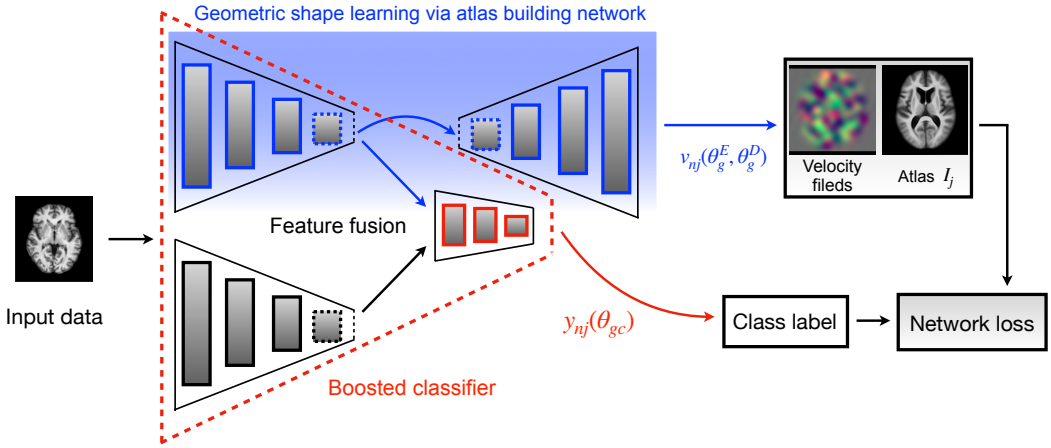

Figure 1: An overview of our proposed Geo-SIC network.

3.1 Alternating Optimization For Network Training

We develop an alternating optimization scheme (28) to minimize the network loss defined in Eq. (5). All parameters are optimized jointly by alternating between the training of geometric shape learning and image classification networks.

Training of geometric shape learning via atlas building. In contrast to current approaches that parameterize deformation-based shapes in a high-dimensional image space (20; 15), our model employs an efficient reparameterization in a low-dimensional bandlimited space (44). This makes the computation of geodesic constraints (Eq. (2) and Eq. (3)) required in the loss function substantially faster in each forward/backward propagation during the training process. The loss of training the sub-module of atlas building network is

$$l_{\text{Geo}}(\theta_g^E, \theta_g^D, I_j) = \sum_{n=1}^{N_j} \sum_{j=1}^{J} [\frac{1}{\sigma_j^2} \| I_j \circ \phi_{nj}^{-1} \left(\tilde{v}_{nj}(\theta_g^E, \theta_g^D) \right) - I_{nj} \|^2 + (\tilde{\mathcal{L}}_j \tilde{v}_{nj}(\theta_g^E, \theta_g^D), \tilde{v}_{nj}(\theta_g^E, \theta_g^D)) + \text{reg}(\theta_g^E, \theta_g^D), \quad s.t. \quad \text{Eq. (2)\& (3)}, \tag{6}$$

Similar to (20; 15), we treat atlas I_j as a network parameter and update it accordingly. To guarantee the network optimization stays in the tangent space of diffeomorphisms, we pull back the network gradient with regard to initial velocity fields by backward integrating adjoint jacobi fields (44) each time after the forward pass. More details are included in the supplementary materials.

Training of boosted image classifier. As highlighted in red box in Fig. 1, Geo-SIC extracts both image and geometric features and then integrate them into a feature fusion block. This boosted classifier is optimized over the loss defined as

$$l_{\text{SIC}}(\theta_{gc}) = -\lambda \sum_{n=1}^{N_j} \sum_{j=1}^{J} \hat{y}_{nj} \cdot \log y_{nj} \left(\theta_{gc}(\theta_g^E, \theta_c^E) \right) + \text{reg} \left(\theta_{gc}(\theta_g^E, \theta_c^E) \right). \tag{7}$$

A summary of our joint learning of Geo-SIC through an alternating optimization is in Alg. 1.

Algorithm 1: Joint learning of Geo-SIC.

Input : A set of images $\{I_{nj}\}$ with class labels $\{\hat{y}_{nj}\}$ and a number of iterations r.
Output : Predicted class label, atlas, and initial velocity fields.

```
for i = 1 to r do
    /* Train geometric learning network                                  */
    Minimize the atlas building loss in Eq. (6);
    Output the atlases {I_j} and predicted initial velocity fields {v_nj} for all image classes;
    /* Train the boosted image classifier                                */
    With learned shape features, Minimizing the boosted classification loss in Eq. (7);
    Output the predicted class labels {ŷ_nj};
end
Until convergence
```

4 Experimental Evaluation

4.1 Data

2D data set. We choose 50000 images (including five classes, circle, cloud, envelope, square, and triangle are shown in Fig 3) of Google Quickdraw dataset (21), a collection of categorized drawings contributed by online players in a drawing game. We run affine transformation within each class as prepossessing and upsample each image from 28×28 to 100×100.

3D brain MRI. For brain data, we include 373 public T1-weighted brain MRI scans from Open Access Series of Imaging Studies (OASIS) (17). All 150 subjects are aged from 60 to 96 with Alzheimer's disease diagnosis (79 cases for dementia and 71 cases for non-demented). All MRIs were all pre-processed as $256 \times 256 \times 256$, $1.25mm^3$ isotropic voxels, and underwent skull-stripped, intensity normalized, bias field corrected, and pre-aligned with affine transformation.

4.2 Experiments

Classification evaluation. We demonstrate the effectiveness of our model on both 2D synthetic data and 3D brain MRI scans. We select four classification backbones (AlexNet (23), a five-layer CNN, ResNet18 (19), and VGG19 (35)) as baseline methods. For a fair comparison, we show the results of Geo-SIC by replacing the backbone in our model with all baselines (named as Geo-SIC:Alex, Geo-SIC:CNN, Geo-SIC:Res, and Geo-SIC:VGG).

To further investigate the advantages of our joint learning, we compare with two-step approaches (Two-step Alex, Two-step CNN, Two-step Res, and Two-step VGG), where the geometric learning network is treated as a preprocessing step for geometric feature extraction. We report the average accuracy, F1-score, AUC, sensitivity, and specificity for all methods. We also show the receiver operating characteristic (ROC) curves by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

To better understand the model interpretability in terms of network attention of all models, we visualize the Grad-CAM (34) of the last neural network layer.

Atlas evaluation. We also evaluate the performance of our newly designed atlas building network, which serves as a core part of Geo-SIC. We compare the estimated atlas of Geo-SIC that achieves the highest AUC with three atlas-building methods: a diffeomorphic autoencoder (20) (LagoMorph), a deep learning based conditional template estimation method (15) (Con-Temp) and a Bayesian atlas building framework with hyper priors (38) (Hier-Baye).

We report the total computational time and memory consumption on 3D real brain images. To measure the sharpness of estimated atlas I, we adopt a metric of normalized standard deviation computed from randomly selected 4000 image patches (24). Given $M(i)$, a patch around a voxel i of an atlas I, the local measure of the sharpness at voxel i is defined as $\text{sharpness}(I(i)) = \text{sd}_{M(i)}(I)/\text{avg}_{M(i)}(I)$, where sd and avg denote the standard deviation and the mean of M_i.

Parameter Setting. We set an optimal dimension of the low-dimensional shape representation as 16^2 for 2D dataset and 32^3 for 3D dataset. We set parameter $\alpha = 3$ for the operator $\tilde{\mathcal{L}}$, the number of time steps for Euler integration in EPDiff (Eq. (3)) as 10. We set the noise variance $\sigma = 0.02$. We set the batch size as 16 and use the cosine annealing learning rate schedule that starts from a learning rate $\eta = 1e-3$ for network training. We run 1000 epochs with the Adam optimizer and save networks with the best validation performance for all models. All networks are trained with an i7, 9700K CPU with 32 GB internal memory. The training and prediction procedure of all learning-based methods are performed on four Nvidia GTX 2080Ti GPUs. For both 2D and 3D datasets, we split the images by using 70% as training images, 15% as validation images, and 15% as testing images.

4.3 Results

4.3.1 Classification on 2D synthetic images

Table. 4.3.1 reports the classification performance across all methods over model accuracy and five micro-averaged evaluation metrics. Our method achieves the highest model accuracy with comparable micro-averaged AUC, F-1 score, precision, sensitivity, and specificity. Compare to all baselines including two-step approaches, Geo-SIC achieves the best classification performance.

Fig. 2 visualizes the micro-averages ROC curve for baselines and our proposed Geo-SIC. For four sets of comparisons, all curves produced by our classifiers are with a larger AUC and are closer to the top-left corner, which indicates a better classification performance.

Fig. 3 (left) displays the heat maps overlaid with 2D QuickDraw data for all methods. Geo-SIC produces more explainable heat maps that are geometrically aligned with the original shapes. It indicates that the latent shape space attract more attention to geometric features that positively contribute to the performance of classifiers. Fig. 3 (right) visualizes a comparison of atlases generated by baselines and Geo-SIC. It shows that Geo-SIC produces image atlases with the best visual quality, e.g., clearer circle/cloud edges, sharper triangle/square corners.

Table 1: Classification performance comparison on 2D synthetic data over six metrics (micro-average).

Models	Accuracy	AUC	F-1 score	Precision	Sensitivity	Specificity
AlexNet	0.880	0.931	0.876	0.914	0.880	0.920
Two-step AlexNet	0.891	0.955	0.891	0.906	0.891	0.905
Geo-SIC: Alex	**0.928**	**0.977**	**0.927**	**0.950**	**0.928**	**0.952**
CNN	0.861	0.928	0.857	0.874	0.861	0.881
Two-step CNN	0.869	0.931	0.866	**0.914**	0.869	**0.918**
Geo-SIC: CNN	**0.897**	**0.954**	**0.898**	0.911	**0.897**	0.910
ResNet18	0.875	0.941	0.877	0.885	0.875	0.883
Two-step Res	0.911	0.972	0.912	0.935	0.911	0.883
Geo-SIC: Res	**0.935**	**0.983**	**0.928**	**0.948**	**0.935**	**0.956**
VGG19	0.883	0.946	0.882	0.904	0.883	0.905
Two-step VGG	0.895	0.951	0.888	0.909	0.895	0.919
Geo-SIC: VGG	**0.927**	**0.980**	**0.924**	**0.957**	**0.927**	**0.960**

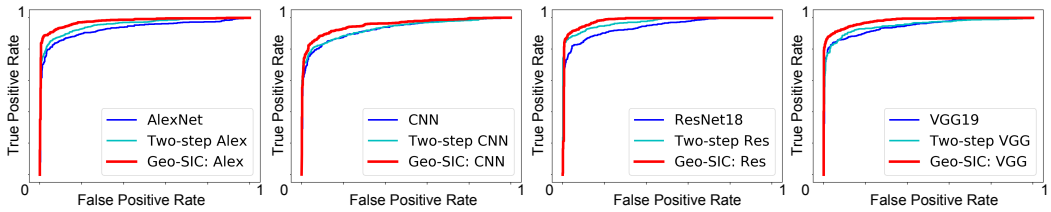

Figure 2: ROC curves of multi-class classification comparison between baselines and the proposed method.

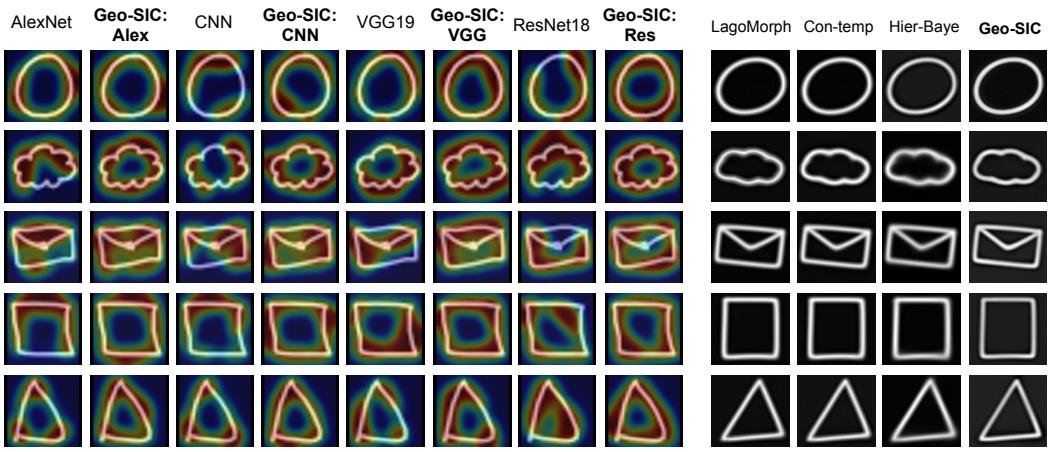

Figure 3: Left: visualization of Grad-CAMs on mutil-atlas building task of five different geometric shapes.; Right: atlas comparison between state-of-the-arts and Geo-SIC.

4.3.2 Classification on 3D real brain MRIs

Table. 4.3.2 reports the model performance for brain images. Our method Geo-SIC outperforms all baselines with comparable accuracy, AUC, F1-score, precision, sensitivity, and specificity. It shows that Geo-SIC substantially improves the classification performance with reduced misclassification rate.

Fig. 4 visualizes the ROC curves for all algorithms. For four sets of comparisons, our boosted classifiers offer curves closer to the top-left corner with higher AUC values. It shows that Geo-SIC has better performance in distinguishing between healthy control and Alzheimer's diseases groups.

Table 2: Classification performance comparison on 3D brain images over six metrics.

Models	Accuracy	AUC	F-1 score	Precision	Sensitivity	Specificity
AlexNet	0.791	0.861	0.796	0.787	0.806	0.775
Two-step AlexNet	0.800	0.887	0.809	0.822	0.796	0.804
Geo-SIC: Alex	**0.835**	**0.917**	**0.827**	**0.823**	**0.831**	**0.839**
CNN	0.779	0.860	0.773	0.804	0.744	0.814
Two-step CNN	0.788	0.883	0.791	0.794	0.787	0.789
Geo-SIC: CNN	**0.824**	**0.902**	**0.828**	**0.833**	**0.822**	**0.825**
ResNet18	0.805	0.887	0.811	0.813	0.809	0.800
Two-step Res	0.833	0.917	0.836	0.804	0.870	0.797
Geo-SIC: Res	**0.873**	**0.933**	**0.874**	**0.874**	**0.874**	**0.872**
VGG19	0.805	0.871	0.813	0.807	0.818	0.790
Two-step VGG	**0.865**	0.880	0.838	**0.822**	0.854	**0.826**
Geo-SIC: VGG	0.852	**0.930**	**0.850**	0.820	**0.881**	0.825

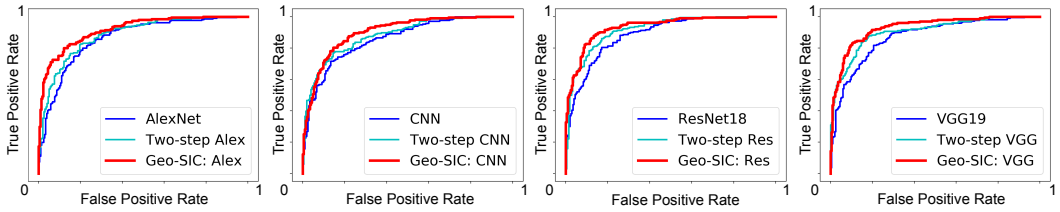

Figure 4: ROC curves of binary classification comparison for baselines, two-step approaches, and Geo-SIC.

Fig. 5 displays the heat maps (produced by Grad-CAM) overlaid with brain MRIs. Our generated heat maps are fairly aligned with brain structures (e.g., ventricle), which are critical to Alzheimer's diseases diagnosis. It profiles the fact that the latent geometric features guide the neural networks to capture the most anatomically meaningful brain regions for distinguishing healthy vs. disease groups. For both dementia and non-dementia cases, the heat maps of Geo-SIC are more explainable. This shows that our model has great clinical potential for better analysis of Alzheimer's diseases.

Fig. 6 (top) visualizes a comparison of the atlas on real brain MRI scans. Our model significantly improves the quality of the atlas with sharper details. More specifically, Geo-SIC offers better brain atlas with clearer anatomical structures, e.g., ventricle, grey and white matter. Fig. 6 (bottom left) quantitatively reports the sharpness metric of all methods. Fig. 6 (bottom right) shows the comparison of computational time and memory consumption across all methods. Although Con-Temp is slightly faster than Geo-SCI due to a very different parameterization of velocity fields (stationary velocity rather than the time-dependent velocity in other methods), it achieves a less sharper atlas and still requires larger memory consumption than our model. Comparing with the other methods (Lagomorph and Hier-Baye) that employs time-dependent velocity fields, Geo-SIC substantially reduces the time consumption and memory consumption.

5 Conclusions

This paper presents a novel deep learning model, named as Geo-SIC, that incorporates deformable geometric shape learning into image classification tasks. We jointly learn a boosted classifier with an unsupervised shape learning network based on atlas building. We defined a joint loss function for network training with an efficient alternating optimization scheme. Our experimental results on both 2D synthetic and 3D brain MRI scans demonstrate the effectiveness of Geo-SIC. Our model promises the improvement of classification performance for different backbones while producing sharper atlas with better image quality. Compared with the state-of-the-arts, our model is more explainable towards interpreting the attention of classification neural networks focusing on geometric features. The theoretical tools developed in this paper are generic to a variety of combinations of shape representations and classification backbones. Geo-SIC not only has a great potential to impact clinically diagnostic routine, such as Alzheimer's disease detection, or post-treatment for

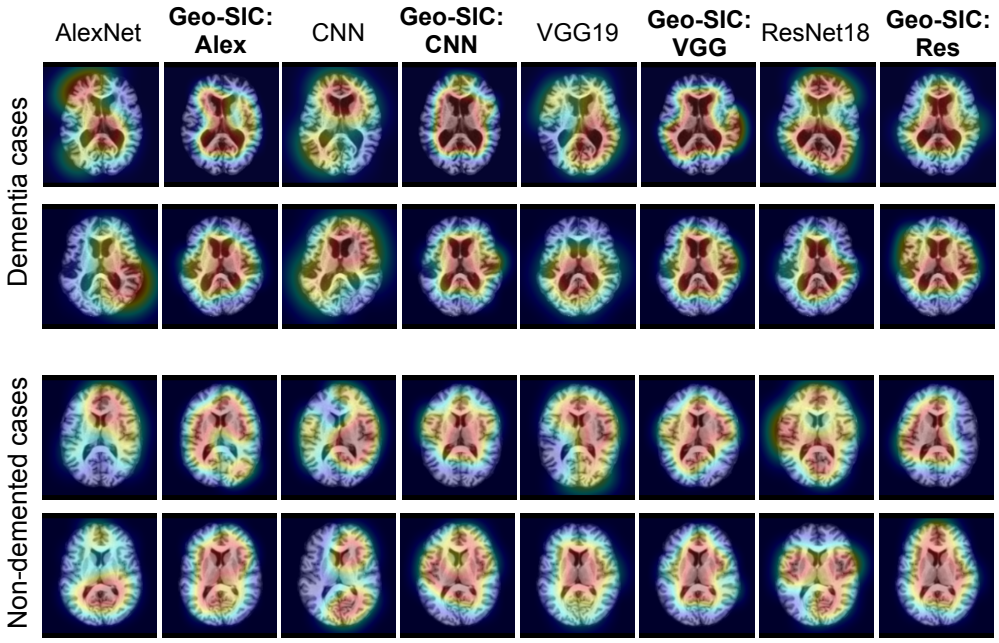

Figure 5: Visualization of Grad-CAMs on single-atlas building of dementia and non-dementia. Left to right: Grad-CAM heatmaps generated by AlexNet, Geo-SIC:Alex, CNN, Geo-SIC:CNN, ResNet18, Geo-SIC:Res, VGG19 and Geo-SIC:VGG.

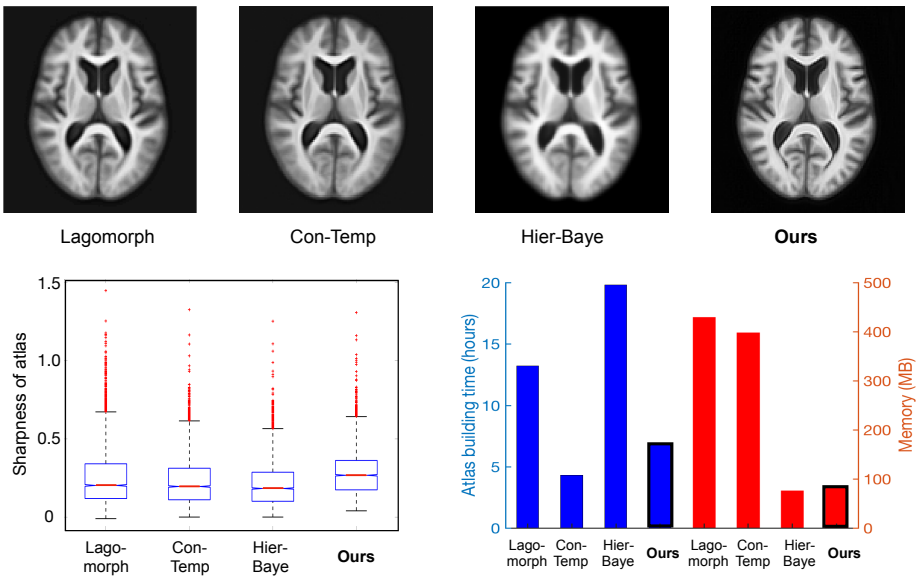

Figure 6: Top: atlas comparison between state-of-the-arts and Geo-SIC. Bottom left: sharpness metric evaluation of atlas (the higher the better).The mean values of the sharpness metric of three baselines and **Geo-SIC** are, 0.259, 0.235, 0.218, and **0.2867**; bottom left: average time and memory consumption comparison for atlas building.

patient care, but also bridges the gap between the developed deformable shape learning theories and classification-based applications. A promising future work is to incorporate images with missing data values that are caused by corrupted signals, or appearance changes such as brain tumor resection, into our framework.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] .
   (b) Did you describe the limitations of your work? [No]
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

References

[1] Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L.: Learning representations and generative models for 3d point clouds. In: International conference on machine learning. pp. 40–49 (2018)

[2] Arnol'd, V.I.: Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier **16**, 319–361 (1966)

[3] Auger, D.A., Bilchick, K.C., Gonzalez, J.A., Cui, S.X., Holmes, J.W., Kramer, C.M., Salerno, M., Epstein, F.H.: Imaging left-ventricular mechanical activation in heart failure patients using cine dense mri: Validation and implications for cardiac resynchronization therapy. Journal of Magnetic Resonance Imaging **46**(3), 887–896 (2017)

[4] Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)

[5] Baker, N., Lu, H., Erlikhman, G., Kellman, P.J.: Deep convolutional networks do not classify based on global object shape. PLoS computational biology **14**(12), e1006613 (2018)

[6] Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision **61**(2), 139–157 (2005)

[7] Bhalodia, R., Elhabian, S.Y., Kavan, L., Whitaker, R.T.: Deepssm: a deep learning framework for statistical shape modeling from raw images. In: International Workshop on Shape in Medical Imaging. pp. 244–257. Springer (2018)

[8] Bilchick, K.C., Auger, D.A., Abdishektaei, M., Mathew, R., Sohn, M.W., Cai, X., Sun, C., Narayan, A., Malhotra, R., Darby, A., et al.: Cmr dense and the seattle heart failure model inform survival and arrhythmia risk after crt. JACC: Cardiovascular Imaging **13**(4), 924–936 (2020)

[9] Bookstein, F.L.: Morphometric tools for landmark data: geometry and biology. Cambridge University Press (1997)

[10] Brechbühler, C., Gerig, G., Kübler, O.: Parametrization of closed surfaces for 3-d shape description. Computer vision and image understanding **61**(2), 154–170 (1995)

[11] Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine **34**(4), 18–42 (2017)

[12] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)

[13] Christensen, G.E., Rabbitt, R.D., Miller, M.I.: A deformable neuroanatomy textbook based on viscous fluid mechanics. In: 27th Ann. Conf. on Inf. Sciences and Systems. pp. 211–216 (1993)

[14] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer vision and image understanding **61**(1), 38–59 (1995)

[15] Dalca, A., Rakic, M., Guttag, J., Sabuncu, M.: Learning conditional deformable templates with convolutional networks. Advances in neural information processing systems **32** (2019)

[16] Ding, Z., Fleishman, G., Yang, X., Thompson, P., Kwitt, R., Niethammer, M., Initiative, A.D.N., et al.: Fast predictive simple geodesic regression. Medical image analysis **56**, 193–209 (2019)

[17] Fotenos, A.F., Snyder, A., Girton, L., Morris, J., Buckner, R.: Normative estimates of cross-sectional and longitudinal brain volume decline in aging and ad. Neurology **64**(6), 1032–1039 (2005)

[18] Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R.: Deformation analysis for shape based classification. In: Biennial International Conference on Information Processing in Medical Imaging. pp. 517–530. Springer (2001)

[19] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[20] Hinkle, J., Womble, D., Yoon, H.J.: Diffeomorphic autoencoders for lddmm atlas building (2018)

[21] Jongejan, J., Rowley, H., Kawashima, T., Kim, J., Fox-Gieg, N.: The quick, draw!-ai experiment. Mount View, CA, accessed Feb **17**(2018), 4 (2016)

[22] Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23, Supplement1**, 151–160 (2004)

[23] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[24] Legouhy, A., Commowick, O., Rousseau, F., Barillot, C.: Online atlasing using an iterative centroid. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 366–374. Springer (2019)

[25] Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of mathematical imaging and vision **24**(2), 209–228 (2006)

[26] Mohanty, N., Rath, T.M., Lee, A., Manmatha, R.: Learning shapes for image classification and retrieval. In: International Conference on Image and Video Retrieval. pp. 589–598. Springer (2005)

[27] Muneer, A., Fati, S.M.: Efficient and automated herbs classification approach based on shape and texture features using deep learning. IEEE Access **8**, 196747–196764 (2020)

[28] Nocedal, J., Wright, S.J.: Numerical optimization. Springer (1999)

[29] Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision. pp. 1520–1528 (2015)

[30] Pizer, S.M., Fritsch, D.S., Yushkevich, P.A., Johnson, V.E., Chaney, E.L.: Segmentation, registration, and measurement of shape variation via image object shape. IEEE transactions on medical imaging **18**(10), 851–865 (1999)

[31] Rematas, K., Martin-Brualla, R., Ferrari, V.: Sharf: Shape-conditioned radiance fields from a single view. arXiv preprint arXiv:2102.08860 (2021)

[32] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

[33] Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. IEEE transactions on medical imaging **22**(8), 1014–1025 (2003)

[34] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)

[35] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[36] Staib, L.H., Duncan, J.S.: Boundary finding with parametrically deformable models. IEEE transactions on pattern analysis and machine intelligence **14**(11), 1061–1075 (1992)

[37] Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision **97**(2), 229–241 (2012)

[38] Wang, J., Zhang, M.: Bayesian atlas building with hierarchical priors for subject-specific regularization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 76–86. Springer (2021)

[39] Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical image analysis **1**(1), 35–51 (1996)

[40] Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., et al.: Pathological findings of covid-19 associated with acute respiratory distress syndrome. The Lancet respiratory medicine **8**(4), 420–422 (2020)

[41] Yang, J., Duncan, J.S.: 3d image segmentation of deformable objects with shape-appearance joint prior models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 573–580. Springer (2003)

[42] Yoon, S.H., Lee, K.H., Kim, J.Y., Lee, Y.K., Ko, H., Kim, K.H., Park, C.M., Kim, Y.H.: Chest radiographic and ct findings of the 2019 novel coronavirus disease (covid-19): analysis of nine patients treated in korea. Korean journal of radiology **21**(4), 494–500 (2020)

[43] Zhang, C., Zheng, Y., Guo, B., Li, C., Liao, N.: Scn: a novel shape classification algorithm based on convolutional neural network. Symmetry **13**(3), 499 (2021)

[44] Zhang, M., Fletcher, P.T.: Fast diffeomorphic image registration via fourier-approximated lie algebras. International Journal of Computer Vision **127**(1), 61–73 (2019)

[45] Zhang, M., Liao, R., Dalca, A.V., Turk, E.A., Luo, J., Grant, P.E., Golland, P.: Frequency diffeomorphisms for efficient image registration. In: International conference on information processing in medical imaging. pp. 559–570. Springer (2017)

[46] Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: International conference on information processing in medical imaging. pp. 37–48. Springer (2013)

[47] Zhang, M., Wells, W.M., Golland, P.: Low-dimensional statistics of anatomical variability via compact representation of image deformations. In: International conference on medical image computing and computer-assisted intervention. pp. 166–173. Springer (2016)

[48] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11. Springer (2018)