Learning to Transform: Unifying Latent Geometric Shape and Appearance Representations in Healthcare Imaging

Tonmoy Hossain

Copenhaver Graduate Endowed Fellow
Department of Computer Science, University of Virginia
tonmoy@virginia.edu

Abstract

Recent advances in deep neural networks have highlighted the importance of geometric shape in various image analysis and computer vision tasks. However, most current approaches rely on coarse or simplified shape representations, such as binary masks, meshes, or point clouds, that are primarily designed to capture global structures of objects presented in images. While effective for general image and visual understanding, these methods often fail to learn fine-grained geometric information that is critical for accurately modeling complex shapes and subtle anatomical variations. This limitation is particularly consequential in healthcare applications, where understanding fine-grained anatomical shapes and their changes is crucial for accurate disease detection and diagnosis. My research focuses on developing a set of advanced deep learning frameworks that learn robust and complex shape representations from dense image data and integrate them into the current paradigm of image appearance and texture learning.

Introduction

`Geometric shapes` represent one of the most fundamental properties by which humans and machines interpret the visual world. Its effectiveness has been demonstrated across various image analysis tasks, including but not limited to image classification, generation, and segmentation (Hossain and Zhang 2025b; Hossain et al. 2019). To fully leverage this broader applicability, the choice of shape representation remains critical for achieving optimal performance. Unlike methods that leverage landmarks, point clouds, or medial axes to extract geometric features, which often overlook internal structures, deformation-based approaches such as elastic deformations and fluid flows (Rueckert et al. 2003) provide comprehensive shape analysis by capturing detailed anatomical information crucial for clinical assessment. Recent research has increasingly focused on incorporating such features into deep networks due to their inherent robustness to variations in image intensity and texture (Hossain and Zhang 2025a). Such robustness has proven particularly effective in medical imaging applications where geometric variations carry critical diagnostic information.

Deformation-based deep networks have emerged as powerful computational frameworks that model complex spatial transformations, enabling robust analysis of geometric variations, specifically in medical imaging. These networks excel in tasks requiring precise geometric understanding, ranging from image registration to shape analysis and morphological studies. Despite such computational advances enabled by these networks, existing approaches exhibit critical limitations in their methodological formulation and applicability to real-time scenarios, failing to (i) *learn* multi-modal distribution of underlying geometric transformations across group-wise images, (ii) *capture* discriminative geometric shapes and ensure generalizability for real-time deployment, (iii) *maintain* robustness across varying data environments, and (iv) *model* appearance variations that correspond to underlying geometric transformations. *In this thesis, I develop a unified theoretical framework that fundamentally redefines how geometric transformations are learned and leveraged in deep learning for healthcare applications.*

Contributions

This dissertation develops a set of shape-informed deep networks that unifies the learning of complex geometric shape and appearance represnetations within a single framework, enabling robust and interpretable analysis of anatomical and pathological changes in healthcare applications (Fig. 1).

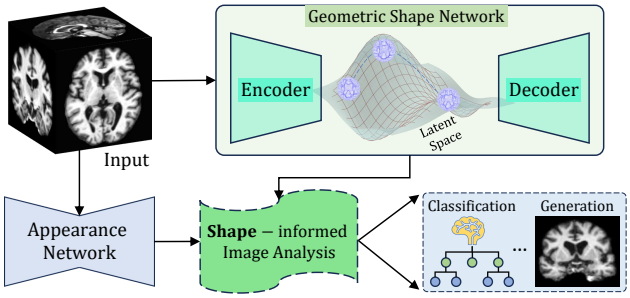

Figure 1: Unified framework for shape-informed representation learning integrating geometric and appearance features.

To address this, I tackle the fundamental challenge that meaningful variations in medical images arises from both

underlying anatomy and observed appearance, which are often treated separately. By embedding structural constraints into the learning process, capturing discriminative variation, and ensuring reliability across diverse environments, the proposed framework jointly learns object geometry and intensity, thereby providing robust and generalized representations of medical images.

Multimodal Geometric Augmentations. We develop a multimodal geometric data augmentation framework, **MGAug**, that learns augmenting transformations in a multimodal latent space of geometric deformations (Hossain and Zhang 2025b). Our generative model learns the distribution of augmented transformations via a mixture of multivariate Gaussians defined as a prior in the tangent space of diffeomorphisms. This captures the naturally clustered geometric variations present in real world clinical datasets while preserving the topological structures. This approach demonstrates consistent improvements in image analysis tasks across diverse imaging datasets, with particular effectiveness in low-data regimes where geometric variations exhibit multiple modes corresponding to distinct anatomical patterns.

Contrastive Learning of Deformable Shapes. To learn discriminative and generalized shape features, we present a contrastive representation learning framework for deformable shapes, **CoRLD**, leveraging a class-aware contrastive objective in latent deformation spaces, promoting proximity among similar classes while ensuring separation of dissimilar groups (Hossain and Zhang 2025a). Our approach further eliminates the requirement for reference images during inference, yielding generalizable shape-aware features that significantly enhance downstream classification tasks.

Invariant Representation Learning. To further strengthen the robustness of image classifiers, we introduce a novel framework that, for the first time, develops invariant shape representation learning, **ISRL** (Hossain et al. 2025). In contrast to existing approaches that derive features in image space, our model is designed to jointly capture invariant features in latent spaces parameterized by deformable and intensity transformations. By embedding deformation-based shape modeling within an invariant feature learning paradigm, this approach yields improved and distributionally robust performance.

In summary, these contributions collectively develop deep networks that unify shape representations with appearance modeling, marking a significant step toward efficient, generalized, and reliable AI-driven healthcare imaging.

Future Works

Representation Learning Under Appearance Changes. While the preceding methods efficiently model geometric deformations in latent spaces, they inherently assume that all anatomical structures can be meaningfully registered through smooth, topology-preserving transformations: an assumption that fails in real-world clinical scenarios where pathological structures such as tumors, lesions, or surgical resections introduce topology-breaking appearance changes. My future research will address this critical limitation by developing frameworks that can distinguish between deformable anatomical variations and non-deformable pathological appearances within a unified learning paradigm. By identifying regions where geometric correspondence exists versus areas where appearance changes that violate topological constraints, these methods will enable spatially-selective deformation modeling that preserves diffeomorphic properties in healthy tissue while appropriately handling pathological structures through learning of separate appearance representations.

Shape-aware Medical Vision Foundation Models. Beyond addressing pathological variations, the ultimate goal of my research is to develop a vision foundation model that unifies deformable shape features from group-wise images with image texture and language representations, enabling comprehensive medical image understanding. Current medical imaging foundation models rely predominantly on image intensities and text alignments, overlooking the rich geometric information encoded in anatomical deformations that capture patient-specific morphological variations, disease progression patterns, and treatment responses. By integrating learned deformable shape features with visual-language embeddings, this foundation model would harness the distinct strengths of each modality, where shape deformations quantify anatomical variations invariant to acquisition protocols, image textures characterize tissue properties, and language grounds findings in clinical terminology and domain knowledge.

These research directions extend shape-based representation learning to complex clinical scenarios with heterogeneous pathologies and bridge vision-language architectures with rich geometric features, establishing a principled framework for clinically grounded AI.

References

Hossain, T.; and Zhang, M. 2025a. CoRLD: Contrastive Representation Learning of Deformable Shapes in Images. In *International Conference on Information Processing in Medical Imaging*, 342–357. Springer.

Hossain, T.; and Zhang, M. 2025b. Mgaug: Multimodal geometric augmentation in latent spaces of image deformations. *Medical Image Analysis*.

Hossain, T.; et al. 2019. Brain tumor detection using convolutional neural network. In *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)*. IEEE.

Hossain, T.; et al. 2025. Invariant shape representation learning for image classification. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE.

Rueckert, D.; et al. 2003. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE transactions on medical imaging*.