Medical Image Analysis 97 (2024) 103260

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A geometric approach to robust medical image segmentation

Ainkaran Santhirasekaram [a,*], Mathias Winkler [b], Andrea Rockall [b], Ben Glocker [a]

[a] Department of Computing, Imperial College London, United Kingdom
[b] Department of Surgery and Cancer, Imperial College London, United Kingdom

ARTICLE INFO

MSC:
41A05
41A10
65D05
65D17

Keywords:
Group Theory
Equivariance
Robustness
Segmentation

ABSTRACT

Robustness of deep learning segmentation models is crucial for their safe incorporation into clinical practice. However, these models can falter when faced with distributional changes. This challenge is evident in magnetic resonance imaging (MRI) scans due to the diverse acquisition protocols across various domains, leading to differences in image characteristics such as textural appearances. We posit that the restricted anatomical differences between subjects could be harnessed to refine the latent space into a set of shape components. The learned set then aims to encompass the relevant anatomical shape variation found within the patient population.

We explore this by utilising multiple MRI sequences to learn texture invariant and shape equivariant features which are used to construct a shape dictionary using vector quantisation. We investigate shape equivariance to a number of different types of groups. We hypothesise and prove that the greater the group order, i.e., the denser the constraint, the better becomes the model robustness. We achieve shape equivariance either with a contrastive based approach or by imposing equivariant constraints on the convolutional kernels. The resulting shape equivariant dictionary is then sampled to compose the segmentation output. Our method achieves state-of-the-art performance for the task of single domain generalisation for prostate and cardiac MRI segmentation. Code is available at https://github.com/AinkaranSanthi/A_Geometric_Perspective_For_Robust_Segmentation.

1. Introduction

A symmetry arises in a transformation applied to an object that retains some of its attributes. Symmetries naturally occur in the field of segmentation which includes, translation, rotation, and reflection along the body's symmetry axis. Generally, patients are positioned within the scanner in a predictable manner (fetal imaging being an exceptional deviation), and any divergence from the average patient alignment is usually slight (often up to 20 degrees). However, with a limited dataset, the observed patient orientations may not adequately reflect the diverse range of poses encountered in routine clinical settings.

An early demonstration of geometric inductive bias is the convolutional layer in a convolutional neural network (CNN). CNNs, crucial for computer vision tasks, owe their success largely to the translation-equivariant convolution operator which arises from using the same convolution kernel across the input signal (LeCun et al., 1998). CNNs can therefore learn feature detection at any location, leveraging inherent translational symmetries in many tasks. Group equivariant CNNs have been introduced to enforce equivariance beyond translational equivariance (Cohen and Welling, 2016a) to include rotation and scale equivariance. It has been shown equivariant models will produce more compact models (Bekkers, 2019).

Magnetic resonance imaging involves a complex acquisition process which differs across subjects and domains. This can lead to varying textural profiles and artefacts which affects the performance of deep learning-based segmentation models. We hypothesise in this work given a segmentation output only contains shape information, we can improve the robustness of segmentation models to textural shifts if we constrain our latent space to a finite number of shape-only components using vector quantisation. We achieve this by firstly enforcing texture invariance in the latent space. Next, we want each element in our shape dictionary to be unique and not contain linear transformations (i.e. rotations) of one another which will lead to a more compact shape codebook. We achieve this by enforcing equivariance to certain groups in the latent space.

The contributions of this work are summarised as follows:

- We propose to constrain the latent space to a dictionary of shape components which is sampled to construct the segmentation output. We impose equivariance of each component in the dictionary

* Corresponding author.
E-mail address: a.santhirasekaram19@imperial.ac.uk (A. Santhirasekaram).

https://doi.org/10.1016/j.media.2024.103260
Received 16 December 2023; Received in revised form 12 June 2024; Accepted 26 June 2024
Available online 29 June 2024

to various discrete continuous groups by either using a contrastive approach or imposing group equivariant constraints on the convolutional kernel itself. We hypothesise this will improve the generalisability of any segmentation model which maps the input space, $\mathcal{X}$ to a lower dimensional embedding space, $\mathcal{E}$ using an encoder, Φ_e before mapping to the segmentation output, $\mathcal{Y}$ with a decoder, Φ_d. As far as we know this work is the first work to propose shape equivariant discrete representation learning.
- We hypothesise and mathematically prove that by increasing the density of the group equivariant constraint by increasing the order of the group, the robustness of a segmentation model improves.
- We evaluate the capability of our method to improve domain generalisability in the task of prostate zonal segmentation with two labels (transitional and peripheral zone) and cardiac segmentation with three labels (Myocardium, Left ventricle and Right Ventricle) when training on a single domain.

2. Background

2.1. Group theory

Throughout this work, group theoretic terminology is prescribed. Therefore, group theory prerequisites are given to provide a basic understanding. This is not an exhaustive coverage of group theory, but rather the fundamental to understand the content in this paper.

Group. A group consists of a set, G of elements and a binary operator $\cdot: G\times G \to G$, termed the group product. This operator dictates how to combine pairs of elements $g_1, g_2 \in G$. To qualify as a group product, the binary operator must adhere to four conditions:

1. Closure: G is closed under $\cdot$: $\forall g_1, g_2 \in G$, $g_1 \cdot g_2 \in G$.
2. Identity: There exists an identity element e such that, $\forall g \in G$, $e \cdot g = g \cdot e = g$.
3. Inverse: For every element, $g \in G$, there is an element $g^{-1} \in G$, s.t. $g \cdot g^{-1} = e$.
4. Associativity: For any set of elements, $g_1, g_2, g_3 \in G$, $(g_1 \cdot g_2)\cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$.

Semi-direct product. A semi-direct product is a generalisation of the direct product which describes a group as the product of more than one subgroup. In this work we view the construction of a group G as the inner semi direct product $\rtimes$, of subgroups $\mathcal{M}$ and $\mathcal{N}$ ($G = \mathcal{N} \rtimes \mathcal{H}$). This intrinsic view enforces $\mathcal{N}$ or $\mathcal{M}$ to be a normal subgroup. Given $\mathcal{N}$ is the normal subgroup, the following statements are then equivalent.

- $\mathcal{N}\mathcal{H} = G$ and $\mathcal{N} \cap \mathcal{H} = \{e\}$.
- Every $g \in G$ can be written uniquely as $g = nh$ with $n \in \mathcal{N}$, $h \in \mathcal{H}$.
- Define $\psi : \mathcal{H} \to G/\mathcal{N}$ in the natural way: $\psi(h) = \overline{h} = h\mathcal{N}$. Then ψ is an isomorphism.

Group Action. A group G can interact with a space made up of a set $\mathcal{X}$. Given a group element $g \in G$, the group operation T_g outlines the effect on any element $x \in X$ when the transformation specified by element g, $T_g(x)$ is applied to it. The action is defined formally as follows:

$$T : G \times X \to X \quad \text{and} \quad T_g : X \to X \tag{1}$$

Group representation. In group theory, a representation ρ is defined as a function that maps each element in the group to an invertible $n \times n$ matrix. Here, n refers to the dimension of the representation, which can be any positive integer, including infinity. A representation ρ of a group G must fulfill the equation $\rho(gg_0) = \rho(g)\rho(g_0)$, where gg_0 represents the composition of two group elements $g, g_0 \in G$, and $\rho(g)\rho(g_0)$ symbolises matrix multiplication. This condition ensures that the matrix multiplication in the representation corresponds to the operation within the group, thereby preserving the group's structure.

Left group representation. Any group action on a set $\mathcal{X}$ implicitly alters functions on $\mathcal{X}$ because it changes the elements of $\mathcal{X}$. This change is typically articulated via a left-regular representations $\mathcal{L}$. Therefore given a function, $f : \mathcal{X} \to \mathbb{R}$, we can determine the nature of a transformed function f' after applying a group element g, by assessing a function's value for $x \in \mathcal{X}$ by reversing the transformed function. For instance, to find the value of f' for a transformed element a', we determine the original value of f for a before g was applied. This is achieved by applying the inverse of g to a'. For the set of functions f on $\mathcal{X}$, the left regular representation for g is shown below.

$$L_g : f \to f' \text{ and for } a' \in T_g(X) : f'(a') = f(T_{g^{-1}}(a')) \tag{2}$$

Equivariance A function is equivariant with respect to a group if it commutes with the action of the group. Therefore a function $\Phi : X \to Y$ is equivariant if the following holds true:

$$\forall g \in G : T_g \circ \Phi = \Phi \circ T_g. \tag{3}$$

We will first describe two finite groups which we use in this thesis.

Cyclic Group. A cyclic group is a group which is generated by a single element X which we describe as the group generator. Cyclic groups are also Abelian. The order of the group is n and the identity $X^n = I$ holds in the cyclic group where I is the identity. An example of the cyclic group is the ring of integers $\mathbb{Z}$ under addition which is an infinite group.

Dihedral Group. For any integer $n \geq 3$, the dihedral group D_n represents the set of symmetry operations for a regular n-gon, with the group operation being the composition of these actions. The group actions consist of rotations and reflections. The group order is 2n. We can formally represent the group as: $\langle x, y | x^2 = 1, y^n = 1, (xy)^2 = 1\rangle$.

We will now formally describe three compact different types of groups which are necessary for this thesis. Firstly, we describe the translation group in two dimensions $R2$ with matrix representation, ρ. This group has a group product and inverse for two elements $x, x_0 \in R2$ shown below.

$$\rho(x) \cdot \rho(x_0) = (x + x_0) \tag{4}$$

$$\rho^{-1}(x) = -x \tag{5}$$

The logarithmic map is given by:

$$\log \rho(x) = x. \tag{6}$$

We also consider the rotation group in two dimensions $SO(2)$ which consists of the set of all continuous rotation transformations around a plane. These rotation matrices are orthogonal matrices R with determinant 1. Its group product and inverse for two elements $R_\theta, R_{0_\theta} \in SO(2)$ is given by:

$$R_\theta R_{\theta_0} = R_{\theta+\theta_0} \tag{7}$$

$$R_\theta^{-1} = R_\theta^T. \tag{8}$$

The final group we consider in our work is the special Euclidean group $SE(2)$ in two dimensions. The 2-dimensional Special Euclidean group represents a series of geometric transformations comprised of rotations and translations in a 2D plane. Every element of the group can be identified using two parameters: the rotation angle (θ) and the translation vector (t), denoted as $g = (t, \theta)$ within G. The group product and inverse for any two elements g, g_θ in SE(2) are defined as follows:

$$g \cdot g_0 = (t, \theta) \cdot (t_0, \theta_0) = (T_\theta(t_0) + t, \theta + \theta_0) \tag{9}$$

$$g^{-1} = (-T_{-\theta}(t), -\theta) \tag{10}$$

In essence, to amalgamate the two elements g and g_0, we first employ the rotation part of g to the translation of g_0, followed by adding it to the translation of g. This is a semi-direct combination of elements. As such, SE(2) is a semi-direct product, made up of the translation group $R2$ and the rotation group $SO(2)$, often notated as $SE(2) = R2 \circ SO(2)$. The above groups are specific to the 2D plane but easily extendable to 3D space.

2.2. Group convolutions

CNNs involve a convolution kernel κ defined on $\mathbb{R}^d$ which modulates an input signal f on the same space. This application yields a function over a d dimensional grid $\mathbb{Z}^d$. Essentially, we first transform the convolution kernel κ for each group element $x \in \mathbb{Z}^d$, creating a set of kernels $\mathcal{L}_x(\kappa)$. Then, we perform the convolution operation on f using these transformed kernels $\mathcal{L}_x(\kappa)$. By maintaining consistent kernel weights across the translation group, we enable learned features to automatically adapt to different spatial positions (LeCun et al., 1995).

We will now go through the processes to derive a group equivariant convolution. There are two types of group CNNs, regular G-CNNs and Steerable G-CNNs. We first construct the recipe for the regular G-CNNs.

We will first introduce the **lifting convolution**. To maintain information about the position of features in the input, an equivariant convolution operation is achieved by elevating a function from the input space to a group's homogeneous space. We are specifically interested in image data that exists on a 2D grid $\mathbb{Z}^2$, so we consider the group of interest H in a semi-direct product with our data's domain; that is, $G = \mathbb{Z}^2 \circ H$. In group convolutions, every transformation in G left-acts on a given kernel, thereby generating a signal in the higher-dimensional space G instead of $\mathbb{Z}^2$. Given the group element $h \in H$ and the position $\tilde{x}$ in the input domain, we formally define the lifting convolution below.

$$(f *_{\text{lifting}} \kappa)(g) = \sum_{\tilde{x} \in \mathbb{Z}^2} f(\tilde{x}) \kappa_h(\tilde{x} - x) d\tilde{x} \tag{11}$$

In the above equation, $g = (x, h)$ and κ_h is the kernel acted upon by the group element $h \in H$ such that, $\mathcal{L}_h[\kappa](x) := \kappa(h^{-1}x)$ (Cohen and Welling, 2016a). The output dimension is the spatial dimension of x plus the group dimension h (Cohen and Welling, 2016a).

We now have the domain transferred to the group space and therefore convolutional operations are now performed over the group space such that, $\kappa : G \rightarrow G$. Given, the Haar measures $d\tilde{g}$ and $d\tilde{h}$ over the group G and H respectively, we define the group convolutional below (Cohen and Welling, 2016a).

$$(f *_{\text{group}} \kappa)(g) = \sum_{G} f(\tilde{g}) \psi(g^{-1} \cdot \tilde{g}) d\tilde{g} \tag{12}$$

$$= \sum_{g \in G} f(\tilde{g}) \mathcal{L}_g \psi(\tilde{g}) d\tilde{g} \tag{13}$$

$$= \sum_{\tilde{x} \in \mathbb{Z}^2} f(\tilde{x}, \tilde{h}) \mathcal{L}_x L_h \psi(\tilde{x}, \tilde{h}) \frac{1}{|h|} d\tilde{x} d\tilde{h} \tag{14}$$

$$= \sum_{\tilde{x} \in \mathbb{Z}^2} f(\tilde{x}, \tilde{h}) \psi(h^{-1}(\tilde{x} - x), h^{-1} \cdot \tilde{h}) \frac{1}{|h|} d\tilde{x} d\tilde{h} \tag{15}$$

Eq. (12) defines a function over G for all elements $g \in G$. We then express Eq. (12) in terms of (x, h) instead of g shown in Eq. (15). Regular group convolutions allow one to force exact lifting and group equivariant convolutions for finite groups such as the D4 group (dihedral group). This is because the sum in Eq. (12) is tractable. For continuous groups, we can either discretise the group H or approximate the group convolution through random sampling (Cohen and Welling, 2016b). We specifically uniformly sample the group elements in the group which are equidistant to each other. Discretisation ensures the network's exact equivariance is only to a discrete subgroup, which results in the network having a bias towards a specific subset of transformations. On the other hand, random sampling produces an unbiased estimate of the continuous group. Also, as one can imagine, a larger sample of the group will produce a better approximation of that group. However, this will increase the number of convolutional operations.

The representations that encapsulate both the group action and the effect of one group element on another are usually reducible (Lang and Weiler, 2020). This means that if the requirement to describe the group action is removed, then they can be simplified. Essentially, we can create decomposable unitary representation matrices composed of smaller block matrices and zero blocks shown in Eq. (16) (Lang and Weiler, 2020). The smaller blocks $\rho_i(g)$ are irreducible, meaning they cannot be further broken down into smaller units and zeros (Lang and Weiler, 2020).

You can also create new representations by combining different irreducible representations such as $\rho_1(g)$ and $\rho_2(g)$ with direct sums i.e. Eq. (16) can also be expressed as $\rho_0(g) \oplus \rho_1(g) \oplus \cdots \oplus \rho_k(g)$.

$$\rho(g) = \mathbf{S}^{-1} \begin{pmatrix} \rho_0(g) & 0 & \cdots & 0 \\ 0 & \rho_1(g) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_k(g) \end{pmatrix} \mathbf{S} \tag{16}$$

These irreps are essential components in constructing more complex representations (Lang and Weiler, 2020). If the group is abelian, then all the irreducible representations are one-dimensional, whereas non-abelian groups are represented as square matrices. In essence, irreducible representations offer a versatile and compact way to represent and understand group structures, providing essential tools for analysis and interpretation.

The count of irreducible representations depends on the nature of the group: finite groups have a finite number of these, while locally compact groups have infinite. These irreducible representations function like orthonormal basis functions in Hilbert spaces (Lang and Weiler, 2020). According to the Peter-Weyl theorem, they can be used to form a complete basis-set for square-integrable L^2 functions on the group (Lang and Weiler, 2020). This transformation is essential since the irreducible representations, which typically output matrices representing g need to be converted to scalars to form a basis for an integrable function (Lang and Weiler, 2020). Integrable functions are derived from our space to our group using a lifting operation. Through this, irreducible representations enable us to express the functions as a direct sum of the coefficients of the irreducible representations. Importantly, the irreducible representations have been comprehensively defined for all groups, allowing us to refer to tables for their values instead of constructing them manually (Lang and Weiler, 2020).

Given a representation of the group denoted, ρ and the equivariant constraint on the kernel denoted in Eq. (17), one can solve the kernel constraint for such representations by simplifying it through the use of irreducible representations (irreps) (Lang and Weiler, 2020).

$$\kappa(gx) = \rho_{\text{out}}(g) \kappa(x) \rho_{\text{in}}(g^{-1}) \forall g \in G, x \in \mathbb{R}^2 \tag{17}$$

This approach hinges on the understanding that any representation of a finite or compact group can be transformed into a collection of irreps through a change of basis. These irreps correspond to invariant subspaces within the representation space and are affected by the action of ρ. By using a change of basis denoted by Q, the representation ρ can be expressed as follows:

$$\rho = Q^{-1} \sum_{i \in I} \psi_i Q \tag{18}$$

Here, ψ_i are the irreducible representations of the group G, and the index set I describes the specific types and quantities of irreps contained within ρ (Lang and Weiler, 2020).

Given, the equivariant constraint imposed on the kernel we can decompose ρ_{in} and ρ_{out} into the following:

$$\kappa(gx) = Q_{\text{out}}^{-1} \left(\sum_{i \in I_{\text{out}}} \psi_i(g) \right) Q_{\text{out}} \kappa(x) Q_{\text{in}}^{-1} \left(\sum_{j \in I_{\text{in}}} \psi_j^{-1}(g) \right) Q_{\text{in}},$$
$$\forall g \in G, x \in \mathbb{R}^n \tag{19}$$

The constraint breaks down into independent constraints through left and right multiplication shown below (Lang and Weiler, 2020; Weiler and Cesa, 2019).

$$\kappa_{ij}(gx) = \psi_i(g) \kappa_{ij}(x) \psi_j^{-1}(g), \quad \forall g \in G, x \in \mathbb{R}^2,$$
$$i \in I_{\text{out}}, j \in I_{\text{in}} \tag{20}$$

Table 1
The angular component basis of SO(2)-steerable kernels that satisfy the irreducible representation kernel constraint. These are constructed for various combinations of input and output field irreducible representations ψ_n.

ψ_m \ ψ_n	ψ_0	$\psi_n, n \in \mathbb{N}^+$
ψ_0	1	$\cos(n\phi)\sin(n\phi)$, $9\sin(n\phi)\cos(n\phi)$
ψ_m, $m \in \mathbb{N}$	$\begin{bmatrix}\cos(m\phi)\\ \sin(m\phi)\end{bmatrix}$, $\begin{bmatrix}9\sin(m\phi)\\ \cos(m\phi)\end{bmatrix}$	$\begin{bmatrix}\cos(m-n)\phi & 9\sin(m-n)\phi\\ \sin(m-n)\phi & \cos(m-n)\phi\end{bmatrix}$, $\begin{bmatrix}9\sin(m-n)\phi & 9\cos(m-n)\phi\\ \cos(m-n)\phi & 9\sin(m-n)\phi\end{bmatrix}$, $\begin{bmatrix}\cos(m+n)\phi & \sin(m+n)\phi\\ \sin(m+n)\phi & 9\cos(m+n)\phi\end{bmatrix}$, $\begin{bmatrix}9\sin(m+n)\phi & \cos(m+n)\phi\\ \cos(m+n)\phi & \sin(m+n)\phi\end{bmatrix}$

We will first solve the irreducible representation constraint shown in Eq. (20) for the SO(2) group. Except for the trivial representation denoted by ψ_0, all the remaining irreps are characterised as 2-dimensional rotation matrices with frequencies that belong to the set of positive natural numbers shown below (Lang and Weiler, 2020; Weiler and Cesa, 2019).

$$\psi^0_{SO(2)}(r\theta) = 1 \tag{21}$$

$$\psi^k_{SO(2)}(r\theta) = \begin{bmatrix}\cos(k\theta) & -\sin(k\theta)\\ \sin(k\theta) & \cos(k\theta)\end{bmatrix} = \psi(k\theta), \quad k \in \mathbb{N} \tag{22}$$

Note, the action of any element from the SO(2) group is norm-preserving which means, $g\|x\| = \|x\|$, $\forall g \in G, x \in \mathbb{R}^2$. The kernel constraint also means the radial part of the kernel is free while the angular component is fixed. It is also important to realise that every irreducible representation in the SO(2) group correspond to a single angular frequency which means one can expand the kernel as a Fourier series with real-valued, radially dependent coefficients (Lang and Weiler, 2020; Weiler and Cesa, 2019). We can then plug in the Fourier expansion of the kernel into Eq. (17) and project onto individual harmonics to constrain the Fourier coefficients which force most coefficients towards 0 (Lang and Weiler, 2020; Weiler and Cesa, 2019). In the context of G's irreducible representations, there are varying dimensional characteristics to consider. They could be 1-dimensional or 2-dimensional. Consequently, there are different possible mappings, such as those between 2-dimensional irreducible representations, those mapping 2-dimensional to 1-dimensional irreducible representations, or vice versa and those between 1-dimensional irreducible representations. Hence, we one must expand the kernel as a Fourier series for each representation mapping, which one then solves to provide solutions to the kernel constraint shown in Eq. (20) (Lang and Weiler, 2020; Weiler and Cesa, 2019). We consider only positive radial parts. The solutions are shown below in Table 1. A detailed derivation of the solutions by using a Fourier expansion of the kernel is shown in Weiler and Cesa (2019).

In Table 1, note each base element is a harmonic corresponding to a singular angular frequency. By expanding the kernel into the Fourier series, we are essentially expressing the kernel as a linear combination of circular harmonics as proposed in Worrall et al. (2017). However, they utilise complex representations and perform actions on complex feature maps. For the O(2) group, the derivations would remain largely unaffected since the real and complex irreps align, only differing by a change of basis. On the other hand, the derivations are simpler in the case of SO(2) as their irreps in complex space are only 1-dimensional. Translating the solution spaces of complex G-steerable kernels back into real values requires careful consideration to avoid an under-parameterised implementation in real space. Additionally, in harmonic networks, while the original complex solution was complete, by shifting the network's operation to the real field, we also incorporate negative frequencies into the feature fields. This permits us to employ an expanded set of 24 steerable kernels without any additional overhead on the system.

In the creation of SO(2) equivariant networks within a computerised environment, the sampling of both the kernels and feature maps on a discrete sampling grid is necessary and can potentially lead to aliasing issues (Lang and Weiler, 2020; Weiler and Cesa, 2019). This demands careful handling, especially with high-frequency filters. Low spatial resolution could exacerbate these issues, particularly with smaller radii where fewer pixels per solid angle are covered near the origin. To mitigate aliasing, a radially dependent angular frequency cutoff is introduced (Weiler and Cesa, 2019). Furthermore, to counterbalance the aliasing effects stemming from the radial part of the kernel basis, a smooth Gaussian radial profile is chosen (Weiler and Cesa, 2019).

3. Related work

3.1. Equivariant neural networks

The exploration of neural networks with equivariance to specific symmetry groups has rapidly evolved (Bekkers et al., 2018; Cohen and Welling, 2016a,b; Dieleman et al., 2016; Hoogeboom et al., 2018; Hy et al., 2018; Kondor et al., 2018; Marcos et al., 2017; Ravanbakhsh et al., 2017; Weiler et al., 2018). These models can be understood at a broader level by looking at two distinct aspects: first, the symmetry group to which they maintain equivariance, and second, the kind of geometric characteristics they utilise (Cohen et al., 2018). We explore both regular and steerable convolutional kernels in this paper.

Equivariant neural networks have shown improved performances in various computer vision tasks. For example, Winkels and Cohen (2018) demonstrated by using regular group convolutions which included the dihedral group, that one can significant improve performance for lung nodule detection on volumetric CT imaging. Specifically, their 3D G-CNN outperformed a similar baseline architecture in false positive reduction for pulmonary nodule detection, showing better sensitivity to malignant nodules and faster convergence. Harmonic networks (Worrall et al., 2017) which use steerable convolutions demonstrated improved parameter efficiency while achieving SOTA classification and boundary detection performance.

Previously, rotation-equivariant UNets were limited to 2D data (Chidester et al., 2019; Linmans et al., 2018; Pang et al., 2022) demonstrating improved sample efficiency. Müller et al. (2021) describes a specific 6 six-dimensional diffusion MRI data application using SE3 equivariant CNN filters for multiple sclerosis segmentation, capturing equivariance in both voxel space and q-space which leads to improved performance. This SE(3)-equivariant volumetric segmentation network is immune to unseen data poses during training, eliminating the need for 3D-rotation-based data augmentation. They also revealed better segmentation results on MRI for both brain tumor and healthy brain structures, with limited training data and increased parameter efficiency. Similarly, Liu et al. (2022a) proposed a SE(3) group equivariant segmentation of diffusion weighted MRI images from the human connectome project. Specifically they employ light-weight regular group convolutions as opposed to the traditional spectral based approaches. However, this work is constrained to only learning group equivariant features which are not necessary shape based features. Our novel contribution compared to this piece of work is to learn a finite set of shape equivariant features to build a shape equivariant dictionary.

3.2. Robustness methods

Recent research reveals that human recognition relies on shape, whereas ImageNet (Deng et al., 2009) pre-trained CNNs depend on texture. To combat this, Geirhos et al. (2018) created Stylized ImageNet (SIN) using the AdaIN (Huang and Belongie, 2017) style transfer algorithm, which replaces natural texture with various random ones. Since a model trained on SIN cannot predict results using local texture, it must consider the input's shape and structure. Experiments in Geirhos et al. (2018) show that a CNN trained on SIN behaves more like human recognition, focusing on shape. This in particularly relevant to segmentation tasks where only shape information is required to perform the task. Kim and Byun (2020) proposed a method to learn texture invariant features to improve the domain generalisability of segmentation models. They diversify synthetic image textures using a style transfer algorithm, preventing segmentation model overfitting to one specific texture. Then, by fine-tuning the model with self-supervised training on the target texture, they achieve state-of-the-art performance. RandConv (Xu et al., 2020) which is perhaps the most related work, attempts to learn textural invariant features by using a randomised convolutional input layer.

The primary methods for improving model robustness include data augmentation, adversarial training, and architectural design. Data augmentation synthesises new training data to make models invariant to various perturbations. In medical imaging, aggressive augmentation schemes like BigAug significantly enhance segmentation performance (Zhang et al., 2020). Recent strategies like CutOut and MixUp encourage generalised feature learning (DeVries and Taylor, 2017; Zhang et al., 2017). Adversarial training manipulates input data with specific objectives; for example, Madry et al. (2017) construct a min–max problem for effective perturbation learning. More sophisticated methods like Meta-learning based Adversarial Domain Augmentation (M-ADA) use meta-learning to create new domains (Qiao et al., 2020). In medical imaging, AdvBias employs adversarial data augmentation to learn bias field deformations (Chen et al., 2020).

A popular architectural design choice to improve the robustness of CNNs is centred around spectral methods which remove high-frequency information. For example, Zhang (2019) introduces a general-purpose computational layer by enforcing anti-aliasing in the max/average pooling and strided convolutional operations. Similarly, Li et al. (2020a) replaces up/down sampling operations in the CNN with a wavelet-based method to improve robustness to noise. Li et al. (2020b) aims to minimise feature redundancies across space and channels by using the discrete cosine transform to learn a series of sparse transforms which they denote as LST. This can be easily integrated into common CNN architectures and enhanced accuracy was demonstrated across all ImageNet-C corruption types after LST is incorporated into the ResNet-50. Compositionality for robust segmentation has also recently been proposed by Liu et al. (2022b). Here, they propose a method where we represent the compositional components, or patterns, of human anatomy using learnable von-Mises-Fisher (vMF) kernels. They decompose image features into these components using composing operations, specifically, the vMF likelihoods. They claim this approach allows to model the intricate structures of human anatomy effectively, even when dealing with data collected from diverse sources.

4. Proposal

There is anatomical consistency across subjects meaning there is reduced spatial variation in the segmentation outputs. Therefore, we propose to constrain the latent space to a dictionary of shape components which is sampled to construct the segmentation output. This is achieved using vector quantisation (Van Den Oord et al., 2017) of the shape equivariant features to create a discrete shape space. We impose equivariance of each of the components in the dictionary to various discrete continuous groups by either using a contrastive approach or imposing group equivariant constraints on the convolutional kernel itself. We hypothesise this will improve the generalisability of any segmentation model which maps the input space, $\mathcal{X}$ to a lower dimensional embedding space, $\mathcal{E}$ using an encoder, Φ_e before mapping to the segmentation output, $\mathcal{Y}$ with a decoder, Φ_d. We also hypothesise that by increasing the density of the group equivariant constraint by increasing the order of the group, the robustness of a segmentation model improves. We assume the dictionary is complete and sufficient to capture the entire distribution of segmentation outputs after composition of the discrete shape space using the decoder. We evaluate the capability of our method to improve domain generalisability in the task of prostate zonal segmentation with two labels (transitional and peripheral zone) and cardiac segmentation with three labels (Myocardium, Left ventricle and Right ventricle) when training on a single domain.

5. Method

5.1. Method 1: Contrastive learning

Fig. 1 highlights the first proposed method to create shape equivariant components in the shape dictionary. This method is broken down into imposing texture invariance, shape equivariance and then quantisation to create the shape dictionary as described below.

5.1.1. Texture invariance

Firstly, to achieve texture invariance we exploit the multiple acquisitions which are used in MRI which leads to differing textural and contrast properties among different sequences. Specifically, we propose a method to learn texture invariant features based on the principle that T2-weighted images and apparent diffusion coefficient (ADC) maps calculated from diffusion-weighted imaging contain similar spatial information and only differ in their textural profiles. Therefore, we impose a contrastive loss such that T2-weighted and ADC imaging are mapped to the same representation. Formally, we start with the image input which is the T2 weighted image $x \in \mathbb{R}^{1\times256\times256\times24}$ and apply an intensity transformation, T_i which is equivalent to acquiring the ADC map.

In cases where there is only a single sequence available we will use a randomised convolutional layer in the first layer of our network to enforce texture invariance as implemented by Xu et al. (2020). The output of random filters retains relative similarities between input patches, as proven in Xu et al. (2020). Thus, locations with similar local textures in the input tend to remain similar in the output, meaning shapes in the output closely resemble those in the input, especially when filter sizes are small compared to typical shape sizes. Essentially, a convolution filter's size dictates the smallest shape it retains. For instance, 1×1 random convolutions act as random colour mapping, preserving single-pixel shapes, while larger filters disrupt shapes smaller than the filter, viewing them as local textures at that scale. In this work, we randomly sample uniform kernel sizes of 1,3 or 5 and we randomly sample weights for a randomised convolution layer from a Gaussian distribution ($\sim \mathcal{N}\left(0, \frac{1}{3k^2}\right)$). We mix the output of the randomised convolutional layer denoted I_r with the original image, I as advised in Xu et al. (2020). Specifically, we randomly sample the mixing parameter, α from a uniform distribution($\alpha \sim U(0,1)$) to linear mix with the original image ($\alpha I + (1-\alpha) I_r$).

5.1.2. Shape equivariance

We also want to impose shape equivariance in our latent before and after quantisation to construct a discrete shape equivariant latent space which we claim improves the robustness of a segmentation model.

We apply a spatial transformation, T_s to the ADC map which involves rotations. Specifically, we apply transformations from the dihedral group. In our work, we will be dealing with 3D imaging and hence we will be using 3D CNN filters. In a three-dimensional CNN, filters take the shape of cubes or cuboids rather than squares. In 3D

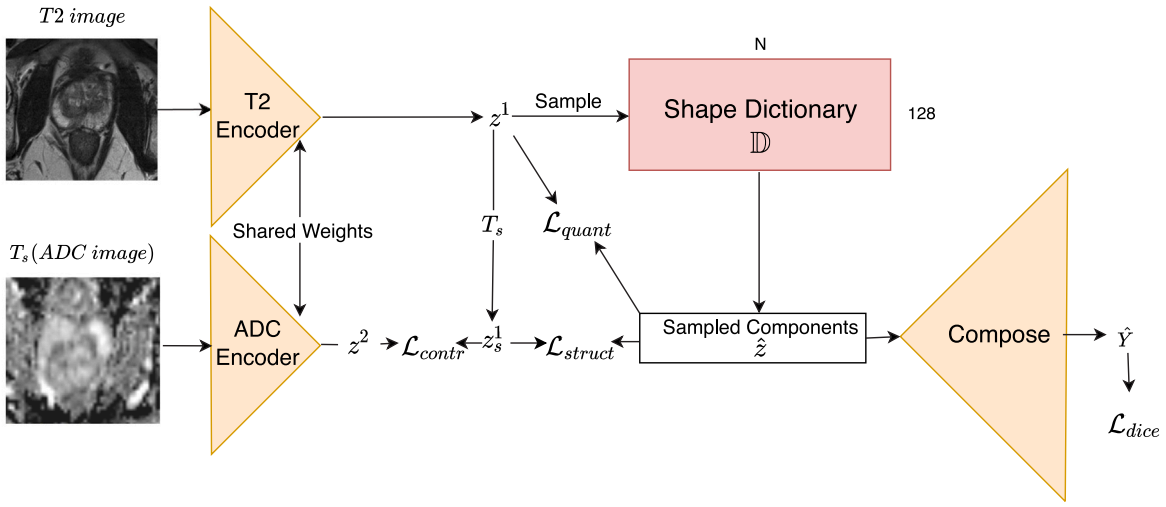

Fig. 1. Overview of our method demonstrating using the ADC map to learn shape equivariant features which is quantised to construct a shape dictionary, $\mathbb{D}$. The shape dictionary is a set of equivariant shape basis which are sampled to construct the segmentation output.

medical imaging methods like CT and MRI, the pixel spacing along the x and y axes may differ from the z axis. In our work, the MRI images we use are heavily anistropic where the thickness of the slices is significantly larger than the axial resolution. Consequently, a filter-sized k × k × k might represent a space resembling a cuboid instead of a perfect cube. Therefore, in this work, we consider symmetries of the cuboid which provides the rationale for why we consider group equivariant convolutions to the dihedral group. We consider rotations which maintain orientation and reflections as well as rotations which do not maintain orientation. In Fig. 2, we show the Cayley diagrams for groups D4 and D4h. In a Cayley diagram, every node represents a symmetry transformation, denoted as $h \in H$. In Fig. 2, each colour signifies the application of a specific generator transformation. By sequentially applying these generators, or tracing the diagram's edges, one can achieve any transformation within the group. We work with dihedral groups Dn which are orientation preserving of varying orders, which include D4, D6, D8, D12 and D16. For example, D4 consists of 90-degree rotations in the z plane and 180 degree rotations in the y plane shown in Fig. 2 by the red arrow and blue line respectively. The order of this group is 8 so we create 8 transformations per sample during training. D12 consists of 30 degree rotations in the z plane and 180 degree rotations in the y plane. The order of this group is therefore 24. The non-orientation preserving dihedral groups of varying orders used in our work consists of D4h, D6h, D8h, D12h and D16h. This group has the additional transformation of a reflection in the Z-axis (green line in Fig. 2) so the order of the D4h and D12h groups is 16 and 48 respectively.

The T2 image and spatially transformed ADC map are passed through an encoder to produce their respective embeddings, z^1 and z^2 as shown in Fig. 1. Shape equivariance and texture invariance are enforced by satisfying Eq. (23).

$$\Phi_e\left(\sum_{i=0}^{n} T_s^i(T_i(x))\right) = \sum_{i=0}^{n} T_s(z) \tag{23}$$

Therefore, we minimise the contrastive loss: $\mathcal{L}_{contr} = \|T_s(z^1) - z^2\|_2^2$. Note, a contrastive loss only theoretically learns equivariance to the spatial transformations applied per sample. It does not constrain the convolutional layers to the group. We assume an approximate equivariance to the group by using our contrastive loss. Note, this contrastive loss only considers positive pairs as this is what is only required for shape equivariance and texture invariance. The justification for the contrastive loss is purely to exploit the advantage of having two sequences which share the same shape information but varying textural profiles for prostate MRI. The texture invariant component of the contrastive loss is not required for the cardiac MRI dataset as we only have one T2-weighted sequences and hence a randomised convolutional layer which is not used for the prostate MRI data. In regards to justification for its use for promoting equivariance, this is purely experimental and we compare to imposing true shape equivariance by group equivariant convolutional kernels described in method 2 in Section 5.2.

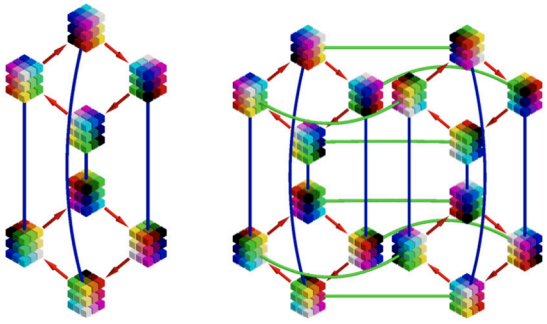

Fig. 2. Cayley Diagram for D4(Right) and D4h(Left). *Source:* This image is taken from Winkels and Cohen (2018).

5.1.3. Quantisation

We quantise $z^1 \in \mathbb{R}^{256\times16\times16\times12}$ and $z^1 \in \mathbb{R}^{256\times18\times18\times12}$ for the prostate dataset and cardiac datasets respectively using vector quantisation. This is carried out by dividing z^1 into 16 × 16 × 12 and 18 × 18 × 12 components for the prostate and cardiac datasets respectively and replacing each component in z^1 denoted z_i^1 with its nearest component $e_k \in \mathbb{D}$ where $k = argmin_j \|z_i^1 - e_j\|_2$. This produces the discrete shape latent space $\hat{z}$ which is inputted into the decoder to construct the segmentation output. The quantisation loss minimises the Euclidean distance between z_i^1 and its nearest component $e_k \in \mathbb{D}$ shown in Eq. (24).

$$\mathcal{L}_{Quant} = \frac{1}{m}\sum_{i=0}^{m-1} \|sg(z_i^1) - e_k\|_2 + \beta\|z_i^1 - sg(e_k)\|_2 \tag{24}$$

We compute the Dice loss between the output, $\hat{y}$ and the T2 segmentation label y. Only T2-weighted images are required as input during inference.

During quantisation, the geometric structure in z_1 may be partially lost after quantisation. This is because it is quite feasible that a large

proportion of the elements in z_1 is mapped to the same codebook vector during quantisation which leads to a many-to-one mapping between spatial elements in z^1 and $\hat{z}$. This is known as codebook collapse and a well-known phenomenon in VQ-VAE (Bie et al., 2022) which leads to invariant rather than desired equivariant properties. Therefore, to preserve geometry after quantisation, we impose a structure-preserving loss shown in Eq. (25). This loss aims to preserve the distance between the neighbouring elements $m_i, m_{i+1} \in z^1$ which are of dimension, $\mathbb{R}^{128\times1\times1\times1}$ after mapping to elements $n_i, n_{i+1} \in \hat{z}$ via quantisation ($n \in \mathbb{R}^{128\times1\times1\times1}$). This loss is applied to all $16 \times 16 \times 12$ elements mapped from z^1 to $\hat{z}$.

We carry ablations to find the minimum of codebook vectors required to not diminish segmentation performance by more than 2 dice points compared to the largest codebook for each experiment. This is because we want learn the most compact codebook possible without affecting segmentation performance. We use 1024, 512, 256, 128 and 64 codebook vectors in the ablations.

$$\mathcal{L}_{Struct} = \sum_{i=0}^{127} \|m_i.m_{i+1} - n_i.n_{i+1}\|_2 \tag{25}$$

The total loss for training our framework in Fig. 1 is $\mathcal{L}_{total} = \mathcal{L}_{Dice}(\hat{y}, y) + \mathcal{L}_{contr} + \mathcal{L}_{Quant} + \mathcal{L}_{Struct}$.

5.2. Method 2: Group convolutional kernels

In our previous method, we used a 2D/3D hybrid UNet to learn equivariance in the discrete latent space to the 3D dihedral groups using a contrastive method. However, in our second method, we constrain the convolutional kernels to group equivariant convolutional kernels of particular groups. Therefore we cannot use a hybrid network but either a 2D or 3D UNet. This is because imposing 2D equivariant constraints on the convolutional kernels in the initial layers of the network followed by 3D equivariant convolutions, will not make the network equivariant to the 3D group but only the 2D groups provided the 2D group is a subgroup of the 3D group. Therefore we use a 2D UNet to impose 2D group equivariant constraints due to the highly anisotropic nature of cardiac and prostate MRI images. We impose texture invariance in the latent space via a contrastive loss ($\mathcal{L}_{contr} = \|(z^1) - z^2\|_2^2$) for the prostate dataset via the representations learnt for the T2-weighted images (z^1) and ADC images (z^2) and an initial randomised convolutional layer for the cardiac dataset. Therefore the method is similar to the contrastive method shown in Fig. 1 except here we remove the spatial equivariant component of the contrastive loss and replace all convolutional kernels with group equivariant convolutional kernels. In this method we want to achieve equivariance to rotation and reflection and so we enforce equivariance to the cyclic, dihedral and SO(2) groups. We also employ SE(2) equivariance which includes translation by adapting the method by Li et al. (2018) which we compare to in our SDG experiments. These groups are useful to demonstrate our theory as one can uniformly increase the order of the group in a gradual manner to assess the correlation between group order and robustness. Additionally these groups are traditionally what is used in the literature for equivariant CNNs (Cohen and Welling, 2016a,b; Cohen et al., 2018)

5.2.1. Dihedral group

We first describe the implementation of the group equivariant convolutions for the dihedral group. In the initial layer, the input feature maps and filters do not have orientation channels. Given n_0 input channels and n_1 filters for each of the n_0 input channels, we modify each of the n_1 filters using every transformation $g \in G$. This creates an expanded filter bank consisting of $n_1 \times |G|$ filters. Therefore each of the n_0 channels is transformed by all the filters in our expanded filter bank.

In the second and higher-order layers, we then reshuffle our orientation channels. In the transformation phase, each filter undergoes modification by each $g \in G$, resulting in $n_{l+1} \times |G|$ adjusted filters which are then applied to each input channel to create as many output channels. The transformations used for the 2D group convolutions are sampled from the 2D dihedral groups which consist of rotations in the 2D plane and reflections in the Y axis. Therefore, the order of the 2D D4 group is 8. The point groups such as D4h do not apply to 2D group convolutions and therefore not used in this method.

5.2.2. SO(2) group

We will also consider the SO(2) group in terms of its cyclic subgroup $Cn \subseteq SO(2)$. We therefore constrain the convolutional kernels in the 2D UNet to either the C4, C6, C8 or C16 cyclical groups. We also use irreducible representation in the form of circular harmonics for the implementation of SO(2) group equivariant convolutions. As previously described, SO(2) can be decomposed into irreducible representations (irreps). We use the escnn library devised by Weiler and Cesa (2019) to build our SO(2) equivariant kernels. We draw inspiration from harmonic networks and build SO(2) equivariant convolutional layers up to an order of 3. Specifically, in our work, we use scalar fields (l=0), vector fields (l = 1) and tensor fields (l=2). A convolutional layer is computed as the direct sum of the irreducible representations up to order 3 ($\oplus_{i=0}^{L}\psi_i$). We use a multiplicity of 1 for each field representation. The feature fields of the SO(2)-equivariant layer are constrained to transform under irreducible representations shown in Table 1.

Every model utilises ELUs for scalar fields and employs norm-ReLUs for higher-order fields. In the convolutional encoder of our models, we subject the scalar-valued feature components to max-pooling. However, when dealing with vector or tensor-valued components, denoted by v, the vector possessing the highest l_2 norm is retained during pooling. The decoding path utilises trilinear or bilinear upsampling, an inherently equivariant operation. Standard instance normalisation (Ulyanov et al., 2016) is applied to scalar features. For vector or tensor-valued features, instance normalisation is carried out by dividing by the average l_2 norm for each instance, expressed as: $\text{norm}(v) := \frac{v}{E(\|v\|)}$. In each convolutional block for each level of the encoder, we use a total of 16, 32, 64, 128 and 256 filters for levels 1, 2, 3, 4 and 5.

The SO(2) group equivariant constraint is the densest constraint imposed on the network as here we are technically imposing infinite constraints in the latent space z_i. Therefore, if we follow Theorem 1 and its corresponding proof 1, this should in theory be the most robust model.

5.3. Equivariance for robustness to domain shifts theory

Here we provide the theory to justify the use of equivariance in the latent space for robust segmentation. We next make the following assumption of a domain shift which occurs due to a change in the acquisition of the image:

Assumption 1. Given convolutional encoder ϕ_e which is texture biased and maps image X to representation z, we assume the mutual geometric information between X and z is corrupted by an acquisition shift in X denoted X_a. In particular, we claim the equivariant relationship between X_a and z is corrupted

The above assumption essentially claims, that shape and texture information is entangled in the neurons of the encoder and therefore a shift in textural information will affect the geometric/shape information extracted by the encoder leading to a semantic shift which is not desired (Islam et al., 2021). This is despite the shape information remaining invariant in an acquisition shift or in other words a textural shift. In fact, local imperceptible textural change in the input space leads to significant semantic shifts which form adversarial examples (Ilyas et al., 2019). The semantics are often largely captured through geometric/shape information (Geirhos et al., 2018). Therefore, as well as minimising the textural information in the latent space as

we proposed previously, we also aim to maximise the mutual geometric/shape information between the input space and latent space. Given the output is a segmentation map with no textural information, we are also thereby maximising the mutual information between the latent space and output space. This leads us to state our first lemma below.

Lemma 1. *By increasing the density of the equivariant constraint on the mapping between X and z we are also increasing the mutual information between Y and z ($I(Y;z)$)*

We next make the following assumptions.

Assumption 2. We assume the decoder, ϕ_d which maps the latent space z to the output Y demonstrates improved performance that surpasses random guessing of each pixel in the output space.

Assumption 3. We also assume instances within our dataset are uniformly distributed across c categories.

We impose a dense constraint on z by constructing n representation of X which are well defined where n is the order of the group. The equivariant constraint is shown in Eq. (23). The transformations are sampled from the matrix representations of all the group elements, $T_s \in \rho(G)$ to construct representations $z_0, z_1 ... z_n$.

We will now first form our first theorem given an acquisition shift $x_a \in X_a$. We state a theorem and prove it based on similar work by Mao et al. (2022) who also showed imposing equivariance constraints in the latent space improves robustness to adversarial attacks.

Theorem 1. *The prediction accuracy given the dense equivariant constraints is bounded as follows:*

$$P(Y|z_0, X_a) \in [b_0, c_0] \tag{26}$$

$$P(Y|z_0, z_1 X_a) \in [b_1, c_1] \tag{27}$$

$$P(Y|z_o, z_1, z_2, X_a) \in [b_2, c_2] \tag{28}$$

$$\vdots \tag{29}$$

$$P(Y|z_0, z_1, \dots, z_n, X_a) \in [b_n, c_n]. \tag{30}$$

Therefore there is an order on the prediction accuracy bounds as follows:

$$b_0 \leq b_1 \leq \cdots \leq b_n \tag{31}$$

and

$$c_0 < c_1 < c_2 < \cdots < c_n. \tag{32}$$

This shows that we can improve the classification accuracy bounds using denser geometric constraints on z.

Proof. Given Lemma 1, it follows that: Given that:

$$I(Y; z_i | X = x^a) > 0 \tag{33}$$

it can be deduced that:

$$I(Y; z_0, z_1, \dots, z_n | X_a) > I(Y; z_0, z_1, z_2 | X^a) \tag{34}$$

$$> I(Y; z_0, z_1 | X^a) > I(Y; z_0 | X^a) = I(Y; X^a) \tag{35}$$

In the next step of our proof, first consider an output space with c classes. Given, our prediction $\hat{Y}$, we define an error bound on our prediction accuracy as follows; $P(\hat{Y} = Y) \geq 1 - \epsilon$. Next using Fano's inequality the following equation holds.

$$H(Y|X^a) \leq H(\epsilon) + \epsilon \cdot \log(n-1) \tag{36}$$

Here, $H(\epsilon) = \epsilon \log \epsilon + (1-\epsilon)\log(1-\epsilon)$. We then add $H(Y)$ to both sides of the equation and rearrange.

$$H(Y) - \epsilon \cdot \log(n-1) \geq H(\epsilon) - I(Y; X^a) \tag{37}$$

We know the mutual information between Y and X_a is defined as: $I(Y; X^a) = H(Y) - H(Y|X^a)$ which we can insert into Eq. (37) to form the following equation:

$$H(\epsilon p) + \epsilon \cdot p \log(n-1) \leq I(Y; X^a) + H(Y) \tag{38}$$

We redefine the left-hand side of Eq. (38) with a new term, $J(\epsilon)$ in Eq. (39).

$$J(\epsilon) \geq -I(Y; X^a) + H(Y) \tag{39}$$

We next multiply both sides of Eq. (39) with the inverse function, J^{-1}.

$$\epsilon \geq J^{-1}(-I(Y; X^a) + H(Y)) \tag{40}$$

Eq. (40) is rearranged to form the following equation.

$$1 - \epsilon \leq 1 - J^{-1}(-I(Y; X^a) + H(Y)) \tag{41}$$

Note we previously placed an error bound on our prediction accuracy as $1 - \epsilon$ and therefore can deduce the following set of equations from Eq. (41).

$$\epsilon \leq c_0 = 1 - J^{-1}(-I(Y; X^a, z_0) + H(Y)), \tag{42}$$

$$\epsilon \leq c_1 = 1 - J^{-1}(-I(Y; X^a, z_0, z_1) + H(Y)), \tag{43}$$

$$\epsilon \leq c_2 = 1 - J^{-1}(-I(Y; X^a, z_0, z_1, z_2) + H(Y)) \tag{44}$$

$$\dots \epsilon \leq c_n = 1 - J^{-1}(-I(Y; X^a, z_0, z_1, z_2 \dots z_n) + H(Y)), \tag{45}$$

The equations above demonstrate that the upper bounds c_i are a function of the mutual information between the latent space z and output space Y. Thus, given $H(Y)$ is a constant, the greater the mutual information between z and Y then the greater the upper bound. In other words the denser the constraint or the larger the order of our group, then the greater the upper bound ($c_0 < c_1 < c_2 < \cdots < c_n$). Simultaneously, an identical proof can be derived to show denser constraints on z can increase the lower bound and thus $b_0 \leq b_1 \leq \cdots \leq b_n$. This concludes our proof.

6. Experiments

Our experiment assesses the single domain generalisability (SDG) of various models. We evaluate segmentation performance when testing a trained segmentation model on an unseen target domain for the prostate and cardiac datasets. We adopt a cross-validation procedure by training a method on a single source domain and holding out the other domains for testing (3 for cardiac and 1 for prostate).

6.1. Datasets and training

We use 2 datasets in our experiments.

Cardiac: This dataset is the M&Ms cardiac imaging 3-class segmentation dataset (Campello et al., 2021) which is divided into 4 domains determined by the MRI scanner vendor. All images were normalised between 0 and 1 and cropped to $288 \times 288 \times 12$ for 3D input in method 1 and 288×288 for 2D input in method 2.

Prostate: We use the NCI-ISBI13 Challenge (Bloch et al., 2015b) dataset. To evaluate the single-domain generalisability of prostate zonal segmentation, we utilise two datasets, each with two labels representing the transitional and peripheral zones. The training set comprises 32 T2-weighted and ADC images from the prostate dataset in the Medical Segmentation Decathlon. These images were acquired from the Radboud University Nijmegen Medical Centre (RUNMC) (Antonelli et al., 2022).

For the test set, we employ 30 T2-weighted scans obtained from the Boston Medical Centre (BMC) as part of the NCI-ISBI13 Challenge (Bloch et al., 2015b; Clark et al., 2013; Bloch et al., 2015a). We normalised images between 0 and 1 and centre cropped to

Table 2
Dice score and Hausdorff distance(HD) ± standard deviation for different order dihedral group shape equivariant models trained via contrastive learning approach. This table shows the results for the prostate dataset.

	Baseline	D4	D6	D8	D12	D16	D4h	D6h	D8h	D12h	D16h
	Transitional Zone										
Dice	0.64 ± 0.13	0.69 ± 0.08	0.70 ± 0.17	0.71 ± 0.11	0.72 ± 0.13	0.72 ± 0.10	0.68 ± 0.15	0.69 ± 0.14	**0.73** ± **0.13**	0.72 ± 0.14	0.71 ± 0.12
HD	24.33 ± 14.12	20.14 ± 3.55	18.93 ± 3.94	18.41 ± 4.50	19.21 ± 4.80	19.01 ± 4.34	19.10 ± 4.11	19.20 ± 5.30	**18.90** ± **4.45**	18.19 ± 5.50	19.07 ± 4.61
	Peripheral Zone										
Dice	0.41 ± 0.09	0.47 ± 0.12	0.49 ± 0.15	0.51 ± 0.10	0.52 ± 0.11	0.54 ± 0.09	0.48 ± 0.07	0.51 ± 0.10	0.53 ± 0.09	**0.55** ± **0.09**	0.53 ± 0.17
HD	41.88 ± 18.72	34.94 ± 15.54	31.74 ± 17.90	29.28 ± 14.72	29.90 ± 17.22	30.82 ± 13.99	32.62 ± 15.72	30.93 ± 12.95	**28.92** ± **9.92**	29.54 ± 12.83	29.10 ± 13.12

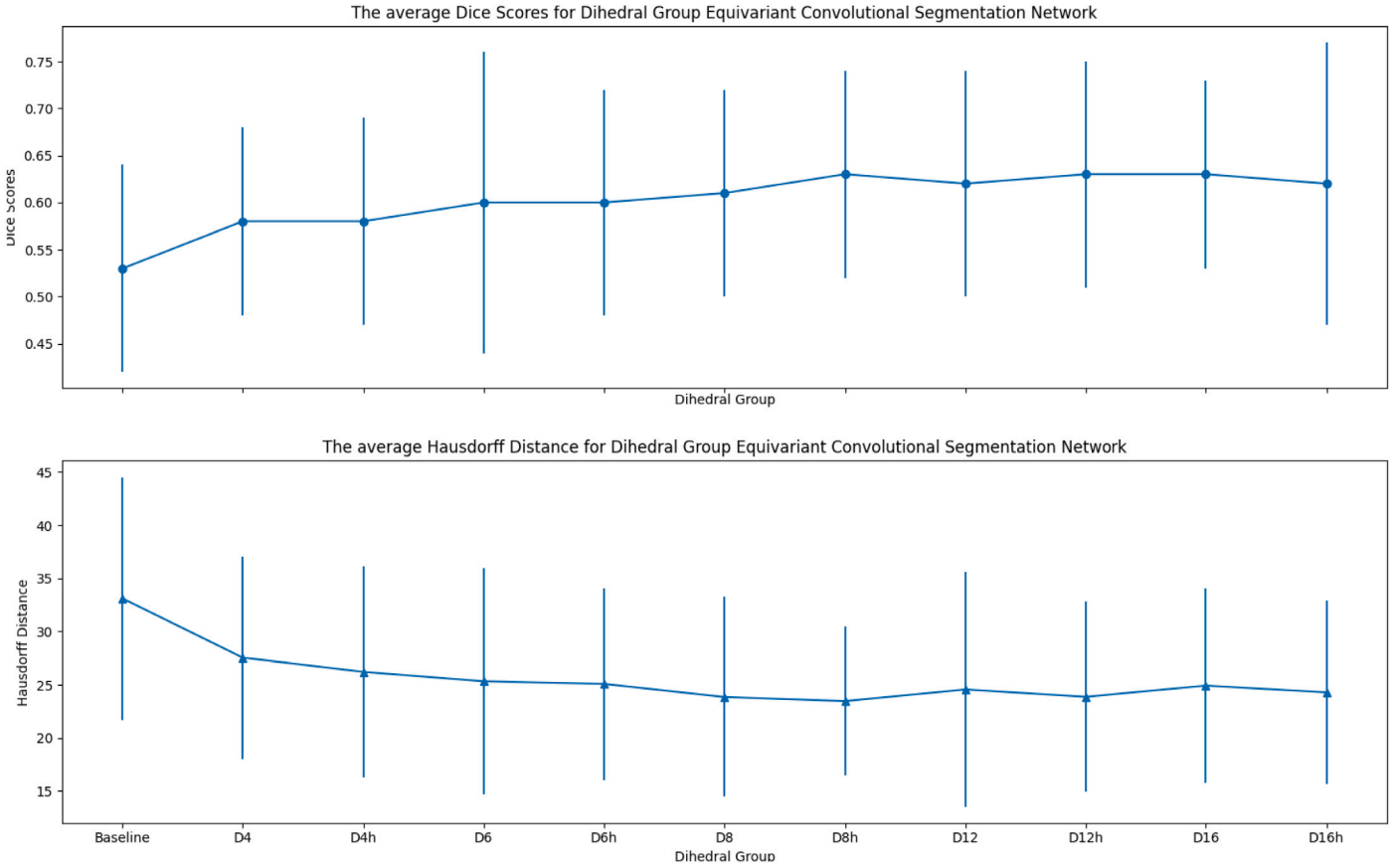

Fig. 3. The average Dice and maximum Hausdorff distance scores ± standard deviation for the respective dihedral group equivariant CNNs trained via contrastive learning and applied to the prostate dataset. The scores are averaged across all classes.

256 × 256 × 12 for 3D input in method 1 and 256 × 256 for 2D input in method 2. The T2-weighted and ADC images are pre-registered.

Training: All models are trained with Adam optimisation (Kingma and Ba, 2014) with a base learning rate of 0.0001 and weight decay of 0.05 for a maximum of 500 epochs on three NVIDIA RTX 2080 GPUs. To prevent over-fitting, we apply a simple augmentation scheme consisting of random rotation and flipping (horizontal and vertical) for training baseline models where our equivariant methods are not incorporated into the model. No augmentation is applied in our group equivariant models. We evaluate model performance with the Dice score and maximum Hausdorff distance (HD) in mm.

6.2. Method 1 experiments

6.2.1. Experimental setup

In the first set of experiments, we first explore method 1 applied to the cardiac and prostate datasets. We use a hybrid 2D/3D UNet as our baseline model to deal with the anisotropic prostate and cardiac MRI images. The encoder and decoder are made up of 5 levels consisting of 2D pre-activation residual blocks in the top 4 levels and a 3D pre-activation residual block in the bottleneck level. The pre-activation residual blocks incorporate group normalisation (number of groups = 8) and ReLU activation. The first level outputs 16 features (channels) and doubles at each level until we have 256 features in the bottleneck. We use 2D max pooling for downsampling and bi-linear interpolation for upsampling operations. Ablations as shown reveal a minimum of 128 codebook vectors are required as shown in the ablation experiments in Section 6.3.7. We use the same encoder and decoder architecture for our approach with various geometric constraints. We compare our method to the following SDG methods applied to the baseline model: CutOut (DeVries and Taylor, 2017), BigAug (Zhang et al., 2020), AdvBias (Chen et al., 2020), RandConv (Xu et al., 2020), Jigen (Carlucci et al., 2019) and vMFNet (Liu et al., 2022b). We further

Table 3
Dice score and Hausdorff distance(HD) ± standard deviation for different order dihedral group shape equivariant models trained via contrastive learning. This table shows the results for the cardiac dataset.

	Baseline	D4	D6	D8	D12	D16	D4h	D6h	D8h	D12h	D16h
	Left Ventricle										
Dice	0.74 ± 0.13	0.78 ± 0.15	0.79 ± 0.12	0.80 ± 0.09	0.82 ± 0.13	0.80 ± 0.09	0.79 ± 0.12	0.80 ± 0.17	0.81 ± 0.11	**0.82** ± **0.14**	0.81 ± 0.12
HD	15.33 ± 3.1	12.61 ± 2.75	11.13 ± 2.58	10.74 ± 3.79	10.91 ± 2.99	10.14 ± 3.13	11.73 ± 3.98	10.18 ± 3.24	**9.93** ± **3.08**	10.72 ± 3.07	10.16 ± 3.82
	Right Ventricle										
Dice	0.72 ± 0.16	0.75 ± 0.20	0.77 ± 0.21	0.79 ± 0.17	0.78 ± 0.16	0.79 ± 0.14	0.74 ± 0.17	0.77 ± 0.10	**0.79** ± **0.15**	0.78 ± 0.18	0.77 ± 0.15
HD	18.76 ± 4.43	14.18 ± 3.74	13.55 ± 2.92	14.01 ± 4.22	12.91 ± 3.08	12.93 ± 4.11	15.08 ± 3.00	14.49 ± 4.65	13.03 ± 2.87	12.82 ± 2.90	**12.11** ± **3.04**
	Myocardium										
Dice	0.57 ± 0.21	0.63 ± 0.17	0.64 ± 0.15	0.62 ± 0.19	**0.66** ± **0.11**	0.65 ± 0.13	0.62 ± 0.15	0.61 ± 0.22	0.64 ± 0.18	0.65 ± 0.13	0.64 ± 0.13
HD	26.55 ± 5.21	20.19 ± 4.87	20.48 ± 4.09	19.27 ± 5.11	18.87 ± 4.76	18.89 ± 5.03	21.10 ± 4.37	20.87 ± 4.88	18.29 ± 4.45	**18.11** ± **3.91**	19.05 ± 5.04

compare our approach to using a SE(3) group convolutional neural network as proposed and designed by Liu et al. (2022a). We also finally compare our method to a standard segmentation architecture in the form of the nnUNet which does not use equivariant convolutions or SDG methods. This is to highlight the advantage of equivariant convolutions to improve single domain generalisability compared to standard convolutions. In the prostate dataset, we have T2 weighted and ADC images and therefore use a contrastive method to impose texture invariance as shown in Fig. 1. In the cardiac dataset, we only have T2-weighted imaging and therefore we incorporate a randomised convolutional layer as our initial layer to impose texture invariance.

6.2.2. Group order results

First, we explore the different order dihedral group constraints imposed on our models. According to Theorem 1, the denser the constraint or the higher the group order then the more robust the model. Therefore the D16h group equivariant model should yield the best performance.

In Tables 2 and 3, we show all the Dice scores and maximum Hausdorff distances for each model trained with different order group equivariant constraints in the latent space. As previously mentioned, here group equivariance is learnt with contrastive learning. We note on average the D8h group equivariant model achieved the best Dice and maximum Hausdorff distance score followed by the D12h group equivariant model for both the cardiac and prostate datasets. For example, the models equipped with D8h and D12h group equivariance in the latent space both achieved an average Dice score of 0.68 for the prostate dataset. In Figs. 3 and 4, we note that as we increase the density of the constraint in the latent space (the order of group increases) there is an overall trend upwards in terms of both the Dice score and maximum Hausdorff distance for the prostate and cardiac datasets which supports our Theorem 1. However, the improvement in the metric scores appears to diminish as the order of the group increases. For example, the Dice scores were 0.75, 0.75, 0.75 and 0.74 for the D12, D12h, D16 and D16h groups respectively in the experiments using the cardiac dataset. This result may arise because by learning equivariance to the lower order groups which are subgroups of the higher order groups, one has already sufficiently learnt equivariance to higher order groups. This means group equivariant constraints provide increased robustness up to a certain order.

6.2.3. SDG results

Next, we compare the D8h and D12h equivariant segmentation models trained via contrastive learning to several established SDG methods in the literature. Tables 4 and 5 showcases that our technique which includes models trained with D8h or D12h equivariant constraints in the latent space, outperforms other SDG methods. This is shown in terms of both the metric scores used for the cardiac and prostate datasets. The notably better maximum HD scores from our method suggest its enhanced ability to yield topologically accurate segmentations. Thus, even without resorting to aggressive augmentation tactics or adversarial training, simply embedding equivariant constraints in a segmentation model can lead to equivalent or superior SDG results. Our method which uses D8h equivariance betters the nnUNet by 7 and 3 dice points on average for the prostate and cardiac dataset respectively highlighting the advantage of incorporating compositional shape equivariance in a segmentation model. Additionally, we observe the superior segmentation performance of our shape equivariant approach compared to a purely compositional method for segmentation in the vMFnet for both the prostate and cardiac dataset. Finally, the SE(3) segmentation model as proposed by Li et al. (2018) performance is significantly worse than our D8h equivariant model by 9 and 4 dice point on average for the prostate and cardiac dataset respectively. This is arising because their approach does not include shape compositionality in the latent space like our method. Furthermore, it is possible that SE(3) equivariance overly reduces the expressivity of the segmentation network leading to poor performance. Our method The efficacy of our approach is visually emphasised in Figs. 5 and 6, where the equivariant models seem to generate more anatomically accurate prostate and cardiac segmentation maps compared to 5 selected SDG methods.

6.3. Method 2 experiments

6.3.1. Experimental setup

The baseline model is a 2D UNet. This model consists of 4 levels with each level of the encoder and decoder consisting of a single 2D pre-activation residual block. The pre-activation residual blocks incorporate group normalisation (number of groups = 8) and ReLU activation for the cyclical and dihedral group equivariant convolutional kernel. The kernels with irreducible SO(2) group equivariant constraints use normalisation and activation functions as prescribed in Section 5.2 The first level outputs 32 features (channels) and doubles at each level until we have 256 features in the bottleneck.

We use the same training setup as for the method 1 experiment and compare it to the same SDG methods as in the method 1 experiment. However, here we apply convolutional kernel constraint to the cyclical and SO(2) groups instead of the point groups.

6.3.2. Group order experiments

The Dice and maximum HD scores for the different order group equivariant segmentation models are shown in Tables 6 (prostate dataset) and 7 (cardiac dataset).

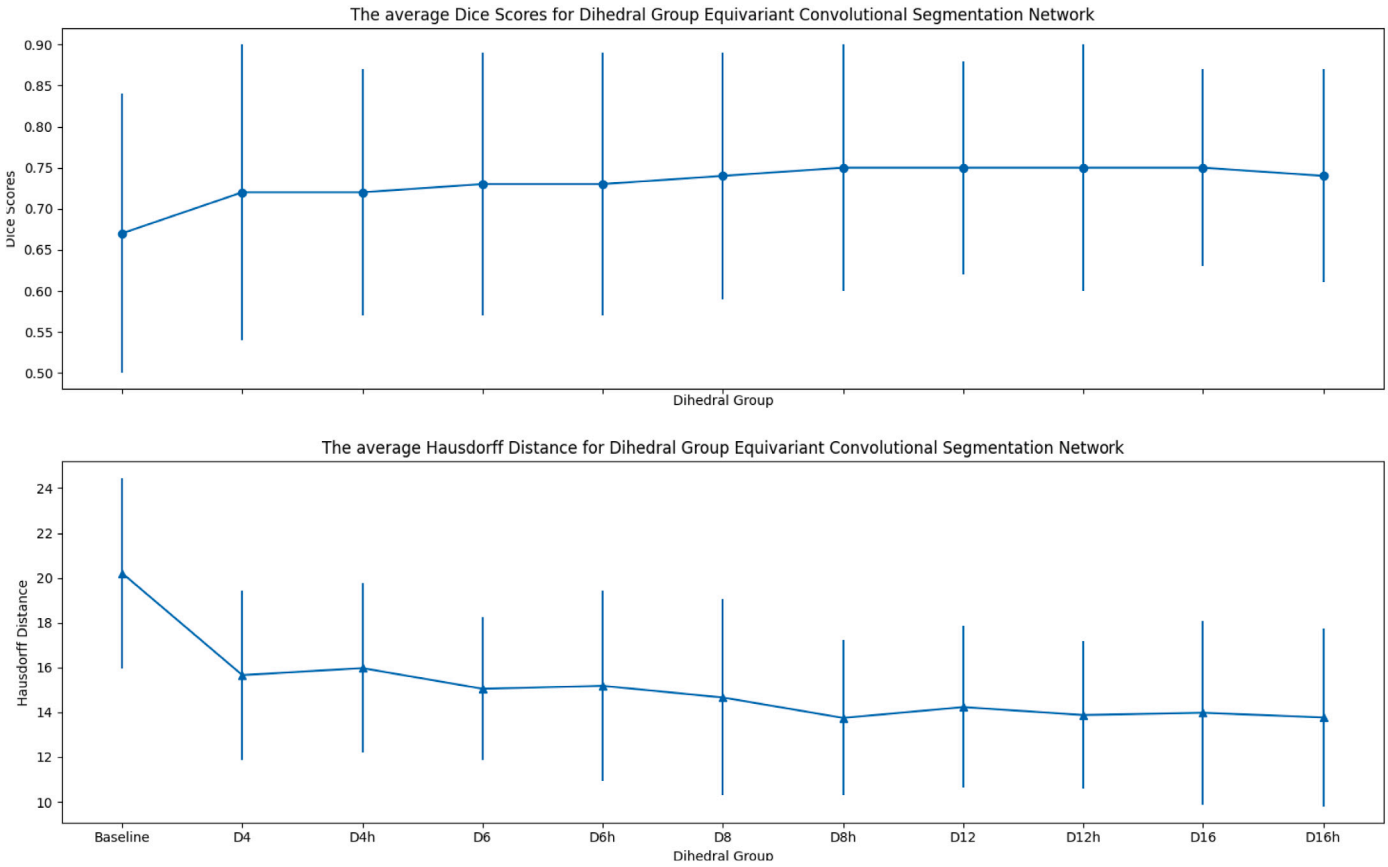

Fig. 4. The average Dice and maximum Hausdorff distance scores ± standard deviation for the respective dihedral group equivariant CNNs trained via contrastive learning and applied to the cardiac dataset. The scores are averaged across all classes.

Table 4
Dice score and Hausdorff distance(HD) ± standard deviation comparing the D8h and D12h group shape equivariant segmentation models to various other SDG methods. This table shows the results for the Prostate dataset.

	Baseline	D8h	D12h	CutOut	BigAug	AdvBias	RandConv	Jigen	3D vMFNet	SE(3)	3D nnUNet
	Transitional Zone										
Dice	0.64 ± 0.13	**0.73 ± 0.13**	0.72 ± 0.14	0.65 ± 0.20	0.72 ± 0.15	0.67 ± 0.17	0.69 ± 0.14	0.67 ± 0.18	0.69 ± 0.13	0.66 ± 0.11	0.67 ± 0.13
HD	24.33 ± 4.12	**18.01 ± 0.08**	18.19 ± 5.00	23.55 ± 4.45	19.81 ± 5.03	21.17 ± 3.98	20.98 ± 4.29	22.98 ± 5.68	18.05 ± 3.29	22.47 ± 4.62	19.93 ± 4.58
	Peripheral Zone										
Dice	0.41 ± 0.09	0.53 ± 0.09	**0.55 ± 0.09**	0.41 ± 0.17	0.53 ± 0.16	0.45 ± 0.14	0.49 ± 0.16	0.41 ± 0.12	0.50 ± 0.10	0.43 ± 0.11	0.46 ± 0.09
HD	56.88 ± 18.72	**28.92 ± 9.92**	29.54 ± 12.83	51.06 ± 20.17	30.94 ± 14.88	44.72 ± 16.83	38.10 ± 12.88	49.87 ± 17.92	32.48 ± 15.73	43.04 ± 15.28	39.82 ± 16.34

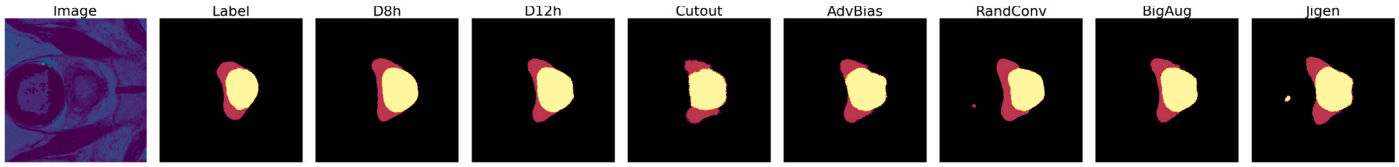

Fig. 5. This figure shows an example prostate segmentation output comparing the D8h and D12h equivariant segmentation model trained using contrastive learning with the 5 other SDG methods.

In this set of experiments, the D16 equivariant segmentation model demonstrated the best segmentation performance for both the cardiac and prostate datasets. This is followed by the C16 equivariant segmentation model. This is noted for all the classes in the cardiac and prostate datasets. Similar to our contrastive approach, we note the segmentation performance improves as we increase the density of the equivariant constraints imposed on the kernel from C4 to C16 (left to right) shown in Fig. 7 for the prostate dataset and Fig. 8 for the cardiac dataset. However, we note a significant drop in performance for the SO(2) group equivariant model for both the prostate (Table 6) and cardiac (Table 7) dataset. This empirically shows the equivariant constraints imposed on a segmentation network to improve robustness have an upper bound

Table 5
Dice score and Hausdorff distance(HD) ± standard deviation comparing the D8h and D12h group shape equivariant segmentation models trained via contrastive learning to various other SDG methods. This table shows the results for the Cardiac dataset.

	Baseline	D8h	D12h	CutOut	BigAug	AdvBias	RandConv	Jigen	3D vMFNet	SE(3)	3D nnUNet
	Left Ventricle										
Dice	0.74 ± 0.13	0.80 ± 0.17	**0.81 ± 0.11**	0.75 ± 0.13	0.78 ± 0.19	0.76 ± 0.11	0.77 ± 0.18	0.75 ± 0.18	0.78 ± 0.14	0.76 ± 0.13	0.77 ± 0.15
HD	15.31 ± 3.1	**9.95 ± 3.08**	10.73 ± 3.07	14.81 ± 4.11	11.11 ± 2.99	12.99 ± 3.88	12.54 ± 3.48	13.99 ± 4.06	11.78 ± 3.92	13.37 ± 4.21	12.33 ± 3.74
	Right Ventricle										
Dice	0.72 ± 0.16	**0.79 ± 0.15**	0.78 ± 0.18	0.72 ± 0.14	0.77 ± 0.13	0.74 ± 0.17	0.75 ± 0.13	0.73 ± 0.18	0.76 ± 0.12	0.73 ± 0.16	0.75 ± 0.11
HD	18.76 ± 4.43	13.03 ± 2.87	**12.82 ± 2.90**	17.05 ± 3.88	14.08 ± 3.34	15.82 ± 4.62	14.19 ± 3.29	16.91 ± 4.50	14.28 ± 4.19	15.43 ± 4.46	14.99 ± 4.72
	Myocardium										
Dice	0.57 ± 0.21	0.64 ± 0.18	**0.65 ± 0.13**	0.59 ± 0.14	0.63 ± 0.18	0.61 ± 0.19	0.61 ± 0.10	0.60 ± 0.19	0.63 ± 0.17	0.61 ± 0.15	0.62 ± 0.16
HD	26.55 ± 5.21	18.29 ± 4.45	**18.11 ± 3.91**	24.88 ± 5.78	20.08 ± 5.19	22.53 ± 4.58	22.08 ± 4.16	23.76 ± 4.74	20.83 ± 4.75	23.95 ± 5.48	21.12 ± 4.68

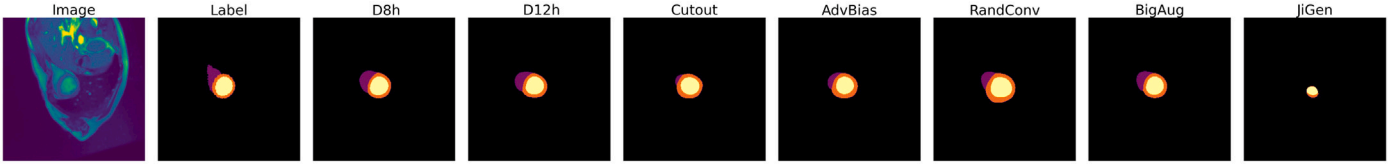

Fig. 6. This figure shows an example cardiac segmentation output comparing the D8h and D12h equivariant segmentation model trained using contrastive learning with the 5 other SDG methods.

Table 6
The average Dice and maximum Hausdorff distance score ± standard deviation for different order dihedral group shape equivariant models trained via the kernel constraint method. This table shows the results for the prostate dataset.

	Baseline	D4	D6	D8	D12	D16	C4	C6	C8	C12	C16	SO(2)
	Transitional Zone											
Dice	0.63 ± 0.16	0.66 ± 0.10	0.67 ± 0.14	0.70 ± 0.08	0.68 ± 0.11	**0.70 ± 0.08**	0.66 ± 0.13	0.65 ± 0.19	0.68 ± 0.11	0.68 ± 0.11	0.69 ± 0.15	0.63 ± 0.18
HD	27.18 ± 5.86	23.88 ± 4.75	21.46 ± 4.29	20.62 ± 4.82	20.71 ± 4.59	**20.18 ± 4.08**	24.47 ± 3.93	22.29 ± 4.23	21.83 ± 4.18	21.14 ± 3.69	21.36 ± 4.33	26.44 ± 5.26
	Peripheral Zone											
Dice	0.38 ± 0.10	0.42 ± 0.17	0.46 ± 0.15	0.44 ± 0.15	0.43 ± 0.08	0.47 ± 0.14	0.43 ± 0.08	0.44 ± 0.11	0.46 ± 0.15	0.48 ± 0.18	**0.49 ± 0.17**	0.37 ± 0.13
HD	45.53 ± 17.98	40.81 ± 16.49	35.39 ± 15.19	33.14 ± 17.82	32.47 ± 16.32	**32.06 ± 13.26**	39.29 ± 18.23	36.38 ± 16.74	33.58 ± 12.62	33.37 ± 16.43	32.99 ± 14.34	44.56 ± 16.85

Table 7
The average Dice and maximum Hausdorff distance score ± standard deviation for different order dihedral group shape equivariant models trained via the kernel constraint method. This table shows the results for the cardiac dataset.

	Baseline	D4	D6	D8	D12	D16	C4	C6	C8	C12	C16	SE2
	Left Ventricle											
Dice	0.71 ± 0.09	0.75 ± 0.13	0.76 ± 0.17	0.76 ± 0.09	0.77 ± 0.12	**0.78 ± 0.09**	0.73 ± 0.12	0.73 ± 0.17	0.76 ± 0.11	0.76 ± 0.10	0.77 ± 0.09	0.73 ± 0.16
HD	17.01 ± 4.09	15.09 ± 3.38	14.38 ± 2.94	12.97 ± 2.12	12.44 ± 1.98	12.03 ± 2.49	14.90 ± 3.15	15.04 ± 3.47	**13.49 ± 3.78**	13.41 ± 2.49	12.73 ± 3.31	15.89 ± 4.10
	Right Ventricle											
Dice	0.70 ± 0.12	0.72 ± 0.15	0.73 ± 0.18	0.75 ± 0.10	0.75 ± 0.13	**0.77 ± 0.12**	0.72 ± 0.11	0.72 ± 0.11	0.74 ± 0.16	0.73 ± 0.16	0.75 ± 0.09	0.70 ± 0.14
HD	19.87 ± 4.63	16.70 ± 4.08	15.90 ± 3.43	15.18 ± 3.89	14.29 ± 3.15	**13.27 ± 3.28**	17.91 ± 4.10	15.98 ± 3.57	15.19 ± 3.29	14.05 ± 3.08	13.37 ± 2.97	18.66 ± 4.19
	Myocardium											
Dice	0.57 ± 0.16	0.61 ± 0.14	0.63 ± 0.16	0.62 ± 0.15	0.63 ± 0.14	**0.65 ± 0.13**	0.59 ± 0.12	0.60 ± 0.19	0.64 ± 0.11	0.62 ± 0.09	0.64 ± 0.13	0.58 ± 0.15
HD	28.16 ± 4.88	23.19 ± 4.87	22.76 ± 4.28	20.81 ± 3.89	20.09 ± 3.39	**19.88 ± 3.27**	24.83 ± 3.86	23.28 ± 3.41	21.18 ± 4.04	21.86 ± 3.38	20.14 ± 4.29	26.29 ± 4.38

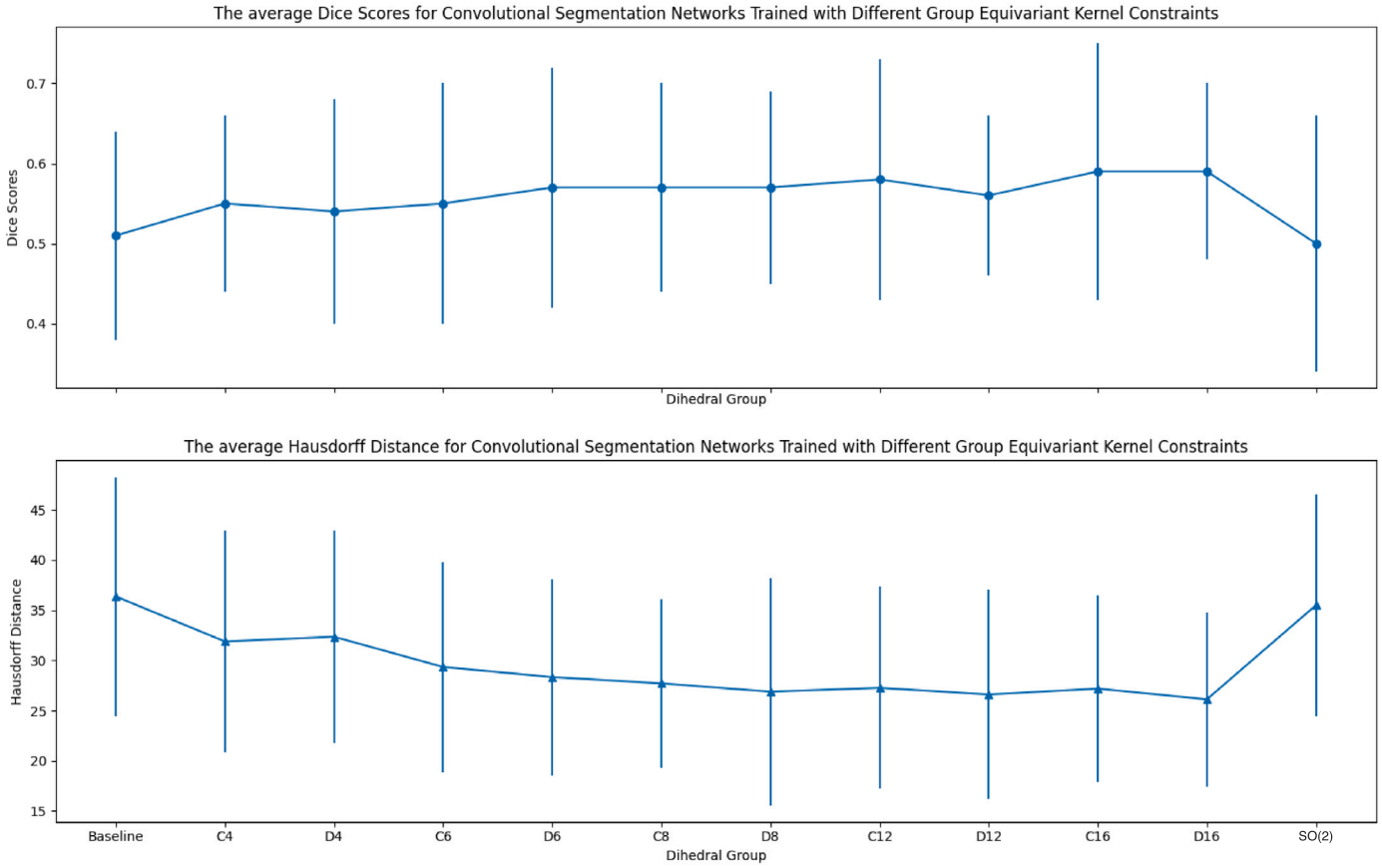

Fig. 7. The average Dice and maximum Hausdorff distance score ± standard deviation for the respective group equivariant CNNs trained via kernel constraints and applied to the prostate dataset. The density of the equivariant constraints imposed on the kernel increases from C4 to SO(2) (left to right). The scores are averaged across all classes.

Table 8
Dice score and Hausdorff distance(HD) ± standard deviation comparing the D16 and C16 group shape equivariant segmentation models trained via the kernel constraint method to various other SDG methods. This table shows the results for the Prostate dataset.

	Baseline	D16	C16	CutOut	BigAug	AdvBias	RandConv	Jigen	2D vMFNet	SE(2)	2D nnUNet
	Transitional Zone										
Dice	0.63 ± 0.10	**0.70 ± 0.08**	0.69 ± 0.11	0.63 ± 0.15	0.68 ± 0.14	0.64 ± 0.10	0.67 ± 0.17	0.63 ± 0.16	0.67 ± 0.10	0.65 ± 0.15	0.66 ± 0.14
HD	27.18 ± 5.86	**20.18 ± 4.08**	21.36 ± 4.33	25.91 ± 5.17	21.28 ± 4.88	23.43 ± 3.77	22.83 ± 4.06	24.90 ± 5.53	22.64 ± 4.11	23.89 ± 5.03	22.91 ± 4.61
	Peripheral Zone										
Dice	0.38 ± 0.10	0.47 ± 0.14	**0.49 ± 0.17**	0.40 ± 0.15	0.46 ± 0.11	0.43 ± 0.18	0.45 ± 0.17	0.41 ± 0.10	0.44 ± 0.17	0.42 ± 0.08	0.45 ± 0.14
HD	45.33 ± 17.98	**32.06 ± 13.26**	32.99 ± 14.34	42.81 ± 12.86	35.38 ± 14.21	37.83 ± 15.05	35.33 ± 15.72	41.74 ± 16.02	37.28 ± 14.78	38.47 ± 14.09	37.48 ± 15.65

which does not fully satisfy Theorem 1. This may arise due to the limited expressiveness of SO(2) equivariant kernels which creates an overly constrained network. We believe expressing the SO(2) group as a direct sum of its irreducible representations up to an order of 3 is likely leading to limited expressivity of the features extracted and therefore reducing performance. Therefore, a method of expressing the SO(2) or SE(2) group which is less restrictive on CNNs could be a way forward to improving performance to level which satisfies Theorem 1.

6.3.3. Comparison with method 1

Interestingly, we demonstrate overall superior performance in our first method where we impose shape equivariance in the latent space via a contrastive loss. For example, the average dice score for the best performing group (D12h) in method 1 was 0.75 and 0.64 for the cardiac and prostate dataset respectively. This is in comparison to the lower average dice scores achieved by the best group in method 2 (D16) which was 0.73 and 0.63 for the cardiac and prostate dataset respectively. The slightly better performance from using a contrastive loss to enforce equivariance as opposed to using group equivariant convolutional kernels may arise because the baseline model in method 1 is a hybrid UNet which demonstrates superior performance to the baseline 2D UNet in method 2. The Dice score averaged across all three classes for the cardiac dataset was 0.67 for the hybrid UNet compared to 0.66 for the 2D UNet. The observed findings may also arise from the stricter constraints imposed in method 2 (see Figs. 9 and 10).

6.3.4. SDG results

Finally, we compare the C16 and D16 equivariant segmentation models trained with kernel constraints to several established SDG methods in the literature. Due to the input being 2D in this set of experiments

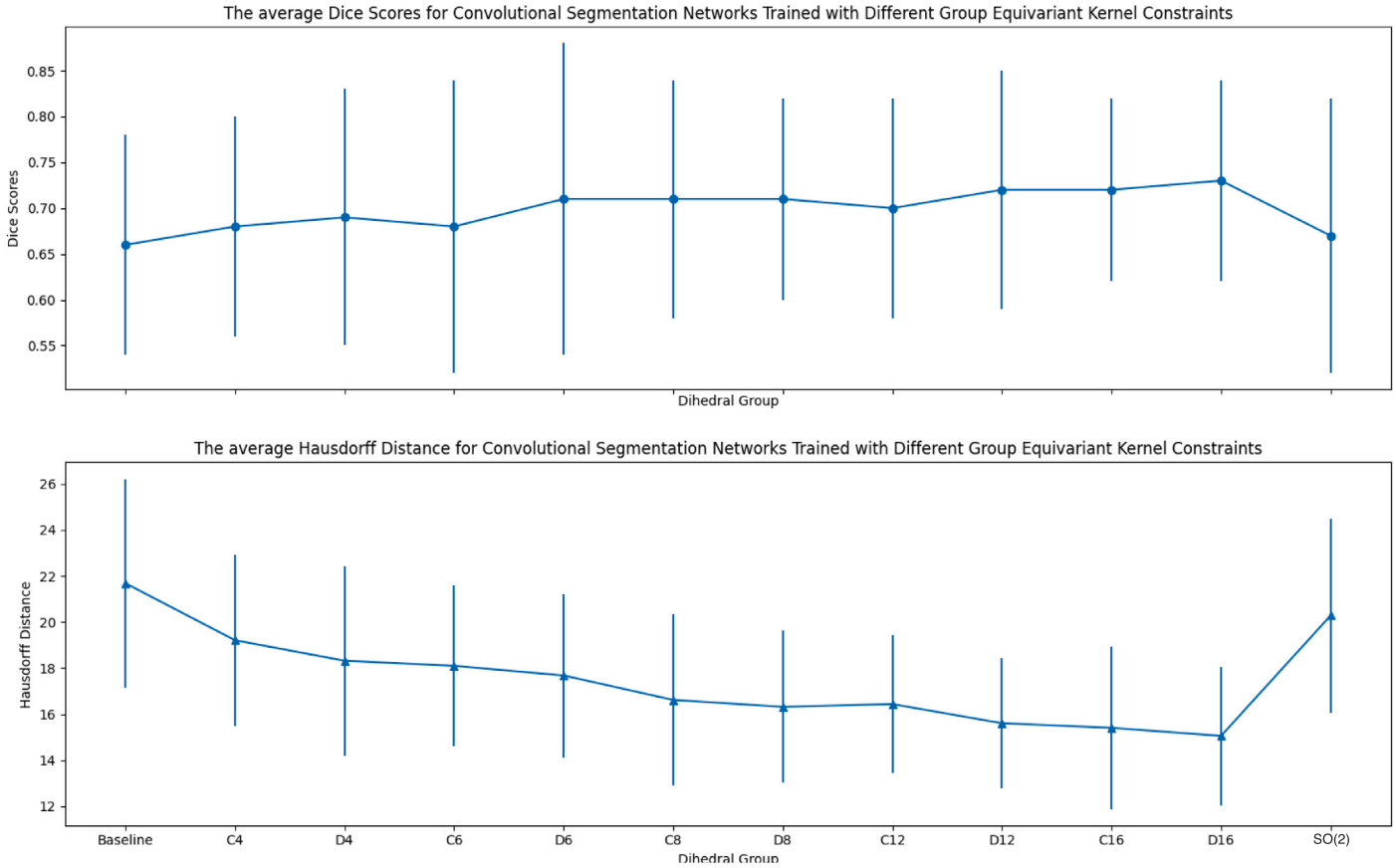

Fig. 8. The average Dice and maximum Hausdorff distance score ± standard deviation for the respective group equivariant CNNS trained via kernel constraints and applied to the cardiac dataset. The density of the equivariant constraints imposed on the kernel increases from C4 to SO(2) (left to right). The scores are averaged across all classes.

Table 9
Dice score and Hausdorff distance(HD) ± standard deviation comparing the D16 and C16 group equivariant segmentation models trained via the kernel constraint method to various other SDG methods. This table shows the results for the Cardiac dataset.

	Baseline	D16	C16	CutOut	BigAug	AdvBias	RandConv	Jigen	2D vMFNet	SE(2)	2D nnUNet
	Left Ventricle										
Dice	0.71 ± 0.09	**0.78 ± 0.09**	0.77 ± 0.09	0.73 ± 0.10	0.77 ± 0.13	0.74 ± 0.08	0.76 ± 0.17	0.73 ± 0.14	0.76 ± 0.10	0.73 ± 0.14	0.75 ± 0.17
HD	17.01 ± 4.09	**12.03 ± 2.49**	12.73 ± 3.31	16.18 ± 4.37	14.06 ± 3.37	15.51 ± 4.19	14.90 ± 3.70	16.20 ± 4.18	13.97 ± 5.02	15.89 ± 5.88	14.21 ± 4.47
	Right Ventricle										
Dice	0.70 ± 0.12	**0.77 ± 0.12**	0.75 ± 0.09	0.70 ± 0.21	0.76 ± 0.10	0.73 ± 0.19	0.75 ± 0.14	0.71 ± 0.11	0.75 ± 0.12	0.72 ± 0.13	0.74 ± 0.15
HD	19.87 ± 4.63	**13.27 ± 3.28**	13.37 ± 2.97	18.84 ± 5.39	15.03 ± 4.17	17.14 ± 3.88	16.48 ± 4.35	17.99 ± 5.38	15.78 ± 4.04	17.02 ± 3.89	16.85 ± 5.03
	Myocardium										
Dice	0.57 ± 0.16	**0.65 ± 0.13**	0.64 ± 0.13	0.59 ± 0.18	0.63 ± 0.20	0.60 ± 0.21	0.61 ± 0.16	0.58 ± 0.14	0.63 ± 0.17	0.60 ± 0.13	0.62 ± 0.11
HD	28.16 ± 4.88	**19.88 ± 3.27**	20.14 ± 4.29	26.09 ± 5.98	22.38 ± 4.45	24.56 ± 5.06	23.70 ± 4.47	25.90 ± 5.86	22.04 ± 4.66	24.29 ± 5.01	22.97 ± 4.51

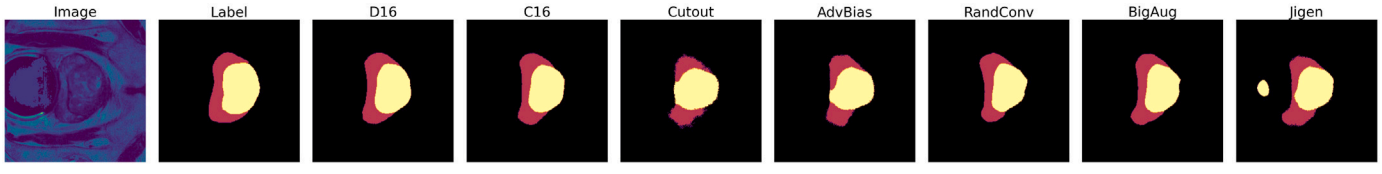

Fig. 9. This figure shows an example prostate segmentation output comparing the C16 and D16 shape equivariant segmentation model trained via the kernel constraint method with 5 other SDG methods.

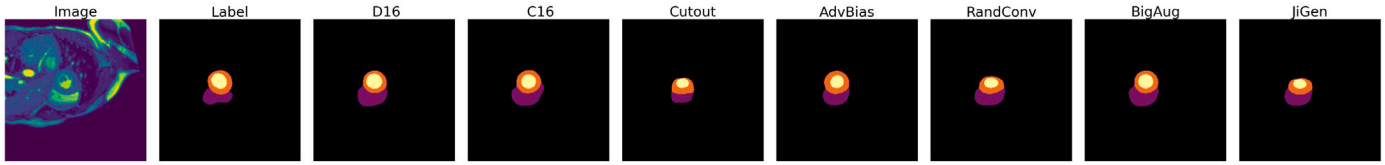

Fig. 10. This figure shows an example cardiac segmentation output comparing the C16 and D16 shape equivariant segmentation model trained via the kernel constraint method with 5 other SDG methods.

Table 10
Time in seconds to train and test a single image for the 3D prostate and cardiac dataset. We compare our contrastive approach (method 1) to various other SDG methods applied the hybrid UNet.

	Baseline	Method 1(D12h)	CutOut	BigAug	AdvBias	RandConv	Jigen	3D vMFNet	SE(3)	3D nnUNet
	Training									
Cardiac	**1.23**	2.06	1.94	2.85	2.11	1.84	3.59	1.95	3.03	2.33
Prostate	**1.09**	2.19	1.74	2.99	1.96	1.77	3.47	1.92	2.95	2.35
	Testing									
Cardiac	0.74	0.82	0.74	0.74	0.78	0.80	0.74	0.80	2.76	1.04
Prostate	0.70	0.80	0.70	0.70	0.74	0.76	0.74	0.81	2.91	2.17

Table 11
Time in seconds to train and test a single image for the 3D prostate and cardiac dataset. We compare our contrastive approach (method 2) to various other SDG methods applied the 2D UNet.

	Baseline	Method 2(D16)	CutOut	BigAug	AdvBias	RandConv	Jigen	2D vMFNet	SE(2)	2D nnUNet
	Training									
Cardiac	**0.53**	1.38	1.01	2.07	1.18	1.06	2.53	1.01	2.19	1.75
Prostate	**0.60**	1.88	1.19	2.12	1.31	1.17	2.50	1.15	2.04	1.72
	Testing									
Cardiac	0.52	1.15	0.52	0.52	0.66	0.58	0.52	0.62	1.54	0.98
Prostate	0.51	1.02	0.51	0.51	0.57	0.57	0.51	0.69	1.38	0.97

we modify the comparison methods to 2D input. Tables 8 and 9 highlight that group equivariant constraints imposed in the kernels of a CNN-based segmentation model, perform competitively with other SDG methods for both the cardiac and prostate datasets. Specifically, we demonstrate that our approach outperforms other SDG methods including aggressive augmentation and purely compositional representation learning strategies in terms of the maximum HD score which was also shown in our contrastive equivariant method. Similar to our contrastive method, the 2D nnUNet on average achieve 3 dice points less than our D16 shape equivariant approach for both the cardiac and prostate dataset. The SE(2) equivariant segmentation network adapted from Li et al. (2018) achieves on average 5 dice points less than our D16 model for the prostate and cardiac dataset. Similar to method 1, the advantage of our method is arising from imposing shape compositionality in the latent space. Once again, our findings are visually emphasised with a prostate segmentation example in Fig. 9 and a cardiac segmentation example in Fig. 10.

6.3.5. Computational speed

In this set of experiment we compare the speed of processing a single training and test image with the other SDG methods. Specifically, we compare computational speed of method 1 applied to the baseline hybrid UNet for 3D input with the SDG methods adapted for 3D input. We then compare method 2 applied to the baseline 2D UNet for 2D input with the SDG methods adapted for 2D input. The average training times denote the duration it took for each method to converge in segmenting the entire training set divided by the product of the number of training examples and training epochs. Conversely, test times indicate the average speed at which a fully-trained model can segment a single image.

In Table 10, we note a significant increase in training time of method 1 compared to the baseline method for both the cardiac (2.06 s vs 1.23 s) and prostate (2.19 s vs 1.09 s) datasets. This is arising from having to compute the contrastive loss and quantise the latent space.

Similarly we note a significant increase in training time of method 2 compared the baseline method for both the cardiac (1.38 s vs 0.53 s) and prostate (1.88 s vs 0.60 s) datasets. The increased compute time for the prostate dataset in Table 11 comes from have to pass the T2 weighted image and ADC image through the encoder separately. This is less computationally expensive than incorporating randomised convolutional filters. This is also demonstrated by faster training times of RandConv (Xu et al., 2020) compared to our method in Tables 10 and 11.

We note in Table 10 that Jigen (Carlucci et al., 2019) has the largest training time for both the cardiac (2.85 s) and prostate (2.99 s) dataset due to the computations required for prior self supervised training which is included in the training time. A similar trend is also noted in Table 11 when we compare to method 2 which uses group equivariant convolutional kernels. We note similar extensive training times for BigAug (Zhang et al., 2020) for both 3D and 2D data in Tables 10 and 11 respectively which arises for their extensive augmentation process. However, at test time, augmentation and self-supervised techniques demonstrates similar compute time to the baseline model as they have the same architecture and augmentation and self-supervision is not used at test time.

Our method 1 10 and is faster than self-supervised and augmentation based techniques during training time but only slightly slower at test time. We also note our contrastive methods in Table 10 is only slightly slower than the baseline model at test time for both prostate (0.70 s vs 0.80 s) and cardiac (0.74 s vs 0.82 s) dataset as we do not need to require to compute the contrastive loss with a second transformed image. Our second method however is still slower than the baseline as well as the self-supervised and augmentation based

Table 12
The average Dice and maximum Hausdorff distance scores ± standard deviation demonstrating each component of our method. This table shows the results for the cardiac dataset.

	Baseline	Discrete Representation	Texture Invariant Discrete Representation	D16 Equivariant/ Texture Invariant Discrete Representation (Contrastive)	D16 Equivariant/Texture Invariant Discrete Representation (Kernel Constraint)
	Left Ventricle				
Dice	0.71 ± 0.09	0.73 ± 0.14	0.75 ± 0.18	**0.78 ± 0.11**	**0.78 ± 0.08**
HD	17.01 ± 4.09	16.00 ± 3.45	14.81 ± 3.19	12.98 ± 2.92	**12.03 ± 2.49**
	Right Ventricle				
Dice	0.70 ± 0.12	0.72 ± 0.18	0.75 ± 0.06	0.77 ± 0.10	**0.78 ± 0.09**
HD	19.87 ± 4.63	16.18 ± 4.04	15.49 ± 3.31	14.02 ± 4.00	**13.27 ± 3.28**
	Myocardium				
Dice	0.57 ± 0.16	0.60 ± 0.10	0.62 ± 0.17	**0.66 ± 0.16**	0.65 ± 0.08
HD	28.16 ± 4.88	23.21 ± 5.02	21.44 ± 4.36	**19.05 ± 4.04**	19.88 ± 3.84

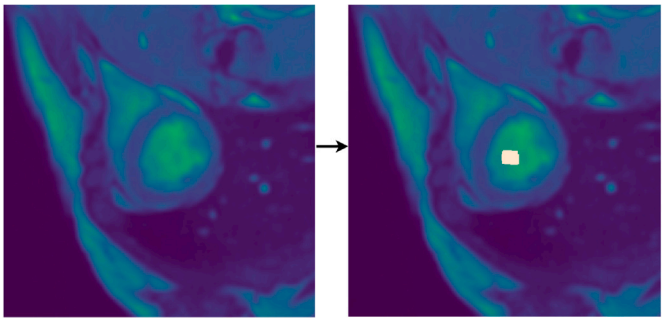

Fig. 11. This figure shows a cardiac shape component (highlighted in beige) in the shape dictionary equivariant to the D16 group trained with method 2. It appears to corresponding to shape filling the central aspect of the left ventricle.

techniques at test time shown in Table 11 due to the extra computation required to enforce group equivariant convolutional kernel which increases with the order of the group. In Tables 10 and 11 we note our method is faster than the SE(3) equivariant segmentation model (Liu et al., 2022a) and nnUNet at training and test time due to fewer computations. However, the vMFNet (Liu et al., 2022b) architecture demonstrates faster training and test times compared to both our methods as shown in Tables 10 and 11. This is because vMFNet do not use extra computation to enforce an equivariant latent space which is texture invariant.

6.3.6. Shape dictionary visualisation

The advantage of constraining the latent space to a finite set of shape components is that one can also visualised each of these shape components. We do this by intervening on the shape component which is sampled from the shape equivariant dictionary by setting this to 0. Note, often the same shape components is sampled multiple times to replace a subset of the features which are all set to 0 in the intervention. We then pass the quantised features with a subset of the features corresponding to the interested shape component set to 0 through the decoder. We next pass the quantised features without intervention through the decoder. The output of the decoder with the intervened quantised features is taken away from the output of the decoder without intervention to visualised the interested shape component. We provide an example in Fig. 11 to visualise a shape component in the shape dictionary equivariant to the D16 group trained with method 2. Here, we see an ill defined shape in the central aspect of the left ventricle corresponding to element 24 in the shape dictionary. Note the compactness of the component which is one single connected component demonstrating it indeed represents a shape.

6.3.7. Ablation experiments

We carried out ablation experiments to emphasise the importance of each component of our method. We specifically show the value of enforcing a discrete latent space, followed by a texture invariant discrete latent space and finally a discrete latent space which is both texture invariant and shape equivariant. Ablations are carried out for the cardiac dataset. The baseline model here is a simple 2D UNet as previously described. We enforce equivariance to the D16 group when adding shape equivariance to our model. This is applied with either a contrastive method or group equivariant convolutional kernels. Texture invariance in enforced with an initial randomised convolutional layer.

In Table 12 we show that each component of our method appears to incrementally improve segmentation performance in terms of the Dice score and maximum Hausdorff distance for all three classes. Specifically, we note the improvement in segmentation performance by first creating a discrete latent space using vector quantisation followed by additionally enforcing a texture invariant discrete latent space which creates a finite shape dictionary. Next, we additionally enforce the shape dictionary to be equivariant to the D16 group via our contrastive method or by using D16 equivariant convolutional kernels shown in the last two columns of Table 12 respectively. We show here the further improved segmentation performance by introducing equivariance of the shape dictionary.

We next perform ablations to demonstrate the effect of codebook size on segmentation performance. We use this to determine the minimum of codebook vectors required without diminishing segmentation performance. The setup is similar to the previous ablation experiments where the baseline model here is a simple 2D UNet. We again enforce equivariance to the D16 group when adding shape equivariance to our model. This is applied with either a contrastive method or group equivariant convolutional kernels. Texture invariance in enforced with an initial randomised convolutional layer. We average the dice scores across all classes.

In Table 13 we show the minimum number of codebook vectors required before the segmentation performance diminishes by more than 2 dice points compared to the largest codebook (1024) is 128 for both cardiac and prostate dataset. This applies to using a contrastive or kernel constraint method to impose equivariance in the discrete latent space. The segmentations performance for all experiments do not appear be significantly affected by the a codebook size between 128 and 1024 codebook vectors.

7. Discussion

7.1. Limitations

A limitation of this study is that components in the dictionary are sampled independently which can lead to anatomically and topologically inaccurate segmentations. In future work, we will therefore explore methods to sample the shape components such that topology is

Table 13
The average Dice scores ± standard deviation for different sized cookbooks used for our method.

Codebook Size	D16 Equivariant/Texture Invariant Discrete Representation (Contrastive)		D16 Equivariant/Texture Invariant Discrete Representation (Kernel Constraint)	
	Prostate	Cardiac	Prostate	Cardiac
64	0.61 ± 0.12	0.72 ± 0.10	0.56 ± 0.14	0.72 ± 0.14
128	0.63 ± 0.10	0.74 ± 0.08	0.59 ± 0.11	0.74 ± 0.12
256	0.63 ± 0.14	0.75 ± 0.12	0.58 ± 0.13	0.74 ± 0.14
512	0.63 ± 0.10	0.74 ± 0.14	0.59 ± 0.14	0.74 ± 0.10
1024	0.64 ± 0.13	0.76 ± 0.11	0.59 ± 0.11	0.75 ± 0.11

preserved using persistent homology. A further limitation we showed is the extra compute time required to enforce both a shape equivariant and texture invariant latent space and there other methods in the literature which are faster during training and test time for improved domain generalisation.

A further limitation of this study is that we only focus on rotation equivariance of the shape components. Scale equivariance of the shape components would further serve as a valuable method to improve robustness of deep learning based segmentation models and will form the basis for future work. This is because different resolutions in the acquisitions can significantly impact the appearance and segmentation of anatomical structures, and a model that does not account for these variations might fail to generalise across different clinical settings.

In this work, we only considered robustness in the task of single domain generalisation. In future work, we aim to explore well defined synthetic corruption such as Gaussian noise in order to provide quantifiable measure of robustness in the form of the relative corruption error.

Our contrastive method only uses positive pairs to impose texture invariance which is a limitation of this method. This could be improved in future work by considering negative samples to construct a triplet contrastive loss which will enforce stronger texture invariance.

Finally, in this work, we have validated our methods on two MRI datasets and in future work we will aim to test our methods on a more diverse range of 3D datasets such as brain MRI and abdominal CT

7.2. Conclusion

In this work, we propose to constrain the latent space for any segmentation model to a finite number of shape equivariant components to improve robustness. We impose these constraints either via a contrastive approach where equivariance is learnt or by imposing fixed constraints on the convolutional kernels themselves. We theorise and prove that the denser the equivariant constraints imposed either in the latent space or the kernels, the more robust the segmentation model. We demonstrate this to be true by improving the domain generalisability of a segmentation model trained on either the cardiac or prostate dataset. We outperform the baseline models by imposing shape equivariance in the model and perform competitively with other SDG methods in the literature.

We anticipate that, over time, equivariant CNNs will be the go-to option for endeavours such as biomedical imaging, where natural symmetries exist. There remains a need for future studies to delve deeper into the myriad of design possibilities for steerable CNNs and further improve performance.

CRediT authorship contribution statement

Ainkaran Santhirasekaram: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mathias Winkler:** Supervision, Funding acquisition. **Andrea Rockall:** Supervision, Funding acquisition. **Ben Glocker:** Writing – review & editing, Writing – original draft, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ainkaran Santhirasekaram reports financial support was provided by Imperial College London. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available freely online references in the manuscript. Code will be released upon acceptance.

Acknowledgements

This work was supported and funded by Cancer Research UK (CRUK) (C309/A28804).

References

Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al., 2022. The medical segmentation decathlon. Nat. Commun. 13 (1), 4128.

Bekkers, E.J., 2019. B-spline cnns on lie groups. arXiv preprint arXiv:1909.12057.

Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A., Pluim, J.P., Duits, R., 2018. Roto-translation covariant convolutional networks for medical image analysis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. Springer, pp. 440–448.

Bie, X., Chen, D., Cun, X., Xi, S., 2022. Learning discrete representation with optimal transport quantized autoencoders.

Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K., 2015a. Cancer imaging archive wiki. http://dx.doi.org/10.7937/K9/TCIA.2015.zF0vlOPv.

Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K., 2015b. NCI-ISBI 2013 challenge: automated segmentation of prostate structures. Cancer Imaging Arch. 370, 6.

Campello, V.M., Gkontra, P., Izquierdo, C., Martin-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., et al., 2021. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. IEEE Trans. Med. Imaging 40 (12), 3543–3554.

Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T., 2019. Domain generalization by solving jigsaw puzzles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2229–2238.

Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., Tarroni, G., Bai, W., Rueckert, D., 2020. Realistic adversarial data augmentation for MR image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 667–677.

Chidester, B., Ton, T.-V., Tran, M.-T., Ma, J., Do, M.N., 2019. Enhanced rotation-equivariant u-net for nuclear segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Dig. Imaging 26, 1045–1057.

Cohen, T.S., Geiger, M., Weiler, M., 2018. Intertwiners between induced representations (with applications to the theory of equivariant neural networks). arXiv preprint arXiv:1803.10743.

Cohen, T., Welling, M., 2016a. Group equivariant convolutional networks. In: International Conference on Machine Learning. PMLR, pp. 2990–2999.

Cohen, T.S., Welling, M., 2016b. Steerable cnns. arXiv preprint arXiv:1612.08498.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
DeVries, T., Taylor, G.W., 2017. Improved regularization of convolutional neural networks with cutout. arxiv 2017. arXiv preprint arXiv:1708.04552.
Dieleman, S., De Fauw, J., Kavukcuoglu, K., 2016. Exploiting cyclic symmetry in convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 1889–1898.
Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W., 2018. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231.
Hoogeboom, E., Peters, J.W., Cohen, T.S., Welling, M., 2018. Hexaconv. arXiv preprint arXiv:1803.02108.
Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510.
Hy, T.S., Trivedi, S., Pan, H., Anderson, B.M., Kondor, R., 2018. Predicting molecular properties with covariant compositional networks. J. Chem. Phys. 148 (24).
Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., Madry, A., 2019. Adversarial examples are not bugs, they are features. Adv. Neural Inf. Process. Syst. 32.
Islam, M.A., Kowal, M., Esser, P., Jia, S., Ommer, B., Derpanis, K.G., Bruce, N., 2021. Shape or texture: Understanding discriminative features in cnns. arXiv preprint arXiv:2101.11604.
Kim, M., Byun, H., 2020. Learning texture invariant representation for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12975–12984.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kondor, R., Son, H.T., Pan, H., Anderson, B., Trivedi, S., 2018. Covariant compositional networks for learning graphs. arXiv preprint arXiv:1801.02144.
Lang, L., Weiler, M., 2020. A wigner-eckart theorem for group equivariant convolution kernels. arXiv preprint arXiv:2010.10952.
LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. Handbook Brain Theory Neural Netw. 3361 (10), 1995.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Li, Q., Shen, L., Guo, S., Lai, Z., 2020a. Wavelet integrated CNNs for noise-robust image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7245–7254.
Li, L., Wang, K., Li, S., Feng, X., Zhang, L., 2020b. Lst-net: Learning a convolutional neural network with a learnable sparse transform. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16. Springer, pp. 562–579.
Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T., 2018. Visualizing the loss landscape of neural nets. Adv. Neural Inf. Process. Syst. 31.
Linmans, J., Winkens, J., Veeling, B.S., Cohen, T.S., Welling, M., 2018. Sample efficient semantic segmentation using rotation equivariant convolutional networks. arXiv preprint arXiv:1807.00583.
Liu, R., Lauze, F., Bekkers, E., Erleben, K., Darkner, S., 2022a. Group convolutional neural networks for DWI segmentation. In: Geometric Deep Learning in Medical Image Analysis. PMLR, pp. 96–106.
Liu, X., Thermos, S., Sanchez, P., O'Neil, A.Q., Tsaftaris, S.A., 2022b. vMFNet: Compositionality meets domain-generalised segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VII. Springer, pp. 704–714.
Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A., 2017. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083.
Mao, C., Zhang, L., Joshi, A., Yang, J., Wang, H., Vondrick, C., 2022. Robust perception through equivariance. arXiv preprint arXiv:2212.06079.
Marcos, D., Volpi, M., Komodakis, N., Tuia, D., 2017. Rotation equivariant vector field networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5048–5057.
Müller, P., Golkov, V., Tomassini, V., Cremers, D., 2021. Rotation-equivariant deep learning for diffusion MRI. arXiv preprint arXiv:2102.06942.
Pang, S., Du, A., Orgun, M.A., Wang, Y., Sheng, Q.Z., Wang, S., Huang, X., Yu, Z., 2022. Beyond CNNs: exploiting further inherent symmetries in medical image segmentation. IEEE Trans. Cybern..
Qiao, F., Zhao, L., Peng, X., 2020. Learning to learn single domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12556–12565.
Ravanbakhsh, S., Schneider, J., Poczos, B., 2017. Equivariance through parameter-sharing. In: International Conference on Machine Learning. PMLR, pp. 2892–2901.
Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.
Van Den Oord, A., Vinyals, O., et al., 2017. Neural discrete representation learning. Adv. Neural Inf. Process. Syst. 30.
Weiler, M., Cesa, G., 2019. General e (2)-equivariant steerable cnns. Adv. Neural Inf. Process. Syst. 32.
Weiler, M., Hamprecht, F.A., Storath, M., 2018. Learning steerable filters for rotation equivariant cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 849–858.
Winkels, M., Cohen, T.S., 2018. 3D G-CNNs for pulmonary nodule detection. arXiv preprint arXiv:1804.04656.
Worrall, D.E., Garbin, S.J., Turmukhambetov, D., Brostow, G.J., 2017. Harmonic networks: Deep translation and rotation equivariance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5028–5037.
Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M., 2020. Robust and generalizable visual representation learning via random convolutions. arXiv preprint arXiv:2007.13003.
Zhang, R., 2019. Making convolutional networks shift-invariant again. In: International Conference on Machine Learning. PMLR, pp. 7324–7334.
Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2017. Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.
Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al., 2020. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Trans. Med. Imaging 39 (7), 2531–2540.