Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **503,** 2369–2379 (2021)
Advance Access publication 2021 February 26

doi:10.1093/mnras/stab530

Fanaroff–Riley classification of radio galaxies using group-equivariant convolutional neural networks

Anna M. M. Scaife[1,2]⋆ and Fiona Porter[1]

[1] *Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL UK*
[2] *The Alan Turing Institute, Euston Road, London NW1 2DB, UK*

Accepted 2021 February 16. Received 2021 February 16; in original form 2021 January 8

ABSTRACT

Weight sharing in convolutional neural networks (CNNs) ensures that their feature maps will be translation-equivariant. However, although conventional convolutions are equivariant to translation, they are not equivariant to other isometries of the input image data, such as rotation and reflection. For the classification of astronomical objects such as radio galaxies, which are expected statistically to be globally orientation invariant, this lack of dihedral equivariance means that a conventional CNN must learn explicitly to classify all rotated versions of a particular type of object individually. In this work we present the first application of group-equivariant convolutional neural networks to radio galaxy classification and explore their potential for reducing intra-class variability by preserving equivariance for the Euclidean group E(2), containing translations, rotations, and reflections. For the radio galaxy classification problem considered here, we find that classification performance is modestly improved by the use of both cyclic and dihedral models without additional hyper-parameter tuning, and that a D_{16} equivariant model provides the best test performance. We use the Monte Carlo Dropout method as a Bayesian approximation to recover epistemic uncertainty as a function of image orientation and show that E(2)-equivariant models are able to reduce variations in model confidence as a function of rotation.

Key words: methods: data analysis – techniques: image processing – radio continuum: galaxies.

1 INTRODUCTION

In radio astronomy, a massive increase in data volume is currently driving the increased adoption of machine learning methodologies and automation during data processing and analysis. This is largely due to the high data rates being generated by new facilities such as the Low-Frequency Array (LOFAR; Van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Beardsley et al. 2019), the MeerKAT telescope (Jarvis et al. 2016), and the Australian SKA Pathfinder (ASKAP) telescope (Johnston et al. 2008). For these instruments a natural solution has been to automate the data processing stages as much as possible, including classification of sources.

With the advent of such huge surveys, new automated classification algorithms have been developed to replace the *'by eye'* classification methods used in earlier work. In radio astronomy, morphological classification using convolutional neural networks (CNNs) and deep learning is becoming increasingly common for object classification, in particular with respect to the classification of radio galaxies. The ground work in this field was done by Aniyan & Thorat (2017) who made use of CNNs for the classification of Fanaroff–Riley (FR) type I and type II radio galaxies (Fanaroff & Riley 1974). This was followed by other works involving the use of deep learning in source classification. Examples include Lukic et al. (2018) who made use of CNNs for the classification of compact and extended radio sources from the Radio Galaxy Zoo catalogue (Banfield et al. 2015), the CLARAN (Classifying Radio Sources Automatically with a Neural Network; Wu et al. 2018) model made use of the Faster R-CNN (Ren et al. 2015) network to identify and classify radio sources; Alger et al. (2018) made use of an ensemble of classifiers including CNNs to perform host galaxy cross-identification. Tang, Scaife & Leahy (2019) made use of transfer learning with CNNs to perform cross-survey classification, while Gheller, Vazza & Bonafede (2018) made use of deep learning for the detection of cosmological diffuse radio sources. Lukic et al. (2018) also performed morphological classification using a novel technique known as capsule networks (Sabour, Frosst & E Hinton 2017), although they found no specific advantage compared to traditional CNNs. Bowles et al. (2021) showed that an attention-gated CNN could be used to perform Fanaroff–Riley classification of radio galaxies with equivalent performance to other applications in the literature, but using ∼50 per cent fewer learnable parameters than the next smallest classical CNN in the field.

Convolutional neural networks classify images by learning the weights of convolutional kernels via a training process and using those learned kernels to extract a hierarchical set of feature maps from input data samples. Convolutional weight sharing makes CNNs more efficient than multilayer perceptrons (MLPs) as it ensures translation-equivariant feature extraction i.e. a translated input signal results in a corresponding translation of the feature maps. However, although conventional convolutions are equivariant to translation, they are not equivariant to other isometries of the input data, such as rotation i.e. rotating an image and then convolving with a fixed filter is not the same as first convolving and then rotating the

⋆ E-mail: anna.scaife@manchester.ac.uk

result. Although many CNN training implementations use rotation as a form of data augmentation, this lack of rotational equivariance means that a conventional CNN must explicitly learn to classify all rotational augmentations of each image individually. This can result in CNNs learning multiple copies of the same kernel but in different orientations, an effect that is particularly notable when the data itself possesses rotational symmetry (Dieleman, De Fauw & Kavukcuoglu 2016). Furthermore, while data augmentation that mimicks a form of equivariance, such as image rotation, can result in a network learning approximate equivariance if it has sufficient capacity, it is not guaranteed that invariance learned on a training set will generalize equally well to a test set (Lenc & Vedaldi 2014). A variety of different equivariant networks have been developed to address this issue, each guaranteeing a particular transformation equivariance between the input data and associated feature maps. For example, in the field of galaxy classification using optical data, Dieleman, Willett & Dambre (2015) enforced discrete rotational invariance through the use of a multibranch network that concatenated the output features from multiple convolutional branches, each using a rotated version of the same data sample as its input. However, while effective, the approach of Dieleman et al. (2015) requires the convolutional layers of a network architecture and hence the number of model weights associated with them to be replicated N times, where N is the number of discrete rotations.

Recently, a more efficient method of using convolutional layers that are equivariant to a particular group of transforms has been developed, which requires no replication of architecture and hence fewer learnable parameters to be used. Explicitly enforcing an equivariance in the network model in this way not only provides a guarantee that it will generalize, but also prevents the network using parameter capacity to learn characteristic behaviour that can instead be specified *a priori*. First introduced by Cohen & Welling (2016), these Group equivariant Convolutional Neural Networks (G-CNNs), which preserve group equivariance through their convolutional layers, are a natural extension of conventional CNNs that ensure translational invariance through weight sharing. Group equivariance has also been demonstrated to improve generalization and increase performance (see e.g. Weiler, Hamprecht & Storath 2017; Weiler & Cesa 2019). In particular, *steerable* G-CNNs have become an increasingly important solution to this problem and notably those steerable CNNs that describe E(2)-equivariant convolutions.

The Euclidean group E(2) is the group of isometries of the plane $\mathbb{R}^2$ that contains translations, rotations, and reflections. Isometries such as these are important for general image classification using convolution as the target object in question is unlikely to appear at a fixed position and orientation in every test image. Such variations are not only highly significant for objects/images that have a preferred orientation, such as text or faces, but are also important for low-level features in nominally orientation-unbiased targets such as astrophysical objects. In principle, E(2)-equivariant CNNs will generalize over rotationally-transformed images by design that reduces the amount of intra-class variability that they have to learn. In effect such networks are insensitive to rotational or reflection variations and therefore learn only features that are independent of these properties.

In this work we introduce the use of G-steerable CNNs to astronomical classification. The structure of the paper is as follows: in Section 2 we describe the mathematical operation of G-steerable CNNs and define the specific Euclidean subgroups being considered in this work; in Section 3 we describe the data sets used in this work and the preprocessing steps implemented on those data; in Section 4 we describe the network architecture adopted in this work, explain how the G-steerable implementation is constructed, and specify the group representations; in Section 5 we give an overview of the training outcomes including a discussion of the convergence for different equivalence groups, validation, and test performance metrics, and introduce a novel use of the Monte Carlo Dropout method for quantitatively assessing the degree of model confidence in a test prediction as a function of image orientation; in Section 6 we discuss the validity of the assumptions that radio galaxy populations are expected to be statistically rotation and reflection unbiased and review the implications of this work in that context; in Section 7 we draw our conclusions.

2 E(2)-EQUIVARIANT G-STEERABLE CNNS

Group CNNs define feature spaces using feature fields $f : \mathbb{R}^2 \rightarrow \mathbb{R}^c$, which associate a c-dimensional feature vector $f(x) \in \mathbb{R}^c$ to each point x of an input space. Unlike conventional CNNs, the feature fields of such networks contain transformations that preserve the transformation law of a particular group or subgroup, which allows them to encode orientation information. This means that if one transforms the input data, x, by some transformation action, g, (translation, rotation, etc.) and passes it through a trained layer of the network, then the output from that layer, $\Phi(x)$, must be equivalent to having passed the data through the layer and then transformed it i.e.

$$\Phi(\mathcal{T}_g x) = \mathcal{T}'_g \Phi(x), \tag{1}$$

where $\mathcal{T}_g$ is the transformation for action g. In the case where the transformation is *invariant* rather than *equivariant* i.e. the input does not change at all when it is transformed, $\mathcal{T}'_g$ will be the identity matrix for all actions $g \in G$. In the case of equivariance, $\mathcal{T}_g$ does not necessarily need to be equal to $\mathcal{T}'_g$ and instead must only fulfil the property that it is a linear representation of G i.e. $\mathcal{T}(gh) = \mathcal{T}(g)\mathcal{T}(h)$.

Cohen & Welling (2016) demonstrated that the conventional convolution operation in a network can be re-written as a group convolution

$$[f * \phi](g) = \sum_{h \in X} \sum_{k} f_k(h) \phi_k(g^{-1}h), \tag{2}$$

where $X = \mathbb{R}^2$ in layer one and $X = G$ in all subsequent layers. Whilst this operation is translationally-equivariant, ϕ is still rotationally constrained. For E(2)-equivariance to hold more generally, the kernel itself must satisfy

$$\phi(gx) = \rho_{\rm out}(g)\phi(x)\rho_{\rm in}(g^{-1}) \;\; \forall \, g \in G, \; x \in \mathbb{R}^2, \tag{3}$$

(Weiler et al. 2018), where g is an action from group G, and $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^{c_{\rm in} \times c_{\rm out}}$, where $c_{\rm in}$ and $c_{\rm out}$ are the number of channels in the input and output data, respectively; ρ is the group representation that specifies how the channels of each feature vector mix under transformations. Kernels that fulfil this constraint are known as *rotation-steerable* and must be constructed from a suitable family of basis functions. As noted above, this is a linear relationship, which means that G-steerable kernels form a subspace of the convolution kernels used by conventional CNNs.

For planar images the input space will be $\mathbb{R}^2$, and for single frequency or continuum radio images these feature fields will be scalar, such that $s : \mathbb{R}^2 \rightarrow \mathbb{R}$. The group representation for scalar fields is also known as the *trivial* representation, $\rho(g) = 1 \;\; \forall \, g \in G$, indicating that under a transformation there is no orientation information to preserve and that the amplitude does not change. The group representation of the output space from a G-steerable convolution must be chosen by the user when designing their network architecture and can be thought of as a variety of hyper-parameter.

However, whilst the representation of the input data is in some senses quite trivial for radio images, in practice convolution layers are interleaved with other operations that are sensitive to specific choices of representation. In particular, the range of non-linear activation layers permissible for a particular group or subgroup representation may be limited. Trivial representations, such as scalar fields, do not transform under rotation and therefore conventional nonlinearities like the widely used ReLU activation function are fine. Bias terms in convolution allow equivariance for group convolutions only in the case where there is a single bias parameter per group feature map (rather than per channel feature map) and likewise for batch normalization (Cohen & Welling 2016).

In this work we use the G-steerable network layers from Weiler & Cesa (2019) who define the Euclidean group as being constructed from the translation group, $(\mathbb{R}, +)$, and the orthogonal group, $\mathrm{O}(2) = \{O \in \mathbb{R}^{2\times2} \mid O^T O = \mathrm{id}_{2\times2}\}$, such that the Euclidean group is congruent with the semi-direct product of these two groups, $\mathrm{E}(2) \cong (\mathbb{R}, +) \rtimes \mathrm{O}(2)$. Consequently, the operations contained in the orthogonal group are those which leave the origin invariant i.e. continuous rotations and reflections. In this work we specifically consider the cyclic subgroups of the Euclidean group with form $(\mathbb{R}^2, +) \rtimes C_N$, where C_N contains a set of discrete rotations in multiples of $2\pi/N$, and the dihedral subgroups with form $(\mathbb{R}^2, +) \rtimes D_N$, where $D_N \cong C_N \rtimes (\{\pm 1\}, *)$, which incorporate reflection around $x = 0$ in addition to discrete rotation. As noted by Cohen & Welling (2016), although convolution on continuous groups is mathematically well defined, it is difficult to approximate numerically in a fully equivariant manner. Furthermore, the complete description of all transformations in larger groups is not always feasible (Gens & Domingos 2014). Consequently, in this work we consider only the discrete and comparatively small groups, C_N and D_N, with orders N and $2N$, respectively.

3 DATA

The data set used in this work is based on the catalogue of Miraghaei & Best (2017), who used a parent galaxy sample taken from Best & Heckman (2012) that cross-matched the Sloan Digital Sky Survey (SDSS; York et al. 2000) data release 7 (DR7; Abazajian et al. 2009) with the Northern VLA Sky Survey (NVSS; Condon et al. 1998) and the Faint Images of the Radio Sky at Twenty centimetres (FIRST; Becker, White & Helfand 1995).

From the parent sample, sources were visually classified by Miraghaei & Best (2017) using the original morphological definition provided by Fanaroff & Riley (1974): galaxies that had their most luminous regions separated by less than half of the radio source's extent were classed as FRI, and those that were separated by more than half of this were classed as FRII. Where the determination of this separation was complicated by either the limited resolution of the FIRST survey or by its poor sensitivity to low surface brightness emission, the human subjectivity in this calculation was indicated by the source classification being denoted as 'Uncertain', rather than 'Confident'. Galaxies were then further classified into morphological sub-types via visual inspection. Any sources which showed FRI-like behaviour on one half of the source and FRII-like behaviour on the other were deemed to be hybrid sources.

Each object within the catalogue of Miraghaei & Best (2017) was given a three-digit classification identifier to allow images to be separated into different subsets. Images were classified by FR class, confidence of classification, and morphological sub-type. These are summarized in Table 1. For example, a radio galaxy that was confidently classified as an FRI type source with a wide-angle tail morphology would be denoted 102.

Table 1. Numerical identifiers from the catalogue of Miraghaei & Best (2017).

Digit 1	Digit 2	Digit 3
0 - FRI	0 - Confident	0 - Standard
1 - FRII	1 - Uncertain	1 - Double-double
2 - Hybrid		2 - Wide-angle tail
3 - Unclassifiable		3 - Diffuse
		4 - Head-tail

We note that not all combinations of the three digits described in Table 1 are present in the catalogue as some morphological classes are dependent on the parent FR class, with only FRI type objects being sub-classified into head-tail or wide-angle tail, and only FRII type objects being sub-classified as double-double. Hybrid FR sources are not considered to have any non-standard morphologies, as their standard morphology is inherently inconsistent between sources. Confidently classified objects outnumber their uncertain counterparts across all classes, and in classes that have few examples there may be no uncertain sources present. This is particularly apparent for non-standard morphologies.

From the full catalog of 1329 labelled objects, 73 were excluded from the machine learning data set. These include (i) the 40 objects denoted as *3 - unclassifiable*, (ii) 28 objects which had an angular extent greater than a selected image size of 150 × 150 pixels, (iii) four objects with structure that was found to overlap the edge of the sky area covered by the FIRST survey, and (iv) the single object in three-digit category 103. This final object was excluded as a minimum of two examples from each class are required for the data set: one for the training set and one for the test set. Following these exclusions, 1256 objects remain, which we refer to as the *MiraBest* data set and summarise in Table 2.

All images in the *MiraBest* data set are subjected to a similar data pre-processing as other radio galaxy deep learning data sets in the literature (see e.g. Aniyan & Thorat 2017; Tang et al. 2019). FITS images for each object are extracted from the FIRST survey data using the Skyview service (McGlynn, Scollick & White 1998) and the `astroquery` library (Ginsburg et al. 2019). These images are then processed in four stages before data augmentation is applied: first, image pixel values are set to zero if their value is below a threshold of three times the local rms noise; secondly, the image size is clipped to 150 by 150 pixels i. e. 270 arcsec by 270 arcsec for FIRST, where each pixel corresponds to 1.8 arcsec. Thirdly, all pixels outside a square central region with extent equal to the largest angular size of the radio galaxy are set to zero. This helps to eliminate secondary background sources in the field and is possible for the *MiraBest* data set due to the inclusion of this parameter in the catalogue of Miraghaei & Best (2017). Finally the image is normalized as:

$$\mathrm{Output} = 255 \cdot \frac{\mathrm{Input} - \min(\mathrm{Input})}{\max(\mathrm{Input}) - \min(\mathrm{Input})}, \tag{4}$$

where 'Output' is the normalized image, 'Input' is the original image and 'min' and 'max' are functions which return the single minimal and maximal values of their inputs, respectively. Images are saved to PNG format and accummulated into a `PyTorch` batched data set.[1]

For this work we extract the objects labelled as Fanaroff–Riley Class I (FRI) and Fanaroff–Riley Class II (FRII; Fanaroff & Riley

[1] The MiraBest data set is available on Zenodo: 10.5281/zenodo.4288837

Table 2. MiraBest data set summary. The original data set labels (MiraBest Label) are shown in relation to the labels used in this work (Label). Hybrid sources are not included in this work, and therefore have no label assigned to them.

Class	No.	Confidence	Morphology	No.	MiraBest Label
FRI	591	Confident	Standard	339	0
			Wide-angle tailed	49	1
			Head-tail	9	2
		Uncertain	Standard	191	3
			Wide-angle tailed	3	4
FRII	631	Confident	Standard	432	5
			Double-double	4	6
		Uncertain	Standard	195	7
Hybrid	34	Confident	NA	19	8
		Uncertain	NA	15	9

Table 3. Data used in this work. The table shows the number of objects of each class that are provided in the training and test partitions for the *MiraBest* data set, containing sources labeled as both Confident and Uncertain, and the *MiraBest** data set, containing only objects labeled as Confident, as well as the mean and standard deviation of the training sets in each case.

	Train		Test			
Data	FRI	FRII	FRI	FRII	μ	σ
MiraBest	517	552	74	79	0.0031	0.0352
MiraBest*	348	381	49	55	0.0031	0.0350

1974) radio galaxies with classifications denoted as Confident (as opposed to Uncertain). We exclude the objects classified as Hybrid and do not employ sub-classifications. This creates a binary classification data set with target classes FRI and FRII. We denote the subset of the full *MiraBest* data set used in this work as *MiraBest**.

The *MiraBest** data set has pre-specified training and test data partitions and the number of objects in each of these partitions is shown in Table 3 along with the equivalent partitions for the full *MiraBest* data set. In this work we subdivide the *MiraBest*$*$ training partition into training and validation sets using an 80:20 split. The test partition is reserved for deriving the performance metrics presented in Section 5.2.

To accelerate convergence, we further normalize individual data samples from the data set by shifting and scaling as a function of the mean and variance, both calculated from the full training set (LeCun et al. 2012) and listed in Table 3. Data augmentation is performed during training and validation for all models using random rotations from 0 to 360 degrees. This is standard practice for augmentation and is also consistent with the G-steerable CNN training implementations of Weiler & Cesa (2019), who included rotational augmentation for their own tests in order to not disadvantage models with lower levels of equivariance. To avoid issues arising from samples where the structure of the radio source overlaps the edge of the field and is artificially truncated in some orientations during augmentation, but not in others, we apply a circular mask to each sample image, setting all pixels to zero outside a radial distance from the centre of 75 pixels.

An example data sample is shown in Fig. 1, where it is used to illustrate the corresponding C_4 and D_4 groups. As noted by Weiler & Cesa (2019), for signals digitised on a pixel grid, exact equivariance is not possible for groups that are not symmetries of the grid itself and in this case only subgroups of D_4 will be exact symmetries with all other subgroups requiring interpolation to be employed (Dieleman et al. 2016).

4 ARCHITECTURE

For our architecture we use a simple LeNet-style network (LeCun et al. 1998) with two convolutional layers, followed by three fully-connected layers. Each of the convolutional layers has a ReLU activation function and is followed by a max-pooling operation. The fully-connected layers are followed by ReLU activation functions and we use a 50 per cent dropout before the final fully-connected layer, as is standard for LeNet (Krizhevsky, Sutskever & Hinton 2012). An overview of the architecture is shown in Table 4. In what follows we refer to this base architecture using conventional convolution layers as the *standard CNN* and denote it $\{e\}$. We also note that the use of *conventional CNN* is used through the paper to refer to networks that do not employ group-equivariant convolutions, independent of architecture.

For the G-steerable implementation of this network we use the `e2cnn` extension[2] to the PyTorch library (Weiler & Cesa 2019) and replace the convolutional layers with their subgroup-equivariant equivalent. We also introduce two additional steps into the network in order to recast the feature data from the convolutional layers into a format suitable for the conventional fully-connected layers. These steps consist of reprojecting the feature data from a geometric tensor into standard tensor format and pooling over the group features, and are indicated in italics in Table 4. Since the additional steps in the G-steerable implementations have no learnable parameters associated with them, the overall architecture is unchanged from that of the standard CNN; it is only the nature of the kernels in the convolutional layers that differ.

For the input data we use the trivial representation, but for all subsequent steps in the G-steerable implementations we adopt the *regular* representation, ρ_{reg}. This representation is typical for describing finite groups/subgroups such as C_N and D_N. The regular representation of a finite group G acts on a vector space $\mathbb{R}^{|G|}$ by permuting its axes, where $|G| = N$ for C_N and $|G| = 2N$ for D_N, see Fig. 1. This representation is helpful because its action simply permutes channels of fields and is therefore equivariant under pointwise operations such as the ReLU activation function, max, and average pooling functions (Weiler & Cesa 2019).

We train each network over 600 epochs using a standard cross-entropy loss function and the Adam optimizer (Kingma & Ba 2014) with an initial learning rate of 10^{-4} and a weight decay of 10^{-6}. We use a scheduler to reduce the learning rate by 10 per cent each time the validation loss fails to decrease for two consecutive epochs. We use mini-batching with a batch size of 50. No additional hyper-

[2] https://github.com/QUVA-Lab/e2cnn

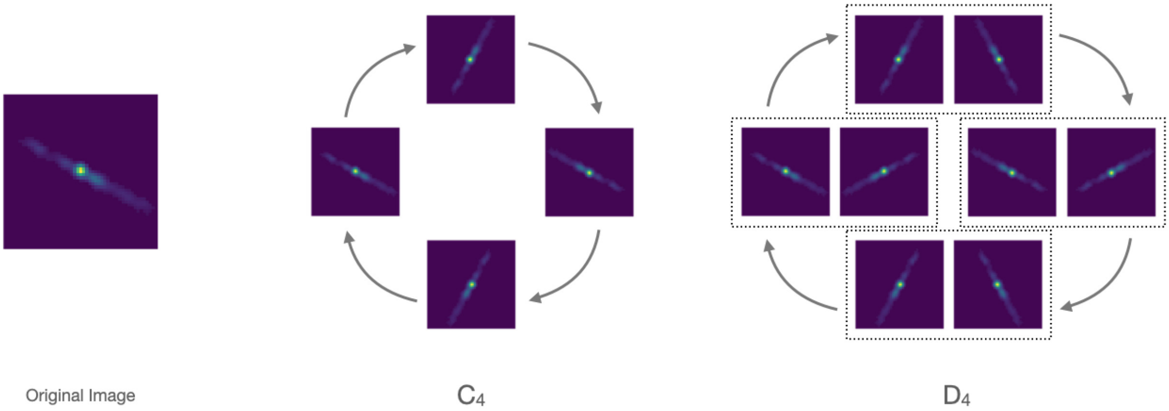

Figure 1. Illustration of the C_4 and D_4 groups for an example radio galaxy postage stamp image with 50 × 50 pixels. The members of the C_4 group are each rotated by $\pi/2$ radians, resulting in a group order $|C_4| = 4$. The members of the D_4 group are each rotated by $\pi/2$ radians and mirrored around $x = 0$, resulting in a group order $|D_4| = 8$.

Table 4. The LeNet5-style network architecture used for all the models in this work. *G*-Steerable implementations include the additional steps indicated in italics and replace the convolutional layers with the appropriate group-equivariant equivalent in each case. Column [1] lists the operation of each layer in the network, where entries in italics denote operations that are applied only in the *G*-steerable version of the network; Column [2] lists the kernel size in pixels for each layer, where appropriate; Column [3] lists the number of output channels from each layer; Column [4] denotes the degree of zero-padding in pixels added to each edge of an image, where appropriate.

Operation	Kernel	Channels	Padding
Invariant projection			
Convolution	5 × 5	6	1
ReLU			
Max-pool	2 × 2		
Convolution	5 × 5	16	1
ReLU			
Max-pool	2 × 2		
Invariant projection			
Global average pool			
Fully-connected		120	
ReLU			
Fully-connected		84	
ReLU			
Dropout ($p = 0.5$)			
Fully-connected		2	

parameter tuning is performed. We also implement an early-stopping criterion based on validation accuracy and for each training run we save the model corresponding to this criterion.

5 RESULTS

5.1 Convergence of G-steerable CNNs

Validation loss curves for both the standard CNN implementation, denoted $\{e\}$, and the group-equivariant CNN implementations for $N = \{4, 8, 16, 20\}$ are shown in Fig. 2. Curves show the mean and standard deviation for each network over five training repeats. It can seen from Fig. 2 that the standard CNN implementation achieves a significantly poorer loss than that of its group-equivariant equivalents. For both the cyclic and dihedral group-equivariant models, the best validation loss is achieved for $N = 16$. Although the final loss in the case of the cyclic and dihedral-equivariant networks is not significantly different in value, it is notable that the lower order dihedral networks converge towards this value more rapidly than the equivalent order cyclic networks. We observe that higher order groups minimize the validation loss more rapidly i.e. the initial gradient of the loss as a function of epoch is steeper, up to order $N = 16$ in this case. Weiler & Cesa (2019), who also noted the same thing when training on the MNIST datasets, attribute this behaviour to the increased generalization capacity of equivariant networks, since there is no significant difference in the number of learnable parameters between models.

Final validation error as a function of order, N, for the group-equivariant networks is shown in Fig. 3. From this figure it can be seen that all equivariant models improve upon the non-equivariant CNN baseline, $\{e\}$, and that the validation error decreases before reaching a minimum for both cyclic and dihedral models at approximately 16 orientations. This behaviour is discussed further in Section 6.4.

5.2 Performance of G-steerable CNNs

Standard performance metrics for both the standard CNN implementation, denoted $\{e\}$, and the group-equivariant CNN implementations for $N = \{4, 8, 16, 20\}$ are shown in Table 5. The metrics in this table are evaluated using the reserved test set of the *MiraBest** data set, classified using the best-performing model according to the validation early-stopping criterion. The reserved test set is augmented by a factor of 9 using discrete rotations of 20° over the interval [0°, 180°). This augmentation is performed in order to provide metrics that reflect the performance over a consistent range of orientations. The values in the table show the mean and standard deviation for each metric over five training repeats. All *G*-steerable CNNs listed in this table use a regular representation for feature data and apply a *G*-invariant map after the convolutional layers to guarantee an invariant prediction.

From Table 5 it can be seen that the best test accuracy is achieved by the D_{16} model, highlighted in bold. Indeed, while all equivariant models perform better than the standard CNN, the performance of the dihedral models is consistently better than for the cyclic models of equivalent order.

For the cyclic models it can be observed that the largest change in performance comes from an increased FRI recall. For a binary

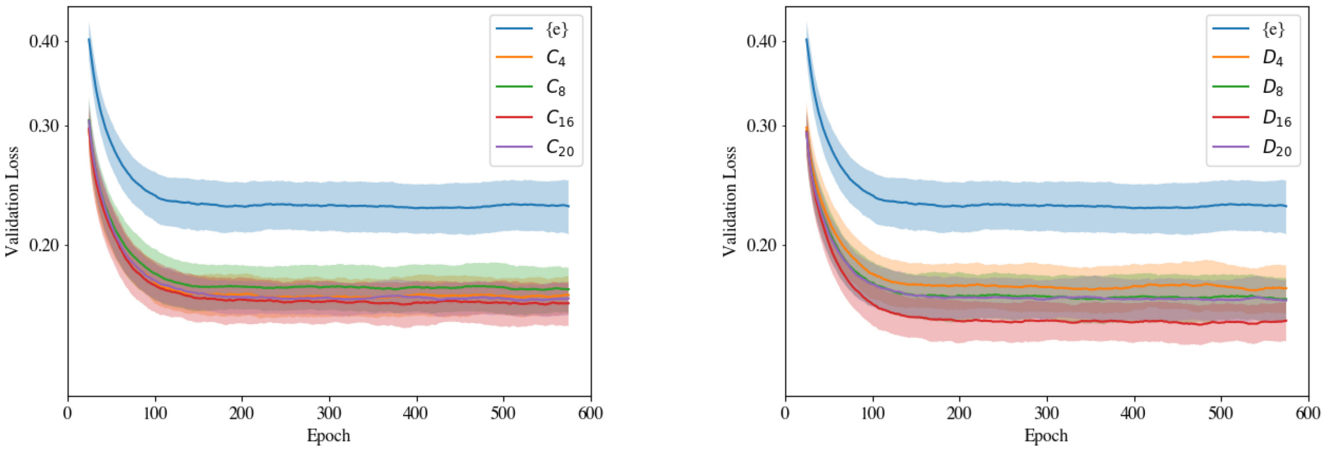

Figure 2. Validation losses during the training of the standard CNN, denoted $\{e\}$, and (i) C_N-equivariant models for the MiraBest* data set (left), and (ii) D_N-equivariant models for the MiraBest* data set (right). Plots show mean and standard deviation over five training repeats. Curves are smoothed over 20 epochs to eliminate small-scale variability.

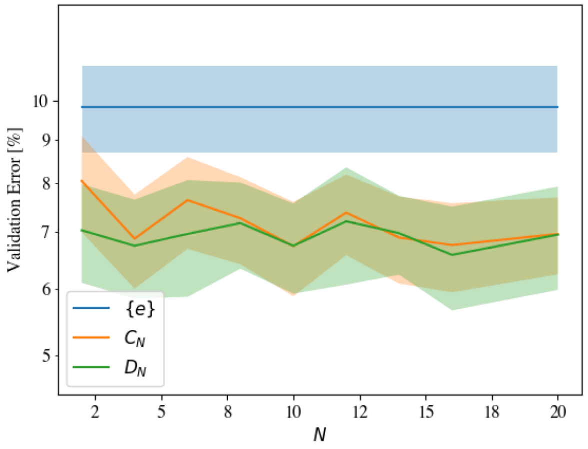

Figure 3. Validation errors of C_N and D_N regular steerable CNNs for different orders, N, for the MiraBest* data set. All equivariant models improve upon the non-equivariant CNN baseline, $\{e\}$.

classification problem, the recall of a class is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

where TP indicates the number of true positives and FN indicates the number of false negatives. The recall therefore represents the fraction of all objects in that class which are correctly classified. Equivalently, the precision of the class is defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{6}$$

Consequently, if the recall of one class increases at the expense of the precision of the opposing class then it indicates that the opposing class is being disproportionately misclassified. However, in this case we can observe from Table 5 that the precision of the FRII class is also increasing, suggesting that the improvement in performance is due to a smaller number of FRI objects being misclassified as FRII. For the cyclic models there is a smaller but not equivalent improvement in FRII recall. This suggests that the cyclic model primarily reduces the misclassification of FRI objects as FRII, but does not equivalently reduce the misclassification of FRII as FRI.

The dihedral models show a more even distribution of improvement across all metrics, indicating that there are more balanced reductions across both FRI and FRII misclassifications. This is illustrated in Fig. 4, which shows the average number of misclassifications over all orientations and training repeats for the standard CNN, the C_{16} CNN and the D_{16} CNN for the reserved test set.

The test partition of the full *Mirabest* data set contains 153 FRI and FRII-type sources labelled as both Confident and Uncertain, see Table 3. When using this combined test set the overall performance metrics of the networks considered in this work become accordingly lower due to the inclusion of the Uncertain sources. This is expected, not only because the Uncertain samples include edge cases that are more difficult to classify but also because the assigned labels for these objects may not be fully accurate. However, the relative performance shows the same degree of improvement between the standard CNN, $\{e\}$, and the D_{16} model, which havepercentage accuracies of 82.59 ± 1.41 and 85.30 ± 1.35, respectively, when evaluated against this combined test set.

We note that given the comparatively small size of the *Mirabest** training set, these results may not generalise equivalently to other potentially larger data sets with different selection specifications and that additional validation should be performed when considering the use of group-equivariant convolutions for other classification problems.

5.3 On the confidence of G-steerable CNNs

Target class predictions for each test data sample are made by selecting the highest softmax probability, which provides a normalized version of the network output values. By using dropout as a Bayesian approximation, as demonstrated in Gal & Ghahramani (2015), one is able to obtain a posterior distribution of network outputs for each test sample. This posterior distribution allows one to assess the degree of certainty with which a prediction is being made i.e. if the distribution of outputs for a particular class is well separated from those of other classes then the input is being classified with high confidence; however, if the distribution of outputs intersects those of other classes then, even though the softmax probability for a particular realisation may be high (even as high as unity), the overall distribution of softmax probabilities for that class may still fill the entire [0,1] range, overlapping significantly with the distributions

Table 5. Performance metrics for classification of the MiraBest* data set using the standard CNN ($\{e\}$) and G-steerable CNNs for different cyclic and dihedral subgroups of the E(2) Euclidean group. All G-steerable CNNs use a regular representation for feature data and apply a G-invariant map after the convolutions to guarantee an invariant prediction.

			FRI			FRII	
MiraBest*	**Accuracy [%]**	Precision	Recall	F1-score	Precision	Recall	F1-score
$\{e\}$	94.04 ± 1.37	0.935 ± 0.018	0.940 ± 0.024	0.937 ± 0.015	0.946 ± 0.020	0.941 ± 0.018	0.944 ± 0.013
C_4	95.24 ± 1.23	0.942 ± 0.018	0.959 ± 0.015	0.950 ± 0.013	0.963 ± 0.013	0.947 ± 0.018	0.955 ± 0.012
C_8	95.96 ± 1.06	0.950 ± 0.020	0.966 ± 0.016	0.958 ± 0.011	0.969 ± 0.013	0.954 ± 0.019	0.961 ± 0.010
C_{16}	96.07 ± 1.03	0.953 ± 0.020	0.964 ± 0.013	0.959 ± 0.011	0.968 ± 0.011	0.958 ± 0.019	0.963 ± 0.010
C_{20}	95.88 ± 1.12	0.951 ± 0.019	0.962 ± 0.013	0.957 ± 0.012	0.966 ± 0.011	0.956 ± 0.018	0.961 ± 0.011
D_4	95.45 ± 1.38	0.948 ± 0.024	0.957 ± 0.017	0.952 ± 0.014	0.962 ± 0.015	0.952 ± 0.023	0.957 ± 0.013
D_8	96.37 ± 0.95	0.960 ± 0.019	0.964 ± 0.014	0.962 ± 0.010	0.968 ± 0.012	0.964 ± 0.018	0.966 ± 0.009
D_{16}	**96.56 ± 1.29**	0.963 ± 0.025	0.965 ± 0.014	0.964 ± 0.013	0.969 ± 0.012	0.966 ± 0.023	0.967 ± 0.012
D_{20}	96.39 ± 1.00	0.959 ± 0.018	0.966 ± 0.015	0.962 ± 0.010	0.969 ± 0.013	0.962 ± 0.017	0.966 ± 0.010

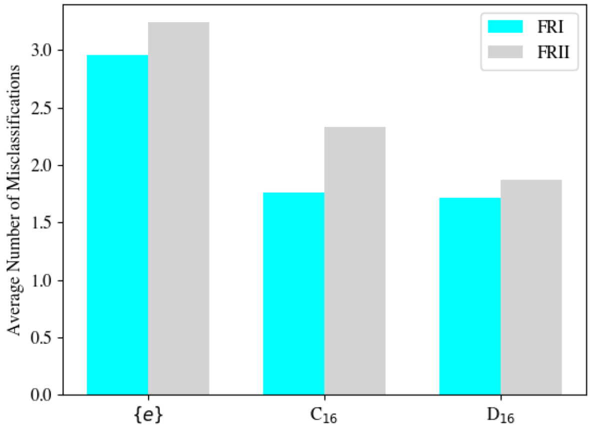

Figure 4. Average number of misclassifications for FRI (cyan) and FRII (grey) over all orientations and training repeats for the standard CNN, denoted $\{e\}$, the C_{16} CNN and the D_{16} CNN, see Section 5.2 for details.

from other target classes. Such a circumstance denotes a low degree of model certainty in the softmax probability and therefore in the class prediction for that particular test sample.

By re-enabling the dropout before the final fully-connected layer at test time, we estimate the predictive uncertainty of each model for the data samples in the reserved *MiraBest** test set. With dropout enabled, we perform $T = 50$ forward passes through the trained network for each sample in the test set. On each pass we recover (x_t, y_t), where x and y are the softmax probabilities of FRI and FRII, respectively. An example of the results from this process can be seen in Fig. 5, where we evaluate the trained model on a rotated version of the input image at discrete intervals of 20° in the range [0°, 180°) using a trained model for the standard CNN (left-hand panel) and for the D_{16}-equivariant CNN (right-hand panel). For each rotation angle, a distribution of softmax probabilities is obtained. In the case of the standard CNN it can be seen that, although the model classifies the source with high confidence when it is unrotated (0°), the softmax probability distributions are not well separated for the central image orientations, indicating that the model has a lower degree of confidence in the prediction being made in at these orientations. For the D_{16}-equivariant CNN it can be seen that in this particular test case the model has a high degree of confidence in its prediction for all orientations of the image.

To represent the degree of uncertainty for each test sample quantitatively, we evaluate the degree of overlap in the distributions of softmax probabilities at a particular rotation angle using the distribution-free overlap index (Pastore & Calcagní 2019). To do this, we calculate the local densities at position z for each class using a Gaussian kernel density estimator, such that

$$f_x(z) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\beta\sqrt{2\pi}}\mathrm{e}^{-(z-x_t)^2/2\beta^2}, \tag{7}$$

$$f_y(z) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\beta\sqrt{2\pi}}\mathrm{e}^{-(z-y_t)^2/2\beta^2}, \tag{8}$$

where $\beta = 0.1$. We then use these local densities to calculate the overlap index, η, such that

$$\eta = \sum_{i=1}^{N_z}\min\left[f_x(z_i),\, f_y(z_i)\right]\delta z, \tag{9}$$

where $\{z_i\}_{i=1}^{N_z}$ covers the range zero to one in N_z steps of size δz. For this work we assume $N_z = 100$. The resulting overlap index, η, varies between zero and one, with larger values indicating a higher degree of overlap and hence a lower degree of confidence.

For each test sample we evaluate the overlap index over a range of rotations from 0° to 180° in increments of 20°. We then calculate the average overlap index, $\langle\eta\rangle$, across these nine rotations. In Fig. 5 the value of this index can be seen above each plot: in this case, the standard CNN has $\langle\eta\rangle_{\{e\}} = 0.30$ and the D_{16}-equivariant CNN has $\langle\eta\rangle_{D_{16}} < 0.01$.

Of the 104 data samples in the reserved test set, 27.7 ± 11.0 per cent of objects show an improvement in average model confidence i.e. $\langle\eta\rangle_{\{e\}} - \langle\eta\rangle_{D_{16}} > 0.01$, when classified using the D_{16}-equivariant CNN compared to the standard CNN, 8.4 ± 2.5 per cent show a deterioration in average model confidence i.e. $\langle\eta\rangle_{D_{16}} - \langle\eta\rangle_{\{e\}} > 0.01$, and all other samples show no significant change in average model confidence i.e. $|\langle\eta\rangle_{\{e\}} - \langle\eta\rangle_{D_{16}}| < 0.01$. Mean values and uncertainties are determined from $\langle\eta\rangle$ values for all test samples evaluated using a pairwise comparison of five training realizations of the standard CNN and five training realizations of the D_{16} CNN.

Those objects that show an improvement in average model confidence are approximately evenly divided between FRI- and FRII-type objects, whereas the objects that show a reduction in model confidence exhibit a weak preference for FRII. These results are discussed further in Section 6.1.

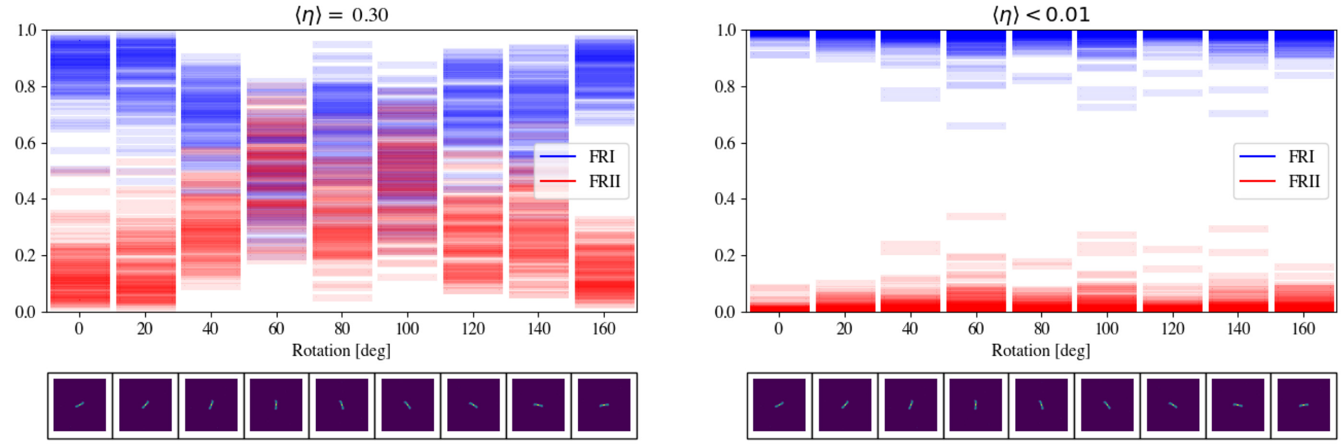

Figure 5. A scatter of 50 forward passes of the softmax output for the standard CNN (left) and the D_{16}-equivariant CNN (right). The lower panel shows the rotated image of the test image. As indicated, the average overlap index for the standard CNN is $\langle\eta\rangle = 0.30$, and $\langle\eta\rangle < 0.01$ for the D_{16}-equivariant CNN.

6 DISCUSSION

6.1 Statistical distribution of radio galaxy orientations

Mathematically, G-steerable CNNs classify equivalence classes of images, as defined by the equivalence relation of a particular group, G, whereas conventional CNNs classify equivalence classes defined only by translations. Consequently, by using E(2)-equivalent convolutions the trained models assume that the statistics of extra-galactic astronomical images containing individual objects are expected to be invariant not only to translations but also to global rotations and reflections. Here we briefly review the literature in order to consider whether this assumption is robust and highlight the limitations that may result from it.

The orientation of radio galaxies, as defined by the direction of their jets, is thought to be determined by the angular momentum axis of the super-massive black hole within the host galaxy. A number of studies have looked for evidence of preferred jet alignment directions in populations of radio galaxies, as this has been proposed to be a potential consequence of angular momentum transfer during galaxy formation (e.g. White 1984; Codis et al. 2018; Kraljic, Davé & Pichon 2020), or alternatively it could be caused by large-scale filamentary structures in the cosmic web giving rise to preferential merger directions (see e.g. Kartaltepe et al. 2008) that might result in jet alignment for radio galaxies formed during mergers (e.g. Croton et al. 2006; Chiaberge et al. 2015). The observational evidence for both remains a subject of discussion in the literature.

Taylor & Jagannathan (2016) found a local alignment of radio galaxies in the ELAIS N1 field on scales $<1^\circ$ using observations from the Giant Metrewave Radio Telescope (GMRT) at 610 MHz. Local alignments were also reported by Contigiani et al. (2017) who reported evidence ($>2\sigma$) of local alignment on scales of $\sim 2.5^\circ$ among radio sources from the FIRST survey using a much larger sample of radio galaxies, catalogued by the radio galaxy zoo project. A similar local alignment was also reported by Panwar et al. (2020) using data from the FIRST survey. Using a sample of 7555 double-lobed radio galaxies from the LOFAR Sky Survey (LoTSS; Shimwell et al. 2019) at 150 MHz, Osinga et al. (2020) concluded that a statistical deviation from purely random distributions of orientation as a function of projected distance was caused by systematics introduced by the brightest objects and did not persist when redshift information was taken into account. However, the study also suggested that larger samples of radio galaxies should be used to confirm this result.

Whilst these results may suggest tentative evidence for spatial correlations of radio galaxy orientations in local large-scale structure, they do not provide any information on whether these orientations differ between classes of radio galaxy i.e. the equivalence classes considered here. Moreover, the large spatial distribution and comparatively small number of galaxies that form the training set used in this work mean that even spatial correlation effects would be unlikely to be significant for the data set used here. However, the results of Taylor & Jagannathan (2016), Contigiani et al. (2017), Panwar et al. (2020) suggest that care should be taken in this assumption if data sets are compiled from only small spatial regions.

In Section 5.1 we found that the largest improvement in performance was seen when using dihedral, D_N, models. We suggest that this improvement over cyclic, C_N, models is due to image reflections accounting for chirality, in addition to orientations on the celestial sphere which are represented by the cyclic group. Galactic chirality has previously been considered for populations of star-forming, or normal, galaxies (see e.g. Slosar et al. 2009; Shamir 2020), as the spiral structure of star-forming galaxies means that such objects can be considered to be enantiomers i.e. their mirror images are not superimposable (Capozziello & Lattanzi 2005). It has been suggested that a small asymmetry exists in the number of clockwise versus anticlockwise star-forming galaxy spins (Shamir 2020). As far as the authors are aware there have been no similar studies considering the chirality of radio galaxies. However, a simple example of such chirality for radio galaxies might include the case where relativistic boosting causes one jet of a radio galaxy to appear brighter than the other due to an inclination relative to the line of sight. Since the dominance of a particular orientation relative to the line of sight should be unbiased then this would imply a global equivariance to reflection. Since the dihedral (D_N) models used in this work are insensitive to chirality, the results in Section 5.1 suggest that the radio galaxies in the training sample used here do not have a significant degree of preferred chirality. Whilst this does not itself validate the assumption of global reflection invariance, in the absence of evidence to the contrary from the literature we suggest that it is unlikely to be significant for the data sample used in this work.

From the perspective of classification, equivariance to reflections implies that inference should be independent of reflections of the input. For FRI and FRII radio galaxy classification, incorporating such information into a classification scheme may be important more generally: the unified picture of radio galaxies holds that both FRI and FRII, as well as many other classifications of active galactic nuclei such as quasars, quasi-stellar objects, blazars, BL Lac objects, Seyfert galaxies etc., are in fact defined by orientation-dependent observational differences, rather than intrinsic physical distinctions (Urry 2004).

Consequently, under the assumptions of global rotational and reflection invariance, the possibility of a classification model providing different output classifications for the same test sample at different orientations is problematic. Furthermore, the degree of model confidence in a classification should also not vary significantly as a function of sample orientation i.e. if a galaxy is confidently classified at one particular orientation then it should be approximately equally confidently classified at all other orientations. If this is not the case, as shown for the standard CNN in Fig. 5 (left), then it indicates a preferred orientation in the model weights for a given outcome, inconsistent with the expected statistics of the true source population. Such inconsistencies might be expected to result in biased samples being extracted from survey data.

In this context it is then not only the average degree of model confidence that is important as a function of sample rotation, as quantified by the value of $\langle\eta\rangle$ in Section 5.3, but also the stability of the η index as a function of rotation i.e. a particular test sample should be classified at a consistent degree of confidence as a function of orientation, whether that confidence is low or high. To evaluate the stability of the predictive confidence as a function of orientation, we examine the variance of the η index as a function of rotation. For the *MiraBest** reserved test set we find that approximately 30 per cent of the test samples show a reduction of more than 0.01 in the standard deviation of their overlap index as a function of rotation, with 17 per cent showing a reduction of more than 0.05. Conversely approximately 8 per cent of test samples show an increase of >0.01 and 4 per cent samples show an increase of >0.05. In a similar manner to the results for average model confidence given in Section 5.3, those objects that show a reduction in their variance i.e. an improvement in the consistency of prediction as a function of rotation, are evenly balanced between the two classes; however, those objects showing a strong improvement of >0.05 are preferentially FRI type objects.

6.2 Comment on capsule networks

The use of capsule networks (Sabour et al. 2017) for radio galaxy classification was investigated by Lukic et al. (2018). Capsule networks aim to separate the orientation (typically referred to as the viewpoint or pose in the context of capsule networks) of an object from its nature i.e. class by encoding the output of their layers as tuples incorporating both a *pose vector* and an activation. The purpose of this approach is to focus on the linear hierarchical relationships in the data and remove sensitivity to orientation; however, as described by Lenssen, Fey & Libuschewski (2018), general capsule networks do not guarantee particular group equivariances and therefore cannot completely disentangle orientation from feature data. It is perhaps partly for this reason that Lukic et al. (2018) found that capsule networks offered no significant advantage over standard CNNs for the radio galaxy classification problem addressed in that work.

In Section 5, we found that not only is the test performance improved by the use of equivariant CNNs, but that equivariant networks also converge more rapidly. For image data, a standard CNN enables generalization over classes of translated images, which provides an advantage over the use of an MLP, where every image must be considered individually. *G*-steerable CNNs extend this behaviour to include additional equivalences, further improving generalization. This additional equivariance enhances the data efficiency of the learning algorithm because it means that every image is no longer an individual data point but instead a representative of its wider equivalence group. Consequently, unlike capsule networks, the equivalence groups being classified by a *G*-steerable CNN are specified *a priori*, rather than the orientations of individual samples being learned during training. Whilst this creates additional capacity in the network for learning intra-class differences that are insensitive to the specified equivalences, it does not provide the information on orientation of individual samples that is provided as an output by capsule networks.

Lenssen et al. (2018) combined group-equivariant convolutions with capsule networks in order to output information on both classification and pose, although they note that a limitation of this combined approach is that arbitrary pose information is no longer available, but is instead limited to the elements of the equivariant group. For radio astronomy, where radio galaxy orientations are expected to be extracted from images at a precision that is limited by the observational constraints of the data, it is unlikely that pose information limited to the elements of a low-order finite group, $G < E(2)$, is sufficient for further analysis. However, given particular sets of observational and physical constraints or specifications it is possible that such an approach may become useful at some limiting order. Alternatively, pose information might be used to specify a prior for a secondary processing step that refines a measurement of orientation.

6.3 Local versus global equivariance

By design, the final features used for classification in equivariant CNNs do not include any information about the global orientation or chirality of an input image; however, this can also mean that they are insensitive to local equivariances in the image, when these might in fact be useful for classification. The hierarchical nature of convolutional networks can be used to mitigate against this, as kernels corresponding to earlier layers in a network will have a smaller, more local, footprint on the input image and therefore be sensitive to a different scale of feature than those from deeper layers that encompass larger-scale information. Therefore, by changing the degree of equivariance as a function of layer depth one can control the degree to which local equivariance is enforced. Weiler & Cesa (2019) refer to this practice as *group restriction* and find that it is beneficial when classifying data sets that possess symmetries on a local scale but not on a global scale, such as the CIFAR and unrotated MNIST datasets. Conversely, the opposite situation may also be true, where no symmetry is present on a local scale, but the data are statistically invariant on a global scale. In this case the reverse may be done and, rather than restricting the representation of the feature data to reduce the degree of equivariance, one might expand the domain of the representation at a particular layer depth in order to reflect a global equivariance.

We investigate the effect of group restriction by using a $D_N|_1\{e\}$ restricted version of the LeNet architecture i.e. the first layer is D_N equivariant and the second convolutional layer is a standard convolution. Using $N = 16$, the loss curve for this restricted architecture relative to the unrestricted D_{16} equivariant CNN is shown in Fig. 6. From the figure it can be seen that while exploiting local symmetries gives an improved performance over the standard CNN, the performance of the group restricted model is significantly poorer than that of the full D_{16} CNN. This result suggests that although local symmetries are present in the data, it is the global symmetries of the population that result in the larger performance gain for the radio galaxy data set.

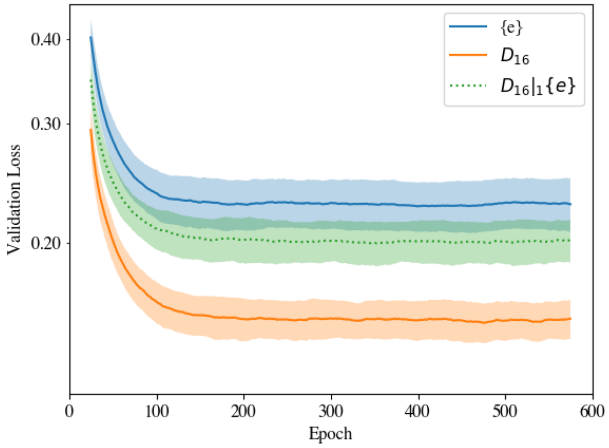

Figure 6. Validation losses during the training of the standard CNN, denoted $\{e\}$ (blue), the D_{16} CNN (orange), and the restricted $D_N|_1\{e\}$ CNN (green; dashed) for the *MiraBest** data set. Plots show mean and standard deviation over five training repeats.

6.4 Note on hyper-parameter tuning

In Section 5 we found that the $N = 16$ cyclic and dihedral models were preferred over the higher order $N = 20$ models. This may seem counter-intuitive as one might assume that for truly rotationally invariant data sets the performance would converge to a limiting value as the order increased, rather than finding a minimum at some discrete point. Consequently, we note that the observed minimum at $N = 16$ might not represent a true property of the data set but instead represent a limitation caused by discretisation artifacts from rotation of convolution kernels with small support, in this case $k = 5$, see Table 4 (Weiler & Cesa 2019). These same discretisation errors may also account in part for the small oscillation in validation error as a function of group order seen in Fig. 3. Consequently, while no additional hyper-parameter tuning has been performed for any of the networks used in this work, we note that kernel size is potentially one hyper-parameter that could be tuned as a function of group order, N, and that such tuning might lead to further improvements in performance for higher orders.

7 CONCLUSIONS

In this work, we have demonstrated that the use of even low-order group-equivariant convolutions results in a performance improvement over standard convolutions for the radio galaxy classification problem considered here, without additional hyper-parameter tuning. We have shown that both cyclic and dihedral equivariant models converge to lower validation loss values during training and provide improved validation errors. We attribute this improvement to the increased capacity of the equivariant networks for learning hierarchical features specific to classification, when additional capacity for encoding redundant feature information at multiple orientations is no longer required, hence reducing intra-class variability.

We have shown that for the simple network architecture and training set considered here, a D_{16} equivariant model results in the best test performance using a reserved test set. We suggest that the improvement of the dihedral over the cyclic models is due to an insensitivity to – and therefore lack of preferred – chirality in the data, and that further improvements in performance might be gained from tuning the size of the kernels in the convolutional layers according to the order of the equivalence group. We find that cyclic models predominantly reduce the misclassification of FRI type radio galaxies, whereas dihedral models reduce misclassifications for both FRI and FRII type galaxies.

By using the MC Dropout Bayesian approximation method, we have shown that the improved performance observed for the D_{16} model compared to the standard CNN is reflected in the model confidence as a function of rotation. Using the reserved test set, we have quantified this difference in confidence using the overlap between predictive probability distributions of different target classes, as encapsulated in the distribution free overlap index parameter, η. We find that not only is average model confidence improved when using the equivariant model, but also that the consistency of model confidence as a function of image orientation is improved. We emphasise the importance of such consistency for applications of CNN-based classification in order to avoid biases in samples being extracted from future survey data.

Whilst the results presented here are encouraging, we note that this work addresses a specific classification problem in radio astronomy and the method used here may not result in equivalent improvements when applied to other areas of astronomical image classification using different data sets or network architectures. In particular, the assumptions of global rotational and reflectional invariance are strong assumptions, which may not apply to all data sets. As described in Section 6.1, data sets extracted from localized regions of the sky may be particularly vulnerable to biases when using this method and the properties of the *MiraBest** data set used in this work may not generalise to all other data sets or classification problems. We note that this is true for all CNNs benchmarked against finite data sets and users should be aware that additional validation should be performed before models are deployed on new test data, as biases arising from data selection may be reflected in biases in classifier performance (see e.g. Wu et al. 2018; Tang 2019; Walmsley et al. 2020). However, in conclusion, we echo the expectation of Weiler & Cesa (2019), that equivariant CNNs may soon become a common choice for morphological classification in fields like astronomy, where symmetries may be present in the data, and note that the overhead in constructing such networks is now minimal due to the emergence of standardized libraries such as `e2cnn`. Future work will need to address the optimal architectures and hyper-parameter choices for such models as specific applications evolve.

ACKNOWLEDGEMENTS

This work makes extensive use of the `e2cnn` library: github.com/QUVA-Lab/e2cnn. AMS gratefully acknowledges support from an Alan Turing Institute AI Fellowship EP/V030302/1. FP gratefully acknowledges support from STFC and IBM through the iCASE studentship ST/P006795/1.

DATA AVAILABILITY

Code for this work is publicly available on Github at the following address: github.com/as595/E2CNNRadGal. The MiraBest data set is available on Zenodo: 10.5281/zenodo.4288837.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Alger M. J. et al., 2018, MNRAS, 478, 5547
Aniyan A. K., Thorat K., 2017, ApJS
Banfield J. K. et al., 2015, MNRAS, 453, 2326

Beardsley A. P. et al., 2019, Publ. Astron. Soc. Aust., 36, 50
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569
Bowles M., Scaife A. M. M., Porter F., Tang H., Bastien D. J., 2021, MNRAS, 501, 4579
Capozziello S., Lattanzi A., 2005, Chirality, 18, 17
Chiaberge M., Gilli R., Lotz J. M., Norman C., 2015, ApJ, 806, 147
Codis S., Jindal A., Chisari N. E., Vibert D., Dubois Y., Pichon C., Devriendt J., 2018, MNRAS, 481, 4753
Cohen T. S., Welling M., 2016, Proceedings of the 33rd International Conference on Machine Learning, 48, 2990
Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693
Contigiani O. et al., 2017, MNRAS, 472, 636
Croton D. J. et al., 2006, MNRAS, 365, 11
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
Dieleman S., De Fauw J., Kavukcuoglu K., 2016, Proceedings of The 33rd International Conference on Machine Learning, 48, 1889
Fanaroff B. L., Riley J. M., 1974, MNRAS, 167, 31P
Gal Y., Ghahramani Z., 2016, Proceedings of The 33rd International Conference on Machine Learning, 48, 1050
Gens R., Domingos P. M., 2014, in Advances in Neural Information Processing Systems. Curran Associates, Inc., Red Hook, New York, p. 2537
Gheller C., Vazza F., Bonafede A., 2018, MNRAS, 480, 3749
Ginsburg A. et al., 2019, AJ, 157, 98
Jarvis M. J. et al., 2016, in 2016 MeerKAT Science: On the Pathway to the SKA, MeerKAT 2016 - Stellenbosch, South Africa
Johnston S. et al., 2008, Exp. Astron., 22, 151
Kartaltepe J. S., Ebeling H., Ma C. J., Donovan D., 2008, MNRAS, 389, 1240
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kraljic K., Davé R., Pichon C., 2020, MNRAS, 493, 362
Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, Advances in Neural Information Processing Systems 25. Curran Associates, Inc., RedHook, New York, p. 1097
LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278
LeCun Y., Bottou L., Orr G., Müller K., 2012, in Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg
Lenc K., Vedaldi A., 2019, International Journal of Computer Vision, 127, 456
Lenssen J. E., Fey M., Libuschewski P., 2018, preprint (arXiv:1806.05086)
Lukic V., Brüggen M., Banfield J. K., Wong O. I., Rudnick L., Norris R. P., Simmons B., 2018, MNRAS, 476, 246
McGlynn T., Scollick K., White N., 1998, in McLean B. J., Golombek D. A., Hayes J. J. E., Payne H. E., eds, New Horizons from Multi-Wavelength Sky Surveys. Springer, Dordrecht
Miraghaei H., Best P. N., 2017, MNRAS, 466, 4346
Osinga E. et al., 2020, A&A, 642, A70
Panwar M., Prabhakar, Sandhu P. K., Wadadekar Y., Jain P., 2020, MNRAS, 499, 1226
Pastore M., Calcagní A., 2019, Front. Psychol., 10, 1089
Ren S., He K., Girshick R., Sun J., 2015, in Advances in Neural Information Processing Systems. Red Hook, New York, p. 91
Sabour S., Frosst N., E Hinton G., 2017, preprint (arXiv:1710.09829)
Shamir L., 2020, PASA, 37, e053
Shimwell T. W. et al., 2019, A&A, 622, A1
Slosar A. et al., 2009, MNRAS, 392, 1225
Tang H., 2019, FR-DEEP, https://github.com/HongmingTang060313/FR-DEEP
Tang H., Scaife A. M. M., Leahy J. P., 2019, MNRAS, 488, 3358
Taylor A. R., Jagannathan P., 2016, MNRAS, 459, L36
Urry C., 2004, in Richards G. T., Hall P. B., eds, ASP Conf. Ser. Vol. 311, AGN Physics with the Sloan Digital Sky Survey. Astron. Soc. Pac., San Francisco. p. 49
Van Haarlem M. P. et al., 2013, A&A, 556, 2
Walmsley M. et al., 2020, MNRAS, 491, 1554
Weiler M., Cesa G., 2019, preprint (arXiv:1911.08251)
Weiler M., Hamprecht F. A., Storath M., 2017, preprint (arXiv:1711.07289)
Weiler M., Geiger M., Welling M., Boomsma W., Cohen T., 2018, preprint (arXiv:1807.02547)
White S. D. M., 1984, ApJ, 286, 38
Wu C. et al., 2018, MNRAS, 482, 1211
York D. G. et al., 2000, AJ, 120, 1579

This paper has been typeset from a TEX/LATEX file prepared by the author.