On Making $\mathrm{SE}(2)$-Invariant Networks Optimal

Tomas Karella[*]
CAI-3: Information Sciences
Los Alamos National Laboratory

Emily Shinkle
CAI-3: Information Sciences
Los Alamos National Laboratory

Alice Elisabeth Anastasia Allen
Machine Learning in Molecular Dynamics
Max Planck Institute for Polymer Research

Pieter J Swart
T-5: Applied Mathematics and Plasma Physics
Los Alamos National Laboratory

Nicholas Lubbers
CAI-3: Information Sciences
Los Alamos National Laboratory

Roxana Bujack
CAI-3: Information Sciences
Los Alamos National Laboratory

Abstract

Respecting a-priori known symmetries of underlying data is a key principle in designing efficient neural architectures. Convolutional neural networks are a canonical example, encoding translation equivariance. Extending this inductive bias to richer symmetry groups such as $\mathrm{SE}(2)$ in the most effective way possible is an ongoing architectural and computational challenge. Building on classical invariant theory, we derive invariant feature maps that are provably optimal: minimal and information-preserving with respect to the symmetry group. Our construction is designed to integrate seamlessly into standard architectures, requiring no modification to nonlinearities or normalization layers. As a concrete implementation, we introduce a plug-and-play module for convolutional neural networks that enforces $\mathrm{SE}(2)$ invariance, capturing both translations and rotations within standard architectures. The resulting models are both efficient and expressive. Experiments show competitive or improved accuracy compared to state-of-the-art equivariant methods at significantly lower computational cost. Finally, we provide a theoretical and empirical analysis of the relationship between equivariance and invariance, offering insight into their roles in controlling robustness and flexibility.

Convolutional neural networks (CNNs) [1] remain a fundamental building block in computer vision, largely due to their strong inductive biases [2]. In particular, translation equivariance enables consistent responses to spatial shifts and has proven crucial in data-limited regimes [3]. More generally, incorporating symmetry priors into neural architectures has emerged as a principled strategy for improving sample efficiency and robustness, especially in scientific and engineering applications where underlying symmetries are well understood [3, 4]. This has motivated a growing body of work extending equivariance beyond translations to other groups, such as rotations and reflections, to capture richer geometric structure in underlying data [5–15].

Many approaches to rotation equivariance construct feature spaces that transform according to irreducible representations of the underlying symmetry group [11, 14–17]. While this provides a principled framework, it introduces significant computational and architectural overhead. In particular, operations on such structured features often scale quadratically with the number of basis functions used to parameterize the feature space [11, 15, 16], leading to increased complexity. Furthermore, standard deep learning components, such as convolutions, nonlinearities, and normalization layers,

[*]Corresponding author: karella@lanl.gov

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

cannot be applied directly and must be redesigned to operate on these structured feature spaces. This complicates implementation and limits compatibility with existing architectures. Consequently, there remains a need for approaches that retain the benefits of symmetry-aware representations while enabling efficient [12, 18] and plug-and-play integration into standard neural network pipelines [19].

An alternative framework for handling symmetry comes from classical pattern recognition [20–22]. In this approach, invariance is achieved by constructing features that are invariant to the transformation group [20], avoiding the need for structured intermediate descriptors and instead operating directly in the invariant space. Early work [23] highlighted that invariance alone is not sufficient: features must remain **discriminative**, i.e., preserve the ability to distinguish patterns up to the group action, avoid redundancy, and remain computationally tractable. Further, Flusser constructed invariant features that are both independent and discriminative on subsets of the input space where the group acts freely (e.g., under conditions such as $m_{1,0} \neq 0$) [24, 25]. More recent work generalizes this result by constructing discriminative invariant maps on broader subsets of the input space [26].

Our contribution is fourfold. First, we derive an **optimal invariant map** grounded in classical invariant theory and prove that it is both **discriminative**—preserving all information up to the group action—and **minimal**. Second, we translate this theory directly to an efficient implementation as a **plug-and-play module** yielding $\mathrm{SE}(2)$-invariant CNNs requiring no modification to standard components and yielding models that are both computationally efficient and expressive. Third, we show that these models achieve **competitive or improved performance** compared to equivariant approaches at lower computational cost on a variety of image datasets. Fourth, we provide a **theoretical and empirical analysis** of the relationship between equivariance and invariance, offering insight in balancing robustness and flexibility.

1 Background: Discriminative and Minimal Invariant Maps

We first review the invariant-theoretic framework underlying our approach. We begin by describing moments as projections onto basis functions with well-defined transformation properties, and then show how such so-called equivariant features can be systematically combined into invariants. This formulation allows for the development of feature maps that are both discriminative and minimal, forming the basis for the neural modules introduced in this work.

1.1 Complex Moments and Irreducible Representations of $\mathrm{SO}(2)$

In pattern recognition, the goal is to distinguish between patterns $p \in \mathcal{P}$ based on observed data, where pattern p is a function defined on compact support $p : \Omega \subset \mathbb{R}^d \to \mathbb{K}$. When symmetries are present, they can be modeled by a group $\mathcal{G}$ acting on $\mathcal{P}$, which defines how patterns p behave under transformation. Two fundamental behaviors arise.

A function $\hat{\phi} : \mathcal{P} \to \mathbb{K}$ is **equivariant** if it transforms consistently with the group action. Typically, the action of $\mathcal{G}$ is realized through representations, so that each $g \in \mathcal{G}$ is associated with a matrix ρ acting on the corresponding space. The group actions on the input and output spaces may be given by different representations, and equivariance is expressed as

$$\hat{\phi}(\rho_{\mathrm{in}}(g)\, p) = \rho_{\mathrm{out}}(g)\, \hat{\phi}(p), \quad \forall g \in \mathcal{G},\, p \in \mathcal{P}. \tag{1}$$

and **invariant** if it remains unchanged under the group action, i.e. $\phi(\rho_{\mathrm{in}}(g)\, p) = \phi(p)$.

To construct informative features of patterns, it is common to project signals onto a set of basis functions with well-defined $\mathcal{G}$ transformation properties. In the context of group symmetries, such bases are often chosen to align with irreducible representations [27] of the underlying group, as formalized by results such as the Peter-Weyl Theorem [28] for compact groups. Classically, these projections are referred to as **moments** [20].

Let $p : \Omega \subset \mathbb{R}^d \to \mathbb{K}$ be a pattern with compact support, and let $\psi_{k,l} : \Omega \to \mathbb{K}$ be a basis function. The **moment** of p associated with $\psi_{k,l}$ is defined as

$$m[\psi_{k,l}](p) := \int_{\Omega} \psi_{k,l}(x)\, p(x)\, dx. \tag{2}$$

Moments are defined with respect to a chosen family of basis functions. For rotational symmetries, widely used basis functions—such as complex or Zernike polynomials [20], or equivariant steerable

bases [15–17]—share a common structure: they decompose into radial and angular components, differing in the radial term. We therefore consider a general family of basis functions of the form

$$\psi_{k,l}(r,\theta) = R_{k,l}(r)\, e^{\mathrm{i}k\theta}, \tag{3}$$

where (r,θ) are polar coordinates, $R_{k,l}$ is a radial function, $k \in \mathbb{Z}$ denotes the angular frequency, and l indexes the radial component. Since $e^{\mathrm{i}k\theta}$ transforms according to an irreducible representation of $\mathrm{SO}(2)$, each basis function $\psi_{k,l}$ induces a moment $m[\psi_{k,l}]$ with a well-defined type, also called spin, weight, character, or eigenvalue, given by k. For notational simplicity, we write $m_{k,l} := m[\psi_{k,l}]$. When the radial index is fixed or irrelevant, we simply write m_k. Formally, for a rotation $g_\theta \in \mathrm{SO}(2)$ by angle θ, a moment of type k satisfies

$$m_k(g_\theta \odot p) \overset{(2)(3)}{=} e^{\mathrm{i}k\theta}\, m_k(p). \tag{4}$$

Rotation invariants [20] can be constructed by combining moments in such a way that their transformation phases cancel. In particular, consider a product of moments

$$\phi(p) := \prod_{j=1}^{n} m_{k_j}(p)^{a_j}, \qquad \text{subject to} \sum_{j=1}^{n} k_j a_j = 0. \tag{5}$$

then ϕ is invariant under rotation g_θ. Indeed, using (4) and (5) and rotation g_θ, we have

$$\phi(g_\theta \odot p) \overset{(4)(5)}{=} \prod_{j=1}^{n} \left(e^{\mathrm{i}k_j\theta} m_{k_j}(p)\right)^{a_j} e^{\mathrm{i}\theta \sum_{j=1}^{n} k_j a_j}\, \phi(p) \overset{(5)}{=} \phi(p).$$

1.2 Flusser Basis as a Minimal and Discriminative Invariant Map

Equation 5 allows for an infinite number of invariants that emerge as a combinatorial explosion. Early approaches, such as Hu's[29] moment invariants, proposed specific targeted subsets of such combinations, but these were proven redundant and did not capture all relevant information. This limitation was addressed by Flusser [25], who proposed a systematic construction of non-redundant invariant representations, known as the **Flusser basis**. The key idea is to normalize moments of arbitrary type k by a reference moment of type 1, thereby canceling the rotational phase. This yields the invariant map

$$\Phi_F(p) := \left\{ m_{k,l}(p)\, m_{1,0}(p)^{-k} \;\middle|\; 1 \le k \le N_k,\ 1 \le l \le N_l,\ m_{1,0}(p) \neq 0 \right\}, \tag{6}$$

where $m_{1,0}$ is a non-zero reference moment of type 1 and $N_k, N_l \in \mathbb{N}^+$ control the maximum angular and radial frequencies. This construction yields a non-redundant set of invariants and is discriminative up to rotation. It can be interpreted as a canonical phase normalization of moments. Our formulation generalizes classical constructions: complex moments [24] are typically indexed as m_{pq} with type $k = p - q$, which corresponds to our notation $m_{k,l}$ up to the choice of radial functions.

The resulting Flusser basis has dimension $\dim(\Phi_F) = \dim(\mathcal{M}) - \dim(G)$ (cf. Appendix K), indicating that it captures all degrees of freedom of the signal up to the group action. Here, $\mathcal{M}$ denotes the manifold induced by the moments.[2]

To formalize properties of **optimal bases** in our setting, we revisit and adapt the notions of minimality and discriminability from classical pattern recognition [23]. Let $\Phi : \mathcal{P} \to \mathbb{K}^n$ denote a map of n invariant features, with components $\phi_j : \mathcal{P} \to \mathbb{K}$. In the context of deep learning, where networks can approximate arbitrary functions [31], invariant maps should be non-redundant, i.e., minimal, meaning that the features do not admit functional dependencies.

An invariant map Φ is **discriminative** if, for all $p_1, p_2 \in \mathcal{P}$,

$$\Phi(p_1) = \Phi(p_2) \quad \Rightarrow \quad \exists\, g \in \mathcal{G} \text{ such that } p_2 = g \odot p_1. \tag{7}$$

An invariant map $\Phi = (\phi_1, \ldots, \phi_m)$ is called **minimal** if it is discriminative and no proper subcollection is discriminative, i.e.,

$$\forall\, \Psi \subsetneq \Phi, \quad \Psi \text{ is not discriminative.} \tag{8}$$

Equivalently, no component ϕ_i can be removed without losing discriminability. Functional independence is sufficient for minimality [24], while discriminability for patterns with $m_{1,0} \neq 0$ follows from completeness and the solution of the inverse problem [25].

[2] In cases where $(m_1 = 0)$, the flexible invariant basis [30] can be employed; see Appendix A for details.

2 Building $\mathrm{SE}(2)$-Invariant Networks with Moment Invariants

Having established the invariant theory, we now introduce our method. We construct $\mathrm{SE}(2)$-invariant CNNs using optimal bases derived from moment invariants. We consider a setting in which $\mathrm{SE}(2)$ acts only on the spatial domain, while feature values remain invariant. In contrast to classical approaches, where invariants are computed from a single pattern p (e.g., at the center of mass), we evaluate invariants at every spatial location of the input feature space $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$. Accordingly, each layer of the network $\hat{\Phi}$ is required to satisfy the equivariance constraint on the spatial domain, while the output feature values themselves remain invariant under rotations ($\rho_{\text{out}} = \mathbb{I}$):

$$\hat{\Phi}(g \odot f)(x) = \hat{\Phi}(f)(g^{-1}x), \quad g \in \mathrm{SE}(2), \tag{9}$$

where $g = (g_t, g_\theta) \in \mathrm{SE}(2)$ denote translation and rotation, and $\odot$ is the action of $\mathrm{SE}(2)$ on the input signal.

Many standard operations used in deep CNNs already satisfy Eq. (9) (up to discretizations errors), including point-wise nonlinearities and normalization layers, as they act independently at each pixel. In the context of CNNs, standard convolutional layers are not equivariant under rotations and therefore must be redesigned to preserve the desired transformation behavior under $\mathrm{SE}(2)$.

2.1 Moment Layer: Projection to a basis

The proposed moment projection layer can be viewed as a form of steerable convolution [16]. In steerable CNNs [11, 15–17], equivariance is enforced by constraining convolutional kernels to transform according to group representations. Our construction follows the same principle: we project the input onto a moment basis at each spatial location. This maps local signals into a moment space whose components transform according to their type. Concretely, each moment m_k satisfies

$$m_k(g \odot f)(x) \overset{(4)}{=} e^{\mathrm{i}k\theta} m_k(f)(g^{-1}x), \quad g = (g_t, g_\theta) \in \mathrm{SE}(2), \tag{10}$$

so that the feature map $(m_{k_1}, \ldots, m_{k_n})$ forms a steerable field transforming under the direct sum of irreducible representations, that is, the combined representation is given by $\rho(g_\theta) = \bigoplus_j e^{\mathrm{i}k_j\theta}$.

Steerable feature fields provide a principled way to achieve equivariance and have been studied in prior work through specialized architectures that preserve the transformation law in Eq. (10) [27]. In these approaches, all layers must be adapted accordingly, and convolutional layers explicitly model interactions between all input and output types. As implemented in frameworks such as escnn [15, 16], steerable convolutions incur a computational scaling of $\mathrm{O}(N_{\text{in}}N_{\text{out}})$ due to these type-to-type interactions.

In contrast, our construction avoids explicit type-to-type interactions by directly projecting all features back into type 0, the invariant type. Each block consists of (i) a moment projection layer and (ii) an invariant layer that combines them into a complete and non-redundant set of invariant features over the quotient space $\mathcal{M}/G$. Our block scales linearly with the number of moments.

Implementation. The moment projection is implemented as a convolution, computing moments at each spatial location. The kernels $\psi_{k,l}(r, \theta)$ form a separable basis in polar coordinates (Eq. (3)), with angular frequency k and radial components $R_{k,l}$. In practice, $R_{k,l}$ can be fixed or learned; following prior work [11, 15, 16], we use a B-spline parameterization. Similar to prior symmetry-aware architectures [15, 32], we also employ larger CNN kernels than in standard CNNs to mitigate discretizations error. Additional details on radial parameterizations and kernel sizes are provided in Appendix E. For stability, moment features are normalized using a shared layer normalization across types, similarly to Equiformer [12]. We observe that performance is not sensitive to the specific normalization choice.

2.2 Optimal Bases: Preserving All Information from Moments

A natural approach to constructing invariant features is to directly apply classical moment invariants, such as the Flusser basis defined in Eq. (6), which combine moments into type-0 quantities. In our setting, these constructions are applied locally at each spatial location. Concretely, one can define pointwise invariants of the form

$$\phi_k(f)(x) \overset{(6)}{=} m_k(f)(x) \left[m_1(f)(x)\right]^{-k}, \tag{11}$$

which cancels the phase and yields a type-0 quantity. For simplicity, we omit the explicit dependence on f and x when it is clear from context, and write m_k and ϕ_k instead of $m_{k,l}(f)(x)$ and $\phi_k(f)(x)$.

Such formulations involve several complexities that need to be addressed for a practical implementation. First, they involve high-order polynomial combinations of moments, resulting in rapidly growing magnitudes and numerical instabilities. They also require explicit manipulation of complex-valued quantities, leading to inefficiency. Our concrete construction handles these difficulties.

Optimal Bases. We leverage the transformation law in Eq. (4) in two ways. (i) The **magnitudes $|m_k|$ are rotation-invariant** and therefore provide type-0 features directly. (ii) Since the phase of each moment changes linearly with its type, relative phases can be normalized using a non-zero reference moment m_1, analogous to classical invariant constructions in Eq. (6). The resulting **phase differences can be represented via cosine or sine functions**. Together, these features capture the invariant degrees of freedom and are sufficient to reconstruct classical moment invariants (Appendix B). We define the invariant feature basis $\Phi_{\mathcal{O}}$ as

$$\Phi_{\mathcal{O}} = \{|m_1|,\ |m_k|,\ \phi_{\mathcal{O}}(m_1, m_k) \mid k = 0, 2 \dots N_k\}, \tag{12}$$

where the magnitude terms provide rotation-invariant amplitudes and $\phi_{\mathcal{O}}(m_1, m_k)$ encodes rotation-invariant relative phase between moments.

To make the phase construction explicit, we express moments in polar form as $m_k = r_k e^{i\theta_k}$. As mentioned, direct constructions such as in Eq. (11) introduce powers of magnitudes, leading to value explosion in deep layers. Although magnitude information is already captured, the phase alone cannot be used directly, as phases of low-magnitude moments are unstable: the phase becomes undefined as the magnitude approaches zero. We therefore construct a single real-valued invariant per pair (m_1, m_k) by taking the real part of the phase-aligned interaction and modulating it by a magnitude-dependent function:

$$\phi_{\mathcal{O}}(m_1, m_k) = \sigma(|m_1|, |m_k|)\ \cos\big(\theta_k - k\theta_1\big).^{3} \tag{13}$$

This suppresses unstable phase contributions near zero-magnitude patterns while preventing large polynomial scaling with the magnitude of the pattern. Different modulation functions are discussed in Appendix H. For choosing mirroring sensitivity ($E(2)$ versus $SE(2)$) there are two options. Using $SO(2)$ kernels [16] then (13) . We use different moments to realize these alternatives. In the main text, we focus on the mirroring-sensitive $SE(2)$ network. Appendix C compares the performance of the two approaches on the benchmarks.

2.3 Plug-and-Play Invariant Moment Layer

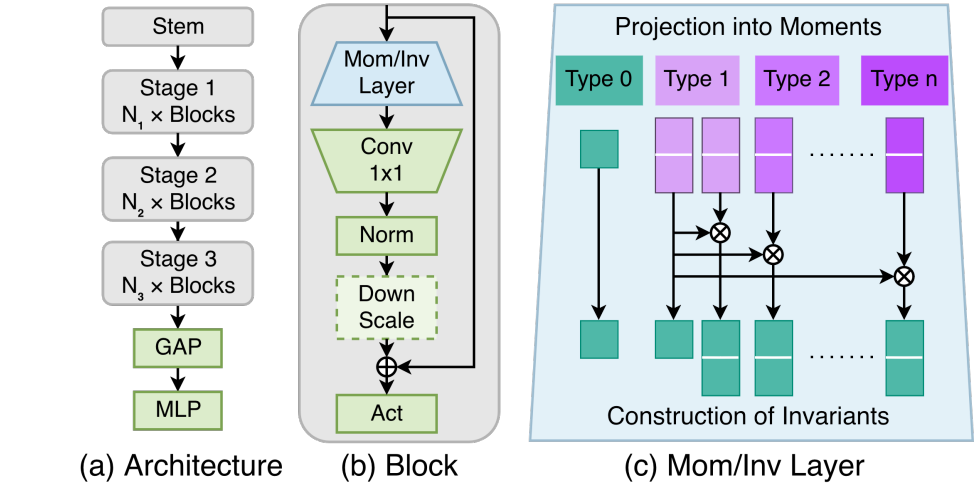
The overall architecture of the CNN block is depicted in Fig. 1. The moment projection is followed by an invariant layer that maps moment features to type-0 representations according to Eq. (13), combining moment magnitudes and relative phase differences with respect to a reference moment illustrated by Fig. 1(c). Together, these components form an invariant convolutional (Mom/Inv) layer that can be integrated into standard CNN architectures.

The output invariant features can be aggregated in a variety of ways. In our implementation, we follow the design illustrated in Fig. 1(b), using a 1×1 convolution followed by normalization and a nonlinearity in a manner similar to MBConv blocks [34]. The proposed blocks can be stacked to form standard CNN architectures; in particular, we adopt a stage-wise design, where multiple blocks are grouped into stages operating at decreasing spatial resolutions and increasing channel dimensions, as illustrated in Fig. 1(a).

We compared the Mom/Inv layer design with alternative invariant maps, including Flusser and the flexible basis (Appendix A). While these approaches are theoretically equivalent in terms of discriminability, the proposed basis proved more numerically stable and computationally efficient in practice. Additional nonlinearities after the invariant projection did not yield consistent improvements, indicating that the construction already provides sufficient nonlinearity, similar to tensor product interactions in equivariant networks such as e3nn [14]. We also considered normalization before or after the invariant layer; as both performed similarly, we apply it only after moment projection.

[3] This follows from complex multiplication in polar form, $c_1 c_2 = r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2))$, and de Moivre's formula [33], $c^w = r^w(\cos(w\theta) + i\sin(w\theta))$. We retain only the real part.

Figure 1: Overview of the proposed invariant convolutional architecture. (a) The overall network architecture obtained by stacking the proposed blocks. (b) The invariant convolution block, combining a moment layer and an invariant layer within a residual structure. (c) Illustration of the invariant layer, showing how moment responses are combined to produce rotation-invariant features.



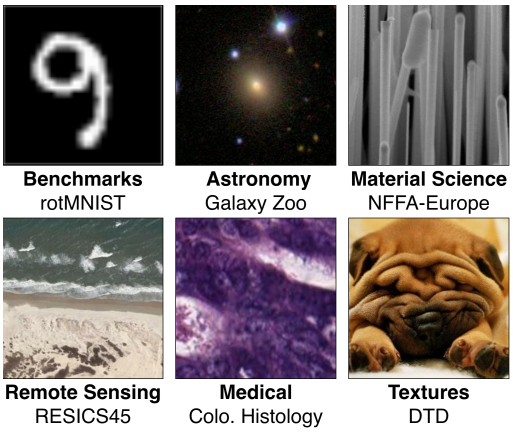


Figure 2: Benchmarks across multiple domains.

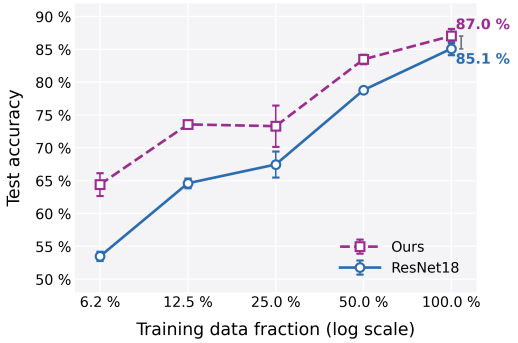


Figure 3: **Sample efficiency on NFFA.** Test accuracy vs. training data fraction. Our model (300k) compared with ResNet18 (11M, aug rot).

3 Experiments

We evaluate our method on benchmarks where rotational symmetries are intrinsic to the data, including satellite imagery, histology, material science, and textures (Fig. 2). In these domains, object orientation is arbitrary or physically irrelevant, providing a natural setting to assess whether rotational invariance improves generalization and data efficiency, particularly in low-data and capacity-constrained regimes. To further study this setting, we construct a material science benchmark based on SEM microscopy images, representing a distinct modality with limited data.

Metrics. In addition to classification accuracy, we measure the rotation consistency index (RCI) by measuring the agreement between predictions for rotated versions of the same input. Given logits for an upright input and its rotated counterpart, we compute the ℓ_2 distance:

$$\mathrm{RCI} = \|\mathbf{z}_{\mathrm{rot}} - \mathbf{z}_{\mathrm{upright}}\|_2. \tag{14}$$

Lower RCI indicates greater consistency under rotation. We found this metric to be particularly informative when comparing invariant and equivariant models. We also considered Lie derivative equivariance metrics [35], but found them less sensitive for comparing models within the class of equivariant architectures.

Table 1: **Quantitative comparison between our Mom/Inv CNN and $E(2)$-CNNs.** Results are reported as mean $\pm$ standard deviation over at least 3 runs. For DTD, we report top-five accuracy, denoted by *.

	Dataset			Ours			E2-CNNs		
Dataset	Reso.	Cls	Size	Acc	Prm.	RCI	Acc	Prm.	RCI
$\text{MNIST}_{\text{RotTest}}$	28	10	50k	$\mathbf{99.42}_{\pm.07}$	119k	$\mathbf{0.01}$	$97.20_{\pm.14}$	362k	0.89
$_{\text{rot}}\text{MNIST}$	28	10	12/50k	$\mathbf{99.08}_{\pm.04}$	119k	$\mathbf{0.01}$	$98.82_{\pm.01}$	362k	0.64
RESISC45	256	45	31.5k	$\mathbf{85.77}_{\pm.47}$	230k	$\mathbf{1.66}$	$84.02_{\pm.27}$	365k	5.13
$\text{EuroSAT}_{\text{rgb}}$	64	10	27k	$\mathbf{95.29}_{\pm.40}$	149k	$\mathbf{0.71}$	$93.90_{\pm.34}$	363k	1.60
DTD*	$\leq$640	47	5k	$67.14_{\pm.27}$	336k	$\mathbf{0.01}$	$67.18_{\pm.81}$	469k	1.97
NFFA	224	10	25k	$87.02_{\pm.87}$	248k	$\mathbf{0.01}$	$87.32_{\pm.30}$	363k	1.23
Galaxy Zoo	424	8	100/31k	$83.30_{\pm.03}$	227k	$\mathbf{0.27}$	$\mathbf{84.39}_{\pm.08}$	363k	0.82
Col. Hist.	150	8	5k	$\mathbf{92.30}_{\pm.40}$	228k	$\mathbf{0.59}$	$91.58_{\pm.41}$	363k	1.16

Datasets. We evaluate our model on a diverse set of datasets covering classical handwritten digits, remote sensing, textures, and scientific imaging. Specifically, we consider MNIST-Rot-Test [32], RotMNIST [36], RESISC45 [37], EuroSAT [38], DTD [39], Lost Oil Wells [40], Galaxy Zoo [41–43], and Colorectal Histology [44].

NFFA. In addition, we construct a dataset based on NFFA-Europe [45], designed to reflect a realistic scientific imaging scenario where data is both scarce and structurally constrained. We manually curate labels and process the images (resizing and cropping to 224×224), resulting in a dataset with a small number of samples and high intra-class variability. This setting is particularly suitable for evaluating $\mathrm{SE}(2)$-invariant architectures in the low-data regime. To further examine this behavior, we analyze performance as a function of the available training data (Fig. 3), highlighting improvements in sample efficiency. Further details are provided in Appendix G.

Baselines. We compare against $E(2)$-CNNs with continuous $\mathrm{SO}(2)$ equivariance implemented via the `e2cnn` framework, a widely used and general implementation of steerable convolutional networks. Additional baselines, including H-Nets [11], standard non-equivariant CNNs [46], and transformer-based models [47] are reported in Appendix I.

3.1 Quantitative results

Across all benchmarks, our approach achieves competitive or better accuracy compared to a baseline while using substantially fewer parameters, as shown in Table 1. In particular, we observe consistent gains on standard rotation benchmarks, as well as on more complex datasets including RESISC45 and EuroSAT. Notably, these improvements are obtained with models that are typically two to three times smaller than their equivariant counterparts.

In addition to higher accuracy, our method yields significantly lower RCI values across all datasets, indicating stronger consistency under rotations. This difference is especially pronounced on benchmarks with explicit rotations, where our invariant representations reduce RCI by an order of magnitude, demonstrating a dramatic improvement in the invariance of the classification function. Even in cases where accuracy is comparable or slightly lower (e.g., Galaxy Zoo), our method maintains substantially lower RCI and reduced parameter count. We attribute this to the invariant construction in Eq. (13), which suppresses unstable phase variations by encoding relative phase differences. We also observe that larger convolutional kernels improve rotational consistency in our method, however, applying larger kernels in `e2cnn` reduced accuracy while further increasing computational cost and parameter count. Collectively, these results show that structured equivariant feature fields are not necessary complexities for invariant classification.

3.2 Ablation Studies

We conduct ablation studies to evaluate key design choices in the proposed invariant moment architecture, including invariant basis selection (ours vs. Flusser), handling of degenerate reference

Table 2: Comparison of invariant representations. Using magnitude and phase-interaction components yields consistently better performance than using real and imaginary parts directly.

Dataset	Flusser	Ours
MnistRotTest	$99.27 \pm .05$	$\mathbf{99.42 \pm .07}$
RotMnist	$98.77 \pm .03$	$\mathbf{99.08 \pm .04}$
RESISC45	$83.33 \pm .21$	$\mathbf{85.77 \pm .47}$
EuroSAT	$93.77 \pm .36$	$\mathbf{95.29 \pm .40}$

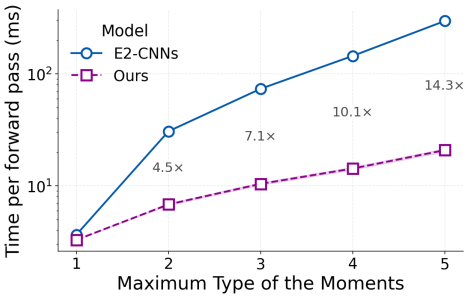


Figure 4: Layer Forward Pass Time

moments (Appendix A), radial function parameterizations (Appendix E), and the addition of an equivariant (type-1) branch (Appendix D). We also report a speed comparison and analyze sample efficiency on the NFFA benchmark, highlighting performance in the low-data regime.

Sample Efficiency. We evaluate sample efficiency by training models with various fractions of data from our NFFA benchmark. As shown in Fig. 3, our model consistently outperforms a ResNet18 baseline trained with standard rotation and flip augmentations, particularly in the low-data regime. The performance gap is most pronounced at smaller data fractions (e.g., 6.2% and 12.5%), where our approach achieves higher accuracy. As the amount of data increases, the gap narrows.

Notably, these results are obtained with a substantially smaller model: our network has approximately 300k parameters, compared to 11M for ResNet18. This further shows how incorporating group theoretical constructions into ML models can improve sample efficiency.

Speed Comparison. We compare the forward-pass runtime of our invariant moment layer against $E(2)$-CNNs as a function of the maximum moment type. Results are shown in Fig. 4. We measure equivalent building blocks: in steerable CNNs (e.g., `e2cnn`), this corresponds to two successive equivariant convolutional layers, whereas our construction uses a moment–invariant block of the form Mom/Inv $\to 1 \times 1 \to$ Mom/Inv. Our method achieves lower runtime; this behavior is consistent with the theoretical complexity. In steerable CNNs, convolutional layers require interactions between all input and output types, resulting in $\mathrm{O}(N_{\text{in}} \cdot N_{\text{out}})$ complexity. In contrast, our construction avoids explicit type-to-type interactions by projecting directly to invariant (type-0) features, leading to linear scaling in the number of moments. The empirical results closely follow this trend, with speedups reaching up to $14\times$ for higher moment orders. A minimal implementation of the invariant construction is provided in Appendix J. For example, on the NFFA benchmark with maximum type $k = 3$, the full model requires 227 seconds per epoch on a single A100 GPU using `e2cnn`, compared to 98 seconds for our method.

Invariant Construction. We analyze the effect of different invariant constructions. As shown in Table 2, representations based on moment magnitudes combined with relative phase interactions (Eq. (13)) consistently outperform those based on real and imaginary components of Flusser basis (6). In practice, we found that the Flusser-based representation requires additional normalization (e.g., a σ function applied to the magnitudes) to prevent numerical instabilities due to large magnitude values.

4 Conclusion and Limitations

We introduced an invariant architecture for constructing efficient $\mathrm{SE}(2)$-invariant networks. To connect classical invariant theory with modern deep learning, we derived a minimal and discriminative invariant map that preserves all information up to the $\mathrm{SE}(2)$ group action while avoiding redundant features. We translated this construction into a plug-and-play CNN module that projects moments directly into invariant type-0 features, enabling seamless integration with standard CNN layers such as pointwise nonlinearities, normalization layers, and 1×1 convolutions.

Empirically, the proposed models achieve competitive or improved accuracy across a range of rotation-symmetric benchmarks while using substantially fewer parameters and with lower runtime than steerable equivariant baselines. These gains are particularly evident in settings involving low

data and scientific imaging, where symmetry priors improve sample efficiency. Overall, the results indicate that invariant maps can provide an effective and computationally efficient alternative to fully equivariant features. Notably, performance remains strong despite enforcing local invariance, indicating that relative orientation can still be inferred from invariant features computed over the continuous spatial domain (Appendix D).

Despite these advantages, the approach is most beneficial for tasks with strong rotational symmetry, where orientation is a nuisance variable. In high-data regimes, similar robustness can often be achieved through data augmentation and increased model capacity. As the theory assumes continuous transformations, invariance is only approximate in discretized implementations. Finally, our invariant map relies on non-zero moment responses; features may become unreliable when moment magnitudes are very small. In Appendix A, we discuss solutions to this issue and show that such cases are rare in practice and can be mitigated by computations over the continuous spatial domain.

Future work may explore adaptive selection of moment types, extensions to broader transformation groups, and applications that combine invariant and equivariant representations. In particular, extending the approach to three-dimensional settings, for example by projecting 3D features through 2D as explored in prior work [18], is a promising direction.

5 Related Work

Moment Invariants and Classical Pattern Recognition. The construction of invariant features has a long history in pattern recognition. Hu introduced one of the first sets of rotation-invariant moments, demonstrating their effectiveness for shape recognition [29]. Complementary to moment-based invariants, steerable filters provided an early framework for handling rotations by expressing filter by linear combinations of a fixed set of basis filters [48]. This idea introduced a classical form of rotation-equivariant feature extraction and later became an important concept to steerable convolutional architectures. Many early hand-designed invariant sets were redundant or incomplete. Flusser and collaborators formalized the notion of invariant feature systems that are both functionally independent and complete, meaning they can generate any invariant [24, 25]. Wood already pointed out that invariant representations should not only remove nuisance redundancy but also retain discriminative information. Subsequent work further explored completeness and independence [20], as well as extensions to more general settings, including tensor fields and flexible invariant bases that avoid degenerate cases [26, 30].

Equivariant Convolutional Networks. A large body of work extends CNNs to encode symmetries beyond translations by enforcing equivariance to transformation groups. Group equivariant CNNs generalize convolutions to discrete and compact groups [5]. Subsequent work extended these ideas to continuous groups, including planar rotations and Euclidean groups, using steerable filters and tools from representation theory. For instance, harmonic networks achieve rotation equivariance by constraining filters to a circular harmonics basis [11]. More general frameworks such as steerable CNNs formalize this approach by constructing feature spaces that transform according to irreducible representations of the underlying group, often referred to as steerable fields [15–17, 49]. Libraries such as `e2cnn` [15] and `escnn` [16] provide practical implementations of these methods. These models typically introduce additional computational and architectural complexity compared to standard CNNs, and are most commonly applied in 2D settings [50–54].

Beyond $\mathrm{SE}(2)$ equivariant CNNs, related work has also explored equivariance in higher-dimensional domains and non-convolutional architectures, including 3D steerable CNNs, tensor field networks, $\mathrm{SE}(3)$-equivariant transformers, and equivariant graph/attention models for point clouds, molecular systems, and physical simulation [6, 8–10, 13, 14, 17, 55–57].

Invariant Features in Deep Learning. An alternative to equivariance is to construct invariant representations directly. Early methods achieved invariance by pooling over transformed inputs, as in TI-Pooling [58], by learning canonicalizing transformations, as in scattering transforms [59]. Other architectures introduced orientation-aware or rotation-equivariant feature extraction within CNNs, including Oriented Response Networks [60] and Rotation Equivariant Vector Field Networks [61]. Recent work has revisited invariance within standard CNNs through analytically defined Bessel kernels [62] and rotation-invariant coordinate convolutions [63]. However, these methods typically

rely on pooling, weight sharing, learned canonicalization, or specific filter parameterizations, and do not explicitly construct complete and non-redundant invariant feature spaces in the classical sense.

Acknowledgemetns We thank Kei Davis for feedback on the manuscript. This work is published under LANL LA-UR-26-23375. We gratefully acknowledge the support of the U.S. Department of Energy through the LANL Laboratory Directed Research Development Program under project number 20250145ER.

References

[1] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[2] Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023. URL `http://udlbook.com`.

[3] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[4] Johann Brehmer, Sönke Behrends, Pim De Haan, and Taco Cohen. Does equivariance matter at scale? *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL `https://openreview.net/forum?id=wilNute8Tn`.

[5] Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `https://proceedings.mlr.press/v48/cohenc16.html`.

[6] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–gordan nets: a fully fourier space spherical convolutional neural network. *Advances in Neural Information Processing Systems*, 31, 2018.

[7] Daniel Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/f04cd7399b2b0128970efb6d20b5c551-Paper.pdf`.

[8] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33:1970–1981, 2020.

[9] Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve e(3) equivariant message passing. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=_xwr8gOBeV1`.

[10] Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.

[11] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5028–5037, 2017.

[12] Yi-Lun Liao, Alexander J. Hoffman, Sabrina C. Shen, Alexandre Duval, Sam Walton Norwood, and Tess Smidt. Equiformerv3: Scaling efficient, expressive, and general se(3)-equivariant graph attention transformers. *arXiv preprint arXiv:2604.09130*, 2026.

[13] Yi-Lun Liao, Brandon Wood, Abhishek Das*, and Tess Smidt*. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=mCOBKZmrzD.

[14] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=KwmPfARgOTD.

[15] Maurice Weiler and Gabriele Cesa. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019. URL https://arxiv.org/abs/1911.08251.

[16] Gabriele Cesa, Leon Lang, and Maurice Weiler. A program to build E(N)-equivariant steerable CNNs. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=WE4qe9xlnQw.

[17] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/488e4104520c6aab692863cc1dba45af-Paper.pdf.

[18] Saro Passaro and C Lawrence Zitnick. Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs. In *International Conference on Machine Learning (ICML)*, 2023.

[19] Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15546–15566. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

[20] Jan Flusser, Barbara Zitova, and Tomas Suk. *Moments and moment invariants in pattern recognition*. John Wiley & Sons, 2009.

[21] William K Pratt. *Introduction to digital image processing*. CRC press, 2013.

[22] Rafael C Gonzalez. *Digital image processing*. Pearson/Prentice Hall, NY, 2009.

[23] Jeffrey Wood. Invariant pattern recognition: A review. *Pattern Recognition*, 29(1):1–17, 1996. ISSN 0031-3203. doi: https://doi.org/10.1016/0031-3203(95)00069-0. URL https://www.sciencedirect.com/science/article/pii/0031320395000690.

[24] Jan Flusser. On the independence of rotation moment invariants. *Pattern Recognition*, 33(9): 1405–1410, 2000. ISSN 0031-3203. doi: https://doi.org/10.1016/S0031-3203(99)00127-2. URL https://www.sciencedirect.com/science/article/pii/S0031320399001272.

[25] Jan Flusser. On the inverse problem of rotation moment invariants. *Pattern Recognition*, 35 (12):3015–3017, 2002. ISSN 0031-3203. doi: https://doi.org/10.1016/S0031-3203(02)00093-6. URL https://www.sciencedirect.com/science/article/pii/S0031320302000936.

[26] Roxana Bujack, Xinhua Zhang, Tomáš Suk, and David Rogers. Systematic generation of moment invariant bases for 2d and 3d tensor fields. *Pattern Recognition*, 123:108313, 2022. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2021.108313. URL https://www.sciencedirect.com/science/article/pii/S0031320321004933.

[27] Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. *Equivariant and Coordinate Independent Convolutional Networks*. 2023. URL https://maurice-weiler.gitlab.io/cnn_book/EquivariantAndCoordinateIndependentCNNs.pdf.

[28] F. Peter and H. Weyl. Die vollständigkeit der primitiven darstellungen einer geschlossenen kontinuierlichen gruppe. *Mathematische Annalen*, 97(1):737–755, Dec 1927. ISSN 1432-1807. doi: 10.1007/BF01447892. URL https://doi.org/10.1007/BF01447892.

[29] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962. doi: 10.1109/TIT.1962.1057692.

[30] Roxana Bujack and Jan Flusser. Flexible moment invariant bases for 2D scalar and vector fields. *Proceedings of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 590, 2017.

[31] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[32] Tomáš Karella, Filip Šroubek, Jan Blažek, Jan Flusser, and Václav Košík. H-next: The next step towards roto-translation invariant networks. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. URL https://papers.bmvc2023.org/0578.pdf.

[33] Abraham De Moivre. Iii. Æquationum quarundam potestatis tertiæ, quintæ, septimæ, nonæ, & superiorum, ad infinitum usque pergendo, in terminis finitis, ad instar regularum pro cubicis quæ vocantur cardani, resolutio analytica. *Philosophical Transactions*, 25(309):2368–2371, 03 1707. ISSN 0260-7085. doi: 10.1098/rstl.1706.0037. URL https://doi.org/10.1098/rstl.1706.0037.

[34] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[35] Nate Gruver, Marc Anton Finzi, Micah Goldblum, and Andrew Gordon Wilson. The lie derivative for measuring learned equivariance. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=JL7Va5Vy15J.

[36] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *ICML*, 2007. URL https://www.cs.toronto.edu/~larocheh/publications/deep-nets-icml-07.pdf.

[37] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 2017. URL https://huggingface.co/datasets/timm/resisc45.

[38] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset for land use and land cover classification. *IEEE J-STARS*, 2019. URL https://huggingface.co/datasets/timm/eurosat-rgb.

[39] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, and Andrea Vedaldi. Describing textures in the wild. *CVPR*, 2014. URL https://www.robots.ox.ac.uk/~vgg/data/dtd/.

[40] Anastasiia Kim, Teeratorn Kadeethum, Christine Downs, Hari S Viswanathan, and Daniel O'Malley. Aerial imagery dataset of lost oil wells. *Scientific Data*, 11(1):1005, September 2024. URL https://www.nature.com/articles/s41597-024-03820-0.

[41] Galaxy zoo dataset, 2024. URL https://huggingface.co/datasets/mrJordi0/galaxy-zoo-dataset.

[42] Joshua Yao-Yu Lin, Song-Mao Liao, Hung-Jin Huang, Wei-Ting Kuo, and Olivia Hsuan-Min Ou. Galaxy morphological classification with efficient vision transformer, 2021. URL https://arxiv.org/abs/2110.01024.

[43] Mike Walmsley, Chris Lintott, Tobias Géron, Sandor Kruk, Coleman Krawczyk, Kyle W. Willett, Steven Bamford, Hugh Dickinson, Lucy Fortson, Yarin Gal, William Keel, Karen L. Masters, Vihang Mehta, Brooke D. Simmons, Rebecca Smethurst, Elisabeth M L Baeten, and Christine Macmillan. Galaxy zoo decals: Detailed visual morphology measurements from volunteers and

deep learning for 314,000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 509(3): 3966–3988, 2022. doi: 10.1093/mnras/stab2093. URL https://doi.org/10.1093/mnras/stab2093.

[44] Colorectal histology dataset, 2024. URL https://huggingface.co/datasets/dpdl-benchmark/colorectal_histology.

[45] Rossella Aversa, Mohammad Hadi Modarres, Stefano Cozzini, and Regina Ciancio. Nffa-europe - 100% sem dataset, October 2018. URL https://doi.org/10.23728/b2share.80df8606fcdb4b2bae1656f0dc6db8ba.

[46] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[47] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

[48] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13:891–906, 1991. URL https://api.semanticscholar.org/CorpusID:29187618.

[49] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2018.

[50] Erik J Bekkers, Maxime W Lafarge, Mitko Veta, Koen AJ Eppenhof, Josien PW Pluim, and Remco Duits. Roto-translation covariant convolutional networks for medical image analysis. In *International conference on medical image computing and computer-assisted intervention*, pages 440–448. Springer, 2018.

[51] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2786–2795, 2021.

[52] Sneh Pandya, Purvik Patel, Jonathan Blazek, et al. E(2) equivariant neural networks for robust galaxy morphology classification. *arXiv preprint arXiv:2311.01500*, 2023.

[53] Arnab Kumar Mondal, Siba Smarak Panigrahi, Oumar Kaba, Sai Rajeswar Mudumba, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models. *Advances in Neural Information Processing Systems*, 36:50293–50309, 2023.

[54] Emanuele Santellani, Christian Sormann, Mattia Rossi, Andreas Kuhn, and Friedrich Fraundorfer. S-trek: Sequential translation and rotation equivariant keypoints for local feature extraction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9728–9737, 2023.

[55] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

[56] Erik J Bekkers, Sharvaree Vadgama, Rob Hesselink, Putri A Van der Linden, and David W. Romero. Fast, expressive $\mathrm{SE}(n)$ equivariant networks through weight-sharing in position-orientation space. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=dPHLbUqGbr.

[57] Alice E. A. Allen, Emily Shinkle, Roxana Bujack, and Nicholas Lubbers. Optimal invariant sets for atomistic machine learning. *npj Computational Materials*, 12(1):75, Jan 2026. ISSN 2057-3960. doi: 10.1038/s41524-025-01948-0. URL https://doi.org/10.1038/s41524-025-01948-0.

[58] Dmitry Laptev, Nikolay Savinov, Joachim M. Buhmann, and Marc Pollefeys. Ti-pooling: Transformation-invariant pooling for feature learning in convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[59] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[60] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[61] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[62] Valentin Delchevalerie, Adrien Bibal, Benoît Frénay, and Alexandre Mayer. Achieving rotational invariance with bessel-convolutional neural networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 28772–28783. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/f18224a1adfb7b3dbff668c9b655a35a-Paper.pdf.

[63] Hanlin Mo and Guoying Zhao. Ric-cnn: Rotation-invariant coordinate convolutional neural network. *Pattern Recognition*, 146:109994, 2024. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2023.109994. URL https://www.sciencedirect.com/science/article/pii/S0031320323006921.

[64] Sergey N. Pozdnyakov, Michael J. Willatt, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Phys. Rev. Lett.*, 125:166001, Oct 2020. doi: 10.1103/PhysRevLett.125.166001. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.166001.

[65] Roxana Bujack, Emily Shinkle, Alice Allen, Tomas Suk, and Nicholas Lubbers. Flexible moment-invariant bases from irreducible tensors, 2025. URL https://arxiv.org/abs/2503.21939. arXiv:2503.21939 [cs.CV], https://arxiv.org/abs/2503.21939.

[66] Chaitanya K Joshi, Cristian Bodnar, Simon V Mathis, Taco Cohen, and Pietro Lio. On the expressive power of geometric graph neural networks. In *International conference on machine learning*, pages 15330–15355. PMLR, 2023.

[67] Godfrey Harold Hardy and Edward Maitland Wright. *An introduction to the theory of numbers*. Oxford university press, 1979.

[68] David Romero, Erik Bekkers, Jakub Tomczak, and Mark Hoogendoorn. Attentive group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 8188–8199. PMLR, 2020.

[69] Gudrun Schwarzer. Development of face processing: The effect of face inversion. *Child development*, 71(2):391–401, 2000.

[70] John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pages 1–29. Springer, 2003.

A Flexible Bases

Flexibility was first formalized by Bujack et al. [30], who observed that the Flusser basis (Eq. (6)) cannot distinguish patterns for which all type-1 moments vanish. This occurs, for example but not exclusively, for n-fold symmetric patterns. For example, in the four-particle configuration in Fig. 5, rotations affect only the phase of $m_4 = \mu(4, 0)$, while the reference moment $m_1 = \mu(1, 0)$ remains zero. Since the Flusser basis uses m_1 as a fixed phase reference, the resulting invariants become degenerate for such patterns. Similar phenomena arise in three-dimensional settings, including atomic structure representations [64, 65] and geometric graph neural networks [66], where distinct configurations may become indistinguishable under insufficiently expressive invariant/equivariant features.

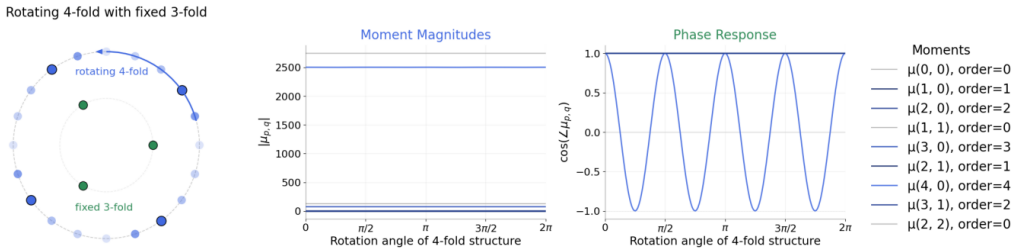


Figure 5: Flexible benchmark and its invariant representation. Left: data generation from rotated three-fold and four-fold structures, with labels determined by their minimal distance. Right: projection into Flusser complex moments, showing the structured embedding induced by invariant features. Inspired by graph neural network benchmarks [66].

This motivates the flexible basis [30]. Rather than fixing m_1 as the reference moment, the flexible construction selects a nonzero moment m_k as the phase reference. This keeps the representation informative when lower-order moments vanish. If the degeneracy structure is unknown, one can consider each nonzero moment as a possible reference, yielding a larger flexible invariant set at the cost of a quadratic increase in moment combinations. The flexible basis by Bujack et al. [30] has two shortcomings. The first is that it requires a-priori knowledge about the input pattern. The second is that it requires keeping multiple frames in case of rotationally symmetric patterns. In the next section, we will show that we can overcome both.

A.1 Synthetic flexibility benchmark

We test this effect in a pixel-based synthetic classification task. Each sample contains a three-fold and a four-fold structure on circles of different radii. The binary label is determined by their minimal distance, split at the median to obtain a balanced dataset. The point configurations are projected onto a regular grid and smoothed with a Gaussian kernel to reduce interpolation artifacts from discretization and rotation.

We evaluate a simple convolutional model in which a single Mon/Inv-layer filter spans the full spatial resolution of the image under valid convolution, producing a 1×1 feature map that is passed to an MLP classifier. With the non-flexible Flusser basis (Eq. (6)) and non-flexible optimal basis (Eq. (13)), the model achieves only chance-level accuracy, approximately 50%. In contrast, applying the flexible construction to the optimal basis increases accuracy to approximately 99%.

Interestingly, adding padding allows the non-flexible bases to recover performance comparable to the flexible bases. Padding introduces boundary context that can break the perfect symmetry seen by the valid convolution, thereby compensating for the degenerate reference moment. Thus, padding acts as an architectural shortcut: it restores empirical performance without resolving the underlying limitation of the non-flexible basis.

A.2 Implications for Dense Invariant CNNs

In our CNNs, replacing the optimal basis with its flexible variant did not lead to measurable improvements in classification accuracy. To understand this, we measured how often candidate reference

moments vanish across channels and spatial locations. Table 3 reports the fraction of near-zero moments for several datasets.

The observed frequencies are extremely small across all datasets and moment orders. This suggests that, in dense pixel or feature-map representations, exact moment degeneracies are rare. Unlike sparse point or graph representations, where highly symmetric configurations can make reference moments vanish exactly, spatially extended convolutional features and image discretization tend to break such degeneracies.

Together with the synthetic benchmark in Fig. 5, these results suggest that flexibility primarily matters in settings with sparse or highly symmetric inputs, where a chosen reference moment may vanish while alternative moments remain available. In contrast, such degeneracies occur only rarely in the dense feature representations produced by our CNNs (tested before and after training). This differs from settings such as geometric graph neural networks [66], where sparsity persists throughout the network and indistinguishable configurations arise more frequently.

Dataset	$\mathbf{m_{1,1}}$	$\mathbf{m_{2,0}}$	$\mathbf{m_{2,1}}$	$\mathbf{m_3}$	$\mathbf{m_4}$
Histology	$8.7\text{e}{-5}$	$2.7\text{e}{-4}$	$2.1\text{e}{-4}$	$7.1\text{e}{-5}$	$5.1\text{e}{-5}$
MNIST	$1.4\text{e}{-6}$	$1.4\text{e}{-6}$	$1.4\text{e}{-6}$	$1.4\text{e}{-6}$	$1.4\text{e}{-6}$
Satellite	$1.8\text{e}{-5}$	$2.2\text{e}{-4}$	$4.9\text{e}{-4}$	$2.3\text{e}{-6}$	$1.4\text{e}{-6}$

Table 3: Conditional probability that another moment is nonzero given that the reference moment $m_{1,0}$ is zero, measured across channels and spatial locations. Larger values indicate cases where flexible reference selection could avoid degeneracies caused by a vanishing reference moment. The consistently small probabilities indicate that situations where an alternative reference moment could be used are exceedingly rare in these CNN feature representations.

A.3 SB–PB comparison.

Table 4 compares the single-reference basis (SB), used in the main text, with the flexible pairwise quadratic basis (PB), using matched parameter counts. SB achieves comparable or better accuracy across these benchmarks while retaining lower computational complexity.

Table 4: Comparison between the single-reference basis (SB) and pairwise quadratic basis (PB). Results are reported as mean accuracy and standard deviation over three runs.

Dataset	Basis	Accuracy
Col. Hist.	PB	$90.43 \pm .56$
	SB	$91.06 \pm .16$
rotMNIST	PB	$98.91 \pm .04$
	SB	$98.97 \pm .01$
EuroSAT	PB	$94.10 \pm .11$
	SB	$95.29 \pm .40$

B Optimal Invariant Descriptors

In this appendix, we show how to derive an optimal map of descriptors, by which we mean that they are discriminative and minimal, i.e., the descriptor is able to distinguish any two inputs that differ by something other than a rotation, and that removing any element breaks that property [23].

After some notation, we will first introduce the **Bézout map**, which can discriminate a known input from every other pattern. Then, we describe the **minimal discriminative map**. We prove that it achieves general discriminability without a-priori knowledge by deriving any Bézout map from its descriptors. We also prove invariance and minimality for both $\mathrm{O}(2)$ and $\mathrm{SO}(2)$.

Note that discriminability refers to a set of admissible input patterns. Generally, we consider patterns only up to a maximal finite order. Flusser's basis can discriminate any pattern that has a non-zero

moment of type one. The Bézout map is more powerful. It can discriminate any pattern as long as we know it a-priori. The flexible map is the most powerful one. It can discriminate any patterns without prior knowledge.

B.1 From Complex to Fourier Moments

Complex moments are typically used with two indices [24]. To make notation more concise and generalize to moments with other radial parts, such as Zernike moments, in this section we split angular from radial parts.

Let $f : \mathbb{R}^2 \to \mathbb{R}$, and write $z = x + iy = re^{i\theta}$. Assume the following moments exist. The complex moments are

$$c_{p,q} = \int_{\mathbb{R}^2} z^p \bar{z}^q f(x, y)\, dx\, dy. \tag{15}$$

In polar coordinates,

$$c_{p,q} = \int_0^\infty \int_0^{2\pi} r^{p+q} e^{i(p-q)\theta} f(r, \theta)\, r\, d\theta\, dr. \tag{16}$$

Thus under rotation $f_\alpha(r, \theta) = f(r, \theta - \alpha)$ the moment $c_{p,q}$ transforms with type

$$k := p - q, \tag{17}$$

and

$$c_{p,q}(f_\alpha) = e^{ik\alpha} c_{p,q}(f). \tag{18}$$

Hence we reduce to Fourier moments

$$m_k := \int_0^{2\pi} f(\theta) e^{ik\theta}\, d\theta, \qquad k \in \mathbb{Z}. \tag{19}$$

For real-valued f,

$$m_{-k} = \overline{m_k}. \tag{20}$$

Under rotation $f_\alpha(\theta) = f(\theta - \alpha)$, the moments transform via

$$m_k(f_\alpha) = \int_0^{2\pi} f(\theta - \alpha) e^{ik\theta}\, d\theta \overset{u=\theta-\alpha}{=} \int_0^{2\pi} f(u) e^{ik(u+\alpha)}\, du = e^{ik\alpha} m_k(f). \tag{21}$$

Under reflection $\tilde{f}(\theta) = f(-\theta)$, the moments transform via

$$m_k(\tilde{f}) = \int_0^{2\pi} f(-\theta) e^{ik\theta}\, d\theta \overset{u=-\theta}{=} \int_0^{2\pi} f(u) e^{-iku}\, du = m_{-k}(f) \overset{(20)}{=} \overline{m_k(f)}. \tag{22}$$

B.2 Bézout Map

In this section, we show that a basis of descriptors can be chosen if the input is known a-priori. For a given set of input moments, we derive a map that discriminates it from any other set of moments using Bézout's identity [67].

Let

$$M = \{m_k \mid k \in \mathbb{N},\ k \le N_k\} \tag{23}$$

be the set of Fourier moments of a real-valued function

$$f : S^1 \to \mathbb{R} \tag{24}$$

up to order $N_k \in \mathbb{N}$. In case all moments are zero, we define all descriptors to be zero as well. From now on, we assume at least one moment is non-zero.

Let

$$S := \{\, k \mid m_k \neq 0,\ k \in \mathbb{N},\ k \le N_k \,\} \tag{25}$$

be the set of indices with non-zero moments, and let

$$d = \gcd(S) \tag{26}$$

be their greatest common divisor. Then all non-zero moment indices are multiples of d,

$$\forall\, m_k \neq 0 : \quad d \mid k, \quad \text{i.e.} \quad \frac{k}{d} \in \mathbb{N}. \tag{27}$$

If zero moments are present, we set their Bézout coefficients to zero, i.e., $a_k := 0$. By Bézout's identity there exist integers $a_k \in \mathbb{Z}$ such that

$$\sum_{k=0}^{N_k} a_k\, k = d. \tag{28}$$

This yields a rotational normalizer of type d,

$$m' := \prod_{k=0}^{N_k} m_k^{a_k}, \qquad \text{type}(m') = \sum_{k=0}^{N_k} a_k k = d. \tag{29}$$

To simplify the inverse problem, we normalize m' to unit magnitude:

$$m := \frac{m'}{\sqrt{m'\overline{m'}}}, \tag{30}$$

where the square root denotes the positive one.

The Bézout map is then defined as

$$\Psi := \{\, \psi_k \mid k = 0, \ldots, N_k \,\}. \tag{31}$$

For each k, we define

$$\psi_k = m_k^{\frac{d}{\gcd(d,k)}}\, m^{\frac{-k}{\gcd(d,k)}}. \tag{32}$$

Since $d \mid k$ for all non-zero moments, we have $\gcd(d,k) = d$, and thus

$$\psi_k = m_k^{\frac{d}{d}}\, m^{-\frac{k}{d}} = m_k\, m^{-\frac{k}{d}}. \tag{33}$$

This expression is well-defined since $\frac{k}{d} \in \mathbb{N}$.

B.2.1 Invariance

Theorem 1. *The Bézout map is invariant to rotations.*

Proof. We know from (21) that under rotations $\theta \mapsto \theta - \alpha$

$$m_k \mapsto e^{ik\alpha} m_k, \qquad m \mapsto e^{ig\alpha} m. \tag{34}$$

Hence

$$\psi_k = m_k m^{-k/d} \mapsto e^{ik\alpha} m_k \left(e^{id\alpha} m\right)^{-k/d} = e^{ik\alpha} m_k e^{-ik\alpha} m^{-k/d} = \psi_k. \tag{35}$$

□

The Bézout map is not invariant to reflections. It follows from (22) that the imaginary part of each descriptor changes sign upon reflection while its real part remains unchanged. Therefore, the Bézout map can be easily adapted to be invariant to rotations and reflections by replacing each complex descriptor by its magnitude and real part

$$\Psi_O := \{\, |\psi_k|, \Re(\psi_k) \mid k = 0, \ldots, N_k \,\}. \tag{36}$$

A more flexible/robust choice, though, is to use the magnitude of the moments and the real part. If a pattern with a different support is used than the one for which the Bézout map was created, its normalizer can vanish ($m = 0$), which causes every element of the Bézout map to vanish, except the trivial ones with $k = 0$, because they are all products of m. Instead, using

$$\Psi'_O := \{\, |m_k|, \Re(\psi_k) \mid k = 0, \ldots, N_k \,\} \tag{37}$$

keeps all the magnitudes usable, maintaining half the information.

B.2.2 Discriminability

Theorem 2. *The Bézout map is discriminative.*

Proof. The full Fourier transform is injective on the function class considered. Since we only use moments up to order N_k, it suffices here to prove discriminability of the truncated moment vectors up to the rotation action.

We solve the inverse problem analogously to Flusser [25] by working in the canonical gauge, i.e., in the chosen standard position where $m' \in \mathbb{R}^+$. Then, we immediately know that $m = 1$ and the reconstruction of the moments becomes trivial

$$m_k = \psi_k m^{k/d} = \psi_k. \tag{38}$$

□

B.2.3 Minimality

The Bézout Map is not in general independent or minimal. Assume that there exists an index l such that $a_l = 1$, then we can replace the complex ψ_l with only its norm $|\psi_l|$.

We construct ψ_l in the canonical gauge. There the normalized Bézout normalizer satisfies $m' = |m'| \in \mathbb{R}^+$, which leads to $m = 1$ and

$$\psi_k \overset{(33)}{=} m_k m^{\frac{-k}{d}} = m_k. \tag{39}$$

Using (29) and $|\psi_l|$, we can first compute

$$m' = |m'| = \prod_{k=0}^{N_k} |m_k^{a_k}|. \tag{40}$$

Then since $a_l = 1$, we can isolate m_l

$$\psi_l \overset{(39)}{=} m_l \overset{(29)}{=} m' \prod_{\substack{k=0 \\ k \neq l}}^{N_k} m_k^{-a_k} \overset{(39)}{=} m' \prod_{\substack{k=0 \\ k \neq l}}^{N_k} \psi_k^{-a_k}. \tag{41}$$

An analogous result can be derived for $a_l = -1$. In both cases, the map

$$\Psi_{min} := \{\, \psi_0, ..., \psi_{l-1}, |\psi_l|, \psi_{l+1}, ..., \psi_{N_k} \} \tag{42}$$

is minimally discriminative.

B.2.4 Relation to Flusser's Basis

If $m_1 \neq 0$, we have $d = 1$ and can directly choose $a_1 = 1, a_2 = 0, ..., a_{N_k} = 0$, which leads to $m' = m_1$,

$$m = \frac{m_1}{|m_1|} = \left(\frac{\overline{m_1}}{|m_1|} \right)^{-1} \tag{43}$$

and

$$\psi_k \overset{(33)}{=} m_k m^{\frac{-k}{d}} \overset{(29)}{=} m_k \left(\frac{\overline{m_1}}{|m_1|} \right)^k. \tag{44}$$

This is exactly the Flusser basis [24] up to normalization.

Analogously to Sec. B.2.1, we can make Flusser's basis

$$\Phi_F = \{\, |m_0|, |m_1|, \psi_k \mid k = 2, \ldots, o \} \tag{45}$$

invariant to rotations and reflections by

$$\{\, |m_0|, |m_1|, |\psi_k|, \Re(\psi_k) \mid k = 2, \ldots, N_k \} \tag{46}$$

and more flexible and robust via

$$\Psi_{\mathcal{O}} = \{\, |m_0|, |m_1|, |m_k|, \Re(\psi_k) \mid k = 2, \ldots, N_k \}. \tag{47}$$

For a pattern with a $m_1 = 0$, all nontrivial elements of Flusser's basis vanish because they contain powers of m_1. We use the flexible one, which maintains all magnitude information.

Note that this variant is not fully discriminative. However, among descriptors that retain only one phase descriptor per order, it is optimal in the sense that it preserves all available magnitude information while keeping the descriptor cardinality fixed. In the next part, we will see how full discriminability can be achieved by adding more descriptors.

B.3 The Minimal Discriminative Map

The previous subsection assumed that we know the input pattern. In a general machine learning model, this assumption does not hold. In this subsection, we derive a minimal discriminative set for arbitrary, a-priori unknown input patterns. We will show that there is no basis, because all discriminative descriptor sets are algebraically dependent. Still, there is an optimal map of descriptors, as it is the smallest that can discriminate any two inputs.

We construct the minimal flexible set in the following way. Let

$$M = \{m_k \mid k \in \mathbb{Z},\ |k| \leq N_k\} \tag{48}$$

be the set of Fourier moments of an arbitrary real-valued function

$$f : S^1 \to \mathbb{R} \tag{49}$$

up to a given order $N_k \in \mathbb{N}$.

The minimal flexible set consists of magnitudes and all pairwise combinations

$$\Phi := \{\, |m_k| \mid k = 0, \ldots, N_k \,\} \ \cup\ \{\phi(k,j) \mid 1 \leq k < j \leq N_k\}, \tag{50}$$

of the form

$$\phi(k,j) := m_k^{\frac{j}{d_{kj}}}\, m_j^{-\frac{k}{d_{kj}}} \tag{51}$$

with

$$d_{kj} := \gcd(k,j). \tag{52}$$

If $m_j = 0$, we define $\phi(k,j) := 0$. We need to prove invariance, discriminability, and minimality.

B.3.1 Invariance

Theorem 3. *The minimal discriminative map Φ is invariant under rotations.*

Proof. We show that every element of Φ is invariant under planar rotations $f_\alpha(\theta) := f(\theta - \alpha)$. First, the magnitudes are invariant because of

$$|m_k(f_\alpha)| \overset{(21)}{=} |e^{ik\alpha} m_k(f)| = |m_k(f)|. \tag{53}$$

Then, for any two moments $m_k, m_j \in M$ with $k, j \geq 1$, the mixed terms satisfy

$$\begin{aligned}
\phi_\alpha(k,j) &\overset{(51)}{=} m_k(f_\alpha)^{j/d_{kj}}\ m_j(f_\alpha)^{-k/d_{kj}} && (54)\\
&\overset{(21)}{=} \left(e^{ik\alpha} m_k(f)\right)^{j/d_{kj}}\ \left(e^{ij\alpha} m_j(f)\right)^{-k/d_{kj}} && (55)\\
&= e^{ikj\alpha/d_{kj}}\ e^{-ijk\alpha/d_{kj}}\ m_k(f)^{j/d_{kj}} m_j(f)^{-k/d_{kj}} && (56)\\
&\overset{(51)}{=} \phi(k,j). && (57)
\end{aligned}$$

Thus each $\phi(k,j)$ is rotation invariant, and consequently the entire map Φ is invariant under planar rotations. □

Φ is not invariant to reflections. It follows from (22) that the imaginary part of each mixed term changes sign upon reflection while its real part remains unchanged. Therefore, Φ can be easily adapted to be invariant to rotations and reflections by replacing each complex descriptor by its magnitude and real part

$$\Phi_O := \{\, |m_k| \mid k = 0, \ldots, N_k \,\} \ \cup\ \{|\phi(k,j)|, \Re(\phi(k,j)) \mid 1 \leq k < j \leq N_k\}. \tag{58}$$

B.3.2 Discriminability

Theorem 4. *The minimal discriminative map Φ is discriminative.*

Proof. We show that any Bézout invariant ψ_k can be constructed from the flexible set Φ. Then, discriminability of Φ follows from the discriminability of the Bézout map.

First, we derive all $\phi(k,j)$ with $k > j$, which are not part of Φ, from

$$\phi(k,j) \overset{(51)}{=} \phi(j,k)^{-1} \tag{59}$$

if $\phi(j,k) \neq 0$ and zero otherwise.

Then since all magnitudes $|m_k|$ are part of Φ, we know which moments are non-zero and part of the support S (25), so we know the type d (26) and Bézout coefficients a_j (28). For $k \in S$, we obtain

$$\prod_{j\in S} \phi(k,j)^{a_j d_{kj}/d} \overset{(51)}{=} \prod_{j\in S} \left(m_k^{j/d_{kj}} m_j^{-k/d_{kj}} \right)^{a_j d_{kj}/d} \tag{60}$$

$$= m_k^{\sum_{j\in S} a_j j/d} \prod_{j\in S} m_j^{-a_j k/d} \tag{61}$$

$$\overset{(28)}{=} m_k \prod_{j\in S} m_j^{-a_j k/d} \tag{62}$$

$$\overset{(29)}{=} m_k \, (m')^{-k/d}. \tag{63}$$

Since all magnitudes $|m_j|$ are part of Φ, we can derive

$$|m'| \overset{(29)}{=} \left| \prod_{j\in S} m_j^{a_j} \right| = \prod_{j\in S} |m_j|^{a_j}. \tag{64}$$

Together, we obtain

$$\psi_k \overset{(33)}{=} m_k m^{-k/g} \overset{(30)}{=} m_k \left(\frac{m'}{|m'|} \right)^{-k/g} = \left(\prod_{j\in S} \phi(k,j)^{a_j d_{kj}/d} \right) |m'|^{k/d}. \tag{65}$$

Thus every Bézout invariant ψ_k can be constructed from Φ. For $k \notin S$, both m_k and ψ_k vanish. Hence Φ is discriminative. □

B.3.3 Minimality

Theorem 5. *The minimal discriminative map Φ is minimal.*

Proof. We will show that removing any element causes two sets of moments that differ by something other than a rotation to be indistinguishable.

Removing a magnitude $|m_k|$ leads to the moment sets $m_k^{(1)} = 1$, others zero becoming indistinguishable from $m_k^{(2)} = 2$, others zero because except for $|m_k|$, all elements of Φ are zero for both.

Fix an admissible pair (k,j) and consider $\Phi' := \Phi \setminus \{\phi(k,j)\}$. We exhibit the two *different* moment sets. For set 1

$$m_k^{(1)} = 1, \quad m_j^{(1)} = 1, \tag{66}$$

others zero, and for set 2

$$m_k^{(2)} = e^{i\pi d_{kj}/j}, \quad m_j^{(2)} = 1, \tag{67}$$

others zero.

Clearly, all magnitudes agree. For any remaining descriptor $\phi(p,q) \in \Phi'$, if $\{p,q\} \not\subseteq \{k,j\}$, then at least one involved moment is zero in both moment sets, hence both descriptors are zero. Thus all descriptors in Φ' coincide, while $\phi(k,j)$ separates the two moment sets.

We show by contradiction that the moments are not related by rotation $e^{i\alpha}$. Assume a rotation exists. Then for some α,

$$m_k^{(2)} = e^{ik\alpha} m_k^{(1)}, \qquad m_j^{(2)} = e^{ij\alpha} m_j^{(1)}. \tag{68}$$

From the j-moment,

$$1 = e^{ij\alpha}, \tag{69}$$

hence

$$j\alpha = 2\pi r, \qquad r \in \mathbb{Z}, \tag{70}$$

and therefore

$$\alpha = \frac{2\pi r}{j}. \tag{71}$$

Using the k-moment, we require

$$e^{i\pi d_{kj}/j} = e^{ik\alpha}. \tag{72}$$

Substituting α gives

$$e^{i\pi d_{kj}/j} = e^{2\pi i k r/j}. \tag{73}$$

Equality of exponentials implies

$$\frac{\pi d_{kj}}{j} = \frac{2\pi k r}{j} + 2\pi q, \qquad q \in \mathbb{Z}. \tag{74}$$

Dividing by π and multiplying by j yields

$$d_{kj} = 2kr + 2qj, \tag{75}$$

i.e.,

$$2kr \equiv d_{kj} \pmod{2j}. \tag{76}$$

Write

$$k = a d_{kj}, \qquad j = b d_{kj}, \qquad \gcd(a, b) = 1. \tag{77}$$

Then

$$2a g_{kj} r \equiv d_{kj} \pmod{2b d_{kj}}. \tag{78}$$

Dividing by d_{kj} gives

$$2ar \equiv 1 \pmod{2b}. \tag{79}$$

This is impossible because the left-hand side is even while the right-hand side is odd. Hence no such rotation exists.

Thus every descriptor is necessary, and Φ is minimally discriminative. □

B.3.4 Dependence

The set Φ is minimal with respect to discriminability, but not independent. Indeed,

$$|\phi(k, j)|^2 = \frac{|m_k|^{2j/d_{kj}}}{|m_j|^{2k/d_{kj}}}. \tag{80}$$

C $\mathrm{SE}(2)$ vs. $\mathrm{E}(2)$

Table 5 compares our method with E2-CNNs under $\mathrm{SE}(2)$ and $\mathrm{E}(2)$ equivariance. For the $\mathrm{SE}(2)$ setting, our method achieves comparable or higher classification accuracy on all datasets except Galaxy Zoo, while using substantially fewer parameters. $\mathrm{E}(2)$ equivariance, and thus invariance to mirroring, has dataset-dependent effects: it slightly reduces accuracy on MNIST-based datasets, while improving performance on RESISC45, EuroSAT, and colorectal histology. Across the settings for which results are available, our method outperforms E2-CNNs while requiring considerably fewer parameters.

Table 5: **Quantitative comparison between our Mom/Inv CNN and $E(2)$-CNNs.** Results are reported as mean ± standard deviation over multiple runs. Top-five accuracy is denoted by *.

		Dataset		Ours		E2-CNNs	
Dataset	G	Cls	Size	Acc↑	Prm↓	Acc↑	Prm↓
$\text{MNIST}_{\text{RotTest}}$	SE(2)	10	50k	**99.42**±.07	119k	97.08±.18	362k
	E(2)			**98.44**±.08	83k	88.84±3.8	667k
$_{\text{rot}}\text{MNIST}$	SE(2)	10	12/50k	**99.08**±.04	119k	98.85±.04	362k
	E(2)			**98.06**±.07	83k	97.58±.13	667k
RESISC45	SE(2)	45	31.5k	**85.77**±.47	230k	83.59±.25	365k
	E(2)			**87.06**±.46	160k	83.95±.26	669k
$\text{EuroSAT}_{\text{rgb}}$	SE(2)	10	27k	**95.29**±.40	149k	94.21±.25	363k
	E(2)			**95.46**±.25	158k	94.28±.21	667k
DTD*	SE(2)	47	5k	67.14±.27	336k	67.18±.81	469k
	E(2)			65.72±.89	122k	**67.88**±.09	875k
NFFA+aug*	SE(2)	10	25k	**87.02**±.87	248k	85.78±1.4	363k
	E(2)			**86.45**±.20	172k	85.76±1.5	667k
Galaxy Zoo	SE(2)	8	100/31k	83.30±.03	227k	**84.39**±.08	363k
	E(2)			83.28±.18	375k	83.27±.20	667k
Col. Hist.	SE(2)	8	5k	**92.30**±.40	228k	91.58±.41	363k
	E(2)			**92.83**±.57	158k	92.14±.51	667k

D Equivariant Structure

Recent work on equivariant neural networks has emphasized the importance of preserving relative orientation information across features [27, 68]. In particular, Romero et al. argue that fully invariant local representations may discard geometric relationships required for object composition [68]. Fig. 6 illustrates this phenomenon: although the local facial components are identical, disrupting their relative orientations and spatial arrangement destroys the perception of a coherent face.

Motivated by these observations, we investigated whether explicitly retaining equivariant information alongside invariant features improves performance in our setting. In addition to the invariant basis introduced in Eq. (13), we considered variants that preserve a subset of equivariant moment information, such as: (i) retaining the reference moment m_1, which encodes local orientation information, (ii) projecting moment features jointly into type-0 (invariant) and type-1 (equivariant) representations.[4]

[4]Using e2cnn layers for dealing with type-1 features.

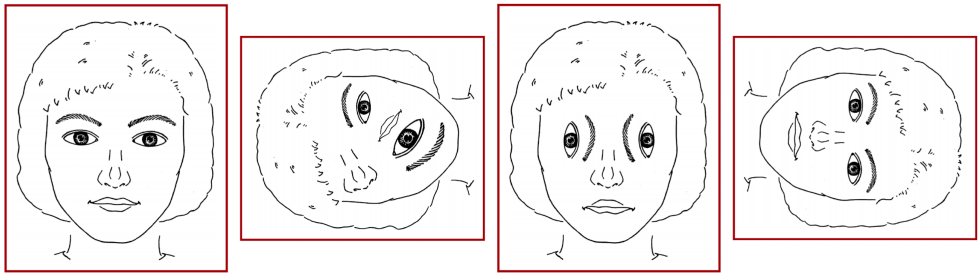


Figure 6: Relative pose relationships are essential for object composition. All panels contain the same local facial components, but only the outer panels preserve a coherent global face configuration. The inner panels disrupt the relative positions, orientations, and scales of the parts, illustrating that independent invariance of local elements can discard information needed to recognize the object as a face. Taken from Romero et al. [68, Fig. 1], based on Schwarzer [69].

Table 6: Comparison between invariant and equivariant variants on RotMNIST. Adding an explicit equivariant branch does not improve performance over the invariant architecture.

Architecture	Accuracy	# Params
E2CNN	$98.82 \pm .01$	362 K
Mom/Inv CNNs	$\mathbf{99.00 \pm .03}$	362 K
Equivariant Branch	$98.82 \pm .06$	363 K

Table 7: Comparison of different radial basis parameterizations. Performance remains consistent across variants and simplified radial constructions.

Variant	Test Accuracy	Val Accuracy
Monomials	98.82	99.25
Legendre polynomials	98.80	98.97
B-splines	**99.07**	99.74

Empirically, we did not observe consistent improvements from including explicit equivariant branches. Across the evaluated benchmarks, models based solely on invariant features achieved comparable or better performance than architectures augmented with equivariant components. This behavior is consistent with the quantitative results in the main text, where the proposed invariant models perform competitively with fully equivariant baselines despite operating entirely in the invariant feature space.

One possible explanation is that, although each local feature is individually invariant, relative orientation information can still be inferred implicitly from spatial arrangements across the continuous feature maps. Thus, global geometric structure may remain accessible without explicitly propagating equivariant feature fields through the network.

To further examine this question, we evaluated a variant of the architecture that augments the invariant branch with an additional type-1 equivariant branch. Results on RotMNIST are summarized in Table 6. The proposed invariant architecture achieves the highest accuracy while maintaining essentially the same parameter count. Adding an explicit equivariant branch did not improve performance and produced results comparable to the equivariant baseline.

E Radial Basis Design

The proposed moment layer depends on the choice of radial basis functions $R_{k,l}(r)$ used in the separable kernel construction of Eq. (3). We therefore investigate the sensitivity of the architecture to different radial discretizations and parameterizations.

We first study the effect of the number of radial rings used to discretize the radial component. Results are reported in Table 8. Performance remains stable across a broad range of ring counts, with a slight optimum observed around four rings. Next, we compare three radial basis parameterizations: B-splines, monomials used in classical pattern recognition [20], and orthogonal Legendre polynomials. The results are summarized in Table 7.

Table 8: Effect of the number of radial rings. Performance is stable across a range of values, with a slight optimum at 4 rings.

Ring Count	Test Accuracy	Train Accuracy	Val Accuracy
4	**99.15**	99.93	99.87
7	99.15	99.98	**99.89**
3	99.13	99.89	99.80
6	99.13	99.96	99.89
5	99.09	99.95	99.83
8	99.02	99.77	99.22

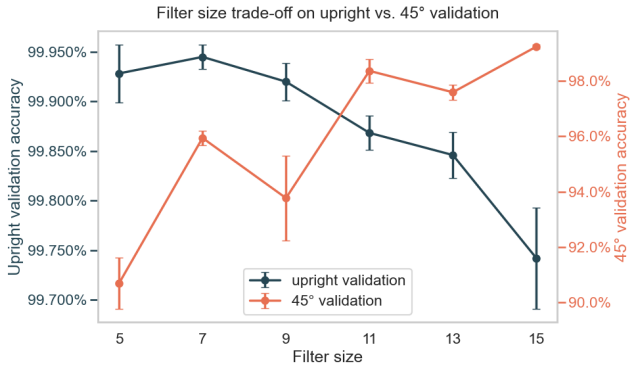


Figure 7: Effect of filter size on validation performance without data augmentation. The left y-axis shows validation accuracy on upright digits, while the right y-axis shows validation accuracy on digits rotated by $45°$. Larger filters preserve very high upright accuracy while substantially improving performance on rotated inputs, illustrating the trade-off between local precision and rotational robustness. Error bars indicate one standard deviation across runs.

F Filter Size Trade-off

We further analyze the influence of filter size on rotational generalization. Results are shown in Fig. 7, where we evaluate validation accuracy on both upright and rotated inputs without data augmentation.

As the filter size increases, performance on upright numerals remains consistently high, while accuracy on rotated samples improves substantially. This suggests that larger receptive fields better capture rotational structure by integrating information over a wider spatial neighborhood. In contrast, smaller filters are more sensitive to local orientation changes and therefore generalize less effectively to unseen rotations.

These results highlight a trade-off between local precision and rotational robustness. While very small filters may better preserve fine-grained local structure, larger filters provide improved stability under rotations with only minimal impact on standard upright performance.

G NFFA-Europe

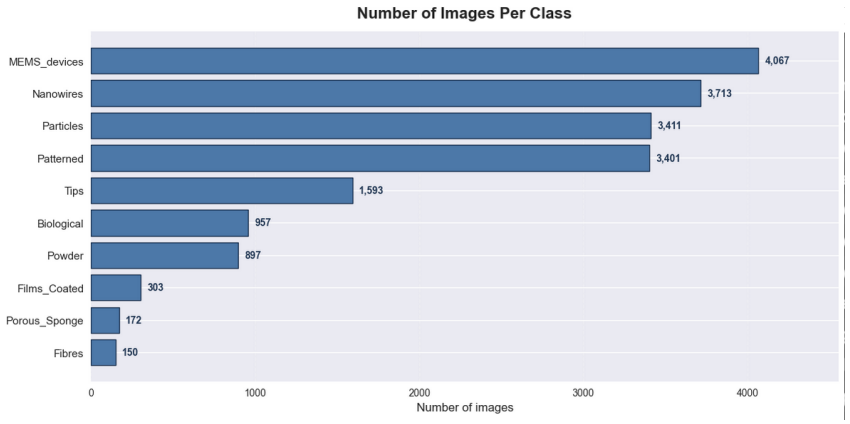


Figure 8: Distribution of SEM images across classes in the NFFA-Europe dataset. The dataset exhibits significant class imbalance, with some categories containing substantially fewer samples than others. This makes the benchmark particularly relevant for evaluating symmetry-aware architectures in low-data regimes.

To evaluate the proposed invariant architecture in a realistic scientific-image setting, we use the NFFA-Europe SEM dataset [45], "a large-scale collection of scanning electron microscopy (SEM)

images acquired at CNR-IOM (Trieste, Italy) within the NFFA-Europe project. The dataset contains approximately 25k grayscale SEM images organized into 10 semantic categories, including biological samples, nanowires, MEMS structures, patterned materials, and other nanoscale structures. Classification labels were validated by groups of nanoscientists through the online annotation platform `http://sem-classifier.nffa.eu`, and only images with strong annotation agreement were retained in the final dataset." [45]

The dataset therefore provides a challenging benchmark for representation learning with a limited number of samples and substantial intra-class variability. Images are resized and center-cropped to a fixed spatial resolution before training. The dataset is particularly well-suited for studying rotational invariance. Object orientation in SEM imaging is typically arbitrary and does not carry semantic meaning, while many classes exhibit strong rotational or texture-like symmetries. At the same time, the dataset remains relatively small compared to standard natural-image benchmarks and exhibits substantial class imbalance, making it a suitable testbed for evaluating sample efficiency and inductive biases.

Fig. 8 shows the distribution of images across categories. While some classes contain several thousand samples, others are comparatively scarce, highlighting the importance of architectures that can generalize effectively in low-data regimes.

H Magnitude-function ablation.

Table 9 compares several magnitude functions using smaller models than those in Table 1. Thus, these results should be interpreted as a controlled comparison of magnitude functions rather than as a direct comparison with the main results. In Table 1, we use

$$\sigma\big(m = |m_1|, n = |m_k|\big) = \frac{m \cdot n}{1 + m \cdot n}.$$

Table 9: Accuracy (%) for different basis functions. Results are reported as mean and standard deviation over runs.

Dataset	$m \cdot n$	$\sqrt{m \cdot n}$	$\frac{m \cdot n}{1+m \cdot n}$
rotMNIST	$98.95 \pm .06$	$99.01 \pm .05$	$99.01 \pm .04$
MNIST RotTest	$99.34 \pm .10$	$99.39 \pm .08$	$99.38 \pm .04$
EuroSAT	$95.19 \pm .14$	$95.04 \pm .39$	$95.25 \pm .06$
DTD Top1	$33.09 \pm .96$	$32.92 \pm .51$	$31.53 \pm .30$

I Additional Quantitive Study

In addition to $E(2)$-CNNs, we compare against several widely used equivariant, non-equivariant architectures, including Harmonic Networks (H-Nets) [11], ResNets [46], and Vision Transformers [47]. These baselines were selected because they represent complementary approaches to incorporating inductive biases while remaining the closest standard architectures to our setting. To ensure a fair comparison, all models were evaluated in compact, low-parameter regimes comparable to our architecture size. For H-Nets, increasing model scale did not yield measurable accuracy improvements, and we therefore report results using a smaller configuration. In addition, scaling to larger benchmarks was computationally challenging, with runtimes often exceeding those of our method by a factor of 5–20; consequently, we omit these experiments from the main comparison table. Overall, the comparisons highlight that our invariant moment architecture achieves strong performance while maintaining substantially lower parameter counts than equivariant and standard non-equivariant alternatives.

J Minimal Implementation

We provide a minimal implementation of the invariant feature construction used in our experiments to facilitate reproducibility. The code illustrates how invariant features are computed from moments, abstracted from framework-specific details. For the moment projection, one can directly use existing implementations of complex moments or steerable convolutions (e.g., `escnn` [15, 16]).

Table 10: **Quantitative comparison of our CNN** to others. Entries are given in the format Acc. ± Std.Dev.(Param.Count)

Dataset	Ours	E2-CNNs	ResNet	H-Nets	ViT
$\text{MNIST}_{\text{RotTest}}$	**99.42**±.07 (119k)	97.20±.14 (362k)	46.28±.26 (373k)	99.05±.03 (33k)	42.08±.64 (853k)
$_{\text{rot}}\text{MNIST}$	**99.08**±.04 (119k)	98.82±.01 (362k)	98.55±.01 (373k)	98.74±.03 (33k)	88.72±.99 (853k)
RESISC45	**85.77**±.47 (230k)	84.02±.27 (365k)	84.32±.11 (377k)	53.06±.02 (37k)	63.18±.55 (928k)
$\text{EuroSAT}_{\text{rgb}}$	**95.29**±.40 (149k)	93.90±.34 (363k)	79.50±.31 (377k)	86.28±1.13 (45k)	76.74±.46 (924k)
DTD*	67.14±.27 (336k)	67.18±.81 (469k)	56.66±.23 (377k)	38.24±.87 (36k)	37.95±.45 (928k)
NFFA	87.02±.87 (227k)	87.32±.30 (363k)	85.08±.79 (11M)	–	74.20±.63 (875k)
Galaxy Zoo	83.30±.03 (227k)	**84.39**±.08 (363k)	82.01±.07 (373k)	–	77.86±.12 (924k)
Col. Hist.	**92.30**±.40 (228k)	91.58±.41 (363k)	91.36±.58 (377k)	88.39±.20 (90k)	85.83±.08 (915k)

```python
@torch.compile
def construction_of_invariants(nt: torch.Tensor,
                               nt_types: Torch.Tensor,
                               eps: float = 1e-7):
    """
    nt: Non-trivial moments (type>=1)
        of shape [Ba, Ch, #Types, 2 (Real/Imag), H, W]
    nt_types: Types of nt [#Types]

    The first of nt is expected to be the reference type-1 [:, :, 0, ...]
    """
    mag = torch.linalg.vector_norm(nt, dim=-3)
    safe_mag = mag.clamp(min=eps)

    real, imag = nt.unbind(dim=-3)
    # NOTE: Torch atan2 is not gradient safe.
    angle = atan2(imag / safe_mag, real / safe_mag, eps)

    n_angle = angle[:, :, 0:1] * -nt_types[:, None, None]
    n_mag = mag[:, :, 0:1]
    nt_mag = mag[:, :, 1:]
    nt_angle = angle[:, :, 1:]

    scaled_mag = magnitude_fn(n_mag, nt_mag)
    angular_inv = scaled_mag * torch.cos(nt_angle + n_angle)
    return torch.cat([mag, angular_inv],dim=2)
```

K Dimension of Quotient

Flusser's construction Eq. (6) of invariant features assumes $m_1 \neq 0$, which removes rotational ambiguity. Under this condition, the group action is free on the corresponding subset of the moment space. As a consequence, the invariant representation can be interpreted as a parametrization of the quotient space whose dimension is given by $\dim(\mathcal{M}) - \dim(G)$ according to the Quotient Manifold Theorem.

Theorem (Quotient Manifold Theorem). Suppose G is a Lie group acting smoothly, freely, and properly on a smooth manifold M. Then the orbit space M/G is a topological manifold of dimension equal to $\dim M - \dim G$, and has a unique smooth structure with the property that the quotient map $\pi : M \to M/G$ is a smooth submersion. [70]

L Experimental Settings

Dataset-specific training hyperparameters, including resolutions, batch sizes, learning rates, and preprocessing steps, are summarized in Table 12. Shared architectural hyperparameters for both our Mom/Inv models and ESCNN baselines are provided in Table 11.

Parameter	Optimal	ESCNN
Stem kernel size	15	7
Stage kernels	$[11, 11, 11]$	$[5, 5, 5]$
Stage depths / blocks	$[1, 2, 2]$	$[1, 2, 2]$
Classifier hidden size	64	64
Dropout	0.2	–
Maximum type	3 (4 for DTD)	3 (4 for DTD)
Block type	Mom/Inv Block	GatedBlock
Pooling	AveragePool	PointwiseMaxPool

Table 11: Shared architecture hyperparameters.

M Computational Resources

All experiments were conducted using a single GPU due to the compact size of the proposed models. Training was performed on NVIDIA A100 GPUs on a shared compute cluster, as well as NVIDIA RTX 5090 GPUs for local experimentation and development. All models were implemented in PyTorch.

N 360-Degree Augmentation Experiment

We conduct a experiment to assess whether exhaustive rotational augmentation can recover the benefits of equivariant models. We use 10% of the NFFA dataset, with disjoint training and test sets, and apply a circular mask to all training, validation, and test images. For augmentation, each training image is augmented with 360 rotated copies, corresponding to one rotation per degree. This substantially increases the storage requirements, from approximately 170 MB for the original training set to approximately 70 GB for the augmented set.

All models are trained for 60 epochs with a batch size of 64, using a randomly shuffled training set as shown in Table 13. Our $\mathrm{SE}(2)$ and $\mathrm{E}(2)$ models achieve the highest accuracies while using substantially fewer parameters than ResNet-18. In particular, the $\mathrm{SE}(2)$ model with flip augmentation obtains an accuracy of 74.37%, compared with 73.34% for ResNet-18 with flip augmentation. The equivariant models also achieve markedly lower RCI values, indicating that their class representations are substantially more consistent under rotations.

Exhaustive 360-degree augmentation is computationally expensive. While our models require approximately 16–17 seconds per epoch without exhaustive augmentation, training ResNet-18 with 360 rotated copies requires approximately 1,000 seconds per epoch. After one epoch of this 360-degree ResNet-18 experiment, corresponding to roughly six times as many gradient steps as one epoch of our model, the accuracy reached only 68%. These preliminary results suggest that exhaustive augmentation alone does not match the accuracy, rotation consistency, or computational efficiency obtained by incorporating equivariance directly into the model.

O Visualization of Moment Channels and Learned Invariants

Figure 9 provides a detailed visualization of the invariant layer construction illustrated in Figure 1(c) in the main text. The top row shows different channel types representing equivariant moment responses of orders $k = 0, 1, 2, 3$ extracted from example scenes, with color-coded phases visualizing their equivariant transformations under rotation. The bottom panel displays the rotation-invariant features Φ_O learned by the network across multiple views and transformations.

Table 12: Dataset-specific training hyperparameters.

Dataset	Resolution	Batch	Optimal LR	ESCNN LR	Preprocess
RotMNIST	64^2	16 / 32	1.525e−3	10^{-3}	Upscaled to 64
MNIST-rot-test	64^2	64	1.5e−3	10^{-3}	–
DTD	96^2	16	1.5e−3	1.5e−3	Circular Mask
EuroSAT	64^2	32	1.5e−3	10^{-3}	–
Galaxy Zoo	96^2	64	1.5e−3	1.5e−3	Circular Mask
RESISC45	96^2	8	1.5e−3	1.5e−3	–
NFFA	224^2	64	10^{-3}	1.525e−3	Grayscale, Aug. Flips

Table 13: Results on 10% of NFFA. Lower RCI indicates greater consistency of the class representation under rotations.

Model	Augmentation	Accuracy	Parameters	RCI	Time per Epoch
SE(2)	None	0.72590	218k	0.00093	∼17 s
SE(2)	Flips	0.74366	218k	0.00080	∼16 s
E(2)	None	0.74072	152k	0.00084	∼17 s
ResNet-18	Flips	0.73336	11M	5.92	∼1,000 s

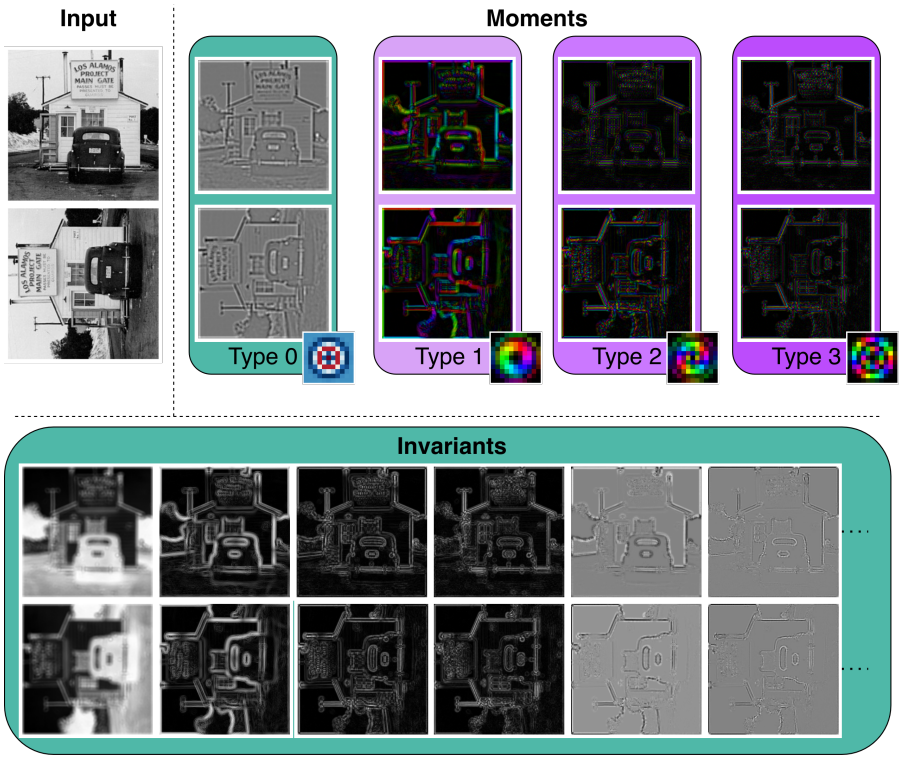


Figure 9: Multi-channel moment representations and learned invariant features. **(Top row)** Different channel types extracted from input scenes: Type 0 (grayscale intensity), Type 1 (equivariant moments of order $k = 1$), Type 2 (moments of order $k = 2$), and Type 3 (moments of order $k = 3$). Each moment type transforms predictably under rotation ($m_k \to e^{ik\theta} m_k$), with phases color-coded to visualize their equivariant responses. **(Bottom panel)** Invariant features Φ_O learned by the network across multiple views and transformations of the scenes, demonstrating rotation-invariant representations extracted via the optimal basis construction $|m_k|$ and $\phi_O(m_1, m_k) = \sigma(|m_1|, |m_k|) \cos(\theta_k - k\theta_1)$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction are consistent with the theoretical results, empirical evaluation, and stated limitations of the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper includes a dedicated "Conclusion and Limitations" section discussing the main assumptions and limitations of the proposed approach. Additional discussion, analyses, and ablation studies are provided throughout the appendix.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes].

   Justification: The paper states the assumptions underlying the theoretical results, and proofs and derivations are provided in the appendix with appropriate references from the main text.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes].

   Justification: The paper provides sufficient details about the architecture, datasets, baselines, evaluation metrics, and experimental setup to reproduce the main results. Additional implementation details and ablation studies are included in the appendix.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No].

   Justification: Due to institutional policy, the software cannot be publicly distributed at submission time without the associated copyright and licensing information, which would compromise double-blind review. Nonetheless, the paper and appendix provide detailed methodological and implementation descriptions, including a minimal code example illustrating the core invariant layer central to the proposed approach. The authors are able to provide additional implementation details to reviewers upon request. The full codebase will be released upon acceptance, and institutional approval for this release has already been obtained.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes].

   Justification: The paper describes the datasets, baselines, evaluation metrics, and architectural setup in the main text, with additional training details and implementation specifics provided in the appendix.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes].

Justification: Experimental results are reported with mean and standard deviation over multiple runs, and error bars are included where applicable and described in the text and figure captions.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes].

   Justification: Details about the compute resources used for the experiments, including hardware and execution settings, are provided in the appendix.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: The authors have reviewed the NeurIPS Code of Ethics and confirm that the research conforms to its guidelines.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A].

    Justification: The paper presents foundational methodological research and does not involve applications or deployment scenarios with direct societal implications that the authors are aware of.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A].

    Justification: The paper presents foundational methodological research and does not release models or datasets that pose a high risk of misuse requiring specific safeguards.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes].

    Justification: Existing datasets, libraries, and prior methods used in the paper are properly cited and credited. Relevant usage details and references are provided in the main text and appendix.

13. **New assets**

    Answer: [No].

    Justification: Due to institutional policy, the processed benchmark assets and accompanying documentation cannot be publicly distributed at submission time without associated attribution and licensing information, which would compromise double-blind review. Nevertheless, the paper and appendix describe the dataset construction, preprocessing, and label validation procedures in detail. The benchmark is derived from the publicly available NFFA-Europe SEM dataset using straightforward processing steps. The curated benchmark assets and documentation will be released upon acceptance, and institutional approval for this release has already been obtained.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing experiments or research with human subjects.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing or research with human subjects requiring IRB approval or equivalent review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A].

Justification: The core methodology, experiments, and scientific contributions of the paper do not involve LLMs as an important, original, or non-standard component.