GenHarris-ResNet: A Rotation Invariant Neural Network Based on Elementary Symmetric Polynomials

Valentin Penaud–Polge, Santiago Velasco-Forero, Jesus Angulo

GENHARRIS-RESNET: A ROTATION INVARIANT NEURAL NETWORK BASED ON ELEMENTARY SYMMETRIC POLYNOMIALS

Valentin Penaud--Polge, Santiago Velasco-Forero and Jesus Angulo
Center for Mathematical Morphology
Mines Paris, PSL University
Fontainebleau, France
{valentin.penaud_polge, santiago.velasco, jesus.angulo}@minesparis.psl.eu

ABSTRACT

In this paper, we propose a rotation invariant neural network based on Gaussian derivatives. The proposed network covers the main steps of the Harris corner detector in a generalized manner. More precisely, the Harris corner response function is a combination of the elementary symmetric polynomials of the integrated dyadic (outer) product of the gradient with itself. In the same way, we define matrices to be the self dyadic product of vectors composed with higher order partial derivatives and combine the elementary symmetric polynomials. A specific global pooling layer is used to mimic the local pooling used by Harris in his method. The proposed network is evaluated through three experiments. It first shows a quasi perfect invariance to rotations on Fashion-MNIST, it obtains competitive results compared to other rotation invariant networks on MNIST-Rot, and it obtains better performances classifying galaxies (EFIGI Dataset) than networks using up to a thousand times more trainable parameters.

***K*eywords** Elementary Symmetric Polynomials · Structure Tensor · Harris Corner Detector · Gaussian Derivatives · Neural Network · Galaxy Morphologies.

1 Introduction

Deep neural networks have been successfully applied across a large number of areas, including but not limited to computer vision [1], remote sensing [2], biology [3]. In many of these areas the data contains geometric properties or symmetries that can be exploited when designing neural networks. Prior to the Deep Learning (DL) era, *rotation invariance* in computer vision were mainly inspired by the Harris Corner Detector (HCD) [4], through local features extraction [5]. The main idea behind the HCD is to extract features based on the local geometry of the image. To do so the trace and the determinant of a matrix called *structure tensor*, defined to be the locally averaged outer product of the gradient with itself, are used to define a feature called the corner response function. Then this feature function is used to detect corners and edges depending where the function takes its local extrema. In DL, the use of intrinsic properties of invariance and equivariance from the data is an active area of research [6, 7, 8, 9]. Two of the main motivations to study equi/in-variance in neural networks are quite intuitive: i) Objects in natural images are not always oriented, scaled or localized in the same way unless the environment of acquisition is highly constrained. ii) It has already shown outstanding results with the example of *translation equivariance* whose most known representatives are Convolutional Neural Networks (CNN). In order to fix the principal notions of the paper, we recall the definition of *equivariance* and *invariance*. Let $f(\cdot)$ be a function and $g(\cdot)$ a transformation. f is equivariant, respectively invariant, to g if and only if $g(f(\cdot)) = f(g(\cdot))$, respectively $g(f(\cdot)) = f(\cdot)$. If f is equi/in-variant to all the transformations of a group $\mathcal{G}$, then f is said to be equi/in-variant to $\mathcal{G}$. In this paper, we focus on the invariance to the group of rotations. Our contribution in this paper is threefold: First, we generalize the corner response function to higher orders of derivation using structure tensor with higher derivatives, using the entire set of elementary symmetric polynomials and not only

the determinant and the trace. Second, we propose and validate the use of top-k max/min pooling as feature extractor. Third, we show that the proposed approach can be used in the context of DL with the *GenHarris-ResNet*.

2 Related Work

The structure tensor is well known in the image processing community [10]. Its famous application remains the HCD [4]. Several generalizations of this structure tensor have been proposed. An approach involves applying the complete HCD method to higher dimensional data, for instance in [11], a space-time interest point detector is proposed for moving corners, object collisions and more generally changes in movements. A generalization of the structure tensor to higher-order tensors is proposed in [12],where the outer product of the gradient vector is applied several times instead of only once. With this high-order tensor, a contrast function is defined whose extrema is used to detect edges and junctions. Another generalized structure tensor was proposed in [13] which corresponds to the classical structure tensor but in a harmonic system of coordinates. In [14], a theory allowing local characterisation of multiple orientations was presented. It supposes that an image is a linear combination of several images carrying, each of them, an orientation. It can be noted that the order of derivation used in the generalized structure tensor is directly linked to the number of orientations to be estimated. This generalization of the structure tensor is the closest to ours and the only difference comes from factors weighting the components of the tensor which allow us to obtain interesting mathematical properties. Rotation invariance has been addressed through various paradigms in DL. The most known approaches are spatial transformer networks [15] and group equivariant convolutional networks [9]. Spatial transformers are used to predict a transformation applied to the input tensor in order to use only a properly transformed section of it. The second approach is intrinsically, by its mathematical construction, invariant to a group of rotations. The principle of group equivariant convolutional networks is to extend convolution to other types of transformations and not only translation. Thus, the filters of a group equivariant convolutional network not only runs through all the possible translation in the input tensor but also through a (restricted) discrete group of rotations. Nevertheless, this approach has the advantage of data efficiency as the learned filters are common to all the rotations of the chosen group. In order to reduce the computation resources, a modified but similar approach has been proposed in [16]. Instead of keeping all the rotated copies of the tensor from a layer to another, a pooling operation along the rotation axis is applied to the output tensor of each layer. In order to keep the orientation information, the pooling does not only return the values of the tensor but also the angle value of the orientation map it has been pooled from. In [17], a CNN using constrained filters is proposed to obtain rotation invariance as the filters used in this network are steerable. But a discretization of the angles is still mandatory even though the angle sampling can be finer than [9, 17] without interpolation. Finally, a recently published paper [6] proposes to use linear combinations of differential invariants as the filters. The used differential invariants are derived using the moving frame method and consist in normalized polynomials of first and second order Gaussian derivatives which makes this network the closest to ours among the mentioned ones.

3 Mathematical Preliminaries

In this section we present the mathematical foundation of the Gaussian Elementary Symmetric Polynomials (ESPs) Layer, key element of the architecture of a GenHarris-ResNet. We will first recall basic knowledge about Gaussian derivative kernels. Then we will present a simple generalization of the structure tensor to higher order of derivation, which we call *Generalized Structure Tensor for Higher Derivatives* (GSTHD). Finally we will recall how to compute the ESPs through the power sums and the Girard-Newton's identities without the direct use of the eigenvalues.

3.1 Gaussian Partial Derivatives

The one-dimensional centered Gaussian kernel and its p^{th} derivative are respectively given by

$$G\left(x,\sigma\right)=\frac{1}{\left(2\pi\sigma^{2}\right)^{\frac{1}{2}}}e^{\frac{-x^{2}}{2\sigma^{2}}}\;;\;G_{p}\left(x,\sigma\right)=\frac{\partial^{p}G\left(x,\sigma\right)}{\partial x^{p}},\tag{1}$$

where σ represents the standard deviation (or scale). Gaussian derivatives of any order can be expressed as a Hermite polynomial multiplied by a Gaussian kernel, and as Hermite polynomials are easily expressible, we will use them to formulate the isotropic Gaussian derivative kernels without any use of derivation operations. The one-dimensional Hermite polynomial of order p is given by the following equation:

$$H_{p}\left(x\right)=\sum_{i=0}^{\lfloor p/2\rfloor}\frac{\left(-1\right)^{i}p!}{i!\left(p-2i\right)!}\left(2x\right)^{p-2i}.\tag{2}$$

The relation between the one-dimensional Gaussian derivative of order p and the Hermite polynomial of order p is given by

$$G_p(x,\sigma) = \left(-\frac{1}{\sqrt{2}\sigma}\right)^p \sqrt{2\pi}\sigma H_p\left(\frac{x}{\sqrt{2}\sigma}\right) G\left(\frac{x}{\sqrt{2}},\sigma\right)^2. \tag{3}$$

This relation between Hermite polynomials and Gaussian derivatives has been used several times in papers applying Gaussian derivative kernels to CNNs [18, 19]. From this, we can define the two-dimensional isotropic separated Gaussian derivative kernel of order $p+q$:

$$G_{p,q}(x_1,x_2,\sigma) = G_p(x_1,\sigma)\, G_q(x_2,\sigma). \tag{4}$$

The convolution of a Gaussian derivative kernel $G_{p,q}(\cdot,\cdot,\sigma)$ with an image I with be denoted

$$L_{p,q}^{(\sigma)}(\cdot) := G_{p,q}(\cdot,\sigma) * I(\cdot) \tag{5}$$

3.2 Generalized Structure Tensor for Higher Derivatives

The structure tensor is defined to be the integrated dyadic product of the Gaussian gradient with itself. Before defining the proposed GSTHD, we first recall the definition of the dyadic (outer) product and an associated, simple but useful, lemma which is the key to the rotational equivariance property.

Dyadic Product and Orthogonal matrices Given two vectors u, v n-dimensional in $\mathbb{R}^n$ with $u = (u_1, ..., u_n)$ and $v = (v_1, ..., v_n)$, the dyadic product of u with v, written as $u \otimes v$, is defined by

$$\forall i,j \in [\![1,n]\!],\ (u\otimes v)_{i,j} = u_i v_j \tag{6}$$

The following lemma will be taken into consideration to define the vectors used to construct the GSTHD as it informs us how the dyadic product of a vector with itself behaves when an orthogonal transformation is applied to the vector. We will denote by $\mathcal{O}^n(\mathbb{R})$, the set of real orthogonal $n \times n$ matrices, i.e., $\forall P \in \mathcal{O}^n(\mathbb{R})$, $P^T = P^{-1}$.

Lemma. *$\forall u, v \in \mathbb{R}^n$, if $\exists P \in \mathcal{O}^n(\mathbb{R})$ s.t. $u = Pv$, then*

$$u \otimes u = (Pv)\otimes(Pv) = P(v\otimes v)P^{-1}.$$

Therefore if we define the vector used to compute the GSTHD in a way that a rotation applied to the input image (or tensor) implies an orthogonal transformation of the vector, then the ESPs can be used as rotational equivariant features as they are similarity invariants.

Generalized Structure Tensor for Higher Derivatives The two main requirements to properly define the GSTHD are: i) to be a generalization of the well known tensor structure when considering first order derivatives and ii) to take into account Lemma 3.2. For an order $k \in \mathbb{N}$, we propose to use the following vector defined at every coordinate of the image:

$$\mathcal{L}_k^{(\sigma)} = \left(\sqrt{\binom{k}{p}} L_{k-p,p}^{(\sigma)}\right)_{p\in[\![0,k]\!]} \tag{7}$$

where $\binom{k}{p} = \begin{cases} \frac{k!}{p!(k-p)!} & \text{if } 0 \le p \le k \\ 0 & \text{otherwise} \end{cases}$.

We can immediately observe that the case $k = 1$ gives the gradient as required by the first requirement. The GSTHD of order k, scale σ and integrated at scale σ_{int}, denoted as $\mathcal{M}_k^{(\sigma)}$ is, pointwise (or pixelwise) defined as follow:

$$\mathcal{M}_k^{(\sigma)} = G(\cdot,\sigma_{int}) * \left(\mathcal{L}_k^{(\sigma)} \otimes \mathcal{L}_k^{(\sigma)}\right) \tag{8}$$

Before taking into account the second requirement stated above, let us define the notations when a rotation is applied to the input image $I : \mathbb{R}^2 \to \mathbb{R}$. We will denote as (x_1, x_2) the system of coordinates attached to I. If a rotation of angle θ is applied to I, the coordinates also transform themselves into a rotated ones (x_1', x_2') through the relation

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{9}$$

In the rest of this subsection, every mathematical object expressed in this rotated system of coordinates will hold an apostrophe. As an example $\mathcal{L}_k'^{(\sigma)}$ corresponds to $\mathcal{L}_k^{(\sigma)}$ expressed with the rotated coordinates. The operator describing a rotation of angle θ applied to an image will be denoted $R_\theta(.)$, i.e., $I' = R_\theta(I)$.

The following result allows us to use Lemma 3.2, i.e., answers to the second requirement.

Proposition. *Let I be the input image, $\forall \theta \in [-\pi, \pi], \forall k \in \mathbb{N}$:*

$$I' = R_\theta(I) \implies \exists \mathcal{P}_k(\theta) \in \mathcal{O}^{k+1}(\mathbb{R}) \mid \mathcal{L}'^{(\sigma)}_k = \mathcal{P}_k(\theta)\mathcal{L}^{(\sigma)}_k \text{ and } \mathcal{M}'^{(\sigma)}_k = \mathcal{P}_k(\theta)\mathcal{M}^{(\sigma)}_k\mathcal{P}_k(\theta)^T.$$

More precisely $\forall i, j \in [\![0, k]\!]$,

$$(\mathcal{P}_k(\theta))_{i,j} = \frac{\sqrt{\binom{k}{i}}}{\sqrt{\binom{k}{j}}} \sum_{r=max(0,i+j-k)}^{min(i,j)} \binom{i}{r}\binom{k-i}{j-r}(-1)^{i-r}\sin(\theta)^{i+j-2r}\cos(\theta)^{k+2r-i-j}. \tag{10}$$

3.3 Elementary Symmetric Polynomials

This section relies on basic mathematical knowledge [20]. We showed in Proposition 3.2 that applying a rotation to the input image/tensor only results in an orthogonal change of basis of $\mathcal{M}^{(\sigma)}_k$ at each point of the image. This means that using elementary symmetric polynomials, which are similarity invariants of these matrices, leads to rotational equivariant features. In this subsection we recall the definition of the ESPs and how to compute them without directly knowing the spectrum of the studied matrices.

Let $M \in M_n(\mathbb{R})$ be a diagonalizable real matrix and let $\Lambda = (\lambda_1, ..., \lambda_n)$ be its eigenvalues. For all k in $[\![1, n]\!]$, the Elementary Symmetric Polynomials are defined as

$$e_k(\lambda_1, ..., \lambda_n) = \sum_{1 \leq i_1 < ... < i_k \leq n} \prod_{j=1}^{k} \lambda_{i_j}. \tag{11}$$

The two mostly used ESPs are the determinant (e_n) and the trace (e_1) of a matrix as it is the case, for example, in the Harris corner detector [4] where the corner response is defined to be $R = e_2\left(\mathcal{M}^{(\sigma)}_1\right) - \beta\left(e_1\left(\mathcal{M}^{(\sigma)}_1\right)\right)^2$. As we propose a generalization to the Harris corner detector for deep learning by using matrices of higher dimensions than two, we will use the entire set of ESPs of each matrix. Nevertheless, as it is not suitable to define a layer computing directly the eigenvalues of some matrices, we propose to use Girard-Newton's identities to avoid this problem. To this end, we recall the definition of the power sums of Λ of order k:

$$p_k(\lambda_1, ..., \lambda_n) = \sum_{i=1}^{n} \lambda_i^k. \tag{12}$$

These quantities have two major advantages: First, they are easy to compute in the case of a diagonalizable matrix, $p_k(\Lambda) = \text{trace}\left(M^k\right)$, and second, they can be used to compute the ESPs using the Girard-Newton's identities:

$$e_k(\Lambda) = \frac{1}{k}\sum_{i=1}^{k}(-1)^{i-1} e_{k-i}(\Lambda) p_i(\Lambda). \tag{13}$$

Finally, it can be highlighted that using a Gaussian kernel to integrate the matrices $\mathcal{M}'^{(\sigma)}_k$ does not alter the result of Proposition 3.2 as the matrices of each pixel of a rotated image are all transformed using the matrix $\mathcal{P}_k(\theta)$.

Algorithm 1 ESP Layer Algorithm

Given an input tensor I and the hyperparameters derivation order N, a list of scale values Σ and an integration scale σ_{int}

1. Create empty list E
2. For σ in Σ
   (a) For $i = 0$ to N
      i. Compute the $\mathcal{L}^\sigma_i$ channelwise.
      ii. Compute the outer product of $\mathcal{L}^\sigma_i$ with itself.
      iii. $\mathcal{M}^\sigma_i \leftarrow$ Convolution of each element of the resulting matrix across all the channels with $G(\cdot, \sigma_{int})$
      iv. Compute the homogenized absolute ESPs $\{|e_1|, |e_2|^{\frac{1}{2}}, ..., |e_{i+1}|^{\frac{1}{i+1}}\}$ of $\mathcal{M}^\sigma_i$ using the power sums and the Girard-Newton's identities.
      v. Add $|e_1|, |e_2|^{\frac{1}{2}}, ..., |e_{i+1}|^{\frac{1}{i+1}}$ to E
3. Return E stacked into a tensor.

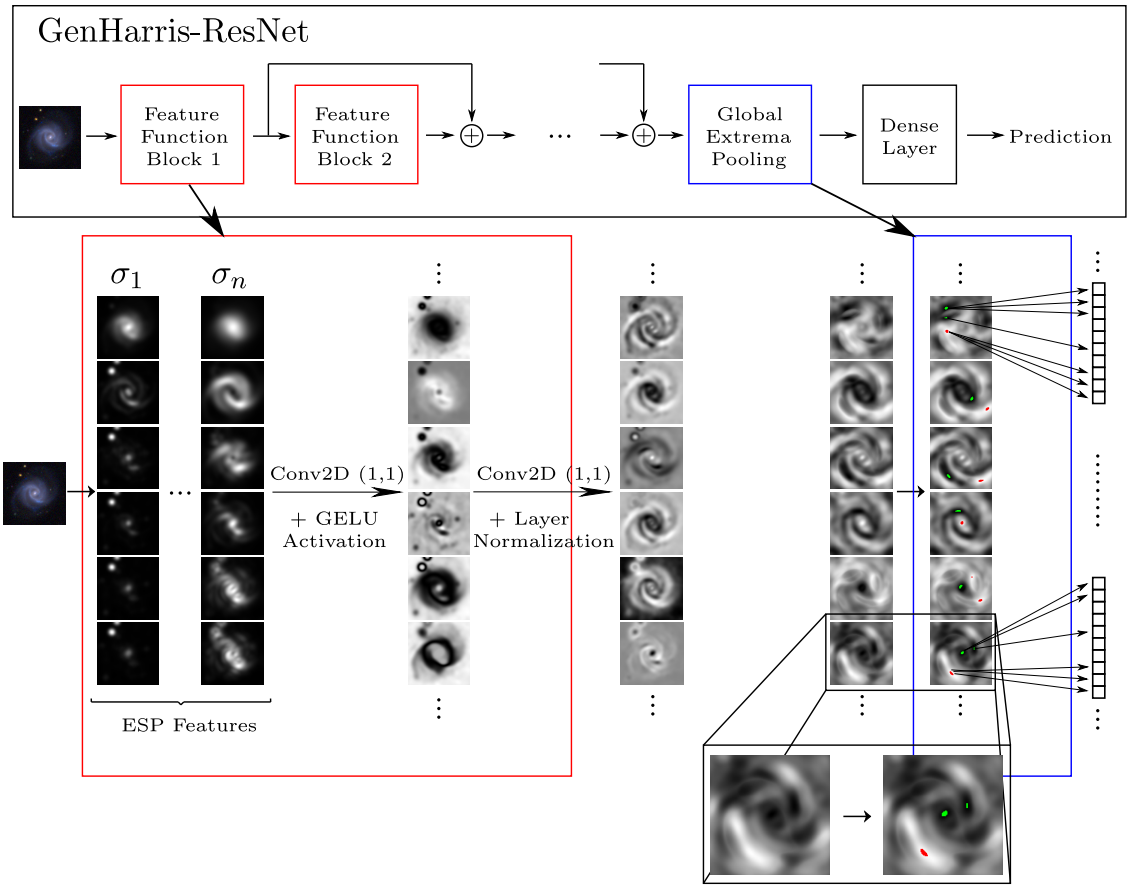


Figure 1: GenHarris-ResNet architecture: The first part of the proposed network is equivariant to rotation and is composed of several feature function blocks arranged in a ResNet configuration. Then, the feature maps are given to the global extrema pooling layer which brings the invariance property. A dense layer ends the network.

4 GenHarris-ResNet

As stated in Section 1, the GenHarris-ResNet aims to generalize, in a deep learning approach, the HCD. As highlighted above, the HCD can be divided into two main stages: the computation of a feature function (a corner response function) and the selection of interest points using the feature function as a decision criterion. We present in this section how we propose to adapt these two processes to deep learning: one with the definition of a feature function block, composed of an ESP layer, two linear combinations and a normalization. And the other with a global extrema pooling layer.

In the same way Harris and Stephens used the trace and the determinant of the structure tensor to obtain the corner response function, we propose to use the elementary symmetric polynomials of the GSTHD and to combine them. More precisely, the ESP layer has three hyperparameters: a maximal order of derivation N, a list of scale values Σ and an integration σ_{int}. Given an input tensor I, the ESP Layer performs Algorithm 1. It can be highlighted that no parameters are learned in the ESP layer, it only computes ESPs in a differentiable manner. Once the ESPs are computed, two convolutional layers using filters of size 1×1 and a layer normalization are used to combine them. The combination of these layers constitutes a feature function block and several blocks in cascade with residual connections constitute the feature function part of the GenHarris-ResNet. After computing the corner response function, Harris and Stephens determine the interest points (corners and edges) through a MaxPooling and a directional MinPooling. In other words, Harris and Stephens proposed to use the local extrema of the corner response function. In our GenHarris-ResNet, we propose to use a global pooling layer which returns the values of the top $2K$ global maxima and global minima, i.e., the K highest values and the K lowest values. Note that global pooling changes the rotation equivariance property of the feature function part of the network into rotation invariance [8]. Finally, a dense layer ends the GenHarris-ResNet in the case of a classification task. The overall architecture of the GenHarris-ResNet is illustrated in Figure 1.

Table 1: Test accuracy, rotation invariance and number of parameters of several configurations of the GenHarris-ResNet trained on Fashion-MNIST

Network	Acc_0	$\mathbb{E}\left(\lvert\text{Acc}_0 - \text{Acc}\left(\theta\right)\rvert\right)$	$\mathbb{E}\left(\text{Cons Error}\left(\theta\right)\right)$	nb param
Vanilla CNN	91.88%	68.31%	76.21%	12.4k
$L = 2$, GlobalMaxPooling	87.76%	0.99%	5.63%	3.8k
$L = 2$, $K = 1$	87.92%	1.10%	5.55%	4.1k
$L = 2$, $K = 2$	88.40%	1.27%	5.15%	4.7k
$L = 2$, $K = 10$	89.03%	0.99%	4.94%	9.9k
$L = 2$, $K = 20$	88.62%	0.83%	4.58%	16.3k
$L = 3$, $K = 2$	89.70%	0.82%	4.62%	6.5k
$L = 3$, $K = 10$	90.26%	1.00%	3.94%	11.6k
$L = 3$, $K = 20$	89.54%	1.04%	4.23%	18k

5 Experiments

This section contains three parts: i) A rotation invariance study using Fashion-MNIST. ii) A comparison with previous works on MNIST-Rot. iii) The application of the GenHarris-ResNet to a real-images problem: the classification of galaxy morphologies using the EFIGI Dataset [21]. An overall ablation study is realized in the first and last experiments.

5.1 Rotation Invariance

Fashion-MNIST is used to study the effects of the number of feature function blocks used in the GenHarris-ResNet and of the value taken by K in the global extrema pooling layer. We also test the invariance to rotations of the proposed model. We trained several GenHarris-ResNets on Fashion-MNIST without any data augmentation. The considered models are composed of either two or three feature function blocks (denoted as $L = 2$ or $L = 3$ in Table 1). The feature function blocks had a maximal order of derivation N values two and a list of scales values $[1.0, 2.0, 3.0]$. Several values of K for the global extrema pooling were tested. The rotation invariance of the trained GenHarris-ResNets has been tested on rotated versions of the entire test dataset of Fashion-MNIST using bilinear interpolation. The rotations applied to the test dataset have been discretized with an angle step value of 10°, giving 36 rotated versions of the samples. Three quantities have been used to evaluate the networks: i) The accuracy at $\theta = 0$, denoted Acc_0, to evaluate the classification performance of the networks. ii) The mean accuracy deviation $\mathbb{E}\left(\lvert\text{Acc}_0 - \text{Acc}\left(\theta\right)\rvert\right)$. iii) The *mean consistency error*, denoted $\mathbb{E}\left(\text{Cons Error}\left(\theta\right)\right)$. We recall that the consistency is the percentage of predictions that did not change after applying the transformation, even the wrong predictions are taken into account. What is called here the consistency error is in the contrary, the percentage of predictions that changed. Table 1 shows the performances of the GenHarris-ResNets compared to the performances of a classical three layers CNN as a reference for comparison. The results show that despite the fact that the CNN slightly outperforms the GenHarris-ResNets when no rotation is applied to the test dataset, the proposed networks show a nearly perfect invariance to rotations even though no transformed data has been used during training. It should be highlighted that at 90°, 180° and 270°, the consistency values of all the GenHarris-ResNets are above 99.9%, implying that the non perfect invariance may come from the interpolation method used to rotate the images, the low resolution of the images and the discretization of the Gaussian derivative kernels. It is observable that for a fixed value of K, using three feature function blocks increases the performance of the GenHarris-ResNet, when applied to Fashion-MNIST. The better performances obtained when using the global extrema pooling compared to global max pooling show that using only one value to describe one image is not the best way to summarize feature maps to discriminate properly the classes. The increase of K from one to ten in the global extrema pooling layer leads to a small increase of the performance of the network until it reaches $K = 20$ and decreases. This suggests that an optimal number of values exists to summarize the feature maps when using extreme values. Using less values than the optimal one may not be enough to summarize correctly the feature maps and using more extreme values may cause too much redundancy while unnecessarily increasing the number of trainable parameters.

5.2 Comparison on MNIST-Rot

In order to compare the performance of the GenHarris-ResNet to other methods dealing with rotation invariance in deep learning, we tested it on MNIST-Rot [22] as it represents the standard dataset for this task. Nevertheless, we believe that evaluating the invariance of a model to a transformation should be done without including transformed samples in the training dataset to truly appreciate the intrinsic invariance property. We chose the configuration of the

Table 2: Comparison of the test error obtained on MNIST-Rot when training on MNIST with the five best results from the paper [23]. Networks: RP_RF_1_32 [24], Spherical CNN [25], ORN-8 (ORAlign) [26], RED-NN [23], Covariant CNN [27].

Network	RP_RF_1	Spherical CNN	ORN-8	RED-NN	Covariant CNN	Ours
% Test error	12.20	6.00	16.24	2.05	17.21	3.83
# parameters	1M	68k	969k	42k	7k	11.6k

GenHarris-ResNet that obtained the best accuracy in the previous experiment and we trained it from scratch on the training set of MNIST without any data augmentation and we used 2k images of the MNIST-rot training dataset for validation. We compare the performance of our model with networks of the literature trained in similar conditions. The test error obtained on MNIST-Rot is presented in Table 2 together with published results. Even though the GenHarris-ResNet does not have the lowest error value, it challenges the state of art methods while having a low number of trainable parameters.

5.3 Galaxy Morphologies

This last experiment aims to apply the GenHarris-ResNet to a classification task on real images. As the feature function block uses the local structure of images to compute its feature maps, we applied the GenHarris-ResNet to the classification of galaxy morphologies to demonstrate its ability to capture the geometry of objects in images. It also serves to complete an overall ablation study of the layer, through the study of the effect of N, the maximal order of derivation in the ESP Layer. We used the EF-4 configuration, named after its number of classes, of the EFIGI dataset [29]. See Figure 3b for examples of images of each class. The EF-4 dataset was randomly split in a training/validation set using 90% of the images and a testing set using the remaining samples. Two GenHarris-ResNet composed of three feature function blocks have been used for this task: one using only feature function blocks with $N = 2$ and another using $N = 3$ in the last two blocks. We found better performances using $K = 30$ for the second order GenHarris-ResNet and $K = 50$ for the third order GenHarris-ResNet. The test accuracy values obtained are shown in Figure 2 together with the performances obtained by other networks [28, 29]. As the random split of the EF-4 dataset is surely not the same between our networks and the others, we also give the accuracy values obtained on the training set to show that the results obtained by our networks are not due to a lucky composition of the test set. The results presented

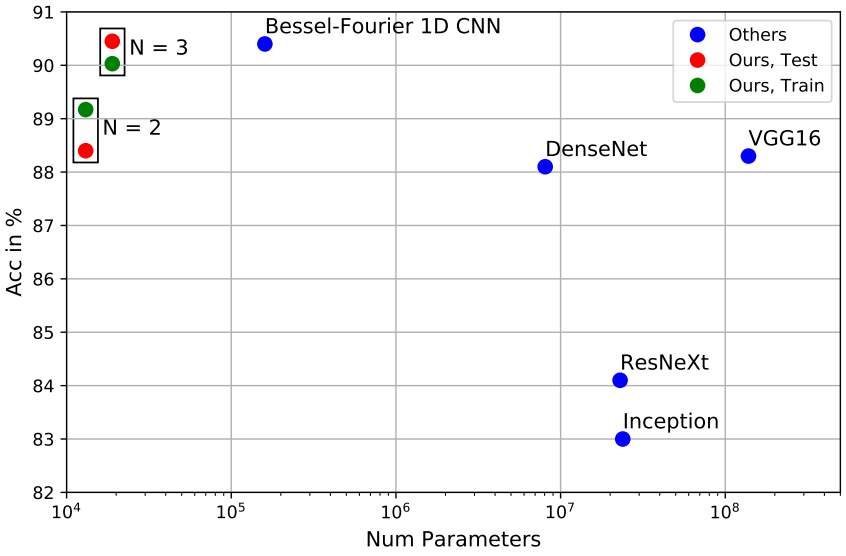
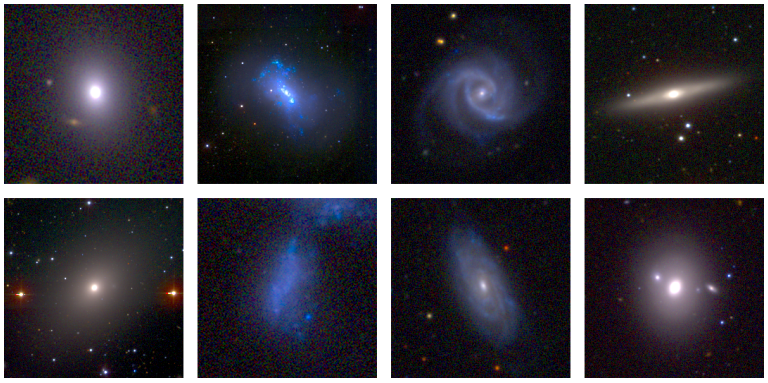


Figure 2: Test accuracy values of related work networks on EF-4: blue points. Test (respectively training) accuracy values of the proposed networks: red (respectively green) points. It can be pointed out that the model 'Bessel-Fourier 1D CNN' [28] has similar performances to the proposed model but does not hold invariance to translations due to the use of Bessel-fourier moments (its use is mainly restricted to images displaying centered objects).

		Predicted labels			
		0	1	2	3
True labels	0	27	0	1	1
	1	0	21	0	4
	2	11	0	36	7
	3	2	7	9	314

(a) Confusion matrix of the test predictions obtained for the model $N = 3$ with the following label indexing:
0: Elliptical, 1: Irregular,
2: Lenticular, 3: Spiral.



(b) Samples from the four classes of EF-4. From left to right: Elliptical, irregular, spiral, lenticular.

in Figure 2 show that the GenHarris-ResNets have competitive performances compared with the other networks while using 10 to 1000 times less trainable parameters. The results also show that, in the case of this dataset, using a maximal order of derivation equals to three allows to obtain better performances than using $N = 2$. We highlight that this result might not hold in general because the use of $N = 3$ requires a value of σ_{int}, in (8), high enough so that the matrices $\mathcal{M}_k^\sigma$ are well defined. Finally, the confusion matrix of the third order GenHarris-ResNet obtained on the test dataset is given in Table 3a for future comparisons.

6 Conclusion

We proposed a low-parameter neural network architecture which allows us to introduce the paradigm of the Harris Corner Detector to DL. The motivation was to define a new network using feature maps that are equivariant to rotations in a rigorous mathematical way, i.e, the invariance property is intrinsic and does not depend on data augmentation: the generalization to orientations not considered during training is total. The proposed approach extended the structure tensor to higher Gaussian derivatives and used a larger set of ESPs. We have explored its interest in a low complexity ablation study which revealed a nearly perfect invariance to rotations. The GenHarris-ResNet also showed competitive performances with related approaches for rotation invariance when evaluated on MNIST-Rot and outperformed networks using millions of trainable parameters on the task of galaxy classification based on their morphology.

6.1 Code, Data and Proofs Availability

The code, the EF-4 dataset splits and detailled proofs (based on [30, 31, 32]) are available at `https://github.com/Penaud-Polge/GenHarrisResNet`.

Acknowledgments

This work was granted access to the HPC resources of IDRIS under the allocation 2022-[AD011013367] made by GENCI.

References

[1] Di Feng, Christian Haase-Schütz, et al. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2020.

[2] Xiao Xiang Zhu, , et al. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[3] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Nuclei segmentation in histopathology images using deep neural networks. In *14th ISBI*, pages 933–936. IEEE, 2017.

[4] Chris Harris, Mike Stephens, et al. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, pages 23.1–23.6, 1988.

[5] Krystian Mikolajczyk, Tinne Tuytelaars, et al. A comparison of affine region detectors. *International journal of computer vision*, 65(1):43–72, 2005.

[6] Mateus Sangalli, Samy Blusseau, Santiago Velasco-Forero, and Jesus Angulo. Moving frame net: SE(3)-equivariant network for volumes. In *NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, pages 81–97. PMLR, 2023.

[7] Tony Lindeberg. Scale-covariant and scale-invariant Gaussian derivative networks. *Journal of Mathematical Imaging and Vision*, 64(3):223–242, 2022.

[8] 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.

[9] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

[10] Thomas Brox, Joachim Weickert, Bernhard Burgeth, and Pavel Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.

[11] Ivan Laptev and Tony Lindeberg. Space-time interest points. In *IEEE CVPR*, volume 1, pages 432–439, 2003.

[12] Thomas Schultz, Joachim Weickert, and Hans-Peter Seidel. A higher-order structure tensor. In *Visualization and processing of tensor fields*, pages 263–279. Springer, 2009.

[13] Josef Bigun, Tomas Bigun, and Kenneth Nilsson. Recognition by symmetry derivatives and the generalized structure tensor. *IEEE PAMI*, 26(12):1590–1605, 2004.

[14] Matthias Mühlich and Til Aach. A theory of multiple orientation estimation. In *European Conference on Computer Vision*, pages 69–82. Springer, 2006.

[15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28:2017–2025, 2015.

[16] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *IEEE CVPR*, pages 5048–5057, 2017.

[17] Simon Graham, David Epstein, and Nasir Rajpoot. Dense steerable filter CNNs for exploiting rotational symmetry in histology images. *IEEE Transactions on Medical Imaging*, 39(12):4124–4136, 2020.

[18] Jorn-Henrik Jacobsen, Jan Van Gemert, Zhongyu Lou, and Arnold WM Smeulders. Structured receptive fields in CNNs. In *IEEE CVPR*, pages 2610–2619, 2016.

[19] Valentin Penaud-Polge, Santiago Velasco-Forero, and Jesus Angulo. Fully trainable Gaussian derivative convolutional layer. In *29th IEEE ICIP*, 2022.

[20] Ian Grant Macdonald. *Symmetric functions and Hall polynomials*. Oxford university press, 1998.

[21] Anthony Baillard, Emmanuel Bertin, Valérie De Lapparent, , et al. The EFIGI catalogue of 4458 nearby galaxies with detailed morphology. *Astronomy & Astrophysics*, 532:A74, 2011.

[22] Hugo Larochelle, Dumitru Erhan, et al. An empirical evaluation of deep architectures on problems with many factors of variation. In *24th ICML*, pages 473–480, 2007.

[23] Rosemberg Rodriguez Salas, Petr Dokladal, and Eva Dokladalova. A minimal model for classification of rotated objects with prediction of the angle of rotation. *Journal of Visual Communication and Image Representation*, 75:103054, 2021.

[24] Patrick Follmann and Tobias Bottger. A rotationally-invariant convolution module by feature map back-rotation. In *2018 IEEE WACV*, pages 784–792. IEEE, 2018.

[25] Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[26] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *IEEE CVPR*, pages 519–528, 2017.

[27] Rosemberg Rodriguez Salas, Eva Dokladalova, and Petr Dokladal. Rotation invariant CNN using scattering transform for image classification. In *2019 IEEE ICIP*, pages 654–658. IEEE, 2019.

[28] César Joel Camacho-Bello, Lucia Gutiérrez-Lazcano, et al. Rotation-invariant image classification using a novel 1D CNN and multichannel accurate Bessel-Fourier moments. *Boletín de Ciencias Básicas e Ingenierías*, 10(3):1–4, 2022.

[29] Ana Martinazzo, Mateus Espadoto, and Nina ST Hirata. Deep learning for astronomical object classification: A case study. In *VISIGRAPP*, pages 87–95, 2020.

[30] Wooram Park, Gregory Leibon, Daniel N Rockmore, and Gregory S Chirikjian. Accurate image rotation using hermite expansions. *IEEE TIP*, 18(9):1988–2003, 2009.

[31] José Luis Silván-Cárdenas and Boris Escalante-Ramírez. The multiscale hermite transform for local orientation analysis. *IEEE TIP*, 15(5):1236–1253, 2006.

[32] Luc MJ Florack et al. Scale and the differential structure of images. *Image and vision computing*, 10(6):376–388, 1992.