A Quaternion Monogenic Layer Resilient to Large Brightness Changes

Abraham Sánchez*, E. U. Moya-Sánchez*, Sebatià Xambó**, Sebastián Salazar-Colores*** and Ulises Cortés**.

*Gobierno de Jalisco, **Barcelona Tech (UPC) & Barcelona Supercomputing Center, ***Centro de Investigaciones en Óptica

Motivation

- The mammalian visual system is resilient to many geometric transformations and contrast variations to which current deep learning (DL) classifiers are not [**hendrycks-dietterich-2019**]
- The learning of an invariance response may fail even with very deep CNNs or large data augmentations in the training. Image from [**dodge-karam-2016**] [**simard-steinkraus-platt-2003**]

Contrast				
Caffe	0.729299	0.635093	0.374961	0.00598902
VGG-CNN-S	0.921437	0.88517	0.678366	0.0301079
GoogleNet	0.970436	0.96616	0.871698	0.349995
VGG16	0.834489	0.742968	0.551311	0.32043

Strategy

- Extent to **other representation** to obtain higher discrimination capacity.

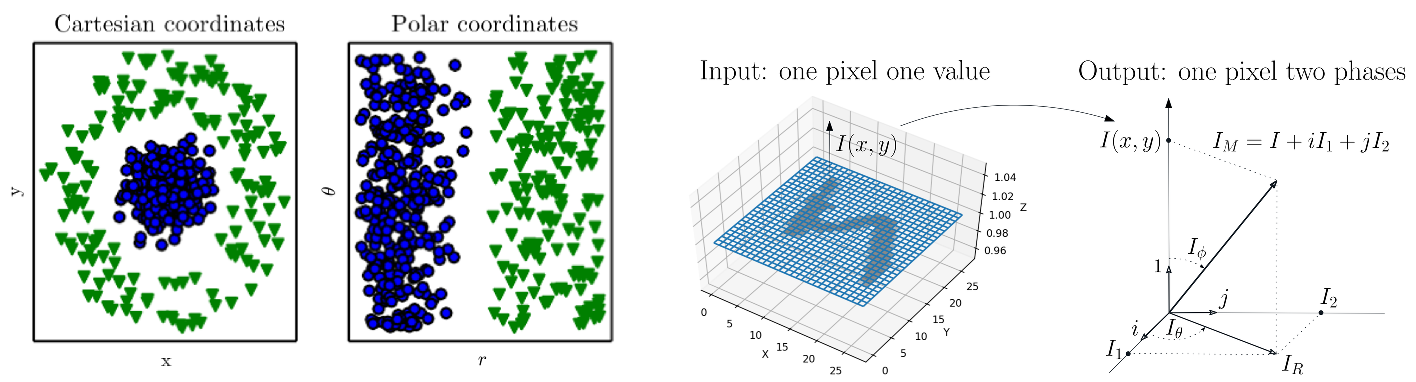
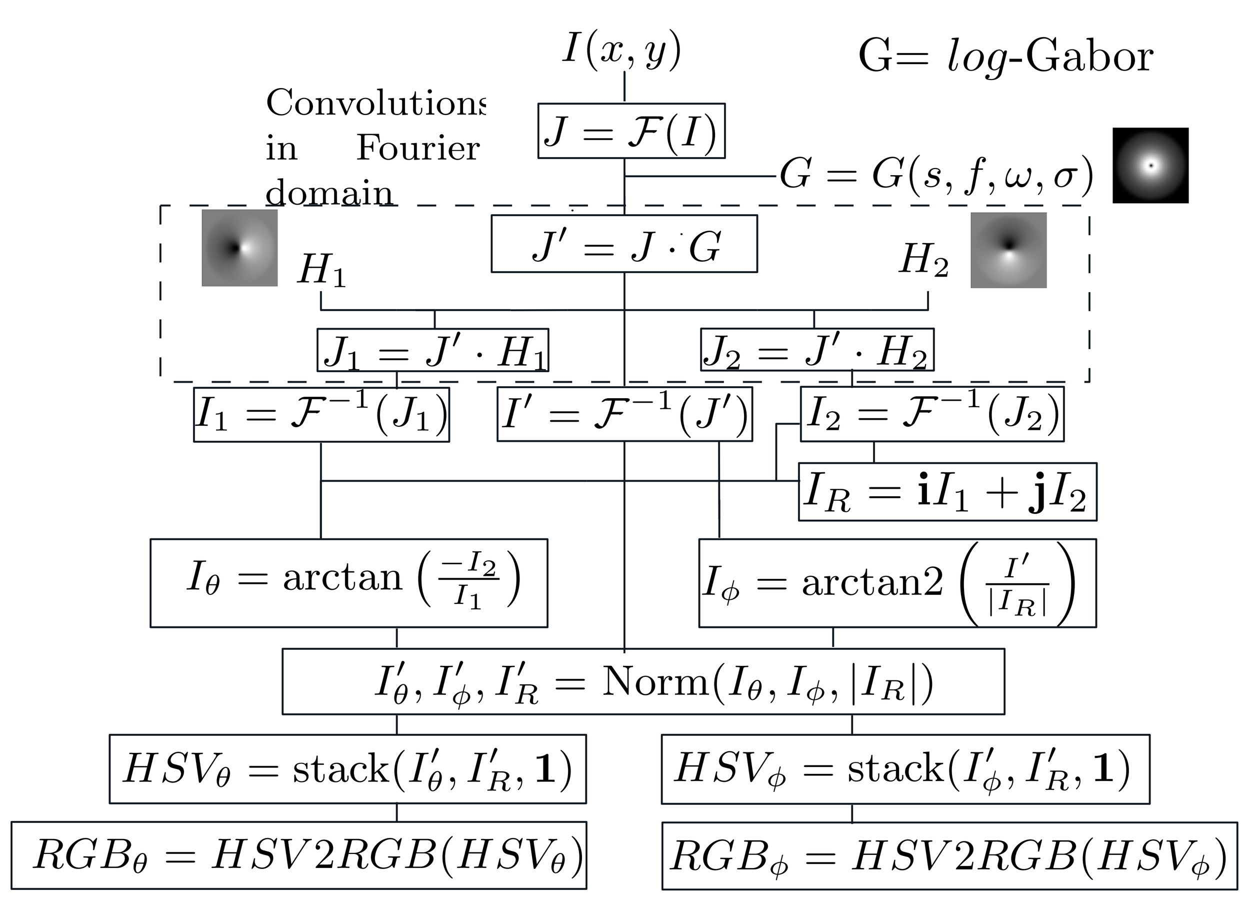
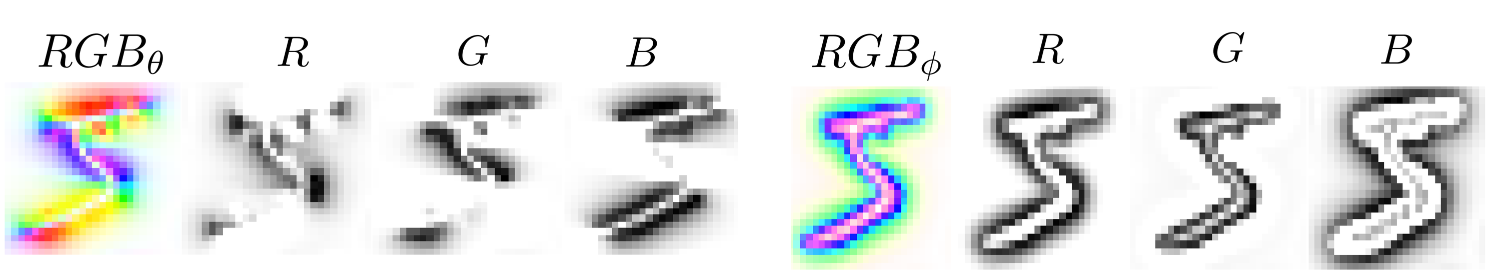


Figure 1: More data Versus representation. Image from DL book.

CNN layer M6

- We use four main bio-inspired tools: the local phase, Log-Gabor, the HSV color space and the artificial neural networks.





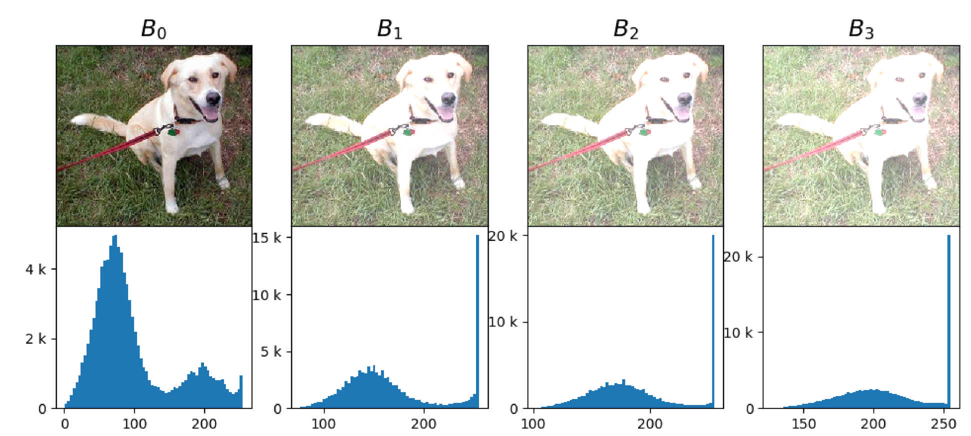
Data and Experimental Setup

- We have used two datasets and CIFAR-10, Dogs vs Cats (DvsC). Table shows how we split the datasets and their main characteristics.

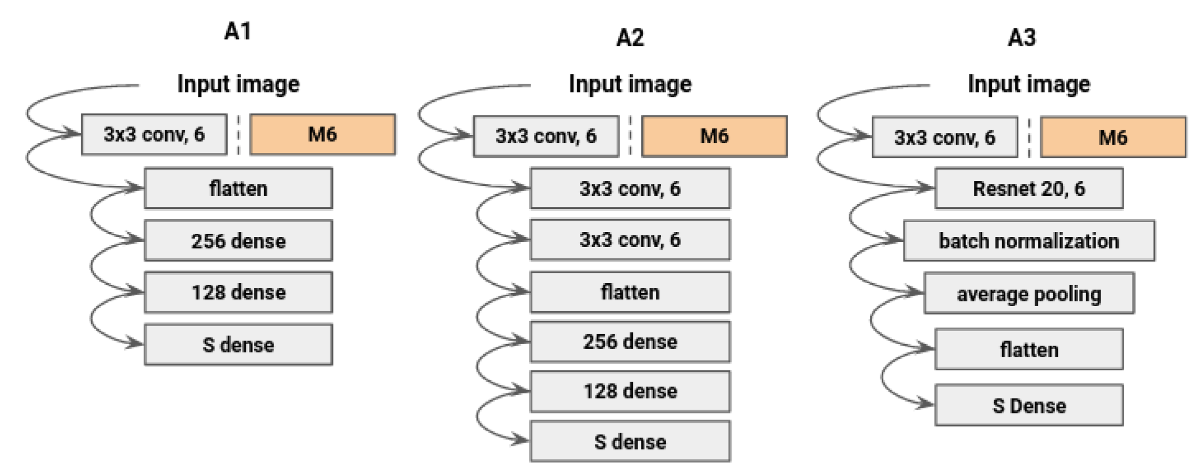
	CIFAR-10	DvsC
Training set	40,000	16,284
Validation set	10,000	3,489
Test set	10,000	3,489
Total	60,000	23,262
Input shape	[32x32x3]	[128x128x3]

- An original image (B_0) and three transformations using Equation 1, B_1, B_2, B_3, where $\alpha = 0, 0.3(255), 0.4(255), 0.5(255)$ respectively. Second row: the histogram of the pixel values of each image.

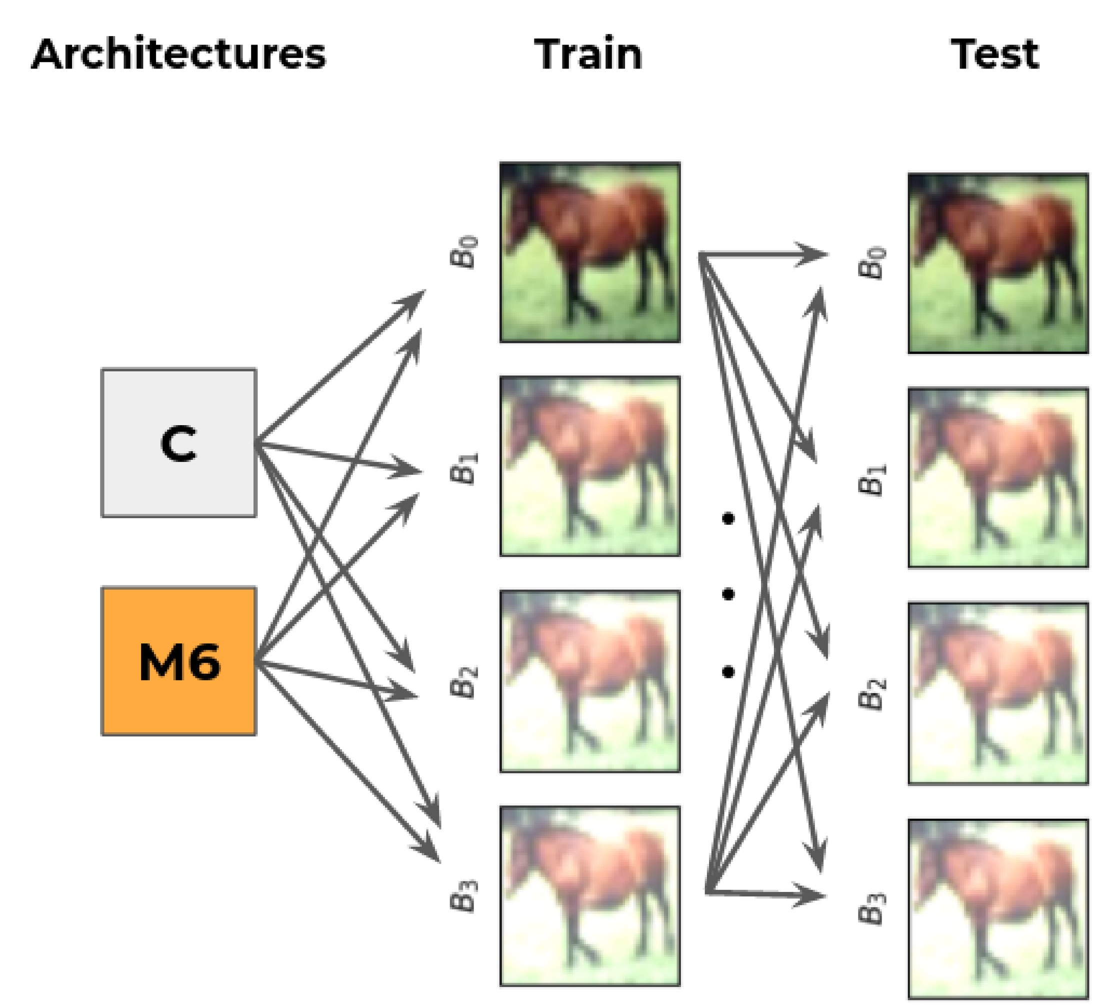
$$I_B(x,y) = \min(I(x,y) + \alpha, 255), \quad (1)$$



- Three architectures used in the experimental setup.



- The experimental setup.

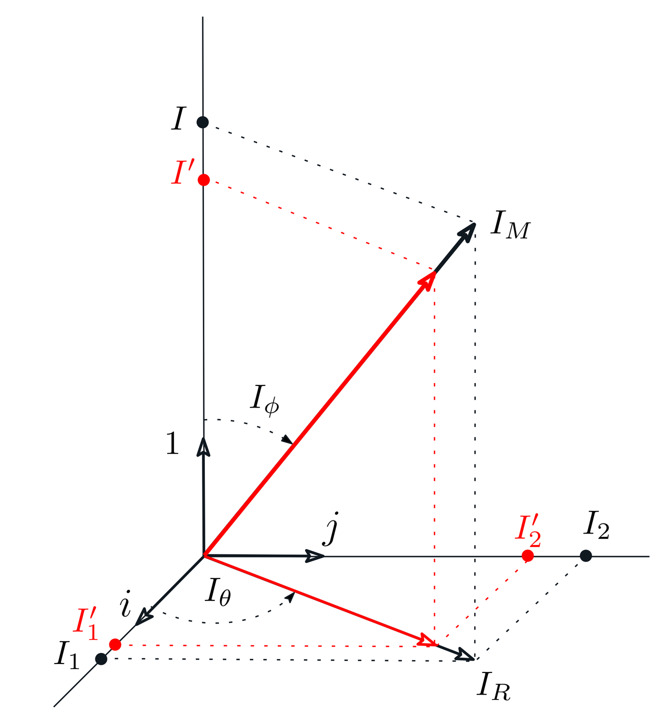


Results

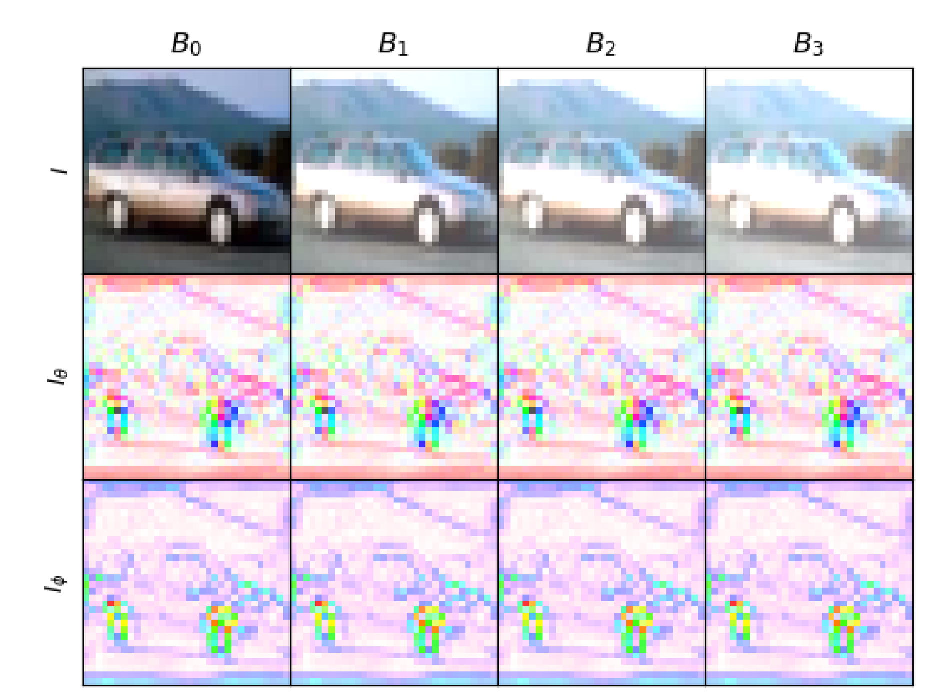
- Accuracy in the test sets using different brightness levels and CNN architectures with C and the M6 layer. Where D=Datasets and, Tr=Train over different brightness levels.

		A_1−C		A_1−M6		A_2−C		A_2−M6		A_3−C		A_3−M6	
D	Tr	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
CIFAR-10	B_0	0.39	9.94E-3	**0.44**	**1.50E-3**	0.44	9.54E-3	**0.50**	**3.85E-4**	0.39	1.96E-2	**0.59**	**1.19E-3**
	B_1	0.44	1.35E-3	**0.47**	**1.44E-4**	0.46	2.45E-3	**0.49**	**1.42E-4**	0.53	2.31E-3	**0.59**	**4.37E-4**
	B_2	0.47	1.15E-3	**0.48**	**1.47E-4**	**0.50**	1.81E-3	0.49	**9.35E-5**	0.52	4.61E-3	**0.56**	**1.05E-4**
	B_3	0.45	2.24E-3	**0.46**	**2.37E-4**	0.50	1.68E-3	**0.51**	**1.08E-4**	0.43	2.21E-2	**0.58**	**3.88E-5**
DvsC	B_0	0.55	3.75E-3	**0.63**	**9.97E-5**	0.61	1.92E-3	**0.69**	**3.29E-5**	0.69	1.00E-2	**0.77**	**1.39E-4**
	B_1	0.62	2.46E-4	**0.68**	**3.52E-5**	**0.69**	1.30E-4	0.62	**3.09E-5**	0.79	2.17E-3	**0.80**	**4.55E-5**
	B_2	**0.62**	**2.97E-5**	0.58	6.33E-5	**0.69**	**7.38E-5**	0.68	7.63E-5	0.75	1.10E-2	**0.78**	**5.92E-5**
	B_3	0.59	2.49E-5	**0.66**	**1.72E-5**	**0.63**	1.36E-4	0.62	**4.16E-5**	0.73	1.00E-2	**0.76**	**3.69E-5**

- We compute a (local) phases I_ϕ I_θ and these values are invariant to the change of the pixel value.



- Activation feature maps with different contrast values. Geometric representation of contrast change.



Conclusions

- We demonstrate that, M6 classifies images in spite of large changes in their brightness.
- The experimental results are consistent with the geometrical observation that the local phase and the local orientation are invariant to variable contrast conditions.
- Among the possible scenarios to use M6, we count self-driving cars under haze conditions, surface glazes in medical images (biopsies), or day-round autonomous video surveillance.