A Quaternion Monogenic Layer Resilient to Large Brightness Changes in Image Classification

Anonymous Author(s)
Affiliation
Address
email

1 Broader Impact

We share the common concern of non-ethical applications of research in machine learning. There is also the risk of possible involuntary collateral effects produced by unexpected responses of layers such as M6 in unfamiliar scenarios. Nevertheless, these concerns do not yet bare on our work. Because of its academic nature, as disclosed by the simple nature of the datasets used, we cannot think about real applications as yet. These will certainly require a bigger research effort. Real world applications will be based on more involved CNNs architectures, as they ought to be capable of performing operations such as image segmentation, object detection, face recognition, etc. in real time and under variable brightness conditions. If we are to consider benefits of our work, the most immediate is for researchers that are in the look for increased resilience of their models in front of brightness variations. Beyond that, we expect that it can be useful for dealing with scenarios in which unpredictable large brightness changes occur.

There are two limitations of our work that have to be mentioned. One is that we found our setup works poorly for $\alpha > 0.6(255)$. The other concern is that our implementation is not optimized, which is the reason why the timings reported for the M6 architectures were greater than those for the C architectures (about 29% for training and 20% for testing).

2 Background

Currently, there is no conventional definition for brightness. In fact, image-processing tools employ several different brightness measurements [1]. Brightness refers to the overall lightness or darkness of the image [2]. In image processing and computer vision, changing brightness of an image is a commonly used point transformation (affecting every pixel in an image). In this transformation, the value of each pixel is increased by a constant. For a one channel image $I = I(x, y) \in \mathbf{R}$ (where $x, y \in U$, U a region of $\mathbf{R}^2$) the relation

$$I_B(x, y) = \min(I(x, y) + \alpha, 255), \tag{1}$$

where $\alpha > 0$ is a constant, defines an image I_B that is brighter than I, the more so the higher the value of α. In Figure 1 we can see an original image B_0 and brighter versions B_i ($i = 1, 2, 3$) corresponding to three values of α. In addition, we display the histogram of the four images. Notice that the contrast also changes and that pixels with $I(x, y) + \alpha \geq 255$ become saturated.

We have changed the brightness of all datasets using equation 1 implemented in Tensorflow 2.1. Figure 2 displays the point transformation from $I(x, y)$ to $I_B(x, y)$. We use degradation labels B_i, $i = 0, 1, 2, 3$, where B_0 corresponds to $\alpha = 0$ (solid line in Figure 2), and B_1, B_2, B_3 to $\alpha = 0.3(255), 0.4(255)$ and $0.5(255)$.

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

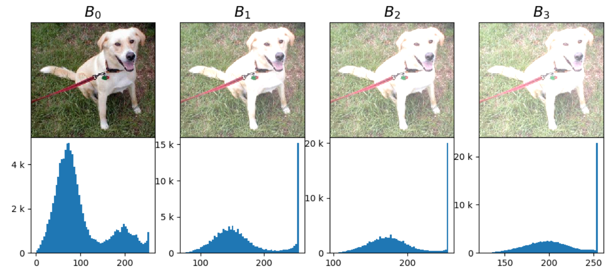

Figure 1: Upper row: an original image (B_0) and three transformations using Equation 1, B_1, B_2, B_3, where $\alpha = 0, 0.3(255), 0.4(255), 0.5(255)$ respectively. Second row: the histogram of the pixel values of each image.

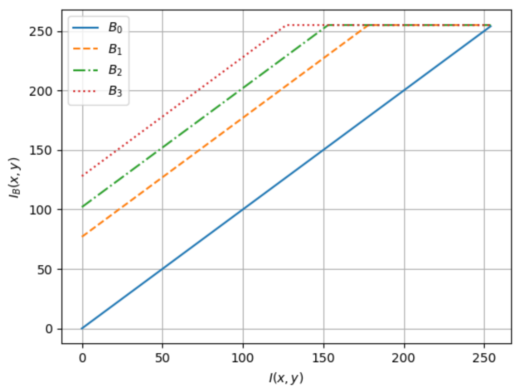

Figure 2: Pixel value transformations of $I(x, y)$ for different levels of brightness B_0, B_1, B_2, B_3

3 Quaternion monogenic layer

In §3.1 we review the general notions about the monogenic signal that are used in §3.2 to describe M6.

3.1 Monogenic signal

We define 1D (resp. 2D) *multivectorial signals* as C^1 maps $U \to \mathcal{G}$ from an interval $U \subset \mathbf{R}$ (a region $U \subset \mathbf{R}^2$) into a *geometric algebra* $\mathcal{G}$ (see [3]). For $\mathcal{G} = \mathbf{R}$ ($\mathcal{G} = \mathbf{C}$, $\mathcal{G} = \mathbf{H}$) we say that the signal is *scalar* (*complex, quaternionic*). For technical reasons, we also assume that signals are in L^2 (that is, the modulus is square-integrable).

The *Riesz-Felsberg transform* maps 2D scalar signals to 2D quaternionic signals. Among the signals obtained in this way, our interest lies in the (quaternionic) *monogenic signals* (see [4] for details). We use a band-pass monogenic signal $I_M = I_M(x, y) \in \mathbf{H}$ associated to an image $I = I(x, y) \in \mathbf{R}$ (where $x, y \in U$, U a region of $\mathbf{R}^2$). The definition of the band-pass I_M is as follows:

$$I_M = I' + I_R, \quad I_R = \boldsymbol{i} I_1 + \boldsymbol{j} I_2, \tag{2}$$

where, $I' = g * I$, $*$ is the convolution operator, $g = g(x, y)$ is a radial (isotropic) bandpass (*Log-Gabor* function), the signals I_1 and I_2 are the *Riesz transforms* (with quadrature filters) of I' in the x and y directions [4]. Note that $I_M \in \langle 1, \boldsymbol{i}, \boldsymbol{j} \rangle \subset \mathbf{H}$. Rewriting equations in Fourier domain we have:

$$I_M = \mathcal{F}^{-1}(J' + J_R), \quad J_R = \boldsymbol{i} J_1 + \boldsymbol{j} J_2, \tag{3}$$

where

$$
\begin{aligned}
&J' = J \cdot G, \quad J_1 = J \cdot H_1 \cdot G, \quad J_2 = J \cdot H_2 \cdot G, \\
&J[u_1, u_2] = \sum_{m_1} \sum_{m_2} I[m_1, m_2] e^{-i2\pi(u_1 m_1 + u_2 m_2)} && (4) \\
&H_1(u_1, u_2) = \frac{u_1}{\sqrt{u_1^2 + u_2^2}}, && (5) \\
&H_2(u_1, u_2) = \frac{u_2}{\sqrt{u_1^2 + u_2^2}}, && (6) \\
&G(u_1, u_2) = \exp\left(- \frac{\log\left(\frac{\sqrt{u_1^2+u_2^2}}{\omega_0^s}\right)^2}{2\log(\sigma)^2} \right), && (7) \\
&\omega_0^s = \frac{1}{\min_w f^{s-1}} && (8)
\end{aligned}
$$

where u_1, u_2 are frequency components, J is 2D Fourier transform $\mathcal{F}$ of I, $\min_{wl}$ is the minimum wavelength, f is a scale factor, $s = 1, 2, \ldots, n_s$ is the current scale.

The *local amplitude signal* $|I_M|$ is defined by $|I_M|(x, y) = |I_M(x, y)|$, where the last expression is the modulus of the quaternion $I_M(x, y)$ [4]. Notice that we have

$$|I_M| = \sqrt{I'^2 + I_R^2}, \tag{9}$$

similarly $|I_R| = \sqrt{I_1^2 + I_2^2}$. The *local phase* I_ϕ and the *local orientation* I_θ associated to I are defined, following [4], by the relations

$$
\begin{aligned}
I_\phi &= \operatorname{atan2}\left(\frac{I'}{|I_R|}\right), && (10) \\
I_\theta &= \operatorname{atan}\left(\frac{-I_2}{I_1}\right), && (11)
\end{aligned}
$$

where the quotients of signals are taken point-wise. For the geometric interpretation of these signals see Figure 3.

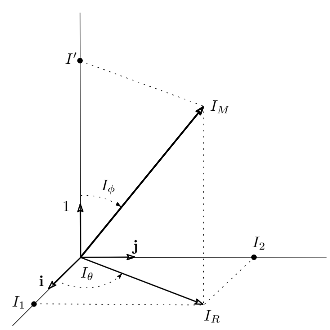

Figure 3: Geometry of the monogenic signal.

3.2 Monogenic layer

The monogenic layer M6 (cf. [5]) is best described by the scheme in Figure 4, where $\mathbf{1}$ in the HSV representation is a ones matrix of $[m, n]$. The Normalization is defined as

$$\text{Normalization}(I) = \frac{I(x, y) - \min(I(x, y))}{\max(I(x, y)) - \min(I(x, y))}, \tag{12}$$

The $(HSV2RGB)$ transforms an HSV image into an RGB image according to the standard color-naming conventions (see page 304 of [6]). See Figure 5 for an illustration of the M6 components

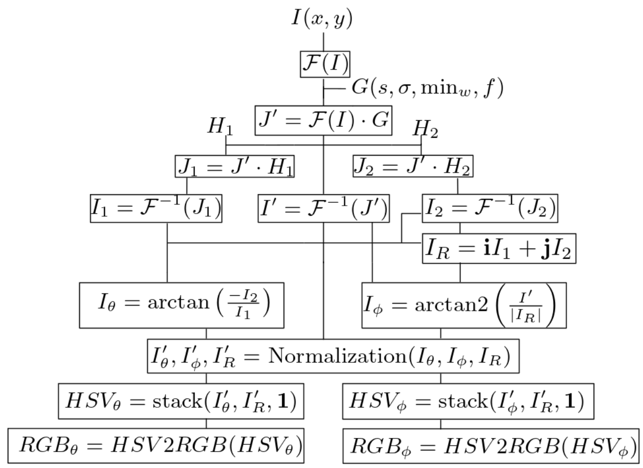

Figure 4: RGB_θ and RGB_ϕ are the outputs of the M6 Layer.

of a simple gray image. Remark that RGB_ϕ enhances lines and edges and RGB_θ enhances the orientation components all over the image. Figure 6 illustrates the six feature maps from image example. In Table 1 we present the main characteristics of a conventional CNN layer C and the M6 layer. Note that the M6 is defined in frequency domain as a result we only have 4 parameters in the layer.

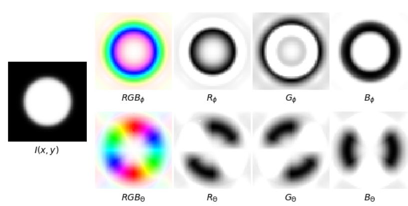

Figure 5: Feature maps of M6 from a circle image input $I(x, y)$.

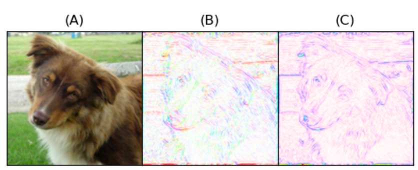

Figure 6: (A) RGB input image. (B) RGB_θ and (C) RGB_ϕ are the output feature maps of the M6 layer.

The implementation of M6 has been coded using Tensorflow 2.1 (TF) and Keras [7, 8].

Appendix 1. Quaternion algebra

The quaternion algebra $\mathbf{H}$ is a four dimensional real vector space with basis $1, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$,

$$\mathbf{H} = \mathbf{R}1 \oplus \mathbf{R}\boldsymbol{i} \oplus \mathbf{R}\boldsymbol{j} \oplus \mathbf{R}\boldsymbol{k} \tag{13}$$

endowed with the bilinear product (multiplication) defined by Hamilton's relations, namely

$$\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -1. \tag{14}$$

As it is easily seen, these relations imply that

$$\boldsymbol{ij} = -\boldsymbol{ji} = \boldsymbol{k}, \quad \boldsymbol{jk} = -\boldsymbol{kj} = \boldsymbol{i}, \quad \boldsymbol{ki} = -\boldsymbol{ik} = \boldsymbol{j}. \tag{15}$$

Table 1: Comparison of the main characteristics of a standard convolutional layer C and the M6 layer.

Characteristics/Name	C	M6
Parameters	$[k_1 \times k_2 \times l]$ kernel shape k_1, k_2, l	4 $s, \sigma, \min_w, f$
Convolution domain	Space	Frequency (Fourier)
Output shape (with input size $[m, n]$, and padding)	$[m, n, l]$	$[m, n, 6]$
Output domain	Space	Space
Nonlinear function	ReLU	$\arctan$, $\arcsin$
Layer position	Any	First hidden
Trainable	Yes	Yes

The elements of $\mathbf{H}$ are named *quaternions*, and $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, *quaternionic units*. By definition, a quaternion q can be written in a unique way in the form

$$q = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}, \quad a, b, c, d \in \mathbf{R}. \tag{16}$$

Its *conjugate*, $\bar{q}$, is defined as

$$\bar{q} = a - (b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}). \tag{17}$$

Note that $(q + \bar{q})/2 = a$, which is called the *real part* or *scalar part* of q, and $(q - \bar{q})/2 = q - a = b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$, the *vector part* of q. Since the conjugates of $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are $-\boldsymbol{i}, -\boldsymbol{j}, -\boldsymbol{k}$, the relations (14) and (15) imply that the conjugation is an *antiautomorphism* of $\mathbf{H}$, which means that it is a linear automorphism such that $\overline{qq'} = \bar{q}'\bar{q}$. Using Hamilton's relations again, we easily conclude that

$$q\bar{q} = a^2 + b^2 + c^2 + d^2. \tag{18}$$

This allows to define the *modulus* of q, $|q|$, as the unique non-negative real number such that

$$|q|^2 = q\bar{q}. \tag{19}$$

Observe that $|qq'| = |q||q'|$. Indeed, $|qq'|^2 = qq'\overline{qq'} = qq'\bar{q}'\bar{q} = q|q'|^2\bar{q} = |q|^2|q'|^2$. Finally, for $q \neq 0$, $|q| > 0$ and $q(\bar{q}/|q|^2) = 1$, which shows that any non-zero quaternion has an inverse and therefore that $\mathbf{H}$ is a (skew) field.

References

[1] S. Bezryadin, P. Bourov, and D. Ilinih, "Brightness calculation in digital image processing," in *International symposium on technologies for digital photo fulfillment*, pp. 10–15, Society for Imaging Science and Technology, 2007.

[2] S. W. Smith *et al.*, "The scientist and engineer's guide to digital signal processing," 1997.

[3] S. Xambó-Descamps, *Real spinorial groups—a short mathematical introduction*. SBMA/Springerbrief, Springer, 2018.

[4] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136–3144, 2001.

[5] E. U. Moya-Sánchez, S. Xambó-Descamps, S. Salazar Colores, A. Sánchez Pérez, , and U. Cortés, "A Quaternion Deterministic Monogenic CNN Layer for Contrast Invariance," in *Systems, Patterns and Data Engineering with Geometric Calculi* (A. Delshams, ed.), ICIAM2019 SEMA SIMAI Springer Series, Springer, 2020. To appear.

[6] M. K. Agoston, *Computer graphics and geometric modeling*, vol. 1. Springer, 2005.

[7] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[8] F. Chollet *et al.*, "Keras." `https://github.com/keras-team/keras`, 2015.