Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts

G. Sfikas[1,3,4] G. Retsinas[2] A.P. Giotis[3] B. Gatos[1]
C. Nikou[3]

[1]CIL / NCSR "Demokritos"

[2]National Technical University of Athens (NTUA)

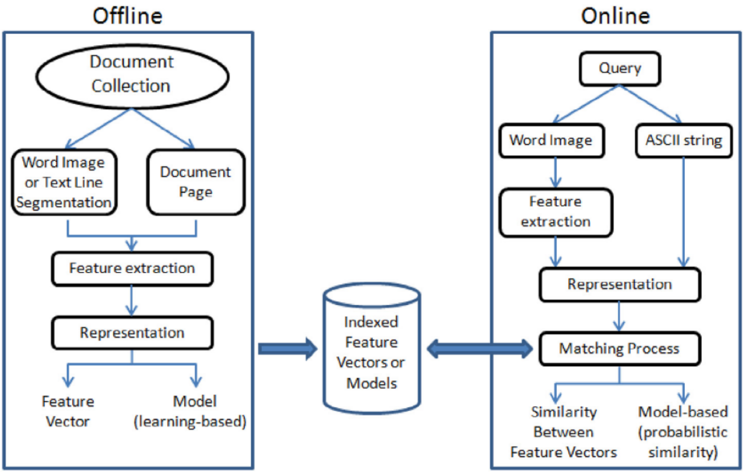
[3]University of Ioannina (UOI)

[4]University of West Attica (Uniwa)

Document Analysis Systems (DAS 2022), May 2022

DAS 2022
May 22-25, La Rochelle

Keyword spotting



Our contribution in a nutshell

- Introduction of a Quaternionic version of the ResNet block
- Testing on KWS
- **Result: As effective as vanilla version but much "lighter"**

Basic form

Quaternions $q \in \mathbb{H}$ share the basic form:

$$q = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$$

where $a, b, c, d \in \mathbb{R}$ and $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are independent imaginary units.

Generalizations of complex numbers

- Two ways to see this:
  - 2 extra imaginary units $(\boldsymbol{j}, \boldsymbol{k})$. Real and complex numbers can be regarded as quaternions with $b, c, d = 0$ or $c, d = 0$ respectively.
  - OR: $a + b\boldsymbol{j}$ with $a, b \in \mathbb{C}$ (Cayley-Dickson form)

Historical note

- William Rowan Hamilton (1805-1865) discovered quaternions, initially in his search of a way to multiply triplets (elements of $\mathbb{R}^3$)
- He discovered that a fourth element is "necessary" ($\mathbb{R}, \mathbb{C}$ and $\mathbb{H}$ are the only possible commutative division algebras [Frobenius, 1878])

Basic "peculiarity": Multiplication is non-commutative

$$pq \neq qp \;\text{ for }\; p, q \in \mathbb{H}$$

Algebraic structure

Quaternions are mathematical objects that form a skew-field $\mathbb{H}$, *i.e.* quaternion addition and multiplication are defined with all the properties of a field, except that of multiplication commutativity

Multiplication rule (Hamilton product)

$$
\begin{aligned}
pq = &(a_p a_q - b_p b_q - c_p c_q - d_p d_q) + &(1)\\
&(a_p b_q + b_p a_q + c_p d_q - d_p c_q)\boldsymbol{i} + &(2)\\
&(a_p c_q - b_p d_q + c_p a_q + d_p b_q)\boldsymbol{j} + &(3)\\
&(a_p d_q + b_p c_q - c_p b_q + d_p a_q)\boldsymbol{k}, &(4)
\end{aligned}
$$

where $p = a_p + b_p\boldsymbol{i} + c_p\boldsymbol{j} + d_p\boldsymbol{k}$ and $q = a_q + b_q\boldsymbol{i} + c_q\boldsymbol{j} + d_q\boldsymbol{k}$.

We can write the above rule also as:

$$pq = S(p)S(q) - V(p)\cdot V(q) + S(p)V(q) + S(q)V(p) + V(p)\times V(q), \quad (5)$$

Properties

Other elementary relations

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

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

$$|q| = \sqrt{q\bar{q}} = \sqrt{\bar{q}q} = \sqrt{a^2 + b^2 + c^2 + d^2}$$

$$\bar{q} = a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}$$

Caley-Dickson form

$$q = \zeta + \eta \boldsymbol{j}$$

$$\zeta = a + b\boldsymbol{i}, \eta = c + d\boldsymbol{i}.$$

Properties

Euler's identity

Euler's identity extends for quaternions, with μ being an arbitrary unit pure quaternion:

$$e^{\mu\theta} = cos\theta + \mu \sin\theta.$$

Convolution

Left-Right convolution

$$(h_L * f)[n,m] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} h_L[i,j] f[n-i, m-j], \tag{6}$$

$$(f * h_R)[n,m] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[n-i, m-j] h_R[i,j], \tag{7}$$

Biconvolution

$$(h_L \prec f \succ h_R)[n,m] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} h_L[i,j] f[n-i, m-j] h_R[i,j]. \tag{8}$$

In Neural Networks

- *Dense layer*

$$f_{dense}(x; W, h) = Wx + h$$

- *Convolution layer*

$$f_{conv}(m; K) = K * m$$

- *ResNet block*

$$f_{ResBlock} = \alpha(\phi(x) + x)$$

Pros and cons of quaternions

Argumentum ab auctoritate

Quaternions have had proponents for their use as well as against since their inception

The arguments against quaternions..

- $\mathbb{H}$ is non-commutative, which creates all sorts of problems
- Set $\mathbb{H}$ is isomorphic to $\mathbb{R}^4$, which is less complex and better understood (so a quaternion can be replaced by a 4-dimensional vector)

Pros and cons of quaternions

..so why should we bother?

+ In the context of Neural Networks, they lead to "light-weight" layer defitions (=requiring much less parameters)

Why are quaternionic layers less costly?

Let us illustrate this with a simple example. Assume a real-valued linear layer, without bias or activation, that transforms an input comprised of 4 neurons to an output of 4 neurons.

- $x \in \mathbb{R}^4$ is mapped to an output $y \in \mathbb{R}^4$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma & \delta \\ \epsilon & \zeta & \eta & \theta \\ \iota & \kappa & \lambda & \mu \\ \nu & \xi & o & \pi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \tag{9}$$

Why are quaternionic layers less costly?

- Quaternionic version of the same operation

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \alpha & -\beta & -\gamma & -\delta \\ \beta & \alpha & -\delta & \gamma \\ \gamma & \delta & \alpha & -\beta \\ \delta & -\gamma & \beta & \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \tag{10}$$

- Corresponds to Hamilton product $y = wx$
- This paradigm extends to cover any-size inputs and outputs (as long as they are multiples of 4)

Architecture

- QNN defined as a cascade of Quaternion ResNet blocks
- On each block: convolution (stride=1, kernel size=3×3) $\rightarrow$ BN $\rightarrow$ ReLU...
- ...$\rightarrow$ convolution $\rightarrow$ BN (function ϕ)
- Whole block: $ReLU(\phi(x) + x)$
- 7 blocks in total
- "Split-activation" sigmoid activation maps output to PHOC estimates

Datasets

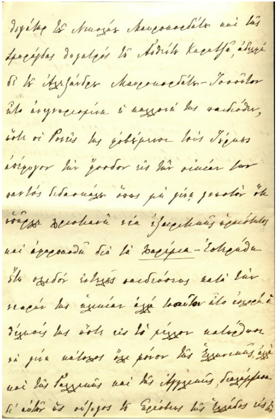
Memoirs

46 manuscripts - segmented into a total of $4,941$ words

PIOP-DAS

22 manuscripts - segmented into a total of $12,362$ words

Datasets



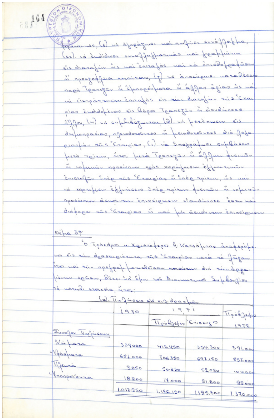


Figure: Sample pages from the two datasets used in this work, "Memoirs" and "PIOP-DAS".

Sample word images

Figure: Word image samples from the “PIOP-DAS” dataset.

Sample word images

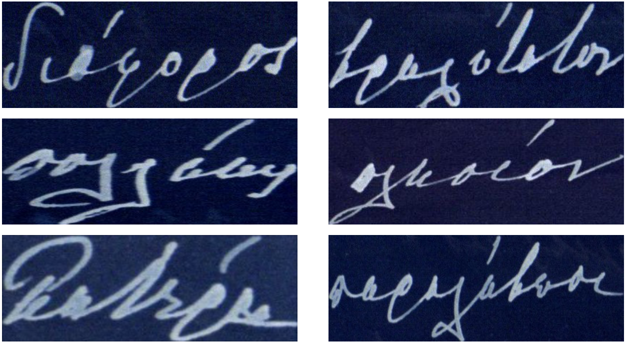


Figure: Word image samples from the “Memoirs” dataset.

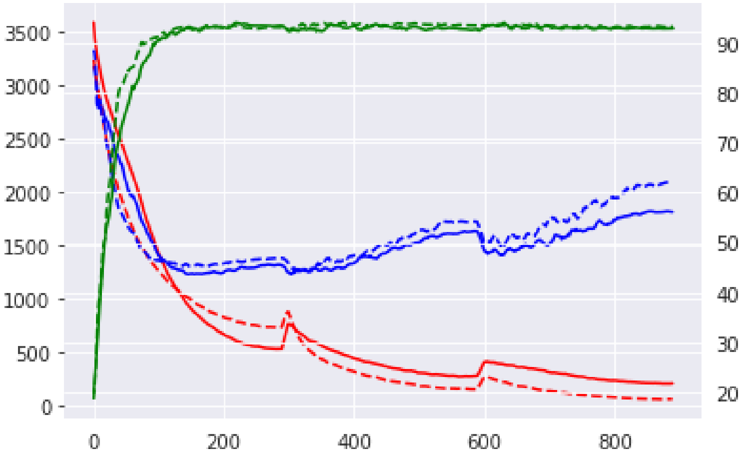


Figure: Results for the "PIOP-DAS" dataset (large models): Training set BCE loss, test set BCE loss and MAP are shown using red, blue and green colors respectively. Solid (dashed) lines correspond to quaternionic (real) models.

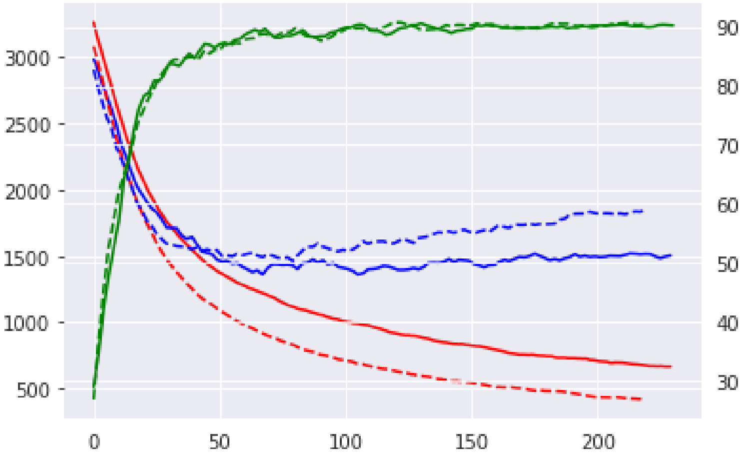


Figure: Results for the "PIOP-DAS" dataset (small models): Training set BCE loss, test set BCE loss and MAP are shown using red, blue and green colors respectively. Solid (dashed) lines correspond to quaternionic (real) models.

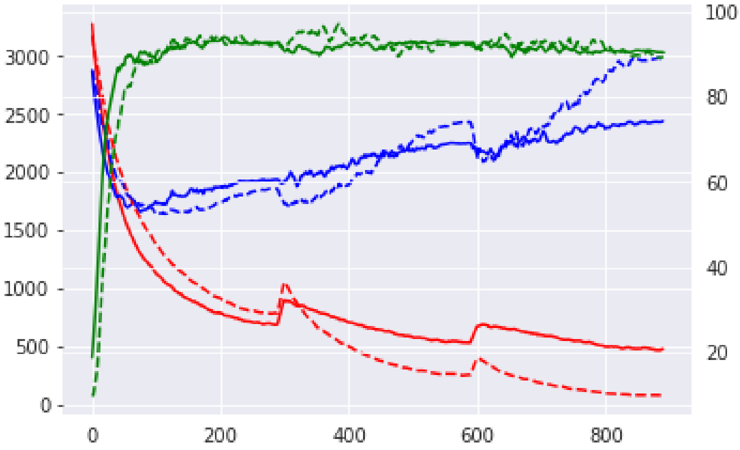


Figure: Results for the "Memoirs" dataset (large models): Training set BCE loss, test set BCE loss and MAP are shown using red, blue and green colors respectively. Solid (dashed) lines correspond to quaternionic (real) models.

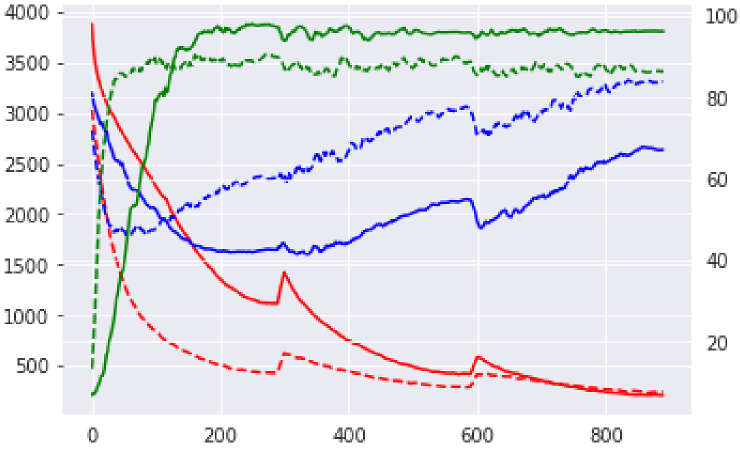


Figure: Results for the "Memoirs" dataset (small models): Training set BCE loss, test set BCE loss and MAP are shown using red, blue and green colors respectively. Solid (dashed) lines correspond to quaternionic (real) models.

Model accuracy comparison

Table: Model accuracy in terms of mean average precision (higher is better). Results correspond to the end of training and best figure reported over all epochs.

MAP%	Memoirs		PIOP-DAS	
	Last epoch	Best epoch	Last epoch	Best epoch
Quaternion / Standard	96.1%	98.6%	92.9%	94.3%
Real / Standard	89.3%	98.6%	93.3%	94.6%
Quaternion / Small	90.2%	94.6%	90.3%	92.8%
Real / Small	86.3	94.5%	90.0%	92.2%

Model size comparison

Table: Model size comparison, in terms of total number of trainable parameters. For the quaternionic variants, the equivalent number of real-valued parameters is reported, to ease comparisons.

Model / Network type	"Small" size	"Standard" size
Quaternionic Resnet	1,317,428	5,946,164
Real-valued Resnet	5,233,840	23,730,864

Summary

- The proposed Quaternionic ResNet works efficiently on the Keyword Spotting task..
- ..and **is on par with its real-valued counterpart.**
- Future work: Parameterized Hypercomplex layer
- Future work: Tests on more datasets, scripts, etc.

Thank you

Thank you for your attention! Questions?

Acknowledgments

This research has been partially co-financed by the EU and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the calls "RESEARCH - CREATE - INNOVATE" (project *Culdile* - code T1EΔK-03785, project *Impala* - code T1EΔK-04517) and "OPEN INNOVATION IN CULTURE" (project *Bessarion* - T6YBΠ-00214).