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

### G. Sfikas<sup>1,3,4</sup> G. Retsinas<sup>2</sup> A.P. Giotis<sup>3</sup> B. Gatos<sup>1</sup> C. Nikou<sup>3</sup>

 $^{1}\text{CIL}$  / NCSR "Demokritos"

<sup>2</sup>National Technical University of Athens (NTUA)

<sup>3</sup>University of Ioannina (UOI)

<sup>4</sup>University of West Attica (Uniwa)

Document Analysis Systems (DAS 2022), May 2022



## Keyword spotting



Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\hfill Contribution Outline$ 

# 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:

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

#### Generalizations of complex numbers

- Two ways to see this:
  - 2 extra imaginary units (j, k). Real and complex numbers can be regarded as quaternions with b, c, d = 0 or c, d = 0 respectively.

• OR: a + bj 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 ℝ<sup>3</sup>)
- ► He discovered that a fourth element is "necessary" (ℝ, ℂ and ℍ are the only possible commutative division algebras [Frobenius, 1878])

イロト イポト イヨト イヨト





### Basic "peculiarity": Multiplication is non-commutative

pq 
eq qp 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)

$$pq = (a_p a_q - b_p b_q - c_p c_q - d_p d_q) +$$
(1)

$$(a_p b_q + b_p a_q + c_p d_q - d_p c_q) \mathbf{i} +$$
(2)

$$(a_pc_q - b_pd_q + c_pa_q + d_pb_q)\mathbf{j} +$$
(3)

$$(a_p d_q + b_p c_q - c_p b_q + d_p a_q) \boldsymbol{k}, \qquad (4)$$

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

We can write the above rule also as:

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

Properties

Other elementary relations

$$i^{2} = j^{2} = k^{2} = ijk = -1$$
$$ij = -ji = k, jk = -kj = i, ki = -ik = j$$
$$|q| = \sqrt{q\bar{q}} = \sqrt{\bar{q}q} = \sqrt{a^{2} + b^{2} + c^{2} + d^{2}}$$
$$\bar{q} = a - bi - cj - dk$$

Caley-Dickson form

$$q = \zeta + \eta \mathbf{j}$$
  
$$\zeta = \mathbf{a} + b\mathbf{i}, \eta = \mathbf{c} + d\mathbf{i}.$$





ヘロト ヘロト ヘヨト ヘヨト

## Properties

### Euler's identity

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

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





### Convolution

Left-Right convolution

$$(h_L * f)[n, m] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} h_L[i, j] f[n - i, m - j], \qquad (6)$$
$$(f * h_R)[n, m] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[n - i, m - j] h_R[i, j], \qquad (7)$$

**Biconvolution** 

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

# In Neural Networks



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

Convolution layer

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

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





## 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 ℍ is isomorphic to ℝ<sup>4</sup>, 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  $y \in \mathbb{R}^4$ 

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





# Why are quaternionic layers less costly?

Quaternionic version of the same operation

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \alpha & -\beta & -\gamma & -\delta \\ \beta & \alpha & -\delta & \gamma \\ \gamma & \delta & \alpha & -\beta \\ \delta & -\gamma & \beta & \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix},$$
(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) → BN → ReLU...
- ▶ ...→ convolution → BN (function  $\phi$ )
- Whole block:  $ReLU(\phi(x) + x)$
- 7 blocks in total
- "Split-activation" sigmoid activation maps output to PHOC estimates





Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\bigsqcup$  Architecture & Results

### 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

Toyathe les Anoyas Augoacchilo and has specielas degalars to Aris's kacella iluci R & Aufarter charge architer Josoilor its is yroccopies is agyone by rachader, iste oi Posis by goliprior lois Tigues isigoyor his goodor is his vicien lor serles Telerviers isos på ging poweler ik its accolesie sie ifaquelaning iquichly wei appendie bis la bagique - loheada The opedin istys sacdionay sale his reach has aquies age leaver its inverti Vincy by into is to sinor actic on si pice wilcours in poror his quereaus, in wai his Payrein's wai his ityyeans , Suggefore Ti color is referres to Excisting his Excises with

164 (2 (3 )3)				
Ser. 9	N		care analy	Jakha
Aspenneds, U	1 - aleela	N.o.	noi boah	
Colo Sa cual	the same line	in the	in Sura	Jeation
		( ) we he		andine
and Josef	and a Nume	anna li	assas agi	an in we
up Canadar	man 2 marks	a as aire	anafin .	ins'ere
pros andres	finan an Que	general as	an in dress of	ince
2220, 60 -	- upleling	m. (8)	herizon	- 63
Information	a mandora	un in frees	Summes	alog and
enoper ac	s'amoins,	and in	grinfines es	places
prese reven	, True Juni	Tenneza.	i iggan	frendy
in reference	-	ie xoeig	un effer	and a
inenfan i	nie ma 'Car	uging 2 to	بسطيد يعذ	the said
in expeription	- ifficien	the spa	- francis	in spece
agazinan 2	and marked	a frequencia	عممتقمعة	Server unit
distant in	S Three grave	ñ nai fie	hands and	incore
Cilia 35				
0.269	egin - Xore	if ago file	areifinas i	oficefori
in a sin des	ampirium	ette "Comple	in and it	Jugar
the way here of	after frances	esan main	and did at	·· àcla.
farme appear,	Weren bit tipe	not kinnen	are server	- Banglion
the second come	the inter:			
	a nagineto	1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	10.	
	1910	Truck		Trailgate
		11000031000	encodio	1935
Rule Var				
Disgo Tropice				
Derezon Trafficer Michaelan Michaelan	mus	412.924	359.300	231.000
Derezon Tenzinan Min Juana Min Juana Min Juana	511400 651400	4/2.450 Tea.350	354.305 491.150	531.000
Denzon Tergine Minfrance Martine Densie Densie	Stress etiana gota	4/2.920 Tes.350 59.550	554.305 693.50 52.450	531.000 531.000
Angen Topico Michan Michan Michan Michan Michan Michan	10000 10000 10000 10000	412.420 To5.350 59.350 (T.408	154.305 691.150 52.450 81.800	231.000 531.000 -0.000 -0.000

(日)

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



Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\bigsqcup_{}$  Architecture & Results

# Sample word images

duagenias diaragen spor upyor 2000 - crefizen Mr

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





э

Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\bigsqcup_{}$  Architecture & Results

# Sample word images



Figure: Word image samples from the "Memoirs" dataset.





イロト イポト イヨト イヨト



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.



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. (additioned) lines correspond to quaternionic (real) models.

ヘロト ヘロト ヘビト ヘビト



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.



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.

# 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%





э

# 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





Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\cap{Conclusion}$ 



- + 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.





Keyword Spotting with Quaternionic ResNet: Application to Spotting in Greek Manuscripts  $\cap{Conclusion}$ 

# 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 T1E $\Delta$ K-03785, project *Impala* - code T1E $\Delta$ K-04517) and "OPEN INNOVATION IN CULTURE" (project *Bessarion* - T6YBIT-00214).



