



UM ESTUDO SOBRE REDES NEURAIIS PROFUNDAS DE VALOR VETORIAL PARA CLASSIFICAÇÃO DE IMAGENS

Palavras-chave: Álgebra, números hipercomplexos, aprendizado profundo, classificação de imagens.

Autores/as:

Carolina Rodrigues Menezes¹ – IMECC/UNICAMP

Prof. Dr. Marcos Eduardo Ribeiro do Valle Mesquita (Orientador) – IMECC/UNICAMP

INTRODUÇÃO:

As redes neurais convolucionais (CNNs) são exemplos de rede neurais profundas que contém um tipo particular de operação baseada na convolução de filtros, sendo bem adequadas para aprender e representar padrões locais (Lecun et al., 2015). Consequentemente, são frequentemente consideradas como base dos modelos de ponta em processamento de imagens e reconhecimento de padrões.

Diferente dos modelos tradicionais, redes neurais de valor vetorial (V-nets) utilizam vetores em vez de números reais para representar as entradas, saídas e os pesos sinápticos (Valle, 2024). Portanto, são adequadas para processar dados multidimensionais como imagens coloridas e hiperespectrais, por se beneficiarem das propriedades geométricas da álgebra utilizada. Vários trabalhos demonstram a vantagem das V-nets, especificamente das redes hipercomplexas, sobre modelos equivalentes com valores reais, especialmente em tarefas envolvendo dados de vários canais, como o processamento de imagens (Breuils et al., 2022; Parcollet et al., 2020; Vieira and Valle, 2022a). Trabalhos recentes também mostram que redes em álgebras hipercomplexas se destacam na redução da complexidade computacional, ao mesmo tempo que fornecem desempenho semelhante ou superior em comparação com modelos com valores reais (Grassucci et al., 2023, 2024; Vieira and Valle, 2022b).

Nesse trabalho, revisamos os conceitos básicos das álgebras não-associativas e estudamos modelos tradicionais e vetoriais de redes neurais convolucionais. Finalmente, investigamos aplicações das V-nets em problemas de classificação de imagens como, por exemplo, no diagnóstico de leucemia linfoblástica aguda (Granero et al., 2021; Vieira and Valle, 2022b).

ÁLGEBRAS NÃO-ASSOCIATIVAS E HIPERCOMPLEXA:

Uma álgebra não-associativa é um espaço vetorial com uma multiplicação de vetores (Schafer, 1961). Como uma operação bilinear, a multiplicação de $x, y \in \mathbb{V}$, denotada por xy , satisfaz

$$(x + y)z = x + yz, \quad z(x + y) = zx + zy, \quad \text{e} \quad \alpha(xy) = (\alpha x)y = x(\alpha y), \quad \forall \alpha \in \mathbb{R} \text{ e } \forall x, y, z \in \mathbb{V}. \quad (1)$$

¹Bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Consideraremos apenas espaços vetoriais de dimensão finita nesse trabalho.

Dada uma base ordenada $\mathcal{E} = \{e_1, \dots, e_n\}$ em $\mathbb{V}$ e $x \in \mathbb{V}$, existe uma única tupla $(x_1, \dots, x_n) \in \mathbb{R}^n$ tal que $x = \sum_{i=1}^n x_i e_i$, onde $x_1, \dots, x_n$ são as coordenadas x relacionadas à base ordenada $\mathcal{E}$. A multiplicação em $\mathbb{V}$ é completamente caracterizada pelo produto dos elementos da base $\mathcal{E} = \{e_1, \dots, e_n\}$. Precisamente, seja

$$e_i e_j = \sum_{k=1}^n p_{ijk} e_k, \quad \forall i, j = 1, \dots, n \quad (2)$$

o produto entre dois elemento de $\mathcal{E}$. Esse produto pode ser organizado na chamada tabela de multiplicação que caracteriza completamente a álgebra não-associativa.

Uma álgebra hipercomplexa $\mathbb{H}$ é uma álgebra não-associativa de dimensão finita onde o produto tem uma identidade pelos dois lados (Kantor and Solodovnikov, 1989; Valle, 2024). A base canônica de uma álgebra hipercomplexa $\mathbb{H}$ de dimensão $\dim(\mathbb{H}) = n+1$ é $\tau = \{1, i_1, \dots, i_n\}$, e um número hipercomplexo é dado por $x = x_0 + x_1 i_1 + \dots + x_n i_n$. Assim, as tabelas de multiplicação de números complexos, quatérnios e octônios são dadas em relação às bases canônicas $\tau_{\mathbb{C}} = \{1, i_1\}$, $\tau_{\mathbb{Q}} = \{1, i_1, i_2, i_3\}$, e $\tau_{\mathbb{O}} = \{1, i_1, i_2, i_3, i_4, i_5, i_6, i_7\}$, respectivamente.

REDES NEURAIS CONVOLUCIONAIS DE VALOR VETORIAL:

Ao usar “vetores de vetores” no lugar de vetores multidimensionais de números reais, V-nets naturalmente levam em conta a intercorrelação entre canais de características usando álgebra vetorial, tornando estes modelos menos propensos a pararem em mínimos locais durante o treinamento. Assim, geralmente têm o treinamento mais robusto em comparação a redes tradicionais e tipicamente têm menos parâmetros. Quando trabalhando com dados de valor vetorial, como sinais de áudio ou imagens coloridas em RGB, modelos construídos especificamente para lidar com “vetores de vetores” podem performar melhor do que modelos tradicionais com vetores de números reais (Valle, 2024). O próximo parágrafo descreve brevemente a camada convolucional usada em redes de valor vetorial, como descrito em (Valle, 2024).

Seja x a imagem de entrada com C canais de características de valor vetorial. Denotamos como $x(p, c) \in \mathbb{V}$ o conteúdo do c -ésimo canal de característica localizado no pixel $p \in D_x$, onde D_x é o domínio de x . Os pesos de uma camada convolucional com K filtros são arranjados em um vetor W tal que $W(q, c, k) \in \mathbb{V}$ é o peso do k -ésimo filtro no c -ésimo mapa de características em $q \in D$, onde D é o domínio do filtro. Assim, a convolução de valor vetorial de W e x é dada por

$$(W * x)(p, k) = \sum_{c=1}^C \sum_{q \in D} W(q, c, k) x(p + S(q), c), \quad p \in D_x, \quad \forall k = 1, \dots, K \quad (3)$$

onde $p + S(q)$ é uma translação que pode contemplar *strides*. A saída de uma camada convolucional é obtida ao adicionar um termo de viés e aplicar uma função de ativação da forma $y = \psi(W * x + b)$.

Redes neurais convolucionais tradicionais incluem, além de camadas densas e convolucionais, camadas de *pooling* (Lecun et al., 2015). Enquanto há estudos atualmente buscando a criação de camadas de *pooling* de valor vetorial, podemos usar um método que combina sua versão tradicional com camadas convolucionais e densas de valor vetorial. Esse método aplica de forma eficaz as camadas de *pooling* componente a componente (Valle, 2024). Além disso, V-nets podem incorporar uma camada de saída densa tradicional para problemas de classificação. Uma camada densa de valor real é equivalente a tomar a parte real de uma camada de valor hipercomplexo (Valle, 2024).

EXPERIMENTO COMPUTACIONAL:

Consideramos o banco de dados “ALL-IDB: Acute Lymphoblastic Leukemia Image Database for Image Processing”, desenvolvido por Fabio Scotti e colaboradores no *Department of Information Technology, Università degli Studi di Milano* (Labati et al., 2011). Este banco de dados contém 260 imagens no total, com 130 imagens classificadas como células saudáveis (classificação 0) e 130 imagens classificadas como linfoblastos (classificação 1).

Codificação Hipercomplexa: Complexos, Quatérnios e Octônios

Seja $x_{\mathbb{R}} = [R, G, B]$ os valores de RGB da imagem colorida. As seguintes estratégias de codificação foram usadas na codificação da imagem RGB em imagens de valor complexo e quaterniônico:

- **Codificação Complexa:** O pixel de cores RGB é codificado em dois números complexos pela equação:

$$x_{\mathbb{C}}^{(1)} = [R\mathbf{i}, G + B\mathbf{i}]. \quad (4)$$

- **Codificação Quaterniônica:** O pixel de cores RGB é codificado como um quatérnio pela equação:

$$x_{\mathbb{Q}}^{(1)} = R\mathbf{i} + G\mathbf{j} + B\mathbf{k}. \quad (5)$$

Além da codificação dada por (4) e (5), consideramos também uma codificação parametrizada, dada pelas expressões:

$$x_{\mathbb{C}}^{(2)} = [(\alpha_{1,0}R + \alpha_{2,0}G + \alpha_{3,0}B + \beta_0) + (\alpha_{1,1}R + \alpha_{2,1}G + \alpha_{3,1}B + \beta_1)\mathbf{i}, \quad (6)$$
$$(\alpha_{1,2}R + \alpha_{2,2}G + \alpha_{3,2}B + \beta_2) + (\alpha_{1,3}R + \alpha_{2,3}G + \alpha_{3,3}B + \beta_3)\mathbf{i}]$$

$$x_{\mathbb{Q}}^{(2)} = (\alpha_{1,0}R + \alpha_{2,0}G + \alpha_{3,0}B + \beta_0) + (\alpha_{1,1}R + \alpha_{2,1}G + \alpha_{3,1}B + \beta_1)\mathbf{i} \quad (7)$$
$$+ (\alpha_{1,2}R + \alpha_{2,2}G + \alpha_{3,2}B + \beta_2)\mathbf{j} + (\alpha_{1,3}R + \alpha_{2,3}G + \alpha_{3,3}B + \beta_3)\mathbf{k}$$

Note que, como $x_{\mathbb{C}}^{(1)}$, a imagem de valor complexo $x_{\mathbb{C}}^{(2)}$ tem dois canais. De forma similar, a imagem RGB é codificada em uma imagem de valor octoniônico pela equação

$$x_{\mathbb{O}} = (\alpha_{1,0}R + \alpha_{2,0}G + \alpha_{3,0}B + \beta_0) + \sum_{k=1}^7 (\alpha_{1,k}R + \alpha_{2,k}G + \alpha_{3,k}B + \beta_k)\mathbf{i}_k. \quad (8)$$

Os parâmetros $\alpha_{i,k}$ e β_k são ajustados durante o treinamento dos modelos V-CNN.

Resultados e Análise

Treinamos 10 vezes cada um dos nossos 6 modelos, cujas arquiteturas exibem um número total de parâmetros decrescente. Especificamente, a CNN de valor real tem 250.050 parâmetros, enquanto o modelo de valor octoniônico tem apenas 47.010 parâmetros (contados como números de ponto flutuante).

A Figura 1 ilustra o gráfico *boxplot* da acurácia atingida por cada modelo no conjunto de teste e a Figura 2 apresenta um diagrama de Hasse, onde os modelos de melhor performance são posicionados no topo. Nesse diagrama, uma linha indica que o modelo acima performou melhor que o modelo abaixo, determinado por um teste de hipótese pareado com 95% de nível de significância.

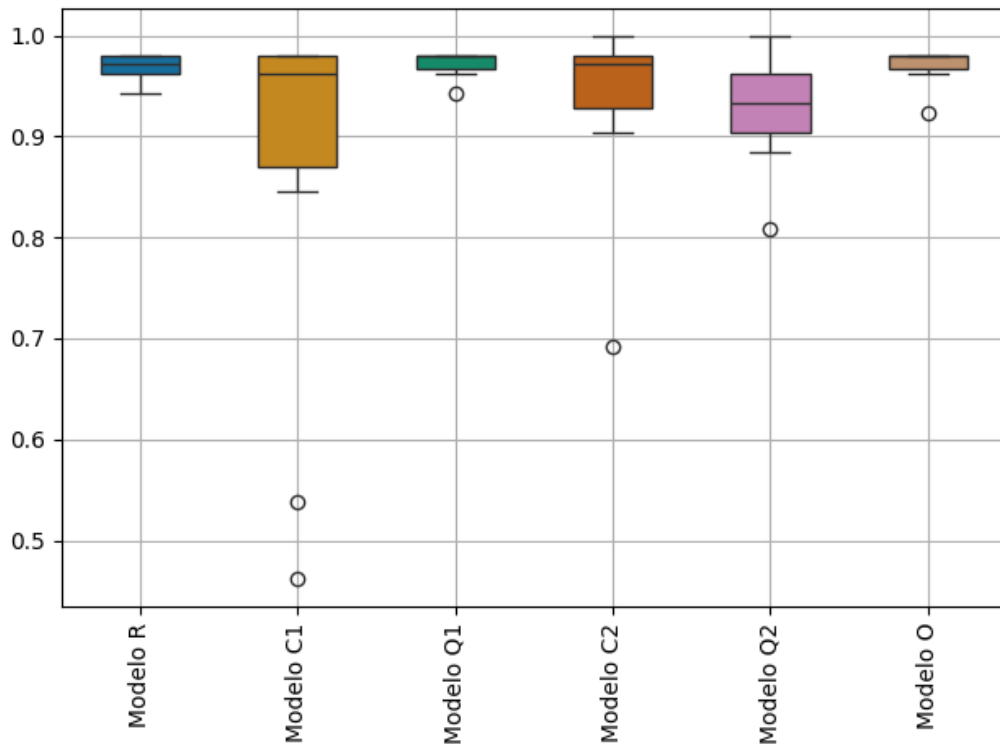


Figura 1: *Boxplot* com a acurácia dos modelos CNN no conjunto de teste.

Os modelos C1, C2 e Q2 exibiram as maiores variações, com os modelos C1 e Q2 tendo também os menores valores de acurácia. Por outro lado, os modelos R, C2, Q1 e O alcançaram as maiores acurácias e demonstraram equivalência estatística. Podemos concluir que V-Nets, especialmente as de valor hipercomplexo discutidas neste relatório, performam tão bem quanto ou mesmo ligeiramente melhor que suas versões de valor real. Isso é conquistado ao mesmo tempo que vemos uma redução significativa do número de parâmetros, e, conseqüentemente, dos recursos computacionais necessários para armazenar estes modelos.

CONSIDERAÇÕES FINAIS:

Neste relatório, revisamos as bases algébricas de redes neurais de valor vetorial e hipercomplexo. Demonstramos que modelos baseados nos números complexos, quatérnios e octônios podem atuar na classificação de imagens coloridas de forma eficaz. Especificamente, avaliamos a performance de redes neurais de valor hipercomplexo usando o banco de dados *ALL-IDB* (Labati et al., 2011). Nossos resultados indicam que modelos de valor vetorial conseguem reduzir o número de parâmetros sem comprometer a acurácia. Essa conclusão é especialmente significativa no contexto de aprendizado profundo leve (*lightweight deep learning*), já que reduz significativamente o número de parâmetros do modelo.

BIBLIOGRAFIA

- S. Breuils, K. Tachibana, and E. Hitzer. New Applications of Clifford's Geometric Algebra. *Advances in Applied Clifford Algebras* 2022 32:2, 32(2):1–39, 2 2022.
- M. A. Granero, C. X. Hernández, and M. E. Valle. Quaternion-Valued Convolutional Neural Network Ap-

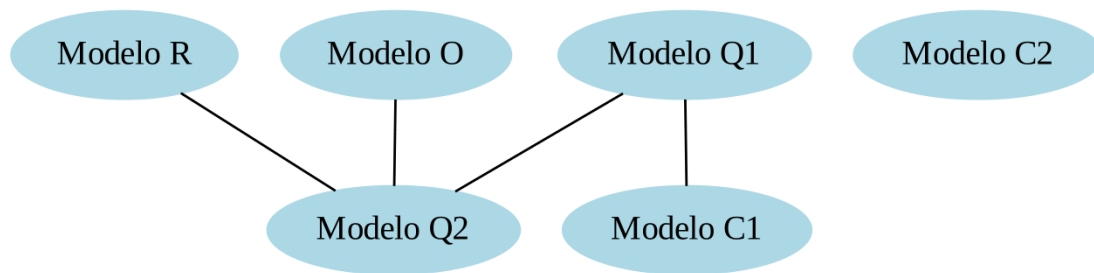


Diagrama Hasse do teste de postos sinalizados de Wilcoxon
(nível de confiança de 95.0%)

Figura 2: Diagrama Hasse comparando os modelos CNN no conjunto de teste.

plied for Acute Lymphoblastic Leukemia Diagnosis. In *Brazilian Conference on Intelligent Systems. BRACIS 2021. Lecture Notes in Computer Science*, pages 280–293. Springer, Cham, 11 2021.

E. Grassucci, G. Mancini, C. Brignone, A. Uncini, and D. Comminiello. Dual quaternion ambisonics array for six-degree-of-freedom acoustic representation. *Pattern Recognition Letters*, 166:24–30, 2 2023.

E. Grassucci, A. Zhang, and D. Comminiello. Phnns: Lightweight neural networks via parameterized hypercomplex convolutions. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):8293–8305, 2024.

I. L. Kantor and A. S. Solodovnikov. *Hypercomplex Numbers: An Elementary Introduction to Algebras*. Springer New York, 1989.

R. D. Labati, V. Piuri, and F. Scotti. ALL-IDB: The acute lymphoblastic leukemia image database for image processing. In *Proceedings - International Conference on Image Processing, ICIP*, pages 2045–2048, 2011.

Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 5 2015.

T. Parcollet, M. Morchid, and G. Linarès. A survey of quaternion neural networks. *Artificial Intelligence Review*, 53(4):2957–2982, 4 2020.

R. Schafer. *An Introduction to Nonassociative Algebras*. Project Gutenberg, 1961.

M. E. Valle. Understanding Vector-Valued Neural Networks and Their Relationship With Real and Hypercomplex-Valued Neural Networks: Incorporating intercorrelation between features into neural networks. *IEEE Signal Processing Magazine*, 41(3):49–58, 5 2024.

G. Vieira and M. E. Valle. A general framework for hypercomplex-valued extreme learning machines. *Journal of Computational Mathematics and Data Science*, 3:100032, 6 2022a.

G. Vieira and M. E. Valle. Acute Lymphoblastic Leukemia Detection Using Hypercomplex-Valued Convolutional Neural Networks. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 7 2022b.