Utilização de Redes Neurais Convolucionais Complexas Para Classificação de Alfabeto de Língua de Sinais

Vitor Boldrin

Instituto de Matemática, Estatística e Computação Cientifica – Universidade Estadual de Campinas (UNICAMP)

Caixa Postal 6065 13083-859 Campinas, SP, Brasil

v245500@dac.unicamp.br

Abstract. *The task of image classification for many years was based on extracting aspects of the image. This technique was revolutionized by the creation of convolutional neural networks that automated and improved this task. In this article, we use this model to classify images of the sign language alphabet. Using not only striking variations like the use of residual learning, normalizations in artificial networks as well as complex algebra. Which in the last decade has led to more accurate networks compared to equivalent networks using real algebra depending on the applications. Finally, we also compare the efficiency of algebras in the classification task using artificial networks. This study hopes to assist in the application of technology to help people of the deaf community. As well as in the study of artificial neural networks of complex values.*

Resumo. *A tarefa de classificação de imagens por muitos anos se baseava na extração de aspectos da imagem. Tal técnica foi revolucionada com a criação das redes neurais convolucionais que automatizaram e aperfeiçoaram essa tarefa. Neste artigo utilizamos este modelo para classificar as imagens do alfabeto da língua de sinais. Utilizando não só variações marcantes como utilização de blocos de resíduos e normalizações nas redes artificiais, como também a álgebra complexa. Que na última década tal escolha tem levado a redes mais precisas comparada com redes equivalentes usando a álgebra real, dependendo da aplicação. Para ao fim também contrastar a eficiência das álgebras na tarefa de classificação utilizando redes artificiais. Este estudo espera auxiliar na aplicação de tecnologia para ajudar pessoas da comunidade surda. Assim como no estudo de redes neurais artificiais de valores complexos.*

1. Introdução

Redes Neurais Artificiais (RNA) são potentes modelos matemáticos inspirados no cérebro humano [Zakaria, AL-Shebany, Sarhan 2014]. Assim como os neurônios do cérebro possuem ligações físicas. Os neurônios artificiais estão conectados entre si por valores. Assim, tais parâmetros ditam a força dessa conexão entre os neurônios impactando na saída da rede. Esses neurônios estão separados em camadas, em que a primeira camada recebe o dado de entrada e processa baseado nos parâmetros dos neurônios. Dessa maneira, a segunda camada recebe a saída da primeira camada, realiza novas operações e passa para a terceira camada. Até o momento em que atingimos a última camada em que é retornado o resultado, podendo ser uma classificação, um valor

ou mesmo uma imagem. Assim, O treinamento das RNA's é realizado alimentando o modelo com dados já conhecidos e ajustando os valores dos neurônios para que a saída corresponda com o que é esperado. Conforme as entradas vão atravessando as camadas, os pesos sinápticos são ajustados para reproduzir a saída esperada utilizando o algoritmo do *backpropagation*.

Redes Neurais Artificiais são extremamente versáteis. Sendo aplicada em diversas áreas como engenharia, medicina, economia e linguagens. Com modelos extremamente precisos, as RNA's já impactam a vida de milhões de pessoas e a economia global [Kachwala 2024].

Das RNA's surge as Redes Neurais Convolucionais (RNC), modelos inspirados no córtex visual dos animais [Aloysius and Geetha 2017]. Usando principalmente para classificação de imagens as RNC hoje também possui outras aplicações por exemplo em detecção de emoções, reconhecimento de texto, classificação de vídeos entre outros. As redes convolucionais são semelhantes às RNA'a a principal diferença é o comportamento das camadas e neurônios que buscam extrair características da imagem de entrada. Tais camadas geram mapas de características que são passados para as camadas mais profundas. Assim, camada por camada é filtrado os principais atributos da imagem de entrada para ao final gerar a classificação da imagem.

Redes neurais artificiais e RNC's realizam suas operações baseadas na álgebra dos números reais. Outras álgebras como os números complexos apresentam uma variedade de soluções para problemas reais. Além de aplicações dos complexos no cálculo para solucionar raízes de polinômios, aplicações na engenharia elétrica na análise das oscilações de correntes elétricas. A álgebra dos complexos pode servir de base para os cálculos de uma rede neural. Tirando proveito das propriedades dos complexos e assim, dependendo da aplicação, podendo atingir resultados superiores a redes de álgebra real. Como no exemplo clássico do *ou exclusivo [*Nitta, Tohru 2003]

2. Classificação do alfabeto de lingua de sinais

A língua de sinais é um meio de comunicação utilizado pela comunidade surda. Tal língua é utilizada entre pessoas com impedimento auditivo permanente. No Brasil, esse grupo representa cerca de 13% dos habitantes [IBGE 2019]. E fora desse grupo, a língua de sinais é de pouco conhecimento da população. Tal fato gera uma exclusão das pessoas com deficiência, marginalizando e diminuindo a qualidade de vida.

Dessa forma, o objetivo desse artigo é gerar uma rede neural convolucional capaz de identificar as letras do alfabeto da língua de sinais. Adicionalmente, será explorado redes neurais complexas e comparado suas performances e resultados.

Para isso, será utilizado um dataset de imagens contendo o alfabeto da língua de sinais *Amerian Sign Language* (ALS) [Khalid 2019], pela vasta quantidade de dados e sua qualidade. As redes neurais utilizadas poderiam ser empregadas na mesma tarefa usando um banco de dados da Língua Brasileira de Sinais (LIBRAS). Dessa forma, analisando o dataset da ALS, as imagens são feitas por pessoas com diversos tipos de fundo nas imagens. O conjunto de dados é amplo e além das classes do alfabeto possui também uma classe adicional contendo fotos de objetos e locais aleatórios. A tarefa será treinar modelos que atinjam a maio precisão possível.

A figura 1 mostra exemplo de dados que será usado no experimento.

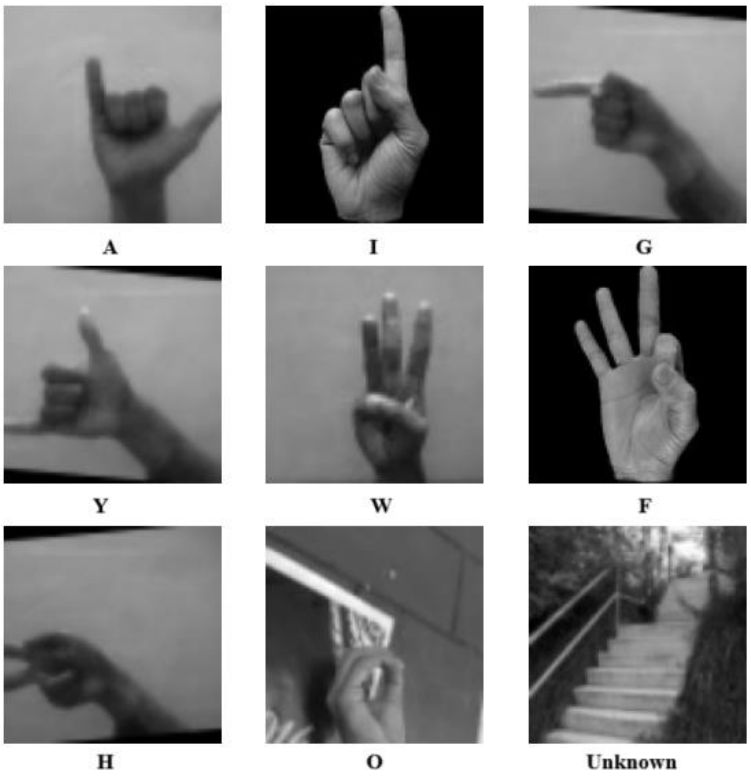

Figura 1. Exemplo de imagens e suas respectivas classes

3. Breve Revisão de Números Complexos e a Transformada de Fourier

Começamos introduzindo o símbolo *i* que representa a raiz quadrada de -1, *i.e.,* $i^2 = -1$. Dessa forma, um número complexo é representado da forma $a+bi$, em que a e b são números reais. Além disso, o número contém 2 partes, a sendo a parte real e b sendo a parte imaginária.

Com está representação os números complexos passam a possuir propriedades diferentes da álgebra real. Uma delas é a fase do número complexo, medido em radianos a fase c é dada por: $c = arctan\ b/a$.

Não obstante, a multiplicação e adição de dois números complexos ficam:

$$(a + bi) + (c + di) = (a + c) + (b + d)i \quad (1)$$

$$(a + bi)(c + di) = ac + adi + bcj + bdi^2 = (ac - bd) + (ad + bc)i \quad (2)$$

Adicionalmente, podemos representar os números complexos na sua forma exponencial. Definindo r = $|c|$, $\theta = c$. Têm-se a forma $c = e^{i\theta}$, também chamada de forma polar.

A transformada de Fourier é uma transformação que expressa uma função em termos de função de base sinusoidal. Também, a transformada passa uma função que opera na dimensão temporal para a dimensão complexa de frequências.

A transformada de Fourier de uma função f é dada por.

$$f(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} d\xi \tag{3}$$

Sendo ξ um número real.

E sua inversa por:

$$f(x) = \int_{-\infty}^{\infty} f(\xi) e^{-2\pi i x \xi} dx \tag{4}$$

Posto isso, uma imagem é um conjunto de pontos de diferentes tons e cores espalhados em diferentes posições. É extremante simples identificar um carro no centro de uma foto, ou uma pessoa, um gato entre outros seres e objetos. Simples, porém, para a visão humana e possivelmente difícil para a visão de um computador. Dessa maneira, é viável utilizar uma ferramenta que extraia informações de uma imagem para facilitar a visão de uma máquina. Para isso, aplicamos a transforma de Fourier em imagens.

No processamento de imagens a transformada é utilizada para passar uma imagem que está no domínio espacial para o domínio complexo da frequência. Decompondo a imagem em componentes de senos e cossenos. Na imagem transformada, cada ponto representa uma frequência particular contida na dimensão espacial da imagem, cores e formatos passam a ser senos e cossenos. Sendo extremamente útil em análises que dependem de cálculos e padrões, diferente da análise de uma pessoa ao se deparar com fotos.

Dessa forma, como uma imagem é composta por um número discreto de pixels e composta por canais, também de valores discretos. Assim, é utilizado a transformada de Fourier discreto, dado por.

$$F_n = \sum_{k=0}^{N-1} f_k e^{-\frac{2\pi i k}{N}} \tag{5}$$

Sendo, $f(\xi) \to f(\xi_k)$ e fazendo $f_k \equiv f(\xi_k)$, em que $\xi_k \equiv k\Delta$, com $k = 0, \dots, N-1$.

4. Modelo

4.1. Batchnormalization

Normalizar os dados de entrada de uma RNA é de extremamente recomendado. A normalização dos inputs faz com que o algoritmo de responsável por aumentar a precisão do modelo, opere de forma mais rápida. Com isso, é obtido a convergência mais rápido ajudando no treinamento do modelo. Nesse sentido, adicionar camadas de normalização dentro do modelo pode aumentar ainda mais a velocidade de treinamento. Já que os valores tendem a aumentar ou diminuir conforme passa por camadas mais profundas.

O *batchnormalization é* uma técnica de normalização que além do que foi posto ajuda na regularização do modelo [Géron 2022]. O que é exatamente importante para fazer com que a rede tenha maior precisão em classificar imagens fora dos dados de

treinamento.

4.2. Resíduos

A técnica da utilização de resíduos foi introduzida pela rede Resnet, tal técnica garantiu a rede o primeiro lugar na competição de classificação de imagens ILSVRC 2015. A inovação da Resnet, ou Residual Network, solucionou o problema do v*anishing/exploding gradient* [He et al. 2016], comum em redes extremamente profundas.

A solução apresentada da Resnet consiste em somar nas camadas mais profundas o valor dos inputs de camadas anteriores, sem passar por nenhuma alteração, além dos dados de saída das camadas prévias. Solucionando o problema e aumentando a velocidade de treinamento e possivelmente melhorando a precisão do modelo.

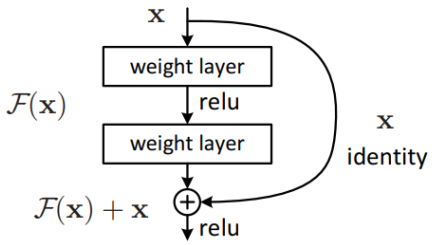

Figura 2. Visualização do Bloco de Resíduo

Dessa forma, a equação do bloco de resíduo é dada por:

$$y = \mathcal{F}(x, \{W_i\}) + x \tag{6}$$

Sendo y e x a saída e a entrada respectivamente. A função $\mathcal{F}(x, \{W_i\})$ representa as transformações das camadas. Usando a figura 2 como exemplo, a função é dada por $\mathcal{F} = W_2\sigma(W_1x + b_1) + b_2$, sendo W_i as matrizes de pesos das camadas convolucionais, σ a função de ativação ReLU e b_i os bias.

4.3. Resíduos Com Camadas

Inspirado na Resnet, foi desenvolvido em uma variação para a soma de resíduos. De forma semelhante, o input de uma camada mais profunda é somado à saída de uma camada prévia antes de prosseguir pela rede. Porém, a técnica de resíduos com camadas aplica camadas convolucionais a esse dado que será para ser somado na camada mais avançada.

Tal técnica melhora a eficiência no treinamento das redes assim como pode aumentar à precisa da RNA [Volova 2021].

Adaptando a equação 6, o bloco de resíduos com camadas é dado por:

$$y = \mathcal{F}(x, \{W_i\}) + \mathcal{H}\left(x, \{W_j\}\right) \tag{7}$$

4.4. Modelo Final

Dessa forma, munido das técnicas das subseções 4.1, 4.2 e 4.3, foi desenvolvido o modelo para a tarefa de classificação do dataset ASL. A estrutura do modelo pode ser conferida

na Figura 3.

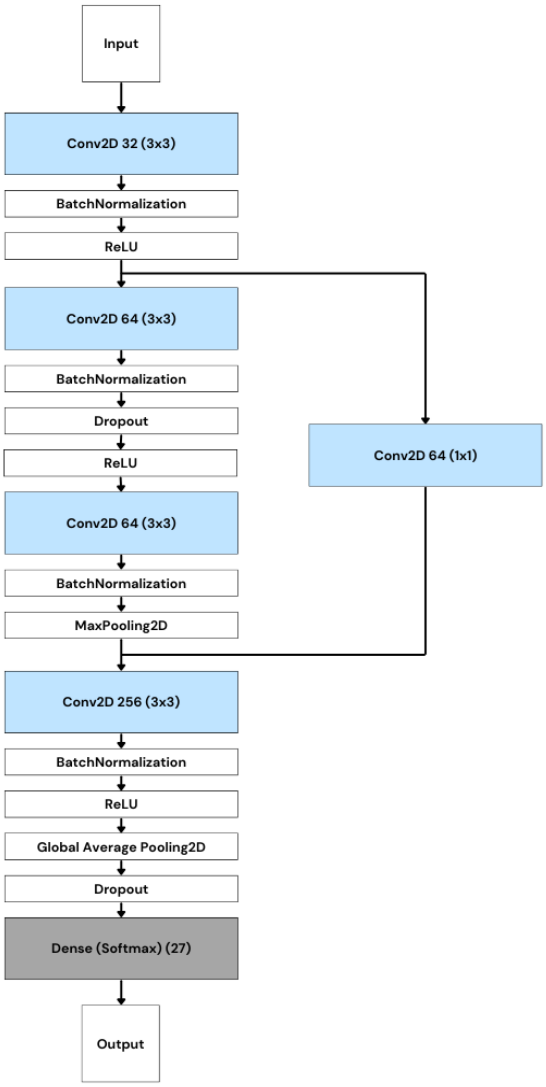

Figura 3. Esquema do modelo

5. Treinamento Rede Real

Utilizando a álgebra real o modelo atingiu uma precisão de 95,4% nos dados de teste (ver tabela 1).

Tabela 1. Precisão do Modelo Real

Treino	Validação	Teste
95,2%	95,1%	95,4%

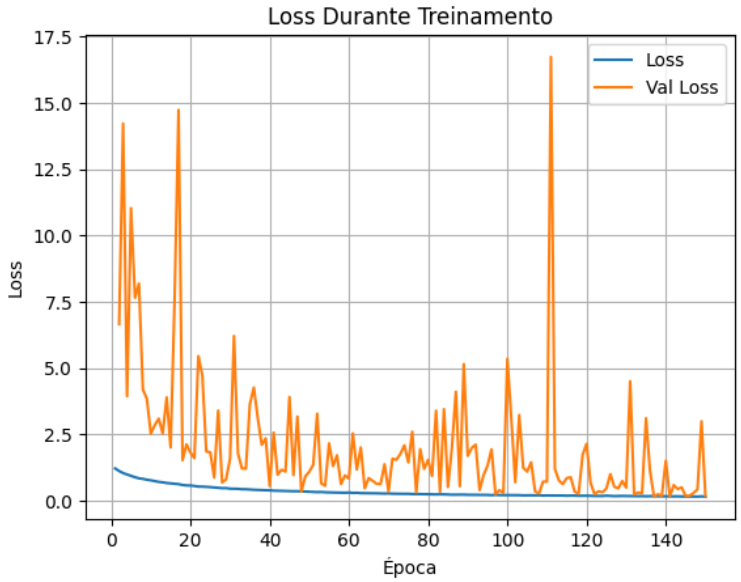

Figura 4. Gráfico da loss do treinamento da rede convolucional real

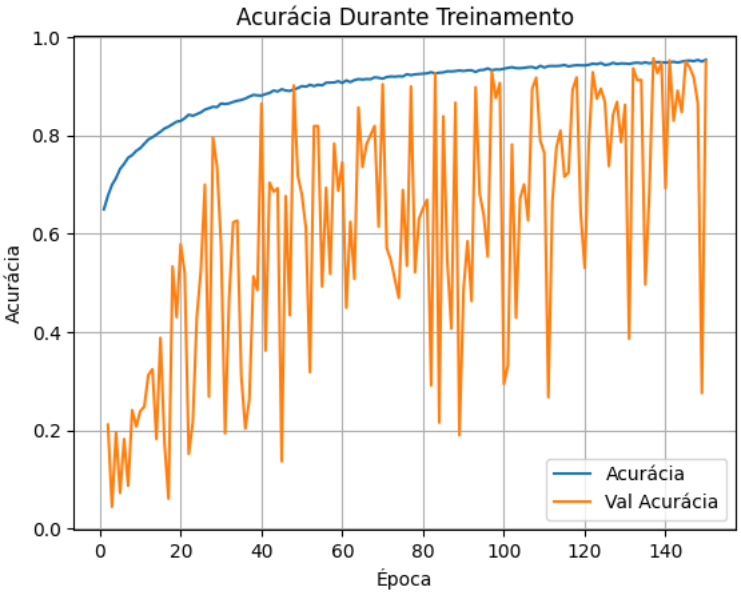

Figura 5. Gráfico da acurácia do treinamento da rede convolucional real

6. Treinamento Rede Complexa

6.1. Camada Convolucional Complexa

Além da adaptação das camadas para a álgebra complexa, foi reduzido o número de neurônios para que a RNC complexa tenha um número de parâmetros equivalente a real.

Posto isso, para a adaptação do modelo para a álgebra do complexa é importante definir a multiplicação de matrizes complexas, além das propriedades da seção 3. Visto que a computação de matrizes é fundamental para as camadas convolucionais. A

multiplicação entre uma matriz $A \in \mathbb{C}^{M\times L}$ e outra $B \in \mathbb{C}^{L\times N}$ resulta na matriz $C \in \mathbb{C}^{M\times N}$. Tal operação é comumente expressada como $C = AB$. Para realizar tal operação em softwares computacionais, é empregado o uso do isomorfismo $\varphi: \mathbb{V} \to \mathbb{R}^n$ e do produto de Kronecker representado pelo símbolo "⊗". A multiplicação então é dada por [Valle 2023]:

$$C = \varphi^{-1}\left(\left(\sum_{k=1}^{n} A_k \otimes P_k^T\right)\varphi(B)\right) \tag{8}$$

Em que P_k são as matrizes de multiplicação da álgebra dos complexos. E por fim, A_k são as matrizes associada às partes reais e imaginárias. Ou seja, A_1 e A_2 são as matrizes das partes reais e imaginárias.

6.2. ℂ**ReLU**

Com a alteração da álgebra, as camadas de ativação ReLU, por conta do isomorfismo $\varphi: \mathbb{V} \to \mathbb{R}^n$, passa a funcionar de acordo com a seguinte equação:

$$\mathbb{C}ReLU(z) = ReLU\big(\Re(z)\big) + iReLU\big(\mathfrak{C}(z)\big) \tag{9}$$

Isto é, a função ReLU é aplicada em cada componente, real e imaginário, separadamente.

6.3. **Resultados**

Apesar de um treinamento mais longo, por conta do maior número de operações, a rede complexa se mostrou mais eficiente nos treinamentos e resultados. Comparando as figuras 3, 4, 5 e 6 conclui-se que o modelo real teve uma pior regularização. Por fim, observando os resultados, a rede de números complexos diminuiu o erro, da rede real, em 17,6%, vide Tabela 2.

Tabela 2. Precisão do Modelo Complexo

Treino	Validação	Teste
96,34%	96,27%	96,18%

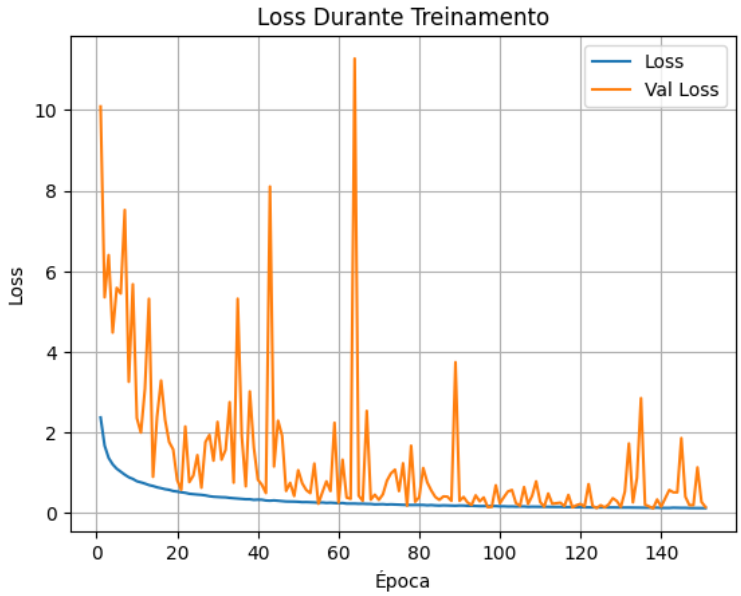

Figura 6. Gráfico da Loss do Treinamento da Rede Convolucional Complexa

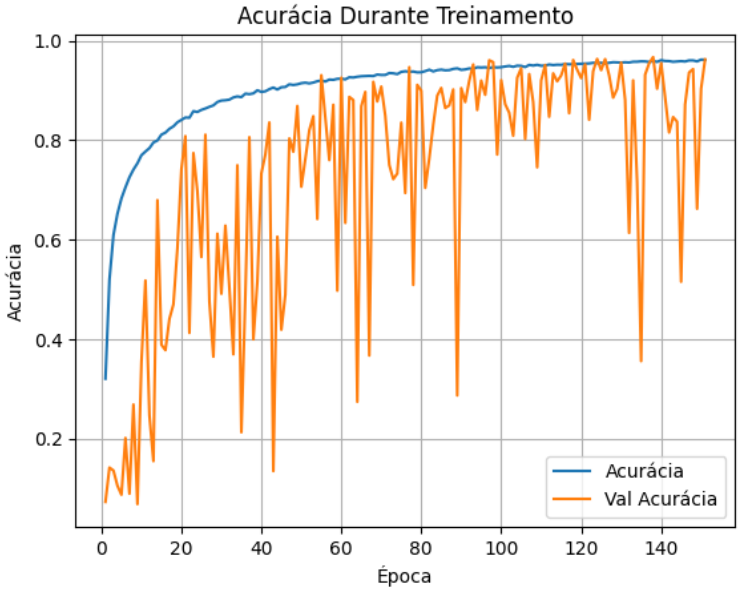

Figura 7. Gráfico da Acurácia do Treinamento da Rede Convolucional Complexa

7. Treinamento Rede Complexo Com Dados Pré Processados

7.1. Transformada de Fourier Discreta

Utilizando a transformada de Fourier discreta, vista na seção 3, as entradas do dataset foram pré processadas. A fim de treinar a rede complexa usando as imagens da dimensão complexa da frequência. Para assim, usufruir por completo as propriedades da álgebra complexa.

Figura 8. Imagem Antes da Transformada de Fourier Discreta

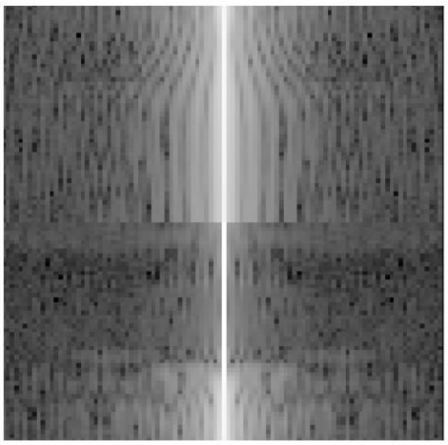

Figura 9. Imagem Após a Transformada de Fourier Discreta

7.2. zReLU

A fim de não alterar a fase do valor complexo ao passar pelas camadas de ativação foi utilizado a função de ativação $zReLU$ [Trabelsi, Chiheb, et al. 2017]. Semelhante à função ReLU, dada por:

$$zReLU(z) = \begin{cases} z \; se \; \theta_z \in \left[0, \frac{\pi}{2}\right] \\ 0 \; c.c. \end{cases} \tag{10}$$

7.3. Resultados

A rede não mostrou bons resultados e aumentou o erro, contrastando com as redes já analisadas, consideravelmente. Assim como na rede complexa, o uso de camadas complexas aumentou o tempo de treinamento. Por último, a RNC desta secção também atingiu a pior regularização durante o ajuste dos parâmetros.

Tabela 3. Precisão do Modelo Complexo Com Transformada

Treino	Validação	Teste
78,95%	67,14%	82,07%

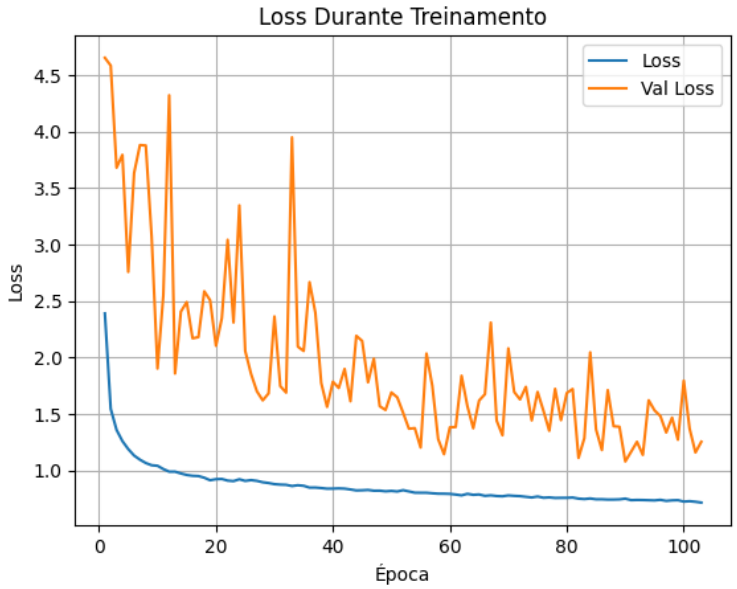

Figura 10. Gráfico da Loss do Treinamento da Rede Convolucional Complexa com Transformada

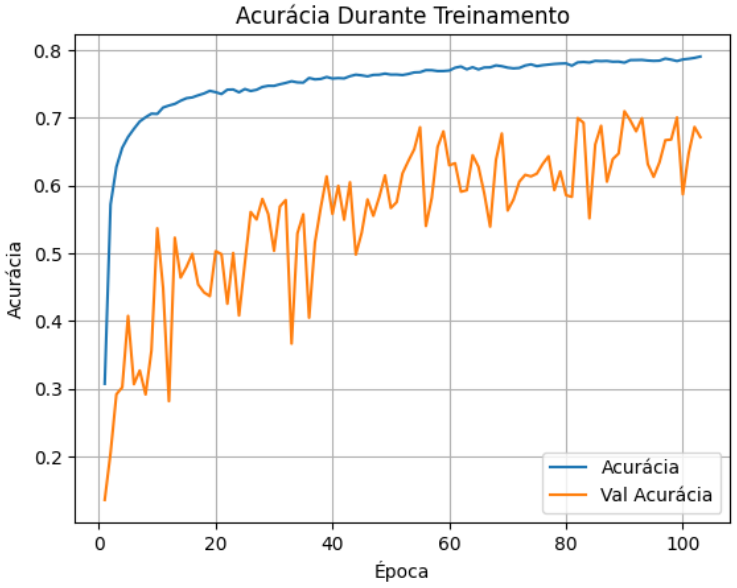

Figura 11. Gráfico da Acurácia do Treinamento da Rede Convolucional Complexa com Transformada

8. Conclusão

O estudo mostrou que aplicando de redes convolucionais no problema de classificação do alfabeto da língua de sinais obteve-se excelentes resultados. Mostrando potencial para a aplicação também em palavras ou até mesmo vídeos da língua de sinais.

Também concluímos que a utilização da álgebra dos números complexos gerou resultados mais acurados que o uso tradicional de redes reais, como pode ser visto na tabela 4. Porém, o uso de imagens transformadas pela transformada de Fourier junto a

utilização de camadas de ativação zReLU, não se mostrou eficiente.

Tabela 4. Comparação dos Modelos em Acurácia

Modelo	Treino	Validação	Teste
Rede Real	95,2%	95,1%	95,4%
Rede Complexa	**96,36%**	**96,27%**	**96,18%**
Rede Complexa com Transformada	78,95%	67,14%	82,07%

References

He, Kaiming, et al. (2016) "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition

Valova, Iren, et al. (2021) "In-between layers modular residual neural network for the classification of images." Procedia Computer Science 185, pages 223-230.

Géron, Aurélien. (2022) “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".

Zakaria, Magdi, Mabrouka AL-Shebany, and Shahenda Sarhan. (2014): "Artificial neural network: a brief overview." neural networks.

Kachwala, Zaheer (2024) Queridinha da IA, Nvidia se aproxima de valor de mercado da Apple. Disponível em: < https://economia.uol.com.br/noticias/reuters/2024/05/28/queridinha-da-ia-nvidia-se-aproxima-de-valor-de-mercado-da-apple.htm >. Acesso em 5 jul. 2024.

Aloysius, Neena, and M. Geetha. (2017) "A review on deep convolutional neural networks." 2017 international conference on communication and signal processing (ICCSP). IEEE.

IBGE (2019). Sidra: Pessoas com deficiência auditiva, por sexo e situação do domicílio (2019). Disponível em: https://sidra.ibge.gov.br/tabela/8217 Acesso em: 5 Jul. de 2024.

Nitta, Tohru. (2003)"Solving the XOR problem and the detection of symmetry using a single complex-valued neuron." Neural Networks 16.8: 1101-1105.

Valle, Marcos Eduardo. (2023). "Understanding Vector-Valued Neural Networks and Their Relationship with Real and Hypercomplex-Valued Neural Networks." arXiv preprint arXiv:2309.07716

Trabelsi, Chiheb, et al. (2017). "Deep complex networks." arXiv preprint arXiv:1705.09792

Muhammad Khalid (2019). “Sign Language for Alphabets”. [banco de dados] disponível em: <https://www.kaggle.com/datasets/muhammadkhalid/sign-language-for-alphabets>. Acesso em 30 jun. 2024.