

Classificação de libras em imagens através de redes neurais convolucionais

Libra classification in images through convolutional neural networks

DOI:10.34117/bjdv8n12-269

Recebimento dos originais: 23/11/2022

Aceitação para publicação: 29/12/2022

Márcio Fabiano Oliveira de Moura Santos

Mestrando do Programa de Pós-Graduação em Engenharia da Computação e Sistemas pela Universidade Estadual do Maranhão (PECS – UEMA)

Instituição: Instituto de Educação Ciência e Tecnologia do Maranhão (IEMA) - Coelho Neto - MA

Endereço: MA 034, CEP: 65620-000, Coelho Neto – MA, Brasil

E-mail: mfomoura@hotmail.com

Cícero Costa Quarto

Doutor em Informática na Educação pelo Programa de Pós-Graduação Informática na Educação pela Universidade Federal do Rio Grande do Sul (PPGIE - UFRGS)

Instituição: Programa de Mestrado Profissional em Engenharia de Computação e Sistemas pela Universidade Estadual do Maranhão (PECS – UEMA) – Campus São Luís - MA

Endereço: Avenida Sarney Filho, Vila Embratel, CEP: 65081-400 São Luís - MA, Brasil

E-mail: cicero.uema@gmail.com

RESUMO

A alfabetização de pessoas surdas, é atualmente um grande desafio que tem gerado muitas discussões no contexto educacional. Apesar da existência da linguagem de sinais (Libras), a capacitação de pessoas para a realização de atividades de ensino nesta área é bastante escassa, e isso tem dificultado bastante o processo de ensino e aprendizagem de indivíduos com ou sem problemas auditivos. Com o crescimento da internet e das tecnologias computacionais, houve a necessidade da criação de aplicações avançadas de inteligência artificial para aprimorar este processo. Um marco bastante importante foi o surgimento da Visão Computacional, que é uma área da inteligência artificial que procura analisar, fazer a interpretação e a extração de determinadas informações úteis de imagens, estudando também a utilização de emoções, reconhecimentos, e análise de movimentos interativos de seres humanos através de sistemas informatizados de I.A. que simulam os pensamentos e ações de seres humanos. O tema deste trabalho aborda uma pesquisa relacionada à aplicação de redes neurais convolucionais (ou profundas), que são ligadas à visão computacional de forma análoga. Foram realizados experimentos através de uma base de dados de Libras que serviu de suporte para o treinamento das imagens através do algoritmo de redes neurais profundas YOLOv5 para a realização da Classificação de alguns grupos de 6 e 7 imagens. Posteriormente, foram comparados os resultados dos testes, e observados quais épocas obtiveram uma melhor performance durante a classificação das imagens de Libras.

Palavras-chave: libras, visão computacional, redes neurais convolucionais, YOLOv5.

ABSTRACT

The literacy of deaf people is currently a major challenge that has generated many discussions in the educational context. Despite the existence of sign language (Libras), the training of people to carry out teaching activities in this area is quite scarce, and this has made the teaching and learning process of people with or without hearing problems very difficult. With the increasing growth of the internet and computer technologies, it was needed to create advanced artificial intelligence applications to improve this process. A very important milestone was the emergence of Computer Vision, which is an area of artificial intelligence that seeks to analyze, interpret and extract certain useful information from images, also studying the use of emotions, recognition, and analysis of interactive movements of human beings through computerized AI systems that simulate the thoughts and actions of human beings. The main subject of this work addresses a research related to the application of convolutional (or deep) neural network, which are similarly linked to computer vision. Experiments were carried out using a Libras database that served as support for the training of images through the YOLOv5 deep neural network algorithm to perform the Classification of some groups of 6 and 7 images. Subsequently, the results of the tests were compared, and it was observed which periods had a better performance during the classification of Libras images.

Keywords: libras, computer vision, convolutional neural network, YOLOv5.

1 INTRODUÇÃO

O uso das tecnologias digitais no ensino de Libras é uma inovação no campo do conhecimento que proporciona novas maneiras de interação e comunicação entre surdos através de computadores. A capacitação de pessoas para o processo de ensino, é realizada por meio de cursos especializados que possuem grande relevância para enfrentar desafios no sentido de desenvolvimento das habilidades para a aplicação de técnicas de ensino visando melhorar os aspectos sociais, educacionais e a alfabetização destes alunos em sala de aula (PACHECO, 2007; SILVA, 2016).

De acordo com BARELLI (2018), uma importante área da inteligência artificial que utiliza redes neurais como base para o seu funcionamento, é a Visão Computacional que se caracteriza como a ciência e a tecnologia das máquinas que enxergam pela forma como o computador distingue à sua volta através da obtenção de informações importantes pela captura de imagens de diferentes dispositivos, pois esta desenvolve as teorias e técnicas para a construção de sistemas artificiais que extraem informações úteis de imagens ou dados multidimensionais. Um dos seus principais aspectos está no fato desta utilizar sistemas de reconhecimento com características semelhantes às de seres humanos como identificações biométricas, de faces, íris dentre várias outras. O reconhecimento de

padrões de objetos é o campo da ciência que estuda técnicas para classificação ou detecção de variados tipos.

Com isso, surge um grande desafio em relação a utilização de complexos algoritmos baseados em Visão Computacional que possam ou não ser aplicados a partir de equipamentos que funcionem a distância ou em tempo real e que identifiquem os movimentos reais das mãos humanas através da visualização de imagens digitais ou vídeos gravados (Gonzaga, 2011). A utilização de aplicações para o ensino de Libras é algo um pouco complexo devido à dificuldade de desenvolvimento destes tipos de softwares computacionais.

Redes Neurais Convolucionais, do inglês Convolutional Neural Networks (CNN), é uma rede biológica que ao ser treinada, consegue aprender características constantes utilizando hierarquias multiníveis de determinados aspectos, extraindo de forma implícita, particularidades de determinados padrões de imagens apresentados como entrada, classificando-os ou detectando-as de forma categórica através de recursos que são extraídos sem depender de escala, iluminação ou pose (Lecun, et al., 2010). Geralmente, são constituídas por arquiteturas de treinamento que se encontram em inúmeros estágios, e cada um destes é composto por três camadas que são uma camada convolucional, uma camada de subamostragem e a terceira camada caracteriza-se por ser totalmente conectada (Kang et al., 2014). A classificação de objetos é atualmente um dos modelos de visão computacional mais conhecidos e com maior versatilidade que os demais, pois buscam fazer análises em determinados tipos de imagens conforme sua relevância. Pode-se citar como exemplo o reconhecimento do uso de máscaras, que é algo importante para a redução do risco do vírus da Covid-19, onde foi utilizado um tipo de reconhecimento de imagens necessário para a realização do treinamento de um determinado modelo de classificação de objetos para que fosse possível identificá-los e classificá-los para um determinado fim (Moura et al., 2022). Um dos sistemas de classificação ou detecção de imagens mais utilizados da atualidade é o algoritmo YOLO (do inglês You only look once), que se caracteriza por ser um algoritmo de detecção de imagens que faz a modelagem do problema de detecção como um problema de regressão, dividindo a imagem em um sistema de grade onde há a predição das células em caixas delimitadoras que utilizam níveis de confiança para mostrar as possibilidades de ocorrência das classes utilizadas para análise em sua estrutura através de inferências. Um outro aspecto interessante deste algoritmo é que este também é capaz de identificar objetos em tempo real, além de possuir muita rapidez e precisão quando equiparado com

outros detectores de imagens, pois ele aplica uma rede neural à imagem completa para poder fazer a classificação dos objetos, e após isso, a rede neural faz o particionamento da imagem em determinadas regiões fazendo a previsão das possíveis ocorrências para cada região. Porém, sob outra perspectiva, outros tipos de algoritmos utilizados para a detecção de objetos, fazem geralmente a execução das redes neurais a uma imagem que se encontra em variados locais e em diferentes proporcionalidades. Resumindo, o algoritmo YOLO primeiramente visualiza toda a imagem, de uma maneira na qual as previsões são notificadas através de uma circunstância abrangente da imagem (Tamang et al., 2022). Atualmente, a ferramenta possui 5 versões (YOLOv1, YOLOv2, YOLOv3, YOLOv4 e YOLOv5). Diante disso, para a identificação da área estudada e do problema relacionado à aprendizagem de Libras através de redes neurais convolucionais, este artigo se norteia com a seguinte questão de pesquisa: A classificação de imagens através da rede neural convolucional YOLOv5 torna produtiva a análise de sinais de libras? Através deste questionamento, tem-se como objetivo geral verificar a eficiência da classificação de Libras através dos resultados obtidos pela análise de imagens realizada pela rede neural convolucional YOLOv5 aplicado-a a uma base de dados desta linguagem no processo de aprendizagem, e como objetivos específicos realizar o treinamento do conjunto de dados de Libras através da arquitetura YOLOv5 em 100, 500 e 1000 épocas, Executar testes de classificação em grupos específicos de imagens, e após isso, analisar os resultados da acurácia de cada classe identificada após a classificação das imagens, investigando os melhores resultados após os testes.

2 TRABALHOS RELACIONADOS

O processo de encontrar soluções para estudo de libras através da utilização de tecnologia da informação aplicada aos conceitos de linguagem de sinais visando uma melhor aprendizagem, é um conhecimento recente, porém já existem alguns trabalhos científicos relacionados à área. A complexidade da construção de redes neurais é certamente um dos principais fatores, pois, dependendo do tipo do problema, o treinamento poderá levar muito tempo, sem contar que após os resultados, a probabilidade da ocorrência de erros poderá ou não ser alta. No entanto, quando os resultados são satisfatórios, a detecção e a classificação de imagens se tornam bastante produtivas, possibilitando assim que a identificação do objeto estudado seja bastante eficiente, fazendo com que os testes sejam de grande produtividade após a execução da aplicação (Alves, 2020).

A ocorrência da dificuldade de aprendizagem de Libras no contexto educacional e social é algo visível, e devido a isso, o surgimento destas técnicas baseadas em redes neurais são úteis para a população (Bastos, 2015). Diversos trabalhos científicos relatam a utilização destas tecnologias computacionais baseadas em aprendizagem de máquina incorporadas ao processo de ensino, que são de extrema relevância para o público surdo e para pesquisadores científicos de tecnologias computacionais de linguagem de sinais, que visem desenvolver aplicações digitais para esta comunidade. Os trabalhos variam de acordo com o contexto no qual as técnicas são aplicadas.

Em Pavan, Ossada, Modesto (2010), é utilizada uma metodologia capaz de realizar o reconhecimento de sinais com o objetivo de extrair características da mão segmentada de imagens dinâmicas para identificação de padrões de Libras pelos usuários, com a utilização de filtros específicos de processamento de imagens, nos quais os dados especificados poderão ser comparados aos modelos pré-estabelecidos na aplicação, reproduzindo em seguida textos ou áudios de letras identificadas pelo mecanismo através da comparação de dados e informações extraídas, possibilitando assim a execução do software em algumas áreas como educação e entretenimento por exemplo, além de melhorar a comunicação dos surdos usuários da aplicação. No entanto este trabalho utilizou imagens em extensões .jpg, que poderiam ou não ser capturadas por uma webcam, e vídeos gerando assim um baixo custo computacional. Já em Santos (2011), foi proposto um sistema de reconhecimento de sinais de mãos humanas através da captura, do processamento e da extração de um vetor de características de imagens de vídeo utilizado por uma Rede Neural Artificial treinada para o reconhecimento do alfabeto de Libras através de uma câmera ligada ao computador. O sistema foi desenvolvido na linguagem de programação C juntamente com a biblioteca OpenCV utilizada na construção de aplicativos com visão computacional. Os resultados apresentaram condições nas quais o algoritmo mostrou bons desempenhos na interação humano-computador, pois os testes foram realizados em ambientes com boas condições nas quais o sistema pôde obter uma ótima performance, com execuções precisas e eficientes sem o auxílio de luvas ou dispositivos que possam facilitar a identificação das mãos humanas, focando sempre nas melhores formas de identificação das imagens com um fundo simples e homogêneo, onde as características são obtidas em quadros capturados em intervalos regulares para que a variação do posicionamento da mão seja reconhecida, e após isso, o intervalo de tempo de captura dos quadros foi analisado para que o sistema obtenha o melhor desempenho possível. O trabalho de Queifer (2019), apresenta uma estratégia para

acionamento de objetos na detecção de sinais manuais utilizando o algoritmo HAAR CASCADE juntamente com redes neurais convolucionais, onde são utilizadas técnicas de visão computacional com internet das coisas que buscam fazer a interpretação de cenários de imagens capturadas através de um Raspberry PI com uma câmera integrada dispositivo computacional. O foco no caso, foi o desenvolvimento de um protótipo para aquisição de sinais emitidos por sinais de mãos através de uma webcam ou imagens de sinais com extensões png, ou .jpg adquiridas ou não através de câmeras digitais para fazer o acionamento e a detecção destes objetos através da utilização do algoritmo, observando qual destes possuirá o melhor desempenho em imagens de vídeo em tempo real.

A utilização do classificador Haar, é uma abordagem que apesar de ser utilizada no contexto de aprendizagem de máquinas, são extremamente lentas se comparadas às novas tecnologias de Visão Computacional atuais, pois o treinamento através deste método é realizado com várias imagens onde são extraídos os recursos desta técnica imagem por imagem, adaptando os classificadores para realizar a detecção dos objetos. Já existem modelos no quais é possível a aplicação de uma única rede neural às imagens por completo para realização de suas previsões (Base app, 2018).

No trabalho de Santos (2020), foi desenvolvido um sistema de reconhecimento de sinais e ações em tempo real baseado em Visão Computacional com o objetivo de investigar e propor alguns mecanismos que pudessem além de conhecer, fazer a antecipação de sinais dinâmicos com ações baseando-se em vídeos em RGB. Foram apresentadas seis propostas principais onde duas destas eram para a representação de movimento em um vídeo, outra para antecipação de ações, a seguinte para antecipar sinais e as duas últimas para reconhecer e antecipar sinais de Libras online em ambientes com uma câmera ou várias câmeras. A validação das propostas foi realizada através de experimentos cujos resultados foram comparados com várias linhas de base, e após os testes, observou-se que modelos que utilizaram somente vídeos RGB também são efetivos quando comparados a dados multimodais para o reconhecimento de sinais de Libras, e que também a escolha empírica da informação de contexto é eficiente na separação entre classes com movimentos ambíguos, principalmente em pequenos conjuntos de dados e que na tarefa de antecipação, a incerteza sobre a predição do teste é um modelo limiar e eficaz e mais confiável do que a probabilidade por ele estimada, com tempos de execução viáveis dos e mais confiável do que a probabilidade estimada. Na pesquisa de Bastos (2015), foram utilizados métodos de reconhecimento de sinais de Libras através de vetores de características resultantes das aplicação de descritores voltados para as formas

e contornos presentes em imagens utilizando redes neurais, onde foi utilizado um classificador Perceptron Multicamada disposto em uma arquitetura na qual o processo de classificação foi dividido em dois estágios, com o auxílio de especialistas em linguagens de sinais e alunos surdos devido à escassez de conjuntos de dados públicos de Libras, fazendo com que a abordagem partisse então da criação de um dataset até a etapa final de classificação dos sinais, onde foram obtidos quase 96,77% de taxa de acerto, sendo que estes resultados foram validados considerando possíveis ameaças à abordagem como a realização de testes com um indivíduo não presente no conjunto treinado do classificador e a aplicação da abordagem em um conjunto público de dados de sinais.

Em Góes (2019), foi abordado o aperfeiçoamento de um sistema de padrões utilizado para a detecção de padrões do alfabeto de Libras utilizando uma luva sensora adicionando um sensor inercial e promovendo melhorias de hardware necessárias para suportar o processamento dos dados do sensor inercial. Para isso foi portado um código existente para o microcontrolador STM32F103C8T6, e houve a implementação de um sensor inercial MPU-6050, e posteriormente foi feita uma coleta de dados para treinamento e validação de um algoritmo de classificação além da utilização de um algoritmo de aprendizagem em Python para realizar o treinamento e as simulações que foram implementadas com esse microcontrolador. Foi realizado também o desenvolvimento de uma interface para visualização dos resultados da classificação e em seguida o sistema foi testado e validado a partir de uma quantidade de usuários, pois a metodologia deste trabalho foi separada na compreensão do funcionamento e na transferência do código para o novo microcontrolador, implementação do sensor inercial, processamento dos sinais, treinamento das redes neurais e implementação das redes já treinadas. Foram utilizados alguns hardwares físicos durante a pesquisa e como resultado, foi obtido um sistema computacional com uma rede neural com uma capacidade de fazer o reconhecimento das letras do alfabeto com um percentual de acerto de 90,67%.

3 CONSIDERAÇÕES SOBRE TRABALHOS RELACIONADOS

A utilização de sinais no processo de comunicação, caracteriza-se como a expressão de significados de palavras através de sinais de ações não verbais utilizando as mãos, dedos, face e braços para interação humana por meio da linguagem de sinais. No decorrer dos anos, alguns trabalhos relacionados a utilização de visão computacional para detecção Libras têm sido publicados devido a elevada aplicabilidade dos conceitos de tecnologias da informação como suporte a alfabetização de surdos e afins. A opção de

escolher quais técnicas utilizar para realização deste processo é bem ampla, podendo haver uma diversificação de qual estratégia utilizar para as pesquisas, pois existem inúmeros trabalhos que se baseiam em modelos estatísticos após a execução de testes.

O que diferencia este trabalho dos demais primeiramente, é que em vez de aplicar o método de detecção de imagens através da utilização de pequenas caixas delimitadoras conhecidas também como bounding boxes, foi utilizada a técnica de classificação de imagens que se baseia na acurácia da probabilidade em um conjunto de dados formado por letras, números e palavras para treinamento através do algoritmo YOLOv5-S.

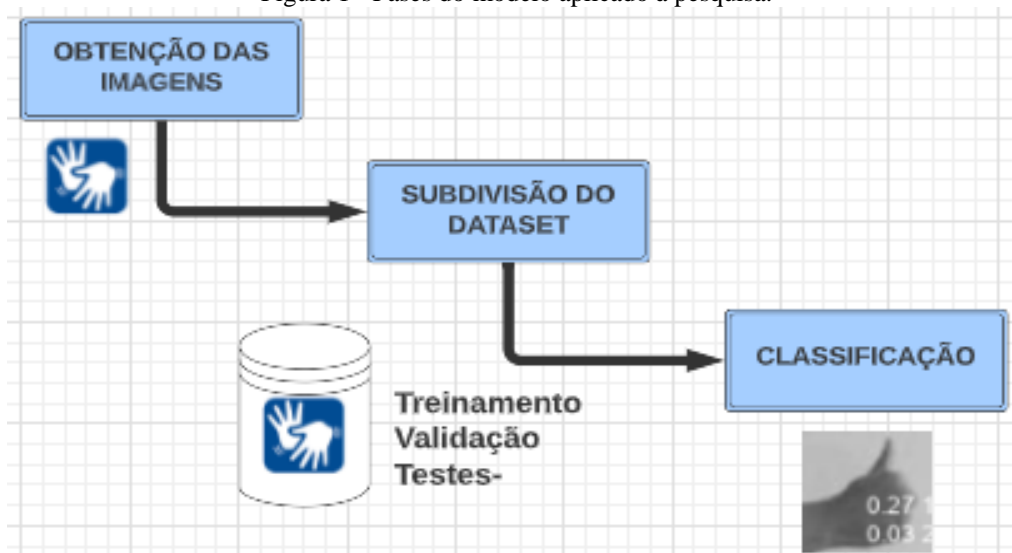
Para passar uma imagem para um modelo de inteligência artificial utilizado para receber a classificação, podemos citar como exemplo um ótimo exemplo o diagnóstico de exames de Raios-X. Através das radiografias de pulmões, um médico consegue fazer a identificação de um câncer no órgão do paciente, e mesmo que o cientista de dados não possua conhecimentos medicinais, no caso da existência de um determinado conjunto de dados específico de imagens de pulmões saudáveis juntamente com seus diagnósticos, é possível treinar o algoritmo do YOLOv5 para fazer esta classificação, e através do ajuste do modelo, é possível analisar a probabilidade de que um exame pulmonar possa apresentar câncer ou não baseado nos milhares de casos já classificados no dataset utilizado, auxiliando nos prognósticos médicos. Este é o principal propósito do método de classificação de imagens. O algoritmo utilizado em vez de extrair informações da imagem por imagem durante o treino para a detecção, como é feita pela maioria das técnicas utilizadas pelos processos de detecções, utiliza um método conhecido como single pass (de passada única). Apesar do YOLOv5 ser famoso por ser um algoritmo de detecção de objetos que utiliza um método single pass através de uma rede neural convolucional que atua como extrator de características, que diferentemente dos algoritmos anteriores de detecção, como R-CNN ou Faster-R-CNN, este precisa olhar apenas uma vez para a imagem para poder enviar suas características para a rede neural convolucional que faz a divisão das imagens em um sistema de grade onde cada célula é responsável por detectar objetos em si, sendo famoso por sua velocidade e precisão. Foram realizados alguns experimentos em algumas épocas de treinamento que serão abordados nas seções posteriores.

4 MÉTODO DE PESQUISA

Nesta seção são apresentadas as principais particularidades da metodologia que foi utilizada para a construção deste trabalho. Inicialmente, foi realizada uma revisão de

literaturas que foram utilizadas para fundamentação desta pesquisa, que é de caráter experimental e descritiva. Posteriormente, selecionada e aplicada uma técnica de classificação de imagens de sinais de Libras extraídas de uma base de dados constituída de letras, números e palavras através de uma Rede Neural Convolutacional, onde o principal objetivo foi identificar o que estava presente na imagem, pois o modelo procura compreendê-la através da atribuição de classificações do conjunto de dados utilizado. A utilização deste método foi dividida em etapas que foram a obtenção das imagens através do site Roboflow (Roboflow, 2021), a subdivisão do dataset, e a realização dos testes Classificação através do YOLOv5.

Figura 1 - Fases do modelo aplicado à pesquisa.



Fonte: O autor (2022).

4.1 OBTENÇÃO DAS IMAGENS

O conjunto de dados das amostras de imagens de Libras foram adquiridas através do site Roboflow, que é uma empresa que auxilia desenvolvedores na construção de determinados modelos rápidos e precisos de visão computacional, além de disponibilizar uma infinidade destas bases de dados gratuitamente. Foi baixado e utilizado o dataset Libras Computer Vision Project com 8.455 imagens que estão divididas em um conjunto de treinamento (65% do total), um conjunto de validação (20% do total) e um conjunto de testes (15% do total). As imagens adquiridas estavam no formato JPEG com resolução 64 x 64 pixels e com composição de cores em preto e branco. O exemplo da figura 8 faz parte das imagens obtidas durante a classificação, e a partir desta é possível visualizar os valores de algumas métricas atingidas.

4.2 SUBDIVISÃO DO DATASET

Para a realização da pesquisa, foi utilizada uma base de dados com imagens de sinais de Libras que foi disponibilizada pelo site Roboflow, que é uma grande referência em suporte a treinamentos e anotações de imagens, dentre vários outros mecanismos importantes para a comunidade científica que estuda redes neurais (Roboflow, 2021). Este dataset serviu como suporte para treinar as diretrizes necessárias para a experimentação científica, pois após este processo, foi possível a realização dos testes a partir da execução da rede neural. As 40 classes utilizadas para o treino são formadas por letras, números e palavras de acordo com a tabela a seguir:

Tabela 1 - Classes do conjunto de dados libras Computer Vision Project.

CLASSES					
LETRAS			NÚMEROS	PALAVRAS	
A	M	V	1	Adulto	Pequeno
B	N	W	2	America	Verbo
C	O	X	4	Aviao	
D	P	Y	5	Casa	
E	Q		7	Gasolina	
F	R		9	Identidade	
G	S			Junto	
I	T			Lei	
L	U			Palavra	
				Pedra	

Fonte: O autor (2022).

4.3 ARQUITETURA UTILIZADA

Após a obtenção das imagens extraídas do site Roboflow com o fluxo de trabalhos de classificação de acordo com a padronização exclusiva do algoritmo de visão computacional aplicado, a versão mais apropriada para a realização dos testes foi a do YOLOv5, pois esta adiciona treinamentos de classificação, validação, previsão e exportação para modelos pequenos pré-treinados. A versão inicial é de simples utilização, além de ser mais rápida e com grande eficiência em relação as demais versões (Alves, 2022). Portanto, foi feita uma análise das competências da base utilizada para poder classificar imagens de sinais de libras através da aplicação do algoritmo citado.

4.4 AMBIENTE UTILIZADO PARA O EXPERIMENTO

Para a execução dos testes, foi utilizado um ambiente virtual de programação de GPUs conhecido como Google Colaboratory (Colab) disponibilizado na nuvem pela Google nas versões gratuita ou paga (com mais recursos) que utiliza um terminal Linux. A arquitetura de rede neural utilizada foi uma versão do YOLOv5-S de classificação de

imagens desenvolvida na linguagem Python pela plataforma site Roboflow (Robloflow, 2021). A plataforma virtual online, possibilitou o desenvolvimento da rede neural, pois foi utilizado o acesso gratuito a uma versão do Jupiter notebook, que é uma importante interface de construção de softwares aplicada ao ensino e aprendizagem de várias linguagens, podendo ser executada simplesmente através de um navegador, com suporte às linguagens Python, R e Julia que são muito utilizadas na área de inteligência artificial. Porém, o objetivo da ferramenta é estender ao máximo sua estrutura para que possa servir como base para vários outros tipos de programações, visando facilitar e difundir o ensino e a pesquisa relacionados à aprendizagem de máquina.

O ambiente do colab também possibilita a criação de uma ou mais sessões nas quais o usuário poderá acessar na nuvem um processador com dois núcleos, 12 GBytes de memória RAM e cache L3 de 40-50 Mbytes, além de disponibilizar ao usuário acessos a GPUs e TPUs e a um terminal Linux que poderá ser configurado de acordo com a necessidade do usuário (Ferreira, Canescho e Carvalho, 2020). Foi utilizada a versão paga do Colab pro devido a eficiência da utilização de GPUs, que são extremamente rápidas e potentes para a realização de treinamento, classificação ou detecção de imagens de redes neurais.

Figura 2 – Informações da GPU utilizada durante o experimento.

```
Sat Oct 8 18:54:45 2022
+-----+
| NVIDIA-SMI 460.32.03    Driver Version: 460.32.03    CUDA Version: 11.2    |
+-----+-----+-----+-----+
| GPU Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC | |
| Fan  Temp  Perf  Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|                               |                 |           |    MIG M. |
+-----+-----+-----+-----+
|   0   Tesla T4            Off | 00000000:00:04.0 Off |             |      0 |
| N/A   41C    P8             9W / 70W |  0MiB / 15109MiB |      0%   Default |
|                               |                 |           |    N/A |
+-----+-----+-----+-----+

+-----+
| Processes: |
| GPU  GI  CI       PID   Type   Process name                      GPU Memory |
|   ID  ID  ID             |                 |                   |      Usage |
+-----+-----+-----+-----+
| No running processes found |
+-----+
```

Fonte: O autor (2022).

4.5 REALIZAÇÃO DO EXPERIMENTO

Durante a experimentação, foram feitos alguns treinamentos da base de dados de Libras com as 40 classes citadas anteriormente através do algoritmo de redes neurais convolucionais **YOLOv5-S** em uma GPU do Google Colab. No primeiro treino foram executadas 100 épocas, posteriormente 500 e 1000.

5 RESULTADOS E DISCUSSÃO

Após a obtenção dos resultados dos testes, foi possível observar através do treinamento das épocas quais destes apresentaram um melhor desempenho de acordo com a acurácia. O principal fator de definição de boas resultâncias de precisão, foi com certeza, a quantidade de épocas realizadas após cada treino, pois foi possível observar que houve uma melhor produtividade na execução de 500 épocas. A **tabela 75** apresenta a soma das acurácias dos testes das sete imagens utilizadas durante a execução de 100, 500 e 1000 épocas.

Tabela 2 - Soma das acurácias dos testes T1 à T7 em 100, 500 e 1000 épocas

Classe	Épocas	T1 Precisão	T2 Precisão	T3 Precisão	T4 Precisão	T5 Precisão	T6 Precisão	T7 Precisão	Total (Precisão)
A	100	0,12	0,34	0,15	0,07	0,07	0,08	0,07	0,90
	500	0,42	0,49	0,51	0,52	0,48	0,51	0,55	3,48
	1000	0,29	0,36	0,37	0,33	0,37	0,33	0,36	2,41
B	100	0,33	0,34	0,32	0,29	0,38	0,08	0,11	1,85
	500	0,44	0,43	0,49	0,46	0,44	0,40	0,41	3,07
	1000	0,24	0,26	0,22	0,23	0,25	0,26	0,24	1,70
C	100	0,38	0,39	0,17	0,22	0,16	0,37	0,19	1,88
	500	0,30	0,38	0,32	0,31	0,41	0,37	0,46	2,55
	1000	0,26	0,24	0,25	0,32	0,28	0,33	0,36	2,04
D	100	0,33	0,36	0,27	0,26	0,31	0,35	0,35	2,23
	500	0,55	0,49	0,56	0,55	0,59	0,55	0,56	3,85
	1000	0,41	0,39	0,37	0,39	0,42	0,44	0,43	2,85
E	100	0,26	0,33	0,29	0,25	0,22	0,31	0,15	1,81
	500	0,43	0,43	0,34	0,42	0,44	0,40	0,45	2,91
	1000	0,25	0,25	0,28	0,26	0,34	0,35	0,34	2,07
F	100	0,47	0,24	0,44	0,44	0,52	0,23	0,13	2,47
	500	0,47	0,46	0,47	0,45	0,48	0,49	0,45	3,27
	1000	0,32	0,30	0,32	0,32	0,34	0,32	0,32	2,24
G	100	0,24	0,22	0,27	0,26	0,12	0,15	0,08	1,34
	500	0,35	0,29	0,40	0,38	0,33	0,29	0,33	2,37
	1000	0,18	0,22	0,18	0,20	0,18	0,21	0,21	1,38
I	100	0,17	0,22	0,20	0,20	0,25	0,21	0,11	1,36
	500	0,24	0,27	0,21	0,29	0,24	0,26	0,24	1,75
	1000	0,28	0,26	0,22	0,25	0,22	0,26	0,23	1,72

L	100	0,22	0,24	0,26	0,29	0,27	0,24	0,30	1,82
	500	0,32	0,32	0,31	0,35	0,32	0,35	0,36	2,33
	1000	0,23	0,27	0,28	0,24	0,33	0,36	0,41	2,12
M	100	0,32	0,43	0,35	0,24	0,41	0,13	0,24	2,12
	500	0,40	0,36	0,40	0,36	0,39	0,30	0,40	2,61
	1000	0,23	0,21	0,24	0,22	0,20	0,26	0,21	1,57
N	100	0,54	0,56	0,51	0,57	0,55	0,54	0,16	3,43
	500	0,52	0,47	0,47	0,48	0,44	0,49	0,49	3,36
	1000	0,29	0,31	0,27	0,32	0,28	0,28	0,29	2,04
O	100	0,27	0,44	0,14	0,20	0,30	0,11	0,10	1,56
	500	0,41	0,36	0,44	0,43	0,44	0,48	0,43	2,99
	1000	0,28	0,32	0,28	0,27	0,26	0,30	0,33	2,04
P	100	0,14	0,14	0,19	0,08	0,10	0,11	0,11	0,87
	500	0,25	0,30	0,33	0,18	0,20	0,23	0,25	1,74
	1000	0,14	0,19	0,23	0,12	0,16	0,19	0,20	1,23
Q	100	0,26	0,28	0,27	0,33	0,22	0,40	0,34	2,10
	500	0,52	0,49	0,47	0,51	0,53	0,49	0,56	3,57
	1000	0,20	0,19	0,19	0,19	0,18	0,18	0,21	1,34
R	100	0,34	0,31	0,28	0,29	0,17	0,18	0,19	1,76
	500	0,30	0,32	0,33	0,32	0,33	0,32	0,32	2,24
	1000	0,25	0,27	0,31	0,27	0,27	0,25	0,31	1,93
S	100	0,33	0,38	0,14	0,42	0,12	0,25	0,31	1,95
	500	0,45	0,51	0,50	0,40	0,52	0,49	0,47	3,34
	1000	0,48	0,50	0,46	0,48	0,46	0,49	0,51	3,38
T	100	0,08	0,10	0,04	0,06	0,04	0,05	0,00	0,37
	500	0,13	0,14	0,13	0,18	0,14	0,17	0,00	0,89
	1000	0,09	0,12	0,16	0,15	0,10	0,14	0,00	0,76
U	100	0,31	0,32	0,31	0,19	0,25	0,21	0,14	1,73
	500	0,49	0,46	0,48	0,47	0,49	0,46	0,48	3,33
	1000	0,40	0,33	0,31	0,37	0,37	0,38	0,00	2,16
V	100	0,13	0,27	0,29	0,28	0,24	0,22	0,25	1,68
	500	0,35	0,32	0,30	0,34	0,36	0,34	0,32	2,33
	1000	0,22	0,22	0,23	0,23	0,20	0,27	0,22	1,59
W	100	0,31	0,27	0,31	0,29	0,28	0,17	0,18	1,81
	500	0,42	0,41	0,39	0,38	0,38	0,41	0,39	2,78
	1000	0,24	0,27	0,24	0,29	0,29	0,28	0,28	1,89
X	100	0,20	0,19	0,19	0,18	0,18	0,20	0,18	1,32
	500	0,31	0,29	0,29	0,30	0,33	0,29	0,31	2,12
	1000	0,22	0,23	0,20	0,20	0,19	0,24	0,21	1,49
Y	100	0,24	0,27	0,22	0,23	0,23	0,21	0,00	1,40
	500	0,33	0,39	0,37	0,37	0,37	0,36	0,34	2,53
	1000	0,14	0,16	0,18	0,15	0,18	0,15	0,19	1,15

Fonte - O autor (2022).

6 CONCLUSÃO E PERSPECTIVAS DE TRABALHO FUTUROS

Através da análise dos resultados obtidos, pôde-se observar que os níveis de precisão variaram bastante de acordo com a quantidade de treinamentos realizados durante a execução do algoritmo YOLOv5. Após o treino de 100 épocas, pôde-se observar que os valores das acurácias dos testes processados nas sete imagens de sinais de libras não alcançaram valores satisfatórios para a maioria das classes. Porém classe “N” atingiu o maior nível, mas ficando abaixo de 60%. Já no treinamento de 500 épocas, houve uma melhora do percentual dos resultados, pois os valores atingidos ficaram entre 28% e 59% com maioria acima de 40%, sendo razoáveis para realização da classificação. E por fim, após treinamento de 1000 épocas, as classes “D” e “S”, foram as que obtiveram uma melhor acurácia, pois os valores ficaram entre 40%, e 51%, um pouco acima da execução de 100 épocas, porém abaixo de 500 épocas.

A maioria dos resultados obtidos pelas demais classes foram menores do que os valores do treino anterior. Portanto, concluindo-se assim que, a precisão após os testes das imagens com o algoritmo YOLOv5 atingiu bons resultados após o treinamento de 500 épocas. Como sugestão de trabalhos futuros, recomenda-se a ampliação do conjunto de dados de libras através da inserção de mais letras e palavras para a realização de testes de classificação ou detecção através do algoritmo YOLOv5 em palavras e textos.

REFERÊNCIAS

- ALMEIDA, S. M. S. **Intérprete da Língua de Sinais – LIBRAS: mediador ou educador**. Trabalho de Conclusão de Curso de Graduação em Letras -
- ALVES, Gabriel. Detecção de Objetos com Yolo: uma abordagem moderna. **IA Expert**, Porto União, SC. 13 de Out. de 2020. Disponível em: <<https://iaexpert.academy/2020/10/13/deteccao-de-objetos-com-yolo-uma-abordagem-moderna>>. Acesso em: 20 de abr. de 2021.
- ALVES, Gabriel. **Simplifying Classification Workflows With YOLOv5**. ultralytics. 30 de Abr de 2022. Disponível em: <<https://ultralytics.com/article/Simplifying-Classification-Workflows-With-YOLOv5-v6.2>>. Acesso em: 20 de Ago. de 2022.
- ALVES, Melo Priscila. **Inteligência Artificial e Redes Neurais**. IPEA, 11 de jun. de 2020. Disponível em: <<https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais>>. Acesso em: 20 de set. de 2021.
- applications invision. In: IEEE. Circuits and Systems (ISCAS)**, Proceedings of 2010 APPOLINÁRIO, F. **Dicionário de metodologia científica: um guia para a produção do conhecimento científico**. São Paulo: Atlas, 2004.
- BALLARD, Dana H; BROWN, Christopher M. ISBN 0-13-165316-5. **Computer Vision**. New Jersey: Editora Prentice Hall, 1982.
- BARELLI, Felipe. **Introdução à Visão Computacional. Uma abordagem prática com Python e OpenCV**. 1º.ed. São Paulo: Casa do Código, 2018.
- BASTOS, I. L. O. **Reconhecimento de sinais de libras utilizando descritores de forma e redes neurais artificiais**. 2015. Dissertação (mestrado) – Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia e Universidade Estadual de Feira de Santana. 2015. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/19374/1/Igor%20Bastos%20-%20Disserta%20a7%20a3o%20VFinal.pdf>>. Acessado em: 10 de fevereiro de 2021.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y.M. Yolov4: **Optimal speed and accuracy of object detection**. arXiv preprintar Xiv: 2004.10934, 2020.
- BORGES, Lucas C.; Obara, Eduardo N.; Leite, Claudia Z.; Rocha, Fabiano G. (2015). **Glossário interativo de Libras para a área de Computação**. Computer on the Beach, Florianópolis/SC.
- BOTELHO, Paula. **Linguagem e letramento na Educação dos Surdos: ideologias e práticas pedagógicas**. Belo Horizonte: autêntica, 2002.
- BRASIL. **Lei nº 10.436, de 24 de abril de 2002, que dispõe sobre a Língua Brasileira de Sinais – Libras**. Brasília: MEC, 2002.
- CASTRO, Alberto Rainha de; CARVALHO, Ilza Silva de. **Comunicação por língua brasileira de sinais: livro básico**. 3ª ed. – Brasília: Senac/DF, 2009.

CHAVES, Magaly. **Recomendações pedagógicas para desenvolvimento de ambientes virtuais de aprendizagem, construcionistas, para surdos, na perspectiva do usuário.** 2019. Dissertação de mestrado - Pós-Graduação em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins. 2019. Disponível em: <<https://repositorio.uft.edu.br/bitstream/11612/1447/1/Magaly%20Liliane%20Chaves%20Campos%20-%20Disserta%C3%A7%C3%A3o.pdf>> . Acessado em: 10 de fevereiro de 2021.

DETECÇÃO de rosto OpenCV vs Detecção de Rosto Yolo. **Base App**, 2018. Disponível em: <<https://www.baseapp.com/deepsight/opencv-vs-yolo-face-detector>>. Acesso em: 12 de janeiro. de 2021.

DIFICULDADES de Aprendizagem em Surdos. **Portal Educação. 2020.** Disponível em: <https://siteantigo.portaleducacao.com.br/conteudo/artigos/idiomas/dificuldades-de-aprendizagem-em-surdos/26871> . Acesso em: 10 de jun. de 2021.

FERREIRA, Ricardo. CANESCHE, Michael; CARVALHO, Westerley. **Minicurso de programação em GPU do ambiente Google Colaboratory.** 2020. Disponível em: <<http://wscad.sbc.org.br/2020/artigos/minicursos/minicurso3-Prog-GPU-Colaboratory-texto.pdf>>. Acessado em: 22 fev. de 2021.

FLECK, Leandro; TAVARES, Maria Hermínia Ferreira; EYNG, Eduardo; HELMANN, Andrieli Cristina; ANDRADE, Minéia Aparecida de Moares. **Redes Neurais Artificiais: Princípios básicos.** Revista Eletrônica Científica Inovação e Tecnologia. 2016. Disponível em: <<https://periodicos.utfpr.edu.br/recit/article/view/4330>>. Acessado em: 10 de julho de 2021.

GÓES, J. W. F. **Aperfeiçoamento de um sistema de reconhecimento de padrões do alfabeto de libras utilizando luva sensora.** 2019. Disponível em: <http://repositorio.utfpr.edu.br/jspui/bitstream/1/24355/1/PB_COENC_2019_1_03.pdf>. Acessado em: 20 de fevereiro de 2021.

GOLDFELD, M. **A criança surda: Linguagem e cognição numa perspectiva sócio interacionista.** 2ª ed. São Paulo: Plexus Editora, 2002.

GONZAGA, A. **Kanguera: Olho Local Mão Distante.** 2011. Disponível em: <<http://iris.sel.eesc.usp.br/weblab/default.html>>. Acesso em 14 de junho de 2021.

HAYKIN, S. S. Neural netwo L. FLECK, M. Tavares, E. Eyng, A. Helmann, and M. Andrade, “Redes neurais artificiais: Princípios básicos,” *Revista Eletrônica Científica Inovação e Tecnologia*, vol. 1, no. 13, pp. 47–57, 2016. rks and Learning Machines. vol. 3. Pearson Upper Saddle River, NJ, USA: 2009

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Brasileiro de 2020.** Rio de Janeiro: IBGE, 2020.

KARNOPP, Lodenir; QUADROS, Ronice Muller de. **Educação infantil para surdos.** In: ROMAN, Eurilda Dias; STEYER, Vivian Edite. (Org.). *A criança de 0 a 6 anos e a educação infantil: um retrato multifacetado.* Canoas, 2001, p. 214-230.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos metodologia científica**. 4.ed. São Paulo: Atlas, 2001.

LECUN, Yann et al. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, v. 86, n. 11, p. 2278-2324, 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. **Convolutional networks and applications in vision**. Paris: IEEE, 2010.

LÉVY, Pierre. **Cibercultura**. São Paulo: Editora 34, 2005.

MEDEIROS, Moisés de. **Seleção de principais atributos para Redes Neurais Artificiais do tipo MLP: um estudo de caso sobre mineração de dados para diagnóstico de dengue**. Trabalho de Conclusão de Curso de Graduação em Sistemas de Informação da Universidade Federal do Rio Grande do Norte, UFRN. Caicó, 2018.

MILANO, Danilo de; HONORATO, Luciano Barrozo. **Visão computacional**. Universidade Estadual de Campinas, 2010.

MOURA, Débora Rodrigues. **Módulo1: introdução à surdez e a Libras**. Enap. 2016. Disponível em: <https://repositorio.enap.gov.br/handle/1/3172> . Acessado em: 22 de abril de 2021.

MOURA, Flávio Rafael Trindade, et al. "**Contamination risk estimation model for respiratory diseases in monitored environments using YOLOv5**." Anais do XLIX Seminário Integrado de Software e Hardware. SBC, 2022.

NEVES, Luiz Antônio Pereira; VIEIRA NETO, Hugo; GONZAGA, Adilson. SBN: 978-85-64619-09-8. **Avanços em Visão Computacional**. Curitiba: Omnipax Editora, 2012.
PACHECO, J. **Caminhos para a inclusão: um guia para o aprimoramento da equipe escolar**. Porto Alegre: Artmed, 2007.

PAVAN, Adilson. OSSADA JUNIOR, J. C. MODESTO, F. **reconhecimento de gestos com segmentação de imagens dinâmicas aplicadas a libras**. Campinas, SP, 2010. Disponível em: <https://repositorio.pgsskroton.com/bitstream/123456789/1240/1/artigo%202023.pdf>. Acessado em: 08 de maio de 2020.

PAVLOVIC, V. I; SHARMA, R.; HUAN, T. S. **Visual Interpretation of Hand Gestures for Human-Computer Interaction**. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 19, no 7, 1997.

PIRES, Vanessa De Oliveira Dagostim. **Andamento Coletivo como Prática de Ensino-Aprendizagem de Língua Portuguesa para Surdos**. 2009. 143 f. Dissertação (Mestrado) - Programa de Pós-graduação em Linguística Aplicada, Universidade do Vale do Rio Dos Sinos, São Leopoldo, 2009.

QUEIFER, H. S. **Reconhecimento de gestos para acionamento de objetos usando CNN e o algoritmo haar cascade**. 2019. Disponível em: <https://dspace.bc.uepb.edu.br/jspui/handle/123456789/21842>. Acessado em: 03 de abril de 2021.

RAUBER, T. W. **Redes neurais artificiais**. Universidade Federal do Espírito Santo, Departamento de Informática, 2005.

REDMON, Joseph. **How computers learn to recognize objects instantly**. Youtube, 17 de ago. 2017. Disponível em: <<https://www.youtube.com/watch?v=Cgxsv1riJhI&t=356s>>. Acesso em: 15 jan. 2021.

RODRIGUES, Diego Alves. **Deep Learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento automotivo**. 2018. Monografia – Universidade Federal da Paraíba (Centro de Informática). Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/123456789/15606/1/DAR20052019.pdf> . Acessado em: 01 de abril de 2021.

Rosa, A. S., & Cruz, C. C. (2001). **Internet: Fator de inclusão da pessoa surda**. Revista Online da Biblioteca Prof. Joel Martins, 2, 38-54. <https://doi.org/10.20396/etd.v2i3.580>.

SACRAMENTO, Gabriel. **Dataset: entenda o que é e conheça 10 fontes confiáveis**. Somostera, 10 de setembro de 2021. Disponível em: <<https://blog.somostera.com/data-science/dataset>>. Acesso em: 20 de agosto de 2022.

SANTOS, Carlos Alexandre Silva dos. **Reconhecimento de Imagens de Marcas de Gato Utilizando Redes Neurais Convolucionais**. 2017. Dissertação de mestrado – Pós-graduação em Engenharia Elétrica - Universidade Federal do Pampa. Disponível em: <<http://dspace.unipampa.edu.br/jspui/handle/rii/2028>> . Acessado em: 12 de março de 2021.

SANTOS, Clebeson Canudo dos. **Sistemas de Reconhecimeto de Gestos e ações em tempo real baseado em visão computacional**. 2020. Tese (Doutorado) – Curso de Engenharia Elétrica, Universidade Federal do Espírito Santo, Centro Tecnológico. 2020.

SANTOS, Rodrigo Neves dos. **Reconhecimento de gestos de mão humana em imagens de vídeo**. São Carlos SP, 2011. Disponível em: <<http://www.tcc.sc.usp.br/tce/disponiveis/18/180450/tce-29032012-085236/?&lang=brDISCENTE>>. Acessado em: 14 de abril de 2021.

SILVA, Carine Mendes da; SILVA, Daniele Nunes Henrique. **Libras na educação de surdos: o que dizem os profissionais da escola? Psicologia Escolar e Educacional**, Maringá, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-85572016000100033&lng=en&nrm=iso>. Acesso em: 5 dezembro. 2020.

SILVA, Ivani Rodrigues. **Línguas em contato e em Conflito: A Trajetória do Aluno Surdo na Escola**. In: II Simpósio Internacional sobre Bilinguismo, 2002, Galicia. Actas/Proceedings II Simposio Internacional Bilinguismo, 2002. p. 1807-1813.

SITE OFICIAL DO ROBOFLOW, Roboflow, 2021. Disponível em: <<https://roboflow.com/>> . Acesso em: 14, mar 2021.

SKLIAR, Carlos. **A surdez: um olhar sobre as diferenças**. Porto Alegre: Editora Mediação, 2015.

SURDEZ no Brasil: acessibilidade ainda é um desafio diário. EuSaúde, 2021. Disponível em: <https://portal.eusaude.com.br/surdez-no-brasil-acessibilidade-ainda-e-um-desafio-diario/#:~:text=De%20acordo%20com%20dados%20do,7%20milh%C3%B5es%20n%C3%A3o%20ouvem%20nada>. Acesso em: 23 de abril de 2021.

TAMANG, Benjamin et al. **Customized Deep Learning Technique for vehicle detection along with speed estimation.** EasyChair, 2022.

TV BRASIL: Apenas 37% dos brasileiros com deficiência auditiva estão empregados. Instituto Locomotiva, 2019. Disponível em: <https://www.ilocomotiva.com.br/single-post/2019/10/01/tv-brasil-apenas-37-dos-brasileiros-com-defici%C3%Aancia-auditiva-est%C3%A3o-empregados>. Acesso em: 18 de maio de 2021.