

# Redes neurais convolucionais aplicada ao problema de classificação do alfabeto da língua de sinais americano

Guilherme Magalhães Júnior

PPGC - Programa de Pós-Graduação em Ciência da Computação

UFOP - Universidade Federal de Ouro Preto

Email: guilherme.junior@aluno.ufop.edu.br

**Abstract**—The abstract goes here.

## I. INTRODUÇÃO

Existem mais de 500 milhões de pessoas surdas no mundo e este número chegará a 1 bilhão até 2050, segundo dados da OMS [1]. A língua de sinais possibilita à estas pessoas com deficiências auditivas, e/ou na fala, de se relacionarem com o mundo e também desempenha papel inclusivo importantíssimo na vida delas ao permitir que as mesmas estejam aptas a viver, por si só, todo tipo de situação em que a comunicação é de fato essencial, como por exemplo, em um atendimento médico.

Toda a comunicação via língua de sinais se dá por meio de gesticulações das mãos e também expressões faciais, ou seja, é uma língua em que a comunicação é estabelecida de forma visual e não sonora. Assim como ocorre com os idiomas, há mais de uma centena de línguas de sinais pelo mundo. No Brasil por exemplo a oficial é a LIBRAS (Língua Brasileira de Sinais) enquanto nos EUA é a ASL (*American Sign Language*).

A língua de sinais americana (ASL) é o quarto idioma mais usado na América do Norte. Ele é utilizado como meio de comunicação por mais de 1 milhão de pessoas dentro e fora dos Estados Unidos da América e em mais de 30 países onde o inglês é o idioma principal [2]. A ASL é formada por 26 sinais gestuais realizados com a mão que representam todo o alfabeto conhecido.

Aprender a língua dos sinais é tão ou mais desafiador quanto aprender um novo idioma falado. É necessário um tempo mínimo de dedicação para memorizar todos as simbologias e também um bom tempo de prática para poder executá-las de forma hábil. Além dos desafios que normalmente ocorrem ao se aprender um novo idioma, há fatores dificultadores como a velocidade com que as mãos são gesticuladas e a semelhança entre algumas simbologias. Sendo assim, tal como já ocorre com quaisquer idiomas que não dominamos, também se faz necessário o estudo e desenvolvimento de um sistema de reconhecimento e tradução de língua de sinais para o idioma local, de forma a permitir ainda mais a integração das pessoas com deficiência auditiva e de fala, com o mundo atual.

Nesse trabalho pretendemos abordar o desenvolvimento de um sistema de reconhecimento de língua de sinais americano

(ASL) utilizando redes neurais convolucionais e assim permitir que qualquer pessoa possa compreender este idioma. A rede será projetada para trabalhar com imagens RGB estática e sem profundidade (2D).

### A. American Sign Language (ASL)

Ao contrário do que muitos possam pensar, existem diferentes tipos de línguas de sinais no mundo. No Brasil, por exemplo, a língua de sinais oficial se chama LIBRAS que é um acrônimo que significa Língua Brasileira de Sinais. Neste trabalho é proposto trabalhar com uma delas a *American Sign Language* (ASL) que é a língua de sinais oficial dos Estados Unidos da América e que é utilizada por todos os países em que a língua inglesa é oficialmente utilizada. Na verdade, neste trabalho iremos nos ater ao alfabeto desta língua como mostrado na figura 1.

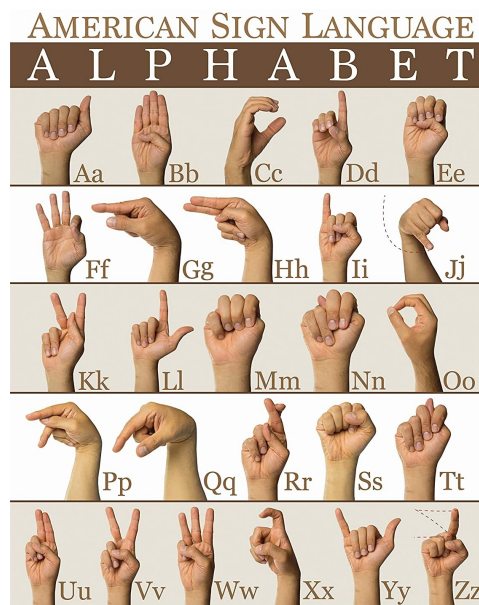


Fig. 1. ASL - American Sign Language Alphabet

## II. OBJETIVO

Usar a tecnologia para traduzir a linguagem de sinais em texto decifrável para que o torne compreensível para o público

dito normal e dessa forma ajudar a (re)inserir na sociedade pessoas que apresentam limitações auditivas e vocais.

### III. TRABALHOS RELACIONADOS

Diferentes abordagens foram propostas na tentativa de resolver o problema de utilizar o aprendizado de máquina para reconhecer o alfabeto a partir da língua de sinais.

Kurdyumov et al. [3] propuseram em seu trabalho usar dois algoritmos, *Support Vector Machine* (SVM) e *k-nearest neighbors* (KNN) para aprender a classificar imagens estáticas de mãos representando o alfabeto da língua de sinais americano (*American Sign Language*). A base de dados utilizada era composta por 1800 imagens coloridas, obtidas a partir de uma webcam, com resolução de 640 x 480 pixels, divididas em 25 classes representando todas as letras do alfabeto com exceção da letra "j". A base de dados sofreu um pré-processamento com as imagens convertidas em escala de cinza e para a resolução de 20 x 20 pixels. Após esta etapa a base de dados foi dividida em 85% para treino e 15% para teste. Os pesquisadores conseguiram uma acurácia de 93.5% com o SVM e um 84.8% com o KNN.

Chuan et al. [4] propuseram um sistema de reconhecimento de língua de sinais americano (ASL) usando um compacto sensor de movimento 3D, conhecido como Leap Motion, para capturar o movimento das mãos e assim gerar a base de dados. Leap Motion consiste em um pequeno dispositivo capaz de captar movimentos dos dedos das mãos do usuário. Novamente foram utilizados os algoritmos SVM e KNN para classificar uma base de dados geradas a partir da captura dos movimentos feitos por dois voluntários que representaram com as mãos as 26 letras existentes na ASL (*American Sign Language*) e assim produziram um banco de dados com 7900 entradas. A base de dados foi dividida em 75% para treino e 25% para teste. A acurácia atingida pelos algoritmos SVM e KNN foram, respectivamente, de 79,83% e 72,78%.

Daroya et al. [5] propuseram em seu trabalho usar uma rede com arquitetura DenseNet. A escolha por essa arquitetura pelos autores foi incentivada principalmente, tal como relatado pelos mesmos, por causa do problema da dissipação do gradiente (*The Vanishing Gradient Problem*) que é comum em redes profundas (*Deep Network*) e menos crítica na arquitetura DenseNet. A base de dados utilizada continha cerca de 50.000 imagens RGB, 120 x 160 pixels, capturadas por uma câmera de um notebook, dividida em 24 classes diferentes que representavam todas as letras do alfabeto com exceção das letras "y" e "z". Novamente a base de dados foi dividida em 90% para treino e 10% para teste e o resultado obtido foi de 90,3% de acurácia.

Já Costa [6] propôs em seu trabalho investigar 18 variações de arquitetura de redes neurais convolucionais (CNN) para descobrir quais destas teria a maior precisão para reconhecer a língua de sinais ASL. Todas as redes foram aplicadas a uma base de dados com 87.000 imagens RGB, 200 x 200 pixels, divididas em 29 classes, sendo 26 destas representando as letras do ASL e as outras 3 representando termos auxiliares ("espaço", "apagar" e "nada"). A rede que alcançou o melhor

resultado foi a "5Conv" que obteve acurácia de 49,77% quando aplicado na base de teste.

Kumar et al. [2] também propuseram utilizar redes neurais convolucionais (CNN) para reconhecer o alfabeto a partir da língua de sinais ASL. A base de dados é composta de 34.627 imagens divididas em 24 classes, sendo as 26 letras do alfabeto com exceção das letras "y" e "z". A base de dados sofreu um pré-processamento com a intenção desta se assemelhar à clássica base de dados MNIST. Dessa forma os dados utilizados para treinamento da rede neural apresentavam imagens com resolução 28 x 28 pixels, em escala cinza com variação de 0-255 e rotuladas de 0 a 25 (com exceção do 9 e 25), representando 24 das 26 letras do alfabeto. A acurácia alcançada pelos autores neste trabalho foi de 99,63%.

Percebe-se, a crescente tendência no uso de redes neurais em trabalhos onde a proposta é a classificação de imagens. Especificamente em trabalhos onde a proposta era de reconhecer imagens de língua de sinais, alguns trabalhos conseguiram acurácia acima de 90% utilizando-se destas redes.

### IV. METODOLOGIA

### V. RESULTADOS

### VI. CONCLUSÃO

### REFERENCES

- [1] "País tem 10,7 milhões de pessoas com deficiência auditiva, diz estudo," Oct. 2019. [Online]. Available: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-10/brasil-tem-107-milhoes-de-deficientes-auditivos-diz-estudo>
- [2] M. Kumar, P. Gupta, R. K. Jha, A. Bhatia, K. Jha, and B. K. Shah, "Sign language alphabet recognition using convolution neural network," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021, pp. 1859–1865.
- [3] R. Kurdyumov, P. Ho, and J. Ng, "Sign language classification using webcam images," 2011.
- [4] C.-H. Chuan, E. Regina, and C. Guardino, "American sign language recognition using leap motion sensor," in *2014 13th International Conference on Machine Learning and Applications*. IEEE, 2014, pp. 541–544.
- [5] R. Daroya, D. Peralta, and P. Naval, "Alphabet sign language image classification using deep learning," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 0646–0650.
- [6] A. COSTA and U. EDU, "American sign language (asl) alphabet image classification using convolutional neural networks."