

Aplicação de Visão Computacional no Reconhecimento de Sinais da Língua Americana de Sinais (ASL)

1st Jamil Soares da Silva Júnior
Instituto de Informatica
Universidade Federal de Goiás
Goiania, Brasil
jamiljunior@discente.ufg.br

2nd Letícia de Cerqueira Xavier
Instituto de Informatica
Universidade Federal de Goiás
Goiania, Brasil
cerqueiraleticia@discente.ufg.br

3rd Victor Guerreiro Pimenta
Instituto de Informatica
Universidade Federal de Goiás
Goiania, Brasil
victorguerreiro@discente.ufg.br

4th Francieli Moreira de Carvalho
Instituto de Informatica
Universidade Federal de Goiás
Goiania, Brasil
francielimoreira@discente.ufg.br

Abstract—Este projeto de visão computacional tem como objetivo aprimorar a comunicação entre pessoas surdas e ouvintes, focalizando na identificação e reconhecimento da Língua Americana de Sinais (ASL). Inicialmente, concentra-se na identificação dos gestos e sinais correspondentes às letras do alfabeto, abrangendo de A a Z. Isso possibilitará aos usuários uma comunicação mais eficiente em atividades básicas de escrita e leitura, promovendo inclusão e acessibilidade. Embora a aplicação inicial não contemple a identificação imediata de outros gestos e palavras da ASL, o principal objetivo é facilitar a compreensão e uso do alfabeto na língua.

Index Terms—ASL, reconhecimento, sistema, alfabeto, acessibilidade.

I. INTRODUÇÃO

A tarefa de classificação desempenha um papel fundamental em diversos aspectos da vida cotidiana, permitindo a categorização eficaz de objetos com base em suas características distintas (SILVA, 2019). A capacidade de desenvolver algoritmos que possam discernir e agrupar objetos tem um impacto significativo, sendo um exemplo notável a identificação de placas de veículos, onde um algoritmo analisa as letras e as associa ao alfabeto, assemelhando-se ao reconhecimento de símbolos.

Além disso, a habilidade de criar algoritmos para identificar símbolos de linguagens também abre portas para aplicações de tradução. Contudo, é essencial compreender a complexidade da classificação da linguagem de sinais, que muitas vezes é desafiadora mesmo para indivíduos proficientes na Língua de Sinais Americana (ASL, do inglês American Sign Language).

Atualmente, há aplicativos que auxiliam na tradução da língua inglesa para ASL por meio de animações 3D de intérpretes. No entanto, existe uma lacuna no que diz respeito a soluções que possam realizar o processo inverso, ou seja,

reconhecer o significado dos gestos por meio de câmeras e traduzi-los.

Este desafio é ainda mais complexo devido à diversidade dos movimentos das mãos, que exige uma distinção precisa das configurações das mãos e dedos. Isso acrescenta complexidade ao reconhecimento do significado do sinal.

Dado que a comunicação é essencial para a socialização, é imperativo explorar abordagens que facilitem a interação entre usuários de ASL e aqueles menos familiarizados com essa língua, a fim de promover a inclusão social. O uso de Redes Neurais Artificiais (RNA) para desenvolver sistemas inteligentes capazes de realizar tais traduções emerge como uma estratégia para aproximar a comunidade surda, que se comunica por meio de ASL, daqueles que não dominam a linguagem.

II. FUNDAMENTOS TEÓRICOS

Este capítulo tem como objetivo realizar uma contextualização de todos os temas necessários para o desenvolvimento deste trabalho. Para tal, esse capítulo se estrutura em seis tópicos, a saber: Língua de Sinais Americana; Visão Computacional; Aprendizado de Máquina; Redes Neurais Artificiais; Aprendizado Profundo e Redes Neurais Convolucionais.

A. Língua de Sinais Americana (ASL)

A ASL é uma língua visual-gestual complexa, vital para a comunicação e identidade da comunidade surda nos EUA e Canadá. Originada no século XIX, a ASL possui gramática própria e papel cultural.

B. Aprendizado de Máquina:

O Aprendizado de Máquina envolve algoritmos que permitem a máquina aprender a partir de dados e melhorar

seu desempenho em tarefas específicas. É uma ferramenta essencial para explorar padrões e tomar decisões inteligentes.

C. Aprendizado Profundo:

O Aprendizado Profundo é uma vertente do Aprendizado de Máquina que utiliza Redes Neurais Artificiais (RNAs) para modelar tarefas complexas. As RNAs, inspiradas em neurônios biológicos, aprendem a partir dos dados, capturando padrões sutis.

D. Redes Neurais Artificiais (RNAs)

As RNAs são estruturas computacionais inspiradas no sistema nervoso biológico. Elas processam informações por meio de neurônios interconectados e são empregadas para resolver problemas complexos de classificação e regressão.

III. METODOLOGIA

A metodologia adotada neste projeto visa desenvolver um modelo de aprendizado profundo capaz de classificar imagens em 29 categorias diferentes, correspondentes aos símbolos da Língua de Sinais Americana (ASL). Utilizamos a técnica de transferência de aprendizagem, empregando a arquitetura MobileNet, conhecida por sua eficiência em dispositivos móveis, como base para o nosso modelo.

A. Modelo de Base

A versão da MobileNet pré-treinada no conjunto de dados ImageNet foi selecionada como nosso modelo base. A eficiência e precisão dessa arquitetura a tornaram uma escolha apropriada para classificar os símbolos da ASL.

B. Personalização do Modelo

O modelo foi adaptado à nossa tarefa específica de classificação dos símbolos da ASL adicionando três camadas densas. As duas primeiras, com 128 unidades e função de ativação ReLU, foram projetadas para extrair características relevantes dos símbolos. A camada de saída, com 29 unidades e ativação softmax, foi configurada para classificar as imagens nas 29 categorias da ASL.

C. Compilação e Treinamento

O modelo foi compilado com o otimizador Adam e uma taxa de aprendizado de 0.001. Utilizamos a entropia cruzada categórica como função de perda e a acurácia como métrica de avaliação. O treinamento foi conduzido usando um conjunto específico de imagens representando os símbolos da ASL, seguido de validação e teste.

D. Aumento de Dados (Data Augmentation)

No desenvolvimento do modelo, foram empregadas técnicas de aumento de dados para aprimorar a capacidade de generalização e robustez. Durante o pré-processamento das imagens, foram aplicadas rotações aleatórias, distorções de cisalhamento, deslocamentos horizontais e verticais, ajustes de zoom e espelhamento horizontal. Essas transformações foram aplicadas de forma aleatória e controlada, visando criar um conjunto de treinamento mais diversificado e representativo.

A utilização dessas técnicas contribuiu significativamente para aprimorar a performance e a capacidade de generalização do modelo final.

IV. RESULTADOS

Após a aplicação das técnicas de aumento de dados e o treinamento da arquitetura escolhida com pesos pré-treinados, constatou-se que a MobileNet alcançou um desempenho superior para o conjunto de dados.

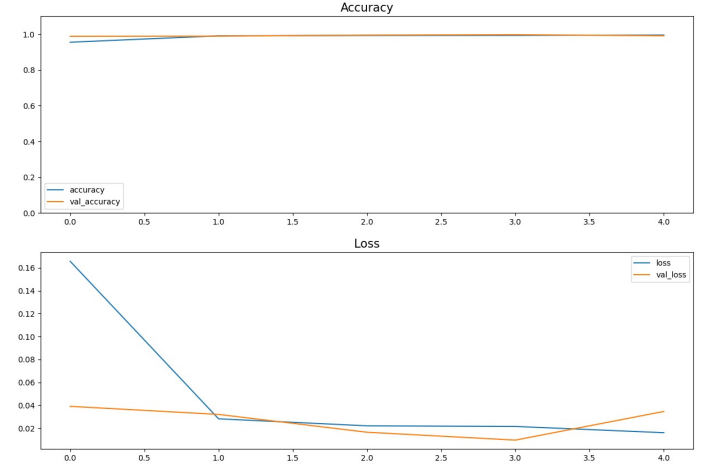


Fig. 1. plot das métricas ao longo das épocas de treinamento

Diante dos resultados obtidos, a MobileNet emergiu como a arquitetura selecionada para o reconhecimento de caracteres ASL, devido à sua performance com uma acurácia notável de 99.54

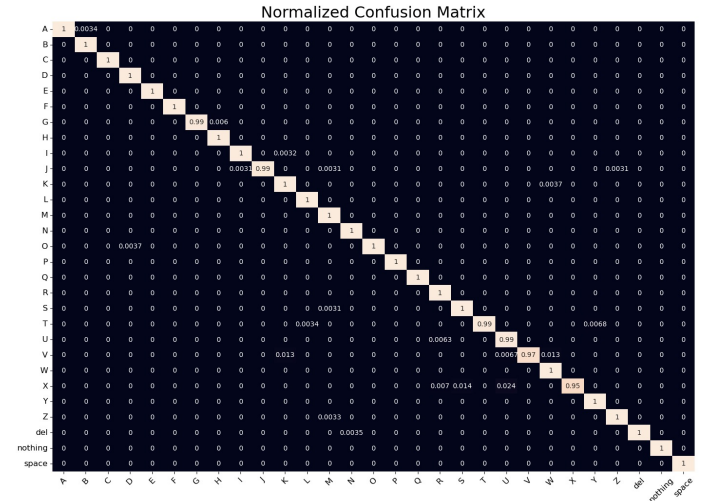


Fig. 2. matriz de confusão normalizada no conjunto de validação

V. CONCLUSÃO

Em síntese, o desenvolvimento desse sistema de reconhecimento de caracteres ASL por meio de técnicas avançadas de aprendizado profundo e aumento de dados contribui substancialmente para a promoção da inclusão social, permitindo

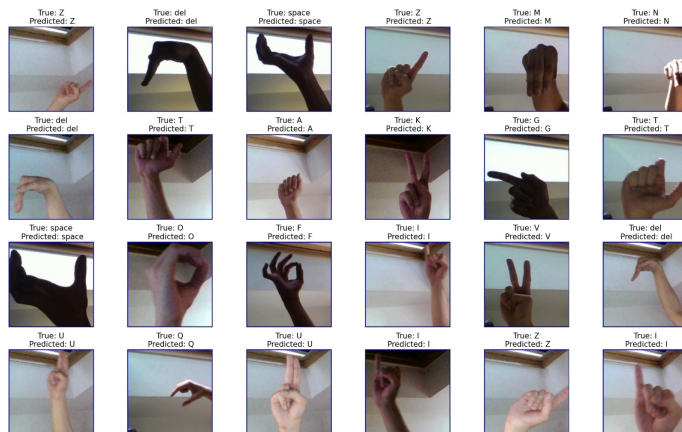


Fig. 3. exemplos de predições no conjunto de validação.

uma comunicação mais eficaz entre a comunidade surda e aqueles que não dominam a língua de sinais. Esse avanço representa um passo significativo em direção a um ambiente mais inclusivo e acessível para todos.

REFERENCES

- SILVA, G. F. d. Tecnologias assistiva e a deep learning: aplicativo com reconhecimento de imagem no auxílio a deficientes visuais. 2019.
- LOCA, A.; RAUBER, T. Uso de uma rede neural convolucional unidimensional para detecção de falhas em processos industriais. In: SBC. Anais da XIX Escola Regional de Computação Bahia, Alagoas e Sergipe. [S.l.], 2019.
- BALLARD, D. H. Ballard d. and brown cm 1982 computer vision. Image, 1982.
- DAVIES, E. R. Computer and machine vision: theory, algorithms, practicalities. [S.l.]: Academic Press, 2012.
- CONCI, A.; AZEVEDO, E.; LETA, F. R. Computação gráfica. [S.l.]: Elsevier, 2008.
- FARIA, A. C. B. Classificação de estilos visuais utilizando aprendizado profundo. 2017.
- MITCHELL, R.; MICHALSKI, J.; CARBONELL, T. An artificial intelligence approach. [S.l.]: Springer, 2013.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, Springer, v. 5.
- MARTINS, V. E. Aplicação de deep learning para detecção de veias em carne suína. 2018.
- CARVALHO, H. M. Aprendizado de máquina voltado para mineração de dados: árvores de decisão. 2014.
- TEIXEIRA, J. de F. Inteligência artificial. [S.l.]: Pia Sociedade de São Paulo-Editora Paulus, 2014.
- LENT, D. M. B.; JR, M. L. P. Detecção de anomalias utilizando redes neurais convolucionais. 2018.
- PONTI, M. A.; COSTA, G. B. P. D. Como funciona o deep learning. arXiv preprint arXiv:1806.07908, 2018.
- LECUN, Y. et al. Comparison of learning algorithms for handwritten digit recognition. In: PERTH, AUSTRALIA. In-

ternational conference on artificial neural networks. [S.l.], 1995. v. 60.

ARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. Proceedings of the xxix conference on graphics, patterns and images. [S.l.], 2016.