

RECONHECIMENTO E TRADUÇÃO DA LINGUAGEM DE SINAIS BRASILEIRA (LIBRAS) UTILIZANDO VISÃO COMPUTACIONAL E DEEP LEARNING

João Pedro do Espírito Santo Almeida, Bolsista graduando em Engenharia de Controle e Automação
Iniciação científica, joaoesalmeida@hotmail.com
Lucas Amparo Barbosa
SENAI CIMATEC, Salvador – BA, lucas.barbosa@fieb.org.br

RESUMO

Este trabalho visa desenvolver um Identificador e Tradutor da Linguagem Brasileira de Sinais (LIBRAS) utilizando tecnologias avançadas como Redes Neurais Convulsionais e Visão Computacional, promovendo a inclusão de pessoas com deficiência auditiva. O projeto é desenvolvido em Python, utilizando as bibliotecas OpenCV para Visão Computacional e PyTorch para Deep Learning. No estágio atual, o sistema é capaz de detectar sinais estáticos, abrangendo números de 0 a 9 e todas as letras estáticas do alfabeto brasileiro em LIBRAS. Os próximos passos incluem a identificação de sinais contínuos em LIBRAS, com planos para implementar modelos de Redes Neurais Recorrentes (RNN) ou Long Short-Term Memory (LSTM) para essa finalidade desafiadora. Após alcançar um desempenho satisfatório com uma pequena base de dados, será aplicada a técnica de aumento de dados (data augmentation) para criar variações, expandindo assim a base de dados e aprimorando a capacidade de reconhecimento dos sinais contínuos em LIBRAS.

PALAVRAS-CHAVE: Deep Learning, RNN, LSTM, LIBRAS, Computer vision, continuous signals.

1. INTRODUÇÃO

Este projeto surge como resposta à escassez de recursos e modelos para identificação da Linguagem Brasileira de Sinais (LIBRAS), focando especialmente na detecção de sinais contínuos. Inspirado em trabalhos prévios (Kelly, McDonald & Markham, 2009; Carneiro, Cortez & Costa, 2021), o projeto inicialmente implementou um modelo de Machine Learning, Random Floresta, e, posteriormente, foi implementado uma Rede Neural Perceptron Multicamada (MLP) para identificação de sinais estáticos, como números e letras do alfabeto em LIBRAS. Contudo, reconhecendo a complexidade adicional dos sinais contínuos, a pesquisa se volta agora para a exploração de Redes Neurais Recorrentes (RNN) e Redes Neurais de Longa Memória de Curto Prazo (LSTM), visando uma identificação mais precisa e eficiente desses sinais.

A abordagem adotada contempla a utilização de tecnologias avançadas, como Redes Neurais Convulsionais e Visão Computacional, a fim de avançar na identificação e tradução dos gestos presentes na LIBRAS. A implementação de técnicas como o data augmentation, que amplia a base de dados e diversifica os sinais existentes, pretende aprimorar ainda mais a precisão na interpretação dos sinais contínuos da LIBRAS, contribuindo para uma comunicação mais inclusiva e acessível para um público diversificado.

Com base em estudos anteriores e referências práticas (Computer Vision Engineer, 2023; Renotte, 2021; Mulla, 2023), a pesquisa busca compreender os fundamentos técnicos necessários para avançar na implementação do projeto. A aplicação desses conhecimentos em conjunto com as tecnologias mencionadas visa alcançar resultados mais robustos, melhorando a capacidade de identificação e tradução dos sinais contínuos em LIBRAS em tempo real, o que pode significar um marco importante na inclusão e acessibilidade para pessoas surdas ou com deficiência auditiva.

2. METODOLOGIA

Para realizar este projeto, foi elaborado uma abordagem cuidadosa e eficiente, utilizando uma metodologia experimental com etapas específicas e métodos bem definidos para coleta e análise de dados, representado pela figura 01.

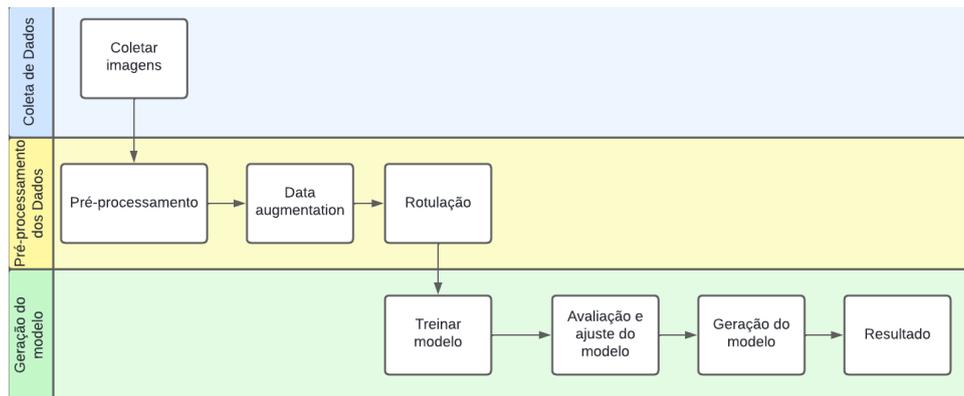


Figura 01: Fluxograma das etapas realizadas

O processo de identificação e tradução da Linguagem Brasileira de Sinais (LIBRAS) teve início utilizando a biblioteca MediaPipe da Google. Essa ferramenta nos permitiu detectar e extrair *landmarks* (pontos de referência) para cada mão nas imagens, os quais foram convertidos em conjuntos de coordenadas normalizadas, figura 02. Esses *landmarks* serviram como base para a construção da nossa base de dados inicial, representando cada conjunto de *landmarks* como um sinal estático em LIBRAS.

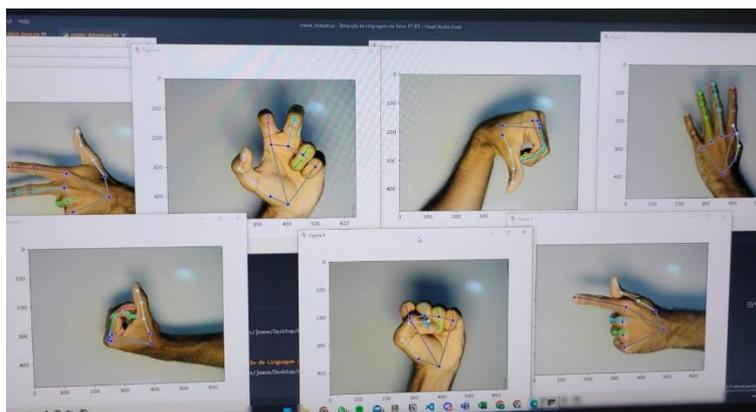


Figura 02: Detecção e registro dos sinais de mãos realizados

No início, foi utilizado o modelo de Machine Learning, Random Forest, com os dados normalizados, utilizando uma divisão dos dados de 75% para treinamento e 25% para teste. Apesar dos resultados aceitáveis, foi implementado uma Rede Neural Perceptron Multicamada (MLP) com duas camadas e 26 neurônios, o que resultou em uma significativa melhoria na precisão da identificação dos sinais estáticos, chegando a valores maiores que 90%.

Apesar da alta precisão atingida, observamos que alguns sinais visualmente semelhantes ainda eram confundidos entre si. Para resolver esse desafio, implementamos a técnica de *data augmentation*, ampliando o tamanho e a diversidade dos dados na base inicial. Esse processo resultou em uma melhoria adicional na precisão da identificação dos sinais estáticos em LIBRAS.

3. RESULTADOS E DISCUSSÃO

O projeto apresentado propõe um Identificador e Tradutor da Linguagem Brasileira de Sinais (LIBRAS) utilizando Visão Computacional e Aprendizado Profundo. Inicialmente, o sistema foi capaz de detectar sinais estáticos, abrangendo números de 0 a 9 e todas as letras estáticas do alfabeto brasileiro em LIBRAS. Essa etapa foi realizada com sucesso, alcançando resultados confiáveis na identificação dos sinais estáticos.

A aplicação da técnica de aumento de dados (*data augmentation*) é uma estratégia promissora para expandir a base de dados inicial, permitindo uma maior variedade de sinais e melhorando a precisão da identificação

dos sinais contínuos em LIBRAS. Através dessa técnica, será possível enriquecer a base de dados existente e aprimorar ainda mais a capacidade do sistema de reconhecer e traduzir uma variedade mais ampla de sinais contínuos em LIBRAS.

Tabela 01: Comparativo entre a precisão dos modelos utilizados

MODELO	PRECISÃO
Random Forest	≈70%
MLP	≈90%

4. CONSIDERAÇÕES FINAIS

O projeto avança na identificação e tradução da LIBRAS usando Visão Computacional e Aprendizado Profundo. Ao detectar sinais estáticos e progredir para sinais contínuos, visa melhorar a acessibilidade para surdos. A implementação de modelos como RNN e LSTM é crucial para lidar com a natureza sequencial e temporal dos sinais contínuos. Com o desenvolvimento contínuo e técnicas como aumento de dados, espera-se que o sistema contribua para a inclusão de pessoas surdas, facilitando a comunicação e promovendo a igualdade de oportunidades. O próximo passo envolve a implementação de RNN ou LSTM para reconhecer e traduzir sinais contínuos da LIBRAS, representando um avanço na identificação desses sinais em tempo real.

5. AGRADECIMENTOS

Agradeço ao professor Marcio Renê Brandão Sousa pela indicação da oportunidade de bolsa e ao empenho de Lucas Amparo Barbosa no CCC, que contribuíram para meu crescimento profissional. Agradeço também ao SENAI CIMATEC pelo financiamento da pesquisa, permitindo avanços na identificação e tradução da LIBRAS.

6. REFERÊNCIAS

Carneiro, A. T. S.; Cortez, P. C.; Costa, R. C. Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu. Laboratório de Engenharia de Sistemas de Computação - LESC, Universidade Federal do Ceará - UFC, Fortaleza, CE, Brasil.

Computer Vision Engineer. (2023, 26 de janeiro). Sign language detection with Python and Scikit Learn | Landmark detection | Computer vision tutorial. YouTube. URL: https://www.youtube.com/watch?v=MJCSjXepaAM&list=PLLCMDwafgqUHoxoTtGf0c2Eiv8bAhvjB_&index=3&t=2930s

Cui, R.; Liu, H.; Zhang, C. Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. Department of Automation, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing, China, 2017.

Google. MediaPipe Documentation [online]. Disponível em: <https://google.github.io/mediapipe/>. Acesso em: 6 mar. 2023.

Kelly, D.; McDonald, J.; Markham, C. Continuous Recognition of Motion Based Gestures in Sign Language. Computer Science Department, National University of Ireland, Maynooth, 2009.

Mulla, R. (2023, 4 de abril). American Sign Language Detection with Python and MediaPipe. YouTube. URL: <https://www.youtube.com/watch?v=L-laQch8KYY>

Renotte, N. (2021, 19 de julho). Sign Language Detection using ACTION RECOGNITION with Python | LSTM Deep Learning Model. YouTube. URL: <https://www.youtube.com/watch?v=doDUihpj6ro>

Sultan, A., Makram, W., Kayed, M., & Ali, A. A. Sign language identification and recognition: A comparative study. 2022.