



II SEMANA DE PEDAGOGIA

EDUCAÇÃO, PESQUISA E ENSINO:
CONSTRUINDO E (RE)CONSTRUINDO SABERES



CAMPUS DE
VITÓRIA DA CONQUISTA

19 A 23 DE AGOSTO DE 2024



IDENTIFICAÇÃO VISUAL EM TEMPO IMEDIATO DA LÍNGUA DE SINAIS BRASILEIRA - LIBRAS

Larissa Santos Oliveira

José Alberto Díaz Amado

Marcelo Meira Alves

Resumo

O presente trabalho tem por objetivo apresentar o desenvolvimento de um tradutor, em tempo imediato, utilizando Visão Computacional e “*Deep Learning*” (“Aprendizado profundo”) que converte o alfabeto manual da Libras em português brasileiro na modalidade escrita. Para isso, optamos pela rede neural PointNet visto a sua eficiência na classificação de nuvens de pontos 3D, utilizando como entrada um novo dataset de imagens da datilologia. A biblioteca MediaPipe extraiu pontos de referência da mão em cada imagem, representando articulações e extremidades dos dedos, que foram normalizados e armazenados em arquivos NumPy para treinamento. Após o processo, a rede foi treinada durante cinco horas, de modo que o sistema classificasse a letra correspondente à imagem da mão capturada em tempo instantâneo proporcionando uma tradução rápida e precisa. A fim de aprimorar a usabilidade, um corretor ortográfico foi incorporado via API corrigindo erros de digitação e tornando a tradução mais fluida e natural. Os resultados demonstram a eficácia da abordagem, com a rede PointNet, alcançando alta precisão na classificação das letras em Libras a partir dos pontos chaves em 3D. A criação do dataset e a integração com o corretor ortográfico contribuíram significativamente para a robustez e usabilidade do sistema.

Palavras-chave: LIBRAS. Rede Neural. Visão Computacional.

Introdução

O reconhecimento da Língua Brasileira de Sinais - Libras é extensa e complexa, e está inerentemente ligada ao cenário sociopolítico do Brasil. Com a chegada do primeiro professor de Língua de sinais ao país, ainda no Segundo Império e, conseqüentemente, a criação da primeira escola para surdos no Rio de Janeiro (1857), a língua de sinais brasileira passou a se desenvolver no final do século XIX. Nas décadas seguintes, a língua sinalizada mudou e evoluiu progressivamente em função de uma série de fatores, como práticas educacionais, pesquisas e, sobretudo, a difusão entre a população surda (Alves; Santos, 2019). Embora exista o reconhecimento legal, por meio da Lei 10.436/02, que reconhece oficialmente a Libras como meio



II SEMANA DE PEDAGOGIA

EDUCAÇÃO, PESQUISA E ENSINO:
CONSTRUINDO E (RE)CONSTRUINDO SABERES



CAMPUS DE
VITÓRIA DA CONQUISTA

19 A 23 DE AGOSTO DE 2024



legal de comunicação e expressão da comunidade surda no Brasil, assegurando o seu uso em diversas esferas sociais, educacionais e profissionais, ainda tem havido um esforço contínuo para que o idioma seja utilizado como um modo de comunicação igualitário - comparando-se com o português, língua majoritária do país (Streichchen et al., 2016).

A riqueza da Libras como língua de modalidade visual-gestual reside em sua estrutura fonológica única. A combinação precisa de configurações, locação e movimentos das mãos, como descrito por Karnopp (2024), permite a codificação de significados e informações gramaticais. A datilologia, por sua vez, surge como uma ferramenta essencial para expressar nomes próprios, termos técnicos e palavras que ainda não possuem sinais específicos, representando as letras do alfabeto através de configurações manuais.

O alfabeto manual em Libras, além de consistir em configurações de mão que representam as letras, também serve como base para a criação de novos sinais. Através da combinação de diferentes configurações de mão, movimentos, expressões faciais e pontos de articulação, novos sinais são formados, enriquecendo a linguagem visual-gestual. Explorar o alfabeto ajuda a compreender melhor a riqueza e a complexidade desta língua.

Neste contexto, torna-se essencial desenvolver ferramentas tecnológicas que promovam a inclusão social e a acessibilidade. Um interpretador da datilologia em Libras para a modalidade escrita do português brasileiro, baseado em Visão Computacional e Deep Learning, atua como um agente de conscientização sobre a importância da inclusão e do respeito à diversidade linguística.

Identificação Visual:

Em uma sociedade cada vez mais digital, a acessibilidade tecnológica é crucial para garantir a inclusão e a justiça social. Pessoas com deficiências enfrentam barreiras significativas na participação social, educacional e profissional quando tecnologias acessíveis não são priorizadas. A falta de atenção a essa questão perpetua a exclusão e limita o acesso a oportunidades. Ferramentas adaptativas são essenciais para derrubar essas barreiras, garantindo que todos, independentemente de suas habilidades, possam participar plenamente da sociedade da



II SEMANA DE PEDAGOGIA

EDUCAÇÃO, PESQUISA E ENSINO:
CONSTRUINDO E (RE)CONSTRUINDO SABERES



CAMPUS DE
VITÓRIA DA CONQUISTA

19 A 23 DE AGOSTO DE 2024



informação. Portanto, investir em acessibilidade tecnológica se trata de uma necessidade ética e social essencial para a construção de um futuro mais justo e igualitário.

Considerando a importância da acessibilidade por meio das tecnologias, e ao mesmo tempo a falta de um conjunto de dados específicos para a Língua Brasileira de Sinais (Libras), no que consiste em um tradutor do alfabeto em Libras para o alfabeto escrito do português, o presente estudo tomou como ponto de partida uma comparação entre a datilologia em Libras com a da American Sign Language (ASL), usando um conjunto de dados preexistentes da ASL. Descobriu-se que das 26 letras do alfabeto, 9 possuem configurações diferentes entre Libras e ASL, quais sejam: f, g, h, m, n, p, q, t, x. Assim, foram necessários dados adicionais para essas letras, completando a base de dados para a datilologia em Libras. Para o treinamento, foram usadas três bases de dados da ASL: "Synthetic ASL Alphabet Dataset", "ASL Alphabet Test Dataset" e "ASL Alphabet Dataset".

As configurações manuais para as letras com diferenças foram capturados em várias posições e ângulos, totalizando 2000 amostras por nova classe. A combinação das três bases da ASL com as novas configurações manuais resultou em 121.700 amostras, com variação de dados entre as letras.

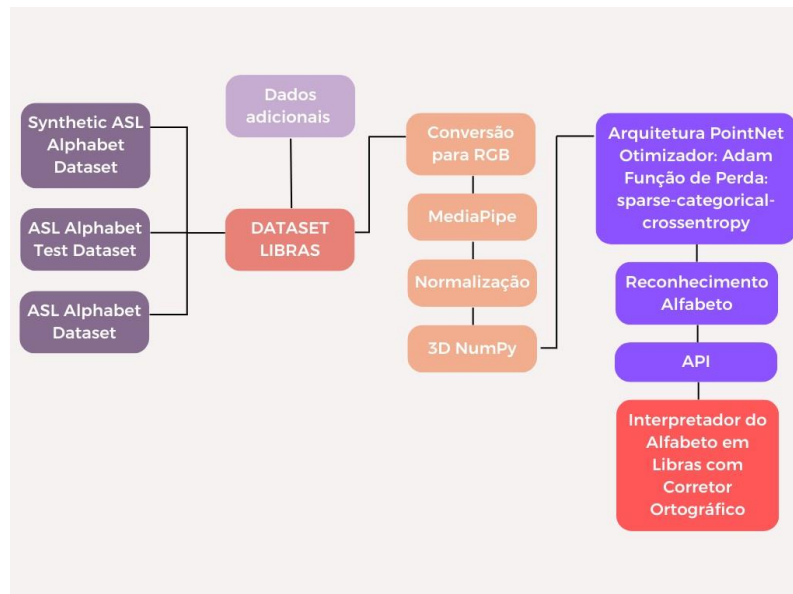
Para ensinar a inteligência artificial a reconhecer os gestos, primeiro foram extraídos imagens das mãos em 3D usando a tecnologia MediaPipe, ferramenta capaz de identificar componentes em imagens e vídeos. Para garantir a qualidade dos dados, foram selecionadas apenas as imagens que mostravam os 21 pontos-chave da mão. Essas informações foram transformadas em um formato que o computador entende, organizando os pontos-chave em um padrão específico. Para ensinar o sistema a reconhecer os diferentes gestos, foi utilizado a arquitetura PointNet, rede projetada para trabalhar com conjuntos de pontos 3D desordenados, conhecidos como nuvens de pontos (Griffiths, 2020). Assim, a cada vez que o modelo recebe a informação de 21 pontos-chave de uma mão, ela consegue identificar qual gesto está sendo feito.

Os dados foram divididos em treino (80%) e teste (20%). O modelo foi compilado com a função de perda "sparse-categorical-crossentropy" e o otimizador Adam, e treinado por 100 épocas, monitorando o progresso com TensorBoard. Este auxiliou no ajuste e na identificação de problemas

durante o treinamento. A acurácia média de validação foi de 98.00% e a perda média de validação foi de 4.29%, indicando um bom desempenho do modelo.

Para evitar erros de digitação, foi integrado o modelo de linguagem da OpenAI ao sistema, este funciona como um corretor automático inteligente, que analisa e corrige as frases digitadas. Sendo programada apenas para focar na correção de digitação, sem alterar o estilo ou o significado do texto original. A Figura 1 ilustra a forma como o sistema foi desenvolvido.

Figura 1: Treinamento do Modelo



Fonte: Próprio Autor

A aplicação combina tecnologias para criar um sistema interativo de interpretação de sinais. Utilizando a captura e processamento de vídeo em tempo real, a aplicação identifica sinais e transmite os resultados diretamente para o usuário por meio de um servidor web. Essa integração permite que o usuário visualize os sinais da datilologia de Libras conforme são reconhecidos, como demonstra a Figura 2, e garante que as transcrições sejam atualizadas instantaneamente, promovendo uma experiência dinâmica e interativa.

Figura 2: Interface do Interpretador



Fonte: Próprio Autor

Essa abordagem integrada de captura de vídeo, processamento em tempo real e comunicação contínua resulta em uma experiência fluida e responsiva. A combinação das tecnologias assegura que o sistema seja eficiente, capaz de gerenciar conexões e oferecer atualizações e interpretações instantâneas.

Considerações Finais

A realização deste trabalho envolveu etapas essenciais, desde a coleta e preparação dos dados até a implementação e treinamento do modelo PointNet para a interpretação de sinais do alfabeto em Libras. Cada fase trouxe desafios específicos que exigiram adaptações.

Iniciou-se com a coleta de dados, criando um novo dataset a partir de imagens de sinais do alfabeto em Libras e complementando um dataset existente encontrado online. A construção desse conjunto de dados foi um desafio significativo, pois foi necessário obter imagens adicionais para representar todas as letras e garantir a qualidade e consistência visual, bem como a anotação precisa dos sinais.

Seguiu-se a detecção de landmarks com o Google MediaPipe, cujos resultados foram armazenados como vetores Numpy 3D e utilizados para treinar o modelo PointNet, adaptado para a classificação de nuvem de pontos. Garantir a precisão na detecção e normalização dos landmarks para criar um dataset coerente foi outro desafio.

A integração de tecnologias, como MediaPipe, TensorFlow e Keras, também foi complexa. No entanto, os resultados foram satisfatórios, com uma perda média de validação de 0.0429 e uma



II SEMANA DE PEDAGOGIA

EDUCAÇÃO, PESQUISA E ENSINO:
CONSTRUINDO E (RE)CONSTRUINDO SABERES



CAMPUS DE
VITORIA DA CONQUISTA

19 A 23 DE AGOSTO DE 2024



acurácia de 98% durante a validação, indicando um bom desempenho na classificação dos sinais. A visualização através do TensorBoard permitiu monitorar e ajustar o treinamento em tempo real, assegurando a precisão e estabilidade do modelo final.

Os objetivos do projeto foram alcançados, confirmando a hipótese de que o modelo PointNet, aliado a técnicas de argumentação e regularização, pode interpretar sinais do alfabeto em Libras com precisão. A aplicação prática, que transmite resultados em tempo instantâneo via um servidor web, demonstrou a viabilidade e eficácia do sistema. O projeto não só atingiu seus objetivos, como também abriu oportunidades para futuras melhorias e aplicações em contextos mais amplos, como a interpretação de palavras e frases em Libras, promovendo a inclusão e acessibilidade para pessoas surdas.

Referências

ALVES, M. M.; SANTOS, W.J. (Des)caminhos da educação bilíngue para surdos no Brasil. In: MACEDO, Y. M.; MAIA, C. B. (Org.). **Educação Especial e Inclusiva: Didáticas, Práticas e Pedagogias em foco**. 1ª ed.. Linhares – Editora Oyá, 2019, 125p.

GRIFFITHS, D. **Point Cloud Classification with PointNet**. 2020. Date created: 2020/05/25, Last modified: 2024/01/09. Disponível em: <<https://keras.io/examples/vision/pointnet/>>.

KARNOPP, L. B. Língua de Sinais Brasileira: Estudos Lingüísticos. In: **Língua de Sinais Brasileira: Estudos Lingüísticos**. [S.l.]: Editora da Universidade Federal do Rio Grande do Sul, 2004. cap. 2.

LEI Nº 10.436, de 24 de abril de 2002. **Dispõe sobre a Língua Brasileira de Sinais – Libras e dá outras providências**. 2002. << http://www.planalto.gov.br/civil_03/LEIS/2002/L10436.htm >>.

LEXSET. **SYNTHETIC ASL ALPHABET**. Kaggle, 2021. Disponível em: << <https://www.kaggle.com/datasets/lexset/synthetic-asl-alphabet> >>.

RASBAND, D. **ASL Alphabet Test**. Kaggle, 2021. Disponível em: <https://www.kaggle.com/datasets/danrasband/asl-alphabet-test>.



II SEMANA DE PEDAGOGIA

EDUCAÇÃO, PESQUISA E ENSINO:
CONSTRUINDO E (RE)CONSTRUINDO SABERES



CAMPUS DE
VITÓRIA DA CONQUISTA

19 A 23 DE AGOSTO DE 2024



SAU, D. **ASL American Sign Language Alphabet Dataset**. Kaggle, 2021. Disponível em: <<
<https://www.kaggle.com/datasets/debashishsau/aslamerican-sign-language-aplphabet-dataset> >>.

STREIECHEN, M. et al. **Pedagogia surda e bilinguismo: pontos e contrapontos na perspectiva de uma educação inclusiva**. Editora da Universidade Estadual de Maringá (Eduem), v. 39, n. 1, p. 91–91, 2016.

Glossário

- API: Interface de Programação de Aplicações, comunicação entre diferentes sistemas ou softwares.
- Dataset: Conjunto de dados utilizado para treinar, validar e testar modelos de aprendizado de máquina.
- Deep Learning: Subcampo do aprendizado de máquina que utiliza redes neurais profundas para modelar e entender dados complexos.
- Keras: Biblioteca de alto nível para construção e treinamento de modelos de aprendizado de máquina.
- MediaPipe: Framework de aprendizado de máquina do Google utilizado para a detecção de landmarks e processamento de vídeo em tempo real.
- NumPy: Biblioteca do Python para computação científica, oferecendo suporte a arrays multidimensionais e operações matemáticas eficientes.
- OpenAI: Organização de pesquisa em inteligência artificial focada no desenvolvimento de tecnologias avançadas de IA.
- Otimizador Adam: Método de otimização usado no treinamento de redes neurais.
- PointNet: Arquitetura de rede neural projetada para processar nuvens de pontos tridimensionais (3D).
- sparse-categorical-crossentropy: Função de perda usada em problemas de classificação multiclasse,
- TensorBoard: Ferramenta de visualização fornecida pelo TensorFlow para monitoramento do treinamento de modelos de aprendizado de máquina.
- TensorFlow: Plataforma open-source de aprendizado de máquina desenvolvida pelo Google, utilizada para construir e treinar redes neurais.