

CLAUDIO FILIPI GONÇALVES DOS SANTOS

Aplicações de Deep Learning para a Língua Brasileira de Sinais (Libras)

Relatório de Pós-doutorado realizado na
Universidade Estadual Paulista (UNESP),
FACULDADE DE CIÊNCIAS DE BAURU,
BAURU - SP.

Supervisor(a): Prof. Dr. JOÃO PAULO
PAPA

BAURU - SP

2026

RESUMO

Este relatório apresenta uma síntese das atividades desenvolvidas durante o estágio de pós-doutorado, cujo tema foi **Aplicações de Deep Learning para a Língua Brasileira de Sinais (Libras)**.

O trabalho compreendeu atividades de pesquisa teórica e aplicada, com foco no reconhecimento contínuo de sinais, na tradução automática de vídeos em Libras para textos em língua portuguesa e no desenvolvimento de representações eficientes para modelos multimodais. Foram analisadas arquiteturas modernas de aprendizado profundo, incluindo Redes Neurais Convolucionais (CNNs), Vision Transformers, modelos baseados em estimativas de poses corporais e abordagens de aprendizado contrastivo.

As atividades desenvolvidas permitiram identificar avanços recentes, desafios científicos e limitações relacionadas ao processamento computacional da Libras. Como resultado, o projeto produziu contribuições voltadas à comunidade científica e ao desenvolvimento de tecnologias de acessibilidade e inclusão para pessoas surdas.

Palavras-chave: Língua Brasileira de Sinais; Deep Learning; Reconhecimento de Sinais; Tradução Automática; Acessibilidade.

1. INTRODUÇÃO

A comunicação entre pessoas surdas e ouvintes ainda enfrenta barreiras significativas em diferentes contextos sociais, educacionais e profissionais. Embora a Língua Brasileira de Sinais seja legalmente reconhecida como meio de comunicação e expressão, a disponibilidade de tecnologias capazes de interpretar e traduzir Libras automaticamente permanece limitada.

Nos últimos anos, modelos baseados em Deep Learning têm apresentado avanços relevantes, tornando possível o desenvolvimento de sistemas para o reconhecimento automático de sinais, a tradução de vídeos em Libras para textos

em língua portuguesa e a geração computacional de movimentos associados às línguas de sinais.

Durante o estágio de pós-doutorado, a investigação desses sistemas mostrou-se relevante para a construção de tecnologias mais acessíveis e inclusivas. O trabalho concentrou-se no estudo da Libras, considerando tanto seus aspectos linguísticos quanto os desafios computacionais relacionados à escassez de conjuntos de dados, à variabilidade entre sinalizantes, às diferenças regionais e à limitada disponibilidade de anotações realizadas por profissionais especializados.

A pesquisa também considerou a natureza multimodal das línguas de sinais, nas quais a comunicação depende da combinação de movimentos das mãos, expressões faciais, orientação corporal, localização espacial e dinâmica temporal. Essas características tornam o reconhecimento e a tradução automática de Libras tarefas complexas, que exigem modelos capazes de representar simultaneamente informações visuais, espaciais, temporais e linguísticas.

2. OBJETIVOS DA PESQUISA

O objetivo geral do projeto foi investigar arquiteturas e metodologias de Deep Learning aplicadas ao reconhecimento e à tradução automática da Língua Brasileira de Sinais.

Para alcançar esse objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- estudar arquiteturas profundas aplicadas ao reconhecimento e à tradução de Libras;
- analisar metodologias modernas baseadas em Transformers, Redes Neurais de Grafos — Graph Neural Networks (GNNs) —, modelos autorregressivos e técnicas de aprendizado autossupervisionado;
- investigar estratégias para reduzir a dependência de glossários e de anotações manuais intermediárias;
- identificar limitações presentes em bases públicas de línguas de sinais e propor direções para pesquisas futuras;

- analisar abordagens baseadas em imagens RGB, estimativas de pose e combinações multimodais;
- consolidar os resultados obtidos em um levantamento de métodos recentes aplicados à Libras e a outras línguas de sinais.

Esses objetivos foram definidos considerando a necessidade de compreender o estado atual das tecnologias de processamento de línguas de sinais e de identificar caminhos para o desenvolvimento de modelos mais robustos, generalizáveis e adequados às particularidades linguísticas e culturais da Libras.

3. MATERIAL E MÉTODOS

A metodologia adotada combinou levantamento bibliográfico, análise comparativa de arquiteturas de aprendizado profundo e avaliação qualitativa das principais abordagens utilizadas em reconhecimento e tradução de línguas de sinais.

Inicialmente, realizou-se um levantamento sistemático da literatura científica publicada entre os anos de 2020 e 2024. Foram analisados trabalhos publicados em periódicos científicos e em conferências relevantes das áreas de Visão Computacional, Processamento de Linguagem Natural e Inteligência Artificial, incluindo eventos como Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), Annual Meeting of the Association for Computational Linguistics (ACL) e AAAI Conference on Artificial Intelligence.

Os trabalhos selecionados foram organizados de acordo com a tarefa abordada, o tipo de representação visual utilizada, a arquitetura computacional, os conjuntos de dados empregados e as métricas de avaliação apresentadas.

As abordagens foram classificadas em três grupos principais:

1. métodos baseados diretamente em vídeos ou sequências de imagens RGB;
2. métodos baseados em estimativas de poses corporais, mãos e expressões faciais;

3. métodos híbridos ou multimodais, que combinam informações visuais, corporais, temporais e linguísticas.

Também foram analisados pipelines voltados ao reconhecimento contínuo de língua de sinais, conhecido como *Continuous Sign Language Recognition* (CSLR), e à tradução automática de línguas de sinais, denominada *Sign Language Translation* (SLT).

Além disso, foram examinados conjuntos de dados amplamente utilizados na literatura, como WLASL, PHOENIX-2014, Chinese Sign Language (CSL) e bases de menor escala desenvolvidas especificamente para Libras. A análise considerou aspectos como quantidade de participantes, número de sinais, diversidade de contextos, disponibilidade de anotações, modalidade de captura e condições de gravação.

4. RESULTADOS E ANÁLISES DA PESQUISA

Os resultados do levantamento demonstraram uma predominância crescente de arquiteturas baseadas em Transformers para o processamento de sequências de sinais. Essas arquiteturas têm sido utilizadas tanto na extração de características visuais quanto na modelagem temporal e na geração de textos em línguas orais.

Observou-se também uma tendência de substituição gradual de modelos que dependem de glosas por abordagens conhecidas como *gloss-free*. Nessas abordagens, o sistema busca realizar diretamente a tradução do vídeo para o texto, reduzindo a necessidade de uma representação intermediária composta por glosas.

Embora as glosas possam auxiliar no alinhamento entre sinais e palavras, sua utilização exige anotações especializadas, apresenta custos elevados e pode não representar integralmente a estrutura linguística das línguas de sinais. Por essa razão, métodos capazes de aprender diretamente a relação entre vídeos e textos vêm recebendo crescente atenção.

Os trabalhos analisados também indicaram o crescimento de técnicas multimodais. Os melhores resultados tendem a ser obtidos quando os modelos combinam diferentes fontes de informação, como imagens RGB, movimentos das mãos, poses corporais, expressões faciais e características linguísticas do texto de destino.

Outro resultado relevante refere-se à importância do alinhamento temporal explícito. Modelos capazes de identificar a correspondência entre segmentos do vídeo e unidades linguísticas apresentam, em geral, melhor desempenho em tarefas de reconhecimento contínuo e tradução automática.

Também foram observados avanços relacionados ao emprego de aprendizado contrastivo e de modelos pré-treinados. Essas estratégias permitem aproveitar grandes volumes de vídeos não anotados e reduzir parcialmente a dependência de conjuntos de dados extensivamente rotulados.

No contexto específico da Libras, foram identificadas limitações importantes nas bases de dados disponíveis. Entre os principais problemas encontram-se o número reduzido de sinalizantes, a baixa diversidade regional, a predominância de sinais isolados, a ausência de contextos naturais de comunicação e a escassez de traduções produzidas ou validadas por especialistas.

Os resultados obtidos foram consolidados em um artigo de revisão, no qual foram sistematizadas as tendências recentes, as principais arquiteturas, os conjuntos de dados utilizados e as oportunidades de pesquisa relacionadas ao reconhecimento e à tradução de línguas de sinais.

5. DISCUSSÃO

Apesar dos avanços observados, o reconhecimento e a tradução automática de Libras ainda apresentam desafios científicos e tecnológicos significativos.

Um dos principais obstáculos é a escassez de corpora amplos, diversificados e adequadamente anotados. A produção de bases de dados para línguas de sinais exige a participação de sinalizantes, intérpretes, pesquisadores e especialistas em linguística, tornando o processo de coleta e anotação complexo e dispendioso.

A variabilidade entre sinalizantes também representa um desafio relevante. Um mesmo sinal pode apresentar diferenças relacionadas à velocidade, amplitude, orientação das mãos, expressões faciais, estilo individual e variações regionais. Modelos treinados com um número reduzido de participantes podem apresentar bom desempenho no conjunto de dados original, mas baixa capacidade de generalização para novos usuários.

Outro problema identificado é a predominância de ambientes controlados de gravação. Muitos conjuntos de dados são produzidos com fundos uniformes, iluminação padronizada e posicionamento previamente definido. Essas condições não representam integralmente situações reais de uso, nas quais podem ocorrer variações de iluminação, oclusões, movimentação de câmera, diferentes ângulos de captura e interferências visuais.

Também se observou uma integração ainda limitada entre pesquisas linguísticas e computacionais. O desenvolvimento de modelos de Deep Learning para Libras não deve considerar apenas a identificação dos movimentos das mãos. Aspectos como expressões não manuais, uso do espaço, simultaneidade, referência e estrutura gramatical exercem papel central na construção de significado.

Futuras soluções devem priorizar métodos de aprendizado autossupervisionado, estratégias multimodais e modelos capazes de aprender com conjuntos de dados parcialmente anotados. Também se mostra necessária a padronização de protocolos de coleta, documentação e avaliação de bases voltadas à Libras.

A participação da comunidade surda em todas as etapas da pesquisa constitui outro aspecto fundamental. A definição das tarefas, a coleta dos dados, a validação linguística e a avaliação das tecnologias devem envolver pessoas surdas e profissionais com experiência em Libras, garantindo que as soluções desenvolvidas sejam linguística, cultural e socialmente adequadas.

6. ORIENTAÇÕES

Durante o estágio de pós-doutorado, foram realizadas atividades de orientação informal de estudantes e de apoio técnico a projetos relacionados à aplicação de

Visão Computacional, Inteligência Artificial e Aprendizado Profundo em iniciativas de acessibilidade digital.

As atividades envolveram discussões sobre definição de problemas de pesquisa, seleção de arquiteturas computacionais, organização de conjuntos de dados, treinamento de modelos e avaliação de resultados.

Também foram realizadas contribuições relacionadas à utilização de modelos de processamento de imagens e vídeos, com atenção especial às aplicações voltadas ao reconhecimento de gestos, movimentos corporais e línguas de sinais.

7. ATIVIDADES DE EXTENSÃO E DISCIPLINAS MINISTRADAS

Durante o período do estágio de pós-doutorado, foram realizadas apresentações, seminários e ações de divulgação científica sobre Inteligência Artificial, Deep Learning, acessibilidade e processamento computacional de línguas de sinais.

As atividades buscaram promover a disseminação do conhecimento científico e ampliar o debate sobre o uso responsável de tecnologias de Inteligência Artificial no desenvolvimento de soluções inclusivas.

Não foram ministradas disciplinas formais durante o período. Entretanto, houve participação em seminários técnicos, discussões acadêmicas e colaborações com estudantes, docentes e grupos de pesquisa relacionados à Visão Computacional, ao Aprendizado Profundo e à acessibilidade digital.