

INTELIGÊNCIA ARTIFICIAL · UECE · 31 DE MAIO DE 2026

Isolated Sign Language Recognition with Gloss-to-Text Smoothing for Assistive Translation

Por Luan Fernandes de França e Josã Everardo Bessa Maia

Fonte: OpenAlex · UECE · [Publicação original](#)

INOVAÇÃO

7/10

APLICABILIDADE

9/10

POTENCIAL
ECONÔMICO

7/10

MATURIDADE

Média

O artigo apresenta uma arquitetura modular de duas camadas para reconhecimento de sinais isolados da Libras, combinando visão computacional (I3D, EfficientNet) para extração de características e Modelos de Linguagem Grande (LLMs) para suavização de glosas em texto natural. A abordagem foca em blocos práticos para sistemas assistivos, alcançando alta precisão ($F1 > 0.97$) e discutindo abertamente limitações de datasets, com potencial escalável para outras línguas de sinais.

NÚCLEO DA ANÁLISE

Ideia de startup ou produto

Criação de uma API de 'Sinal-para-Texto' focada na modularidade proposta (Libras-first), oferecendo SDKs para integradores de EdTechs e plataformas de videoconferência adicionarem camadas de acessibilidade sem precisar treinar modelos do zero, utilizando a suavização de LLM para legibilidade.

Aplicações práticas

Desenvolvimento de sistemas de tradução assistiva em tempo real para reuniões corporativas, ferramentas educacionais para ensino de Libras, legendagem automática para conteúdo audiovisual e interfaces de acessibilidade para atendimento ao público.

Potencial de mercado

Alto, impulsionado por legislações de acessibilidade no Brasil e globalmente, crescimento de iniciativas de D&I nas empresas e demanda por tecnologias assistivas que cubram a lacuna de comunicação entre ouvintes e não ouvintes.

FUNDAMENTO CIENTÍFICO

Problema abordado

A barreira de comunicação enfrentada pela comunidade surda, especificamente a dificuldade de tradução técnica que converte sinais brutos (glosas) em texto fluido e gramaticalmente correto em português, além da escassez de ferramentas que lidem realisticamente com a segmentação de sinais.

Metodologia

Implementação de um pipeline em dois estágios: (1) Reconhecimento de sinais no nível de glosa usando extratores de características espaço-temporais (I3D, EfficientNet, MobileNet) com classificadores leves; (2) Suavização de linguagem natural via LLMs utilizando engenharia de prompt. Avaliação abrangente em datasets de Libras, incluindo testes com injeção de ruído para simular condições realistas.

Principais descobertas

Alta acurácia na classificação de sinais isolados ($F1 > 0.97$). A suavização via LLM demonstrou eficácia na conversão de glosas para texto natural mesmo sob ruído. O estudo revela limitações críticas nos datasets atuais (AUC quase perfeito indicativo de viés), oferecendo uma análise honesta necessária para evolução da área.

LEITURA PARA GESTÃO PÚBLICA

Implementação da arquitetura em portais de governo digital e balcões de atendimento (SAC) para automatizar o atendimento prévio a cidadãos surdos, bem como subsídio para políticas de inclusão digital em escolas públicas com ferramentas validadas academicamente.

PARCERIA UNIVERSIDADE × EMPRESA

Parceria estratégica entre a UECE e empresas de tecnologia de nuvem (AWS/Azure) ou EdTechs nacionais para refinar o modelo de suavização de LLM para o contexto linguístico brasileiro e viabilizar a deployabilidade da solução em escala.

OPORTUNIDADES DERIVADAS

4 direções estratégicas identificadas

STARTUP · IMPACTO ALTO · INTELIGÊNCIA ARTIFICIAL

LibrasFlow API

Startup focada em ofertar a pipeline de reconhecimento e suavização como serviço (SaaS) para desenvolvedores que criam apps de acessibilidade, cobrando por transação processada de vídeo para texto.

POLÍTICA PÚBLICA · IMPACTO ALTO · GOVTECH

Atendimento Acessível Ceará

Projeto piloto para integrar a tecnologia de tradução nos terminais de autoatendimento de órgãos públicos estaduais, facilitando o acesso de surdos aos serviços governamentais.

Pesquisa Aplicada em LLMs Inclusivos

Colaboração entre a UECE e um laboratório de IA para treinar LLMs menores e mais eficientes (Edge AI) focados especificamente na gramática de Libras, rodando em dispositivos móveis para uso offline.

Plugin de Acessibilidade Corporativa

Solução B2B para empresas adicionarem legendagem automática baseada em Libras em suas videoconferências e treinamentos internos, utilizando a arquitetura proposta para garantir fidelidade e fluidez.

ABSTRACT ORIGINAL

This work presents a modular two layer architecture for recognizing isolated sign language video and converting gloss sequences into fluent natural language text. The proposed pipeline comprises: (1) glosslevel sign recognition using spatiotemporal feature extraction (I3D, EfficientNet, MobileNet) with lightweight classifiers, and (2) natural language smoothing via large language models (LLMs) using prompt engineering. Unlike prior claims of continuous translation, this paper explicitly focuses on isolated signs as a practical building block for assistive systems where segmentation is either manually provided or handled by external modules. Comprehensive evaluations on Brazilian Sign Language (Libras) datasets demonstrate high accuracy in isolated sign classification ($F1 > 0.97$ for I3D-RGB + LR). A critical analysis of near perfect AUC scores reveals dataset limitations that are openly discussed. For gloss to text conversion, we evaluate LLM smoothing on both clean and noise injected gloss sequences and report BLEU scores under realistic conditions. The proposed architecture is scalable, modular, and adaptable to other sign languages, advancing accessibility for deaf and hard of hearing communities.

MATÉRIA PARA LEIGOS

Para leigos: Tradução assistiva de Libras para texto com inteligência artificial

O cenário atual

A tecnologia assistiva para surdos e deficientes auditivos é um campo em evolução. Embora existam promessas de tradução contínua de língua de sinais, existem desafios técnicos significativos. Sistemas práticos frequentemente necessitam de componentes confiáveis, como o reconhecimento de sinais isolados, antes de alcançarem uma tradução fluida e automática de conversas inteiras.

O que os pesquisadores fizeram

Os pesquisadores da Universidade Estadual do Ceará (UECE) desenvolveram uma arquitetura modular de duas camadas. O objetivo é reconhecer vídeos de sinais isolados e converter esses sinais (chamados de "gloss") em texto natural fluido. O trabalho foca explicitamente em sinais isolados como um bloco construtivo para sistemas assistivos, diferenciando-se de abordagens que prometem tradução contínua imediata sem dividir as tarefas.

Como funciona na prática

O sistema proposto funciona em um pipeline de duas etapas:

1. Reconhecimento de Sinais: O sistema utiliza técnicas de extração de características espaço-temporais (usando modelos como I3D, EfficientNet e MobileNet) para identificar o sinal específico realizado no vídeo. Isso é feito com classificadores computacionalmente leves.
2. Suavização de Texto: Uma vez identificado o sinal (gloss), Modelos de Linguagem Grande (LLMs) são utilizados com "engenharia de prompt" para transformar a sequência técnica de sinais em uma frase natural e

gramaticalmente correta em português.

Para o funcionamento deste módulo específico, a separação do vídeo em sinais individuais (segmentação) é feita manualmente ou por módulos externos, não sendo resolvida internamente pela arquitetura proposta.

Resultados e evidência

As avaliações foram conduzidas usando datasets da Língua Brasileira de Sinais (Libras). Os resultados mostraram alta precisão na classificação de sinais isolados, atingindo uma pontuação F1 superior a 0,97 ao combinar o modelo I3D-RGB com Regressão Logística.

O paper também apresenta uma análise crítica: pontuações AUC (Área Sob a Curva) quase perfeitas revelaram limitações nos próprios datasets utilizados, uma questão que os autores discutem abertamente. Para a conversão de gloss para texto, a eficiência foi medida através de pontuações BLEU, testando o sistema tanto em sequências limpas quanto em sequências com injeção de ruído para simular condições realistas.

Implicações práticas

A pesquisa avança na acessibilidade para comunidades surdas e com deficiência auditiva. A arquitetura criada é escalável, modular e pode ser adaptada para outras línguas de sinais além da Libras. Ela oferece uma base técnica para o desenvolvimento de ferramentas assistivas que necessitam traduzir sinais específicos para texto de forma compreensível.

Limitações e próximos passos

O estudo esclarece que se limita a sinais isolados, ou seja, não traduz frases contínuas de vídeo espontâneo automaticamente. A dependência de segmentação externa (manual ou automática) é uma limitação atual do método. Além disso, os autores apontam as limitações dos datasets disponíveis como um fator que impacta a avaliação de precisão dos modelos.

QUEM SÃO OS PESQUISADORES

Quem são os pesquisadores

Luan Fernandes de França e José Everardo Bessa Maia são os autores deste estudo, vinculados à Universidade Estadual do Ceará (UECE). O paper não fornece detalhes sobre a formação acadêmica específica, trajetória profissional ou papéis individuais distintos de cada autor na pesquisa além da autoria conjunta e da afiliação institucional.

PALAVRAS-CHAVE

Sign language · Smoothing · American Sign Language · Segmentation · Architecture · Gloss (optics) · Feature extraction · Natural language