

# Reconhecimento de sinais da Libras por Visão Computacional

Alexandre Fuzeti Bertipaglia<sup>1</sup>, Gilberto Fernandes Junior<sup>1</sup>

<sup>1</sup>Departamento de Computação – Universidade Estadual de Londrina (UEL)  
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

ale.bertipaglia@uel.br, gilfernandes@uel.br

**Abstract.** *Human beings are intrinsically social, and the main form of interaction with other individuals is through speech. Through it, we express needs, feelings and opinions. However, there are millions of people around the world who cannot use speech. To this end, languages represented by signs were created so they could express themselves. In Brazil, Brazilian Sign Language (Libras) is used by approximately 2.3 million people, but it still faces a significant barrier, as the majority of listeners do not understand it. The goal of this research is to convert sign language into speech with computer vision, image processing, and convolutional networks. This system will provide a more inclusive environment, facilitating communication between deaf and hearing people. To achieve this objective, current technologies will be investigated and analyzed, developing an innovative solution that recognizes and translates Libras signals from video to audio.*

**Resumo.** *Os seres humanos são intrinsecamente sociais, e a principal forma de interação com outros indivíduos é por meio da fala. Com ela, expressamos necessidades, sentimentos e opiniões. No entanto, há milhões de pessoas em todo o mundo que não conseguem usar a fala. Para isso, foram criadas linguagens representadas por sinais para elas conseguirem se expressar. No Brasil, a Língua Brasileira de Sinais (Libras) é usada por aproximadamente 2,3 milhões de pessoas, mas ainda enfrenta uma barreira significativa, pois a maioria dos ouvintes não a compreende. O objetivo dessa pesquisa é converter linguagem de sinais em fala utilizando um modelo de visão computacional, que emprega técnicas de processamento de imagens e aprendizado profundo. Este sistema proporcionará um ambiente mais inclusivo, facilitando a comunicação entre pessoas surdas e ouvintes. Para alcançar esse objetivo, serão investigadas e analisadas as tecnologias atuais, desenvolvendo uma solução eficiente que reconheça e traduza sinais de Libras de vídeo para áudio.*

## 1. Introdução

A comunicação é uma necessidade humana fundamental, mas para milhões de pessoas com deficiência auditiva, a barreira linguística é uma realidade diária que limita sua interação com o mundo. Dentre as mais de 240 linguagens de sinais diferentes [14], o Brasil utiliza a Língua Brasileira de Sinais (Libras), considerada como segunda língua oficial. No entanto, apesar de seu reconhecimento oficial, a Libras é compreendida por uma parcela muito pequena da população ouvinte, criando um abismo de comunicação entre essas comunidades.

Os esforços para desenvolver tecnologias que facilitem a comunicação entre surdos e ouvintes resultaram em métodos como luvas de dados e aplicativos de tradução.

Este estudo realiza uma revisão sobre os modelos existentes e propõe um sistema que utiliza visão computacional e redes neurais convolucionais para traduzir sinais de Libras em fala audível instantaneamente.

A abordagem proposta avaliará técnicas de processamento de imagem para capturar e interpretar gestos em tempo real, empregando algoritmos avançados, como CNN, 3D-CNN e RNN, que oferecem maior precisão e eficiência. Ao focar na tradução de Libras, o projeto visa criar uma ferramenta que não só melhore a acessibilidade, mas também promova a inclusão social, permitindo que mais pessoas se comuniquem sem barreiras.

O documento está estruturado dessa forma: a Seção 2 apresenta a fundamentação teórico-metodológica e o estado da arte. Em seguida, a Seção 3 detalha os objetivos do projeto. A Seção 4 e 5 descrevem os procedimentos metodológicos e as técnicas que serão utilizadas para atingir os objetivos estabelecidos, junto com os tempos previstos para eles. Por fim, a Seção 6 discute as contribuições e os resultados esperados do estudo.

## 2. Fundamentação Teórico-Metodológica e Estado da Arte

### 2.1. Língua Brasileira de Sinais (Libras)

A Libras é uma linguagem visual-espacial não falada que utiliza as mãos e expressões faciais como meio de comunicação [7]. Ela desempenha um papel fundamental na vida de pessoas surdas, permitindo-lhes interagir, expressar ideias, emoções, e participar plenamente na sociedade. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), aproximadamente 2,3 milhões de brasileiros usam Libras [13], sendo a maioria composta por pessoas com deficiência auditiva. A figura 1 mostra o alfabeto em Libras.

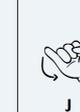
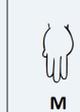
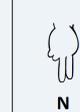
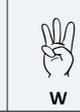
<b>ALFABETO MANUAL</b>				
Link para download do Alfabeto Manual em alta resolução : <a href="http://www.libras.com.br/alfabeto-manual">www.libras.com.br/alfabeto-manual</a>				
				
			<b>A</b>	<b>B</b>
				
<b>C</b>	<b>Ç</b>	<b>D</b>	<b>E</b>	<b>F</b>
				
<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>	<b>K</b>
				
<b>L</b>	<b>M</b>	<b>N</b>	<b>O</b>	<b>P</b>
				
<b>Q</b>	<b>R</b>	<b>S</b>	<b>T</b>	<b>U</b>
				
<b>V</b>	<b>W</b>	<b>X</b>	<b>Y</b>	<b>Z</b>

Figura 1. Alfabeto Manual [4]

Apesar de ser a segunda língua oficial do Brasil desde 2002 [1] [17], Libras e português são bem diferentes em termos de estrutura gramatical e sintaxe. Libras tem regras específicas para ordem, pronúncia e formação das palavras [20]. Por exemplo, a frase em português “A laranja é boa para a saúde por causa da vitamina C” se traduz em Libras como “laranja bom saúde vitamina C”, refletindo uma ordem gramatical que prioriza a essência do significado.

Além disso, assim como em qualquer idioma, Libras possui variações regionais e dialetos que refletem as diversidades culturais do Brasil. Diferentes estados ou regiões podem ter sinais distintos para a mesma palavra ou conceito, similar ao que ocorre com sotaques e gírias no português falado. Essas variações enriquecem a língua, mas também apresentam desafios para a comunicação e o reconhecimento automático dos sinais. Na figura 2 observamos diferenças no sinal “verde”.



**Figura 2. Diferenças regionais no sinal “verde” em Libras. Elaborado pelo autor.**

Libras não é apenas um meio de comunicação, mas também um componente central da identidade cultural da comunidade surda. Ela é um veículo de transmissão de valores, tradições e história, contribuindo para a coesão e o fortalecimento da comunidade surda no Brasil [21].

## **2.2. Processamento de Imagem e Visão Computacional**

O processamento de imagem é um campo da computação que envolve técnicas para analisar, modificar e obter informações úteis de imagens. Essas técnicas são amplamente utilizadas em diversas aplicações, como reconhecimento de padrões, análise de vídeo, diagnósticos médicos, entre outras. No contexto do reconhecimento de linguagem de sinais, o processamento de imagem é usado para extrair características visuais dos sinais capturados por câmeras, como forma, movimento e expressão facial. Métodos comuns incluem a filtragem de imagens, realce de bordas, detecção de contornos, e transformação de Fourier, que ajudam na identificação e segmentação dos componentes relevantes da imagem. [11]

Visão computacional é a área da inteligência artificial que permite o computador “enxergar”, extraindo informações relevantes de imagens de câmeras, scanners, sensores,

entre outros. Essas informações permitem reconhecer, manipular e lidar com os objetos presentes nas imagens [5]. Essa capacidade é essencial para o desenvolvimento de sistemas que necessitam de compreensão visual, como veículos autônomos, vigilância, e interfaces homem-máquina. [22]

Os principais componentes dos sistemas de Visão Computacional incluem:

- **Aquisição de Imagem:** O primeiro passo em um sistema de visão computacional, envolvendo a captura de imagens através de sensores de câmera. Essas imagens podem ser bidimensionais (2D), tridimensionais (3D), ou sequências temporais (vídeos).
- **Pré-processamento:** Antes de extrair informações de uma imagem, são aplicadas técnicas de pré-processamento para melhorar a qualidade e realçar características importantes. Isso pode incluir a remoção de ruído, ajuste de contraste, e a aplicação de filtros para destacar contornos e bordas.
- **Extração de características:** Envolve a identificação de elementos chave em uma imagem, como textura, forma, cor, e movimento.
- **Deteção e segmentação:** Este processo identifica e isola regiões ou objetos específicos dentro de uma imagem, segmentando áreas relevantes para um processamento mais detalhado. Isso é especialmente útil em aplicações de reconhecimento de sinais, onde é necessário distinguir mãos e gestos do fundo.
- **Processamento de alto nível:** Inclui tarefas avançadas como classificação de objetos, reconhecimento de padrões, e interpretação de cenas. Modelos de aprendizado profundo, como redes neurais convolucionais (CNNs), são frequentemente utilizados nesta etapa para melhorar a precisão do reconhecimento.

### **2.3. Aprendizado de Máquina**

Aprendizado de Máquina é um método que permite que os computadores aprendam com os dados e melhorem a si mesmos, gerando uma maior precisão nas previsões sem uma programação explícita. A estrutura do Aprendizado de Máquina consiste em receber dados de entrada, usar análises estatísticas para prever uma saída e ajustar as saídas conforme os dados são obtidos. [8]

O Aprendizado de Máquina busca por padrões não lineares entre parâmetros dependentes e independentes para prever resultados contínuos, chamados de regressão, ou prever classes exatas, conhecidas como classificação. A regressão é a previsão de valores numéricos constantes, enquanto a classificação é a previsão de classes. [8]

As técnicas de Aprendizado de Máquina geralmente são divididas em duas categorias, supervisionadas e não supervisionadas. O objetivo do aprendizado supervisionado é prever uma variável dependente(saída), a partir de um conjunto de variáveis independentes(entrada). Já o aprendizado não supervisionado é bem mais complexo, porque não tem uma variável dependente predefinida, e esses métodos focam na deteção de padrões entre todas as variáveis de interesse no conjunto de dados. [2]

#### **2.3.1. Redes Neurais**

Redes neurais são uma especialização de aprendizado de máquina que consistem em unidades de processamento de informações inspiradas no funcionamento do cérebro humano.

O cérebro é composto por uma rede de bilhões de células interconectadas chamadas neurônios. O estudo das redes neurais busca entender como essa vasta coleção de elementos interconectados pode realizar cálculos complexos, como o reconhecimento de padrões. [3]

Nas redes neurais, a informação é processada em paralelo, com todos os neurônios operando simultaneamente. Uma rede neural pode aprender e não precisa ser reprogramada. O reconhecimento de padrões busca fornecer uma resposta razoável para todas as possíveis entradas e realizar a correspondência "mais provável". [3]

### 2.3.2. Aprendizado Profundo

O aprendizado profundo é uma subcategoria das redes neurais que utiliza algoritmos de aprendizado de máquina para derivar conceitos de alto nível dos dados através de múltiplas transformações não lineares. O termo profundo refere-se ao uso de várias camadas na rede neural, cada uma extraindo diferentes características dos dados de entrada. [6]

O processo de aprendizado profundo possui duas etapas principais: o treinamento e a inferência. Durante a etapa de treinamento, grandes quantidades de dados são rotuladas, identificando características correspondentes. Já a etapa de inferência envolve rotular novos dados usando o conhecimento prévio adquirido durante o treinamento. [9]

Em redes neurais profundas, cada neurônio recebe múltiplos sinais como entrada, combinando-os linearmente com pesos e aplicando uma função não linear para produzir o resultado desejado. Uma das vantagens do aprendizado profundo é a capacidade de extrair automaticamente características relevantes dos dados, minimizando a necessidade de pré-processamento manual dos dados de entrada. [9]

Ao contrário de outros modelos de aprendizado de máquina, onde os dados de entrada precisam ser cuidadosamente pré-processados, o aprendizado profundo utiliza múltiplas camadas ocultas entre as camadas de entrada e saída. Cada camada amplifica características relevantes para a tomada de decisão e suprime as irrelevantes. As camadas são conectadas entre si, permitindo a extração automática de representações apropriadas para fins de classificação ou detecção. [9] A figura 3 demonstra a diferença do funcionamento de Aprendizado de Máquina e Aprendizado Profundo.



**Figura 3. Aprendizado de Máquina e Aprendizado Profundo. Elaborado pelo autor.**

### 2.3.3. Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs) se tornaram um modelo de aprendizado de alto desempenho para tarefas centrais em campos de visão computacional, como reconhecimento de alvos, classificação de imagens, segmentação de imagens e detecção de objetos. Por outro lado, a derivação e aplicação multifacetada das redes neurais convolucionais também estenderam a aplicação do aprendizado profundo para níveis e campos mais diversos [18].

Uma rede neural convolucional padrão geralmente consiste em uma camada de entrada, uma camada convolucional, uma camada de pooling, uma camada totalmente conectada e uma camada de saída.

1. **Camada Convolucional:** A camada convolucional é a camada central da rede neural convolucional [12]. Ela é composta por múltiplos núcleos de convolução (filtros) e funções de ativação, e seu objetivo é extrair informações de características profundas dos dados de entrada por meio de filtragem em várias camadas. O núcleo de convolução cobre os dados da imagem de entrada e desliza em um certo passo, realizando operações de convolução e ativação para gerar mapas de características. O tamanho do mapa de características é determinado por três hiperparâmetros: o tamanho do núcleo de convolução, o tamanho do passo e o preenchimento [16].
2. **Camada de pooling:** A camada de pooling é comumente inserida entre camadas convolucionais consecutivas. Após a extração de características na camada convolucional gerar um mapa de características, a camada de pooling (também conhecida como camada de subamostragem) realiza uma seleção de características para simplificar a complexidade dos dados. Essa camada tem como objetivo preservar as principais informações dos dados de entrada enquanto reduz a quantidade de parâmetros e a dimensão das características, contribuindo para uma melhor generalização do modelo. A operação de pooling se assemelha ao processo de convolução, mas a função de filtro na camada de pooling é ajustada para capturar apenas o valor máximo ou médio da posição correspondente na imagem. O pooling é frequentemente utilizado entre camadas convolucionais e o pooling médio é geralmente aplicado na última camada da rede convolucional. [25]
3. **Camada totalmente conectada:** Após a extração de características pela camada convolucional e pooling, os mapas de características resultantes mantêm informações posicionais, o que pode levar a valores de características semelhantes, mas com classificações distintas dependendo da localização. A camada totalmente conectada serve para combinar as características das camadas anteriores, transformando o mapa de características em um valor único e eliminando as informações de posição, aumentando a robustez do modelo. Cada camada da camada totalmente conectada é composta por múltiplos neurônios e é a camada com maior quantidade de parâmetros em uma CNN. Normalmente, adiciona-se uma ou duas camadas totalmente conectadas ao final da rede convolucional para melhorar a robustez e a capacidade de aprendizado do modelo. [19]

### 2.4. Trajetória do reconhecimento de sinais

As pesquisas no campo de reconhecimento de linguagem de sinais evoluíram significativamente ao longo dos anos e podem ser classificadas principalmente em dois métodos:

sistema baseado em luvas de dados e sistema baseado em visão computacional [15].

#### **2.4.1. Sistemas Baseados em Luvas de Dados**

O sistema baseado em luva de dados requer que o usuário use uma luva com módulos de controle e vários sensores [24]. Esses sensores capturam o ângulo das mãos, as trajetórias dos movimentos, postura, tempo e outras informações. As limitações consistem principalmente no custo dos equipamentos e na falta de praticidade de aplicação do dia-a-dia.

#### **2.4.2. Sistemas Baseados em Visão Computacional**

Esse método utiliza câmeras e algoritmos de processamento de imagens para capturar e interpretar dados de imagens e vídeos. Com o avanço do aprendizado de máquina e das redes neurais, os sistemas de visão computacional se tornaram mais eficientes e acessíveis. Eles podem ser divididos em dois tipos principais: reconhecimento de palavras isoladas e reconhecimento de expressões contínuas.

Quando se trata de palavras separadas, redes neurais convolucionais (CNNs) são utilizadas para analisar profundidade, esqueleto, pontos-chave de postura das mãos e esquemas de otimização para reconhecer bem palavras isoladas. Mesmo com uma boa capacidade de extração de dados, existe a limitação de informações temporais e espaciais dos dados de entrada.

Para lidar com as limitações dos métodos baseados em palavras isoladas, alguns métodos utilizam técnicas avançadas, são os modelos de reconhecimento de expressões contínuas. Alguns exemplos são:

- Modelos de redes neurais convolucionais tridimensionais (3D-CNN) [10]: Capturam informações espaciais e temporais simultaneamente, permitindo a análise de sequências de movimentos.
- Redes neurais recorrentes (RNN) e Memória de curto longo prazo (LSTM): São eficazes na modelagem de dependências temporais, capturando a dinâmica de gestos ao longo do tempo.
- Modelos híbridos: Combinam diferentes técnicas de redes neurais para melhorar a precisão e a eficiência do reconhecimento, como por exemplo o DeepConvLSTM(CNN e LSTM) [23].
- Algoritmos YOLO: Utilizados para detecção de objetos em tempo real, os algoritmos YOLO são conhecidos por sua estrutura simples e processamento rápido, sendo amplamente aplicados em sistemas de visão computacional.

O foco desta pesquisa é explorar os sistemas baseados em visão computacional para o reconhecimento de sinais de Libras, desenvolvendo um modelo que seja preciso, rápido, e prático para uso em tempo real.

### **3. Objetivos**

O objetivo deste projeto é desenvolver um sistema baseado em visão computacional e redes generativas convolucionais para reconhecer e traduzir os sinais de Libras em tempo real de vídeo para voz. Este sistema visa facilitar a comunicação e a inclusão social de

peessoas surdas ou com deficiência auditiva, promovendo um ambiente mais inclusivo. Para alcançar este objetivo principal, os seguintes objetivos específicos foram definidos:

1. Revisão bibliográfica: Estudar a fundamentação teórica, analisando a literatura existente sobre redes convolucionais e seu uso em reconhecimento de padrões, com foco particular em aplicações no reconhecimento de linguagens de sinais.
2. Seleção do Conjunto de Dados: Identificar e selecionar bases de dados de Libras representativas e diversas, considerando variações regionais e contextuais para garantir a cobertura abrangente dos sinais de Libras.
3. Pré-processamento de dados: Se necessário, implementar técnicas para normalizar, limpar e segmentar os dados de imagem e vídeo, preparando-os para o treinamento do modelo e garantindo a consistência e qualidade dos dados.
4. Desenvolvimento e Treinamento da Rede Neural: Projetar e treinar uma rede neural convolucional (CNN) eficaz, explorando arquiteturas e hiperparâmetros para otimizar o desempenho no reconhecimento de sinais de Libras em tempo real.
5. Testes e avaliações: Comparar com outros modelos com verificação de dados como acurácia, precisão, e demais métricas de desempenho que serão estudadas no decorrer do trabalho.

#### **4. Procedimentos metodológicos/Métodos e técnicas**

Primeiro, será conduzida uma revisão abrangente da literatura. A revisão permitirá identificar metodologias eficazes, fornecendo a fundamentação teórica necessária para o desenvolvimento do modelo e aplicação no reconhecimento de sinais.

O próximo passo é escolher um conjunto de dados rico, que será a base para o treinamento e teste do modelo de reconhecimento de sinais de Libras. Os conjuntos de dados serão avaliados quanto à diversidade de sinais, qualidade de imagem e quantidade de dados disponíveis para garantir que sejam abrangentes e representativos.

Após a seleção dos dados, será feita uma preparação dos dados para o treinamento do modelo. As etapas podem variar dependendo da necessidade, e podem incluir normalização para ajustar as escalas de imagem, limpeza para remover ruídos e dados irrelevantes, e segmentação para isolar as partes significativas das imagens que contêm sinais, além da separação de dados de treinamento e validação.

Depois, será desenvolvido o modelo de reconhecimento de sinais centrado em redes neurais convolucionais (CNNs), a arquitetura da rede neural será baseada nos modelos estudados anteriormente, explorando diferentes camadas, funções de ativação, e configurações de hiperparâmetros, como a taxa de aprendizado e o número de filtros nas camadas convolucionais.

Uma vez que o modelo esteja treinado, ele será testado para avaliar seu desempenho. O conjunto de dados de validação, separado durante a fase de pré-processamento, será utilizado para testar o modelo, medindo métricas de desempenho como acurácia, precisão, recall e F1-score. O modelo desenvolvido será comparado com outros modelos e métodos existentes na literatura.

Por fim, a implementação final do sistema será documentada, abordando todas as etapas do desenvolvimento, desde a revisão da literatura até a implementação prática do modelo no TCC.

## 5. Cronograma de Execução

Para que os objetivos sejam alcançados, as atividades serão divididas da seguinte maneira:

Atividades:

1. Revisão bibliográfica;
2. Seleção do Conjunto de Dados;
3. Pré-processamento de dados;
4. Desenvolvimento e Treinamento da Rede Neural;
5. Testes e avaliações;
6. Escrita do TCC;

**Tabela 1. Cronograma de Execução**

	jul	ago	set	out	nov	dez	jan	fev
1	•	•	•					
2	•	•						
3		•	•					
4			•	•	•			
5					•	•	•	
6					•	•	•	•

## 6. Contribuições e/ou Resultados esperados

Este trabalho espera contribuir significativamente para o avanço no campo do reconhecimento de linguagem de sinais, especialmente no contexto da Língua Brasileira de Sinais (Libras). Ao desenvolver um modelo que utiliza redes neurais convolucionais para a tradução em tempo real de sinais de Libras em fala, o projeto visa aumentar a precisão e a eficiência em comparação com as soluções atuais. Este avanço tecnológico tem o potencial de reduzir as barreiras de comunicação enfrentadas por pessoas surdas ou com deficiência auditiva, promovendo um ambiente mais inclusivo.

Além das contribuições técnicas, espera-se que o sistema desenvolvido sirva como uma ferramenta prática e acessível que possa ser utilizada em diversos contextos sociais, educacionais e profissionais. A capacidade de facilitar a comunicação direta entre surdos e ouvintes pode impactar positivamente a integração social, melhorando a qualidade de vida e ampliando as oportunidades de interação para a comunidade surda.

## 7. Espaço para assinaturas

Londrina, 29 de julho de 2024.

---

Aluno

---

Orientador

## Referências

- [1] Elizabete Gonçalves Alves and Silvana Soriano Frassetto. Libras e o desenvolvimento de pessoas surdas. *Aletheia*, pages 211 – 221, 04 2015.
- [2] Trent D. Buskirk, Antje Kirchner, Adam Eck, and Curtis S. Signorino. An introduction to machine learning methods for survey researchers. *Survey Practice*, 11(1), 1 2018.
- [3] Soni Chaturvedi, Rutika N. Titre, and Neha Sondhiya. Review of handwritten pattern recognition of digits and special characters using feed forward neural network and izhikevich neural model. In *2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies*, pages 425–428, 2014.
- [4] Almir Cristiano. Alfabeto manual, 2020. Disponível em: <https://www.libras.com.br/alfabeto-manual>. Acesso em: 28 jul. 2024.
- [5] Luciano Barrozo Honorato Danilo de Milano. Visão computacional. *UNICAMP – Universidade Estadual de Campinas, FT – Faculdade de Tecnologia*, 2010. Focus.
- [6] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari, and Gulshan Kumar. A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27(4):1071–1092, sep 2020.
- [7] Fábio Denilson De Oliveira Feliciano, Amanda De Andrade Briano, Ricardo Bastos Cavalcante Prudencio, and Erison Cavalcante Alves. Recognition of static or dynamic libras words in complex background environments. In *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–4, 2023.
- [8] Mercy Dol and Angelina Geetha. A learning transition from machine learning to deep learning: A survey. In *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, pages 89–94, 2021.
- [9] Mercy Dol and Angelina Geetha. A learning transition from machine learning to deep learning: A survey. In *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, pages 89–94, 2021.
- [10] Rob Fergus Lorenzo Torresani Manohar Paluri Du Tran, Lubomir Bourdev. Learning spatiotemporal features with 3d convolutional networks. *arXiv preprint arXiv:1412.0767*, 2015. Focus.
- [11] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Pearson Education, Upper Saddle River, NJ, USA, 3rd edition, 2008.
- [12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*, volume 1. MIT Press, Cambridge, 2016.
- [13] Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico 2010: Características gerais da população, religião e pessoas com deficiência, IBGE, 2010. Disponível em: <https://censo2010.ibge.gov.br>. Acesso em: 26 jul. 2024.
- [14] R. Kiruba, M. Saranya, V. Radhika, R. Kiruthika, and M. Mubashira. A data driven approach for computer vision based sign language model using convolutional neural network. In *2023 Third International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, pages 372–375, Los Alamitos, CA, USA, sep 2023. IEEE Computer Society.

- [15] Wanbo Li, Hang Pu, and Ruijuan Wang. Sign language recognition based on computer vision. In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 919–922, 2021.
- [16] Yujian Li. *Deep learning: Convolutional neural network from entry to proficiency*. Machinery Industry Press, Beijing, 2019.
- [17] Tamiris Locatelli. LIBRAS: aspectos, desafios e possibilidades proporcionadas pela tecnologia. *REVISTA CIENTÍFICA MULTIDISCIPLINAR*, aug 2018.
- [18] Li Mengyi and Zhu Dingju. Overview of image semantic segmentation methods based on full convolutional networks [j]. *Computer System Applications*, 30(09):41–52, 2021.
- [19] Jin Qiu, Jian Liu, and Yunyi Shen. Computer vision technology based on deep learning. In *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 2, pages 1126–1130, 2021.
- [20] Jampierre Rocha, Jeniffer Lensk, Taís Ferreira, and Marcelo Ferreira. Towards a tool to translate brazilian sign language (libras) to brazilian portuguese and improve communication with deaf. In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–4, 2020.
- [21] A. P. de M. Santana. *Libras e cultura surda: Um estudo sobre a identidade cultural surda*. Editora CRV, Curitiba, Brasil, 2017.
- [22] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, London, UK, 2010.
- [23] So Xue Thong, Eng Lip Tan, and Ching Pang Goh. Sign language to text translation with computer vision: Bridging the communication gap. In *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, pages 215–219, 2024.
- [24] Xuebin Xu, Kan Meng, Chen Chen, and Longbin Lu. Isolated word sign language recognition based on improved skresnet-tcn network. *Journal of Sensors*, 2023:1–10, 07 2023.
- [25] Xiangfu Zhang, Jian Liu, Zhangsong Shi, Zhonghong Wu, and Zhi Wang. Summarization of research on semantic segmentation based on deep learning. *Progress in Laser and Optoelectronics*, 56(15):20–34, 2019.