

# Processo para Reconhecimento e Tradução de Sinais em LIBRAS Utilizando Redes Neurais Artificiais

João Pedro C. Sobrinho  
Centro Federal de Educação Tecnológica de Minas Gerais  
Leopoldina, Minas Gerais  
lejoaoconte@gmail.com

Gabriella C. B. C. Dalpra  
Centro Federal de Educação Tecnológica de Minas Gerais  
Leopoldina, Minas Gerais  
gabriella@cefetmg.br

Lucas O. Pacheco H. da Silva  
Centro Federal de Educação Tecnológica de Minas Gerais  
Leopoldina, Minas Gerais  
lucaspac97@gmail.com

Samuel C. A. Basilio  
Centro Federal de Educação Tecnológica de Minas Gerais  
Leopoldina, Minas Gerais  
samuel@cefetmg.br

## ABSTRACT

Recognized by law, the Brazilian Sign Language (LIBRAS), is the second Brazilian official language and, according to IBGE (Brazilian Institute of Geography and Statistics), Brazil has a large community of hearing-impaired people, with approximately nine million of deaf people. Besides that, most of the non-deaf community cannot communicate or understand this language. Considering that, the use of LIBRAS' interpreters becomes extremely necessary in order to allow a greater inclusion of people with this type of disability with the whole community. However, an alternative solution to this problem would be to use artificial neural network methods for the LIBRAS recognition and translation. In this work, a process of LIBRAS' recognition and translation is presented, using videos as input and a convolutional-recurrent neural network, known as ConvLSTM. This type of neural network receives the sequence of frames from the videos and analyzes, frame by frame, if the frame belongs to the video and if the video belongs to a specific class. This analysis is done in two steps: first, the image is analyzed in the convolutional layer of the network and, after that, it is sent to the network recurrent layer. In the current version of the implemented network, data collection has already been carried out, the convolutional-recurrent neural network has been trained and it is possible to recognize when a given LIBRAS' video represents or not a specific sentence in this language.

## KEYWORDS

Redes Neurais, Língua de Sinais, Visão Computacional

## 1 INTRODUÇÃO

A Língua Brasileira de Sinais (LIBRAS) é uma língua gestual-visual que se diferencia da Língua Portuguesa por utilizar como canal de comunicação movimentos gestuais e expressões faciais percebidos pela visão e, como qualquer outra língua falada, contém sua própria estrutura linguística e conta com um processo gradual de mudanças e variações [1]. Além de ser reconhecida por lei, a LIBRAS é a segunda língua oficial do Brasil.

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), o país possui cerca de nove milhões de surdos, sendo que, desses, dois milhões têm surdez severa [2]. Apesar disto, esta língua ainda é pouco conhecida e utilizada pela comunidade não surda.

A maior parte das pessoas que têm deficiência auditiva não possuem dificuldades linguísticas, ou seja, apesar de não poderem ouvir, têm capacidade de produzir pensamentos linguísticos. Com isso, o surgimento das línguas de sinais é natural e tende acontecer em qualquer lugar onde há pessoas surdas [3].

Considerando que é por meio da linguagem e da língua que ocorrem os processos de socialização, este trabalho tem como objetivo *propor um processo para facilitar a comunicação entre os utilizadores de LIBRAS e os que não possuem um amplo conhecimento desta linguagem através da utilização de redes neurais artificiais*, de forma a permitir uma maior aproximação entre a comunidade surda que utiliza LIBRAS e a comunidade que não compreende esta língua.

Diversos aplicativos para a tradução da Língua Portuguesa (oral e escrita) para LIBRAS estão disponíveis no mercado, tais como o HandTalk<sup>1</sup> e o VLibras<sup>2</sup>. Porém, não foram localizadas soluções que, através do uso de reconhecimento de imagens e aplicação de redes neurais artificiais fossem capazes de realizar a tradução de LIBRAS para o Português.

Este trabalho está organizado da seguinte forma: na próxima seção é apresentada a solução proposta. Na Seção 3 são descritos os resultados preliminares e, por fim, a Seção 4 traz as considerações finais.

## 2 SOLUÇÃO PROPOSTA

A solução proposta é dividida em duas etapas: Levantamento de Dados e Reconhecimento e Tradução

### 2.1 Levantamento dos Dados

Uma das etapas para o desenvolvimento de redes neurais artificiais consiste na coleta de dados relativos ao problema e a sua separação em um conjunto de treinamento e um conjunto de testes. Para o processo proposto, o conjunto de dados a serem utilizados consiste em diferentes vídeos, feitos por diferentes pessoas, falando em LIBRAS.

### 2.2 Reconhecimento e Tradução

O processo de reconhecimento e tradução utiliza dois tipos de redes neurais diferentes: as redes neurais convolucionais e as redes

<sup>1</sup><https://www.handtalk.me/>

<sup>2</sup><http://www.vlibras.gov.br/>

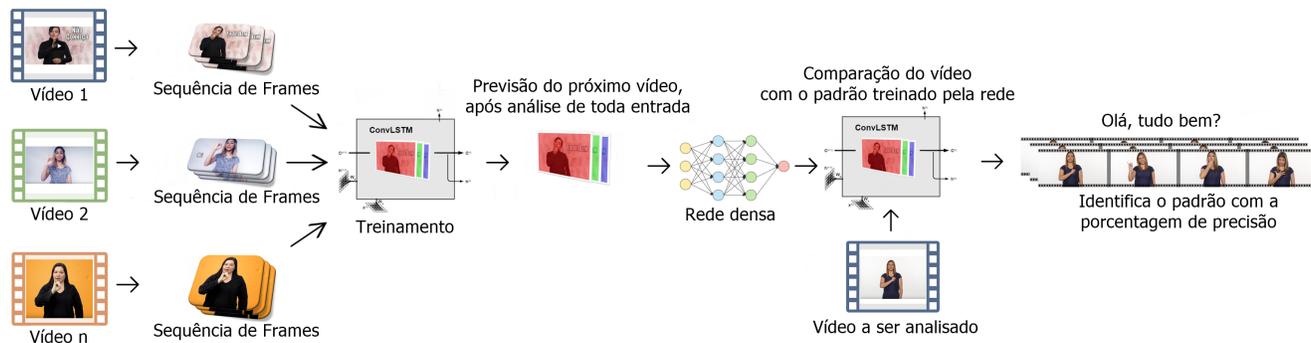


Figure 1: Processo de Reconhecimento e Tradução de Sinais em LIBRAS

neurais recorrentes, que resultaram em um tipo de rede chamado de ConvLSTM, ou *Convolutional Long short-term memory*.

Rede Neurais Convolucionais, do inglês, *Convolutional Neural Networks* (CNNs) são redes neurais muito usadas no reconhecimento de imagens, sendo assim uma aplicação prática muito ampla na área da computação [4, 5]. Por outro lado, uma Rede Neural Recorrente (ou RNN, do inglês *Recurrent Neural Network*) é um tipo de rede neural que produz uma saída a cada interação e que pode ser usada para ser passada para a célula anterior da rede [5]. Portanto, uma rede ConvLSTM tem como entrada uma sequência de imagens e vai analisando uma a uma através da rede convolucional. Após essa análise, a saída é passada para a rede recorrente, que entende essa sequência da série temporal de imagens [6].

A Figura 1 apresenta o processo de reconhecimento e tradução de LIBRAS. Inicialmente, os vídeos são divididos em uma sequência de *frames*, ou seja, cada vídeo é quebrado em várias imagens. Assim, o pacote de vídeos passa ser um pacote de pacotes de imagens. Essas imagens produzem uma matriz de cinco dimensões a ser utilizada como entrada da rede ConvLSTM.

Como saída, a rede ConvLSTM retorna somente um pacote de vídeos que, junto com seus respectivos rótulos, é passado uma rede densa, responsável pela etapa de previsão tradicional de uma rede neural.

De forma simples e sucinta, redes neurais densas ou multicamadas são caracterizadas pelo uso de pelo menos uma camada intermediária de neurônios que se encontram entre a camada de entrada e de saída. São utilizadas para a classificação de padrões, pois retornam como saída se determinado dado pertence ou não ao pacote de classificação [7].

Com isso podemos gerar um modelo e com isso realizar, de maneira inversa, todo processo. O vídeo de entrada é introduzido na rede, da mesma forma, e a mesma nos diz se ele pertence aquele pacote que foi treinado, de uma forma binária.

### 3 RESULTADOS

Para validar o processo proposto, este foi implementado e testado para uma frase específica expressa em LIBRAS ("Olá, tudo bem?"). A etapa de levantamento de dados foi realizada através da gravação de vídeos da frase anterior com quarenta voluntários. Após o tratamento destes com um gerador de dados, estes foram triplicados e o

resultado foi dividida em 20% para testes e 80% para treinamento da rede neural.

Como resultado foi obtido um modelo com um viés alto, o que causou um *underfitting*, ou seja, a rede classificava qualquer vídeo com movimentos parecidos à frase proposta como sendo a mesma, ao invés de classificar somente os vídeos que continham a frase analisada.

Visando solucionar o problema anterior, aumentou-se a quantidade de vídeos de entrada da rede e estes foram colocados em escala de cinza. Após isso, os resultados melhoraram e o modelo ficou mais próximo ao esperado. A precisão encontrada para os dados de teste foi de 70%. Porém, cabe ressaltar que esse valor ainda é baixo para os trabalhos que são desejados no futuro.

### 4 CONSIDERAÇÕES FINAIS

Para que a rede neural utilizada no processo proposto traga resultados satisfatórios, constatou-se que a quantidade de dados de entrada deve ser grande e devem ser utilizados diferentes tipos de vídeos para que a rede possa entender e classificar as frases analisadas.

Como trabalhos futuros, além do levantamento de uma maior quantidade de vídeos para a entrada da rede, os parâmetros da mesma serão cuidadosamente analisados para obter melhores resultados.

Por fim, tem-se como proposta a análise e avaliação de uma nova configuração de rede neural do tipo convolucional de três dimensões. Esse tipo de rede também é usado para reconhecimento de atividades humanas e os resultados obtidos com a mesma serão comparados com a rede utilizada atualmente no trabalho.

### ACKNOWLEDGMENTS

Os autores agradecem o apoio do CEFET-MG e da FAPEMIG durante o desenvolvimento deste trabalho, através da concessão de bolsa de Iniciação Científica.

### REFERENCES

- [1] R. M. e Leite T. A. Stumpf, M. e Quadros. *Estudo da Língua Brasileira de Sinais*. Insular, 2014. ISBN 978-8574747248.
- [2] IBGE. Censo de 2010. URL <https://censo2010.ibge.gov.br/>.
- [3] A. Baker, B. van den Bogaerde, R. Pfau, and T. Schermer. *The Linguistics of Sign Languages: An introduction*. John Benjamins Publishing Company, 2016. ISBN 9789027267344. URL <https://books.google.com.br/books?id=IECEDAAAQBAJ>.

- [4] A. Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2017. ISBN 9781491962268. URL <https://books.google.com.br/books?id=khpYDgAAQBAJ>.
- [5] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2016. ISBN 9780262035613. URL <https://books.google.com.br/books?id=Np9SDQAAQBAJ>.
- [6] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional lstm network: A machine learning approach for precipitation nowcasting. pages 802–810, 2015. URL <http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting.pdf>.
- [7] I.N. DA SILVA, D.H. SPATTI, and R.A. FLAUZINO. *REDES NEURAIS ARTIFICIAIS PARA ENGENHARIA E CIENCIAS APLICADAS - CURSO PRATICO*. ARTLIBER. ISBN 9788588098534. URL <https://books.google.com.br/books?id=w2VHbwAACA AJ>.