EXTRAÇÃO E TRADUÇÃO DE TEXTO EM IMAGENS HISTÓRICAS UTILIZANDO VISÃO COMPUTACIONAL

Patrick Elias do Nascimento[1], Marlon Oliveira[2]

Resumo: A crescente interconexão da sociedade pela internet tem proporcionado um acesso cada vez maior a imagens em diferentes idiomas. Dessa forma, a tradução das imagens torna-se essencial para que aqueles que não dominam o idioma original possam compreendê-las. O objetivo desta pesquisa consiste em extrair os textos das imagens com visão computacional e posteriormente traduzi-los. Foram selecionadas três opções diferentes para a extração de textos, sendo elas, *pytesseract*, *Keras OCR* e *Google Cloud Vision API*, e o *Google Translate API* para a tradução dos mesmos. A melhor opção para a extração de textos foi o Google Cloud Vision, que demonstrou uma precisão de 92,5%.

Palavras-chave: Visão Computacional; Reconhecimento Óptico de Caracteres; Extração de textos; Tradução;

ABSTRACT: The increasing interconnection of society through the internet has led to greater access to images in different languages. Thus, the translation of images becomes essential for those who do not understand the original language to comprehend them. The objective of this research is to extract texts from images with computer vision and subsequently translate them. Three different options were selected for text extraction, namely *pytesseract*, *Keras OCR* and *Google Cloud Vision API*, and *Google Translate API* for their translation. The best option for text extraction was Google Cloud Vision, which demonstrated an accuracy of 92.5%.

Keywords: Computer vision; Optical Character Recognition; Text extraction; Translation

[1]patrickelias1107@unesc.net
[2]marlon.oliveira@unesc.net

1 INTRODUÇÃO

A sociedade em rede, através da mídia e da internet, é o resultado de transformações econômicas, tecnológicas, sociais e culturais que abrangem todo o planeta, fenômenos esses chamados genericamente de globalização. Com a interação entre povos surge a necessidade de entender ou traduzir outros idiomas, a tradução é uma comunicação que repassa conhecimentos entre línguas distintas, facilitando a compreensão do leitor, tendo em vista que o conhecimento de outras línguas chega a ser bastante complexo (Oliveira, 2019).

Na computação a tradução automática surgiu na década de 40 e continua como uma tarefa difícil de ser implementada e aperfeiçoada, devido à necessidade de grande quantidade de conhecimento humano e de mundo para uma tradução perfeita (Gomes; Pardo, 2007). No Brasil, as pesquisas envolvendo tradução automática são ainda modestas. Um dos poucos grupos de pesquisa no Brasil que trabalham nessa área, senão o único, é o Núcleo Interinstitucional de Linguística Computacional (Gomes; Pardo, 2007).

Com o advento da tecnologia, discussões foram criadas sobre o uso da inteligência artificial para a tradução, segundo (Gomes, 2010) a inteligência artificial é um ramo da Ciência da Computação cujo interesse é fazer com que os computadores pensem ou se comportem de forma inteligente. Por ser um tópico muito amplo, inteligência artificial também está relacionada com psicologia, biologia, lógica matemática, linguística, engenharia, filosofia, entre outras áreas científicas.

A visão computacional é um dos subcampos da inteligência artificial, usada em diversos campos atualmente, como reconhecimento de pessoas, de padrões e também no reconhecimento de caracteres. Entre as técnicas que podem ser utilizadas pela visão computacional está o reconhecimento óptico de caracteres (OCR), tecnologia que permite reconhecer caracteres de texto em imagens, transformando-os em texto editável (Mendonça, 2008).

Em vista do cenário apresentado, esse trabalho propõe como objetivo geral extrair os textos das imagens utilizando recursos de visão computacional e traduzi-los para a Língua Portuguesa. Os objetivos específicos consistem em compreender a área de visão computacional; aplicar o reconhecimento óptico de caracteres as imagens; elaborar um *script* capaz de traduzir os textos extraídos das imagens.

Serão apresentados três trabalhos correlatos na Seção 2 que reforçam a relevância de estudos na área de visão computacional. Na metodologia, apresentada na Seção 3, serão testadas três ferramentas diferentes para a extração de textos a fim de quantificar qual extraiu mais textos de forma correta e sem perda de sentido. Após a escolha da melhor ferramenta será utilizado o *Google Translate API* para a tradução dos textos. Na Seção 4 será apresentado os resultados obtidos nesta pesquisa, seguido na Seção 5 será dada uma conclusão.

2 TRABALHOS CORRELATOS

O presente estudo se fundamenta em uma série de pesquisas correlatas que abordam temas similares ou que contribuem para a compreensão do problema em análise. A revisão abrange tanto o cenário nacional quanto o internacional, visando construir uma base teórica sólida.

Bazoni (2022) apresentou um projeto de conclusão de curso no Instituto Federal do Espírito Santo com o objetivo de traduzir textos em imagens de histórias em quadrinhos e gerar novas imagens aplicando o texto traduzido e mantendo o máximo da formatação possível.

Mueller-Gastell, Sena e Tan (2020) desenvolveu um trabalho para a Universidade de Stanford que estuda o reconhecimento de caracteres numéricos escritos a mão em documentos de censos históricos. O objetivo final era implantar um produto que tenha duas funções distintas, a primeira é uma etapa de pré-processamento adaptada ao conjunto de dados que se deseja digitalizar e a segunda é uma arquitetura de deep learning para digitalizar manuscritos com vários caracteres.

Bantupalli e Xie (2018) apresentou um trabalho que teve como principal objetivo desenvolver um aplicativo baseado em visão computacional que oferece tradução da linguagem de sinais para texto. Eles utilizaram dois modelos para classificar os gestos, a primeira foi utilizando previsões da camada *Softmax* que atingiu uma precisão de 93% com 100 gestos, a segunda foi utilizando a saída da camada pool global, que atingiu apenas 58% de precisão com os mesmos 100 gestos.

3 MATERIAIS E MÉTODOS

O presente estudo é uma pesquisa aplicada com o objetivo de solucionar problemas práticos relacionados à tradução de imagens históricas. A aplicação é composta por três partes, sendo elas pré-processamento de imagem, extração de textos de imagens e tradução dos textos extraídos

da imagem. Para a extração dos textos foram testados três ferramentas diferentes, sendo elas: *pytesseract*, *Keras OCR* e *Google Cloud Vision API*.

Os testes foram realizados utilizando o *Google Colaboratory*, em um ambiente com 12,7gb de memória RAM e 107.7gb de armazenamento. Não foi utilizado nenhuma das opções de aceleração de hardware.

Toda a pesquisa foi realizada em um ambiente que utiliza Python 3 e utilizando as bibliotecas *cv2*, *base64*, *requests*, *json*, *os*, *keras ocr*, *pytesseract* e *PIL*.

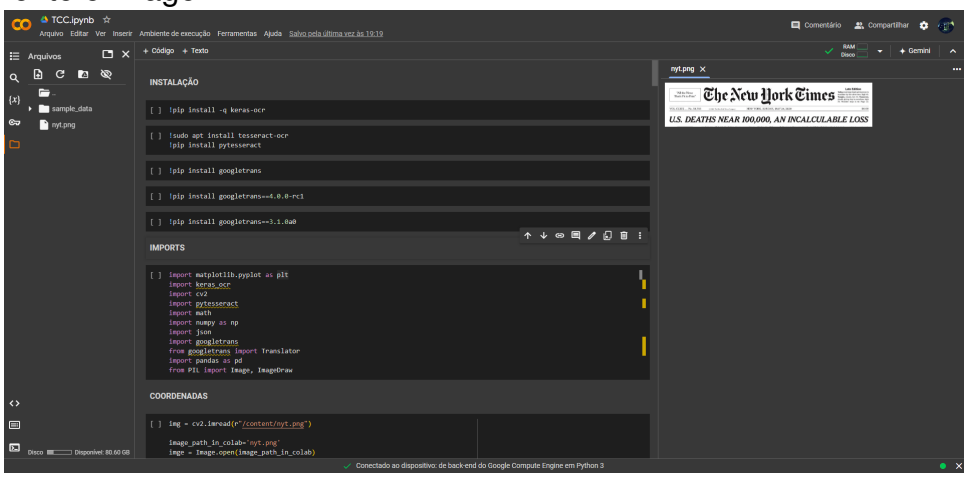
Foram selecionadas aleatoriamente quarenta imagens de diferentes épocas do jornal *The New York Times*, entre elas vinte contém a página inicial do jornal inteira, e as outras vinte apenas a manchete a fim de testar o reconhecimento de caracteres em maior definição. As imagens estão disponíveis no endereço site do jornal[3], e, requerem uma assinatura para ser acessada.

3.1 GOOGLE COLABORATORY

A própria Google descreve o serviço como um serviço de *Jupyter Notebook* hospedado que não requer configuração para uso e oferece acesso gratuito a recursos de computação, incluindo GPUs e TPUs. O Colab é adequado principalmente para aprendizado de máquina, ciência de dados e educação.

A partir deste serviço, foi realizado o desenvolvimento deste trabalho, utilizando as APIs e bibliotecas compatíveis com a ferramenta e a linguagem de programação Python.

Figura 1 - Interface do Google colaboratory, exibindo os arquivos, código fonte e imagem.



Fonte: Elaborado pelo autor.

[3]https://www.nytimes.com/search

A Figura 1 apresenta a interface do *Google Colaboratory*, onde pode-se ver na esquerda uma seção para os arquivos, no meio o código fonte e a direita a imagem que está sendo trabalhada.

3.2 OPENCV

OpenCV é uma biblioteca de código aberto que permite a análise de imagens e vídeos. Inicialmente foi desenvolvida pela Intel para avançar a pesquisa de visão computacional e disseminar o conhecimento. Desde então, muitos programadores já contribuíram com o desenvolvimento da biblioteca (Culjak et al., 2012). A biblioteca conta com uma interface em C++ e mais de 2500 algoritmos otimizados.

3.3 EXTRAÇÃO DOS TEXTOS

Foram testados três formas diferentes para extrair o texto da imagem, sendo elas: *pytesseract*, *keras ocr* e *Google Cloud Vision API*.

3.3.1 Pytesseract

Para extrair o texto com o *pytesseract* foi necessário primeiro processar as imagens. O primeiro passo foi a conversão da imagem para a escala cinza com o comando *cv2.cvtColor* da biblioteca OpenCV. Com esta técnica pode-se reduzir a dimensionalidade dos dados, tendo em vista que as imagens estão em escala de cinza e possuem apenas um canal de cor, diferentemente dos três canais (vermelho, verde e azul) das imagens coloridas. Após isso foi utilizado o *cv2.findContours*, outro comando da biblioteca OpenCV que permite separar o texto nas imagens em caixas facilitando o reconhecimento futuro. Por fim, a imagem separada foi utilizado o *pytesseract* para reconhecer os textos em cada parte da imagem.

3.3.2 Keras OCR

O *Keras OCR* é uma biblioteca disponível em Python que fornece modelos de OCR prontos para uso e permite o treinamento completo para construir novos modelos de OCR. Esta biblioteca já possui um pré-processamento de imagem, não sendo necessário processar as imagens anteriormente como no *pytesseract*.

Foi utilizada a função *pipeline* presente no *Keras OCR*, que é uma sequência automática de etapas pré-definidas para reconhecer o texto. Com essa função foi possível extrair o texto e as coordenadas onde ele fica na imagem.

3.3.3 Google Cloud Vision API

O *Google Cloud Vision API* é um serviço oferecido pela Google que permite o processamento de imagens na nuvem. Ele utiliza-se de visão computacional para permitir que desenvolvedores interajam com seus recursos em suas próprias aplicações.

Uma das funções desta API é o reconhecimento de caracteres em imagens, onde é retornado várias características do texto como as palavras, suas coordenadas dentro da imagem e até o texto completo.

Para a utilização foi necessário criar uma chave na API que permite o uso grátis até um número máximo de requisições, após isso foi realizado o envio da imagem convertida em base64 para a API junto com a opção de *TEXT DETECTION* que sinaliza para a API que a função realizada seria o reconhecimento de texto.

Figura 2: Retorno do Google Cloud Vision API dada uma imagem.

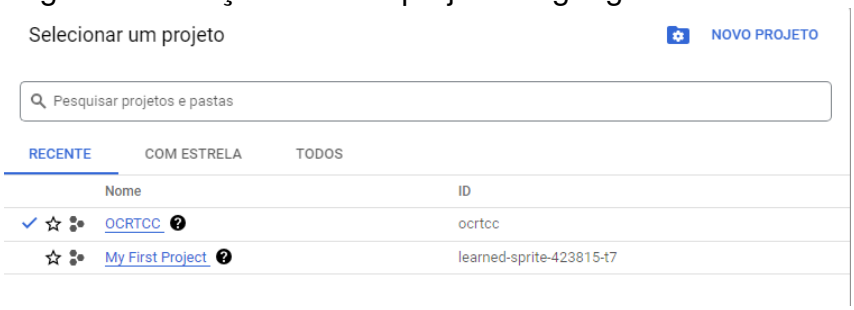
Fonte: Elaborado pelo autor.

A Figura 2 demonstra os textos que foram extraídos pela API, na direita tem-se a imagem que está sendo testada e na esquerda os textos que a API conseguiu identificar na imagem.

3.3.4 Configuração do Google Cloud Vision API

O primeiro passo para a utilização da API é entrar na plataforma do Google Cloud com uma conta Gmail e criar um novo projeto.

Figura 3 - Criação de novo projeto no google cloud.



Fonte: Elaborado pelo autor.

A Figura 3 demonstra a tela para criação de novos projetos no *Google Cloud Vision API*. Após a criação do projeto é possível habilitar os

serviços que serão utilizados, basta clicar em *Cloud Vision API* e ativar o serviço. Na Figura 4 está representada a tela de ativação de API depois que ela foi habilitada.

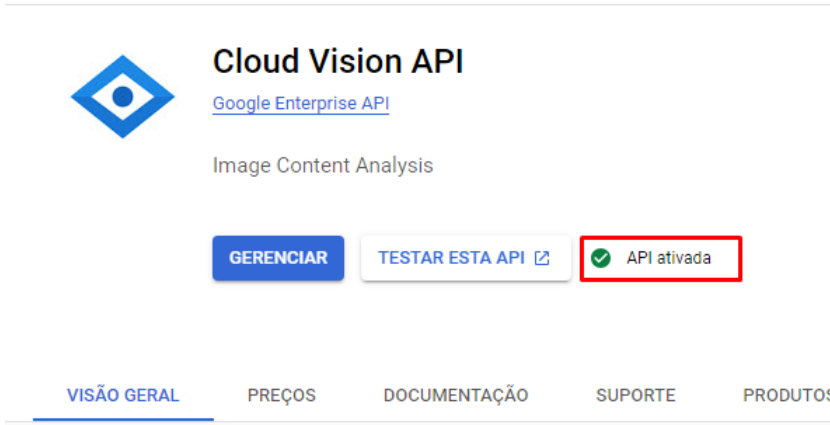
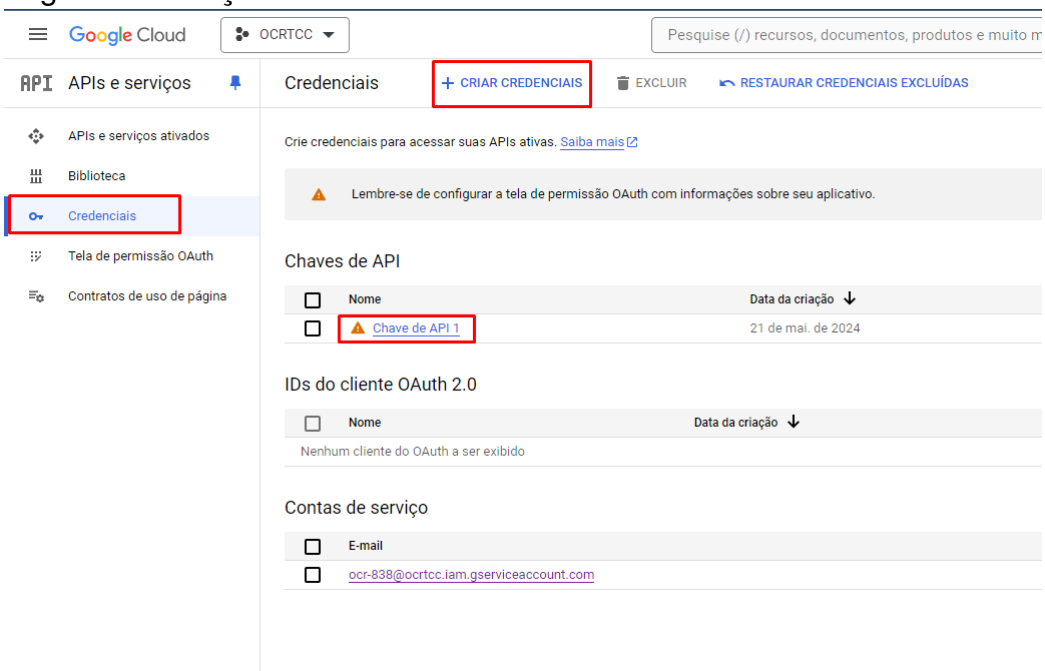


Figura 4 - Ativação da API.
Fonte: Elaborado pelo autor.

O Google Cloud disponibiliza duas opções para a utilização de APIs, a primeira é a utilização de uma chave para fazer as requisições e a segunda é a configuração de um arquivo json que é importado direto no aplicativo. Neste caso foi escolhida a criação de uma chave que será enviada no endereço da requisição. Para criar a chave é necessário ir na aba credenciais e selecionar as opções ”CRIAR CREDENCIAIS” e ”Chave de API”. A Figura 5 exemplifica este passo a passo.

Figura 5 - Criação da Chave de API.



Fonte: Elaborado pelo autor.

3.4 TRADUÇÃO

Após extrair os textos das imagens, vem o processo de tradução, onde o texto no idioma original será convertido em um texto na Língua Portuguesa.

3.4.1 Google Translate API

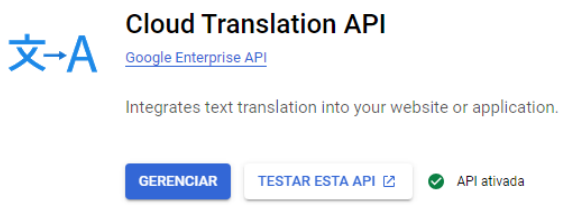
Para realizar a tradução foi escolhido o serviço da Google conhecido como *Google Translate API*, é uma API capaz de detectar o idioma original do texto e traduzir para o idioma escolhido pelo usuário. Assim como o *Cloud vision API* ela também é paga, mas possui um limite de requisições gratuitas que foi utilizado para a realização deste trabalho.

Para utilizar o serviço é realizado uma requisição enviando o texto que será traduzido e o idioma que deseja a tradução, neste caso o português, que retornará o texto traduzido e o texto original para comparação.

3.4.2 Configuração do Google Translate API

Para utilizar o *Google Translate API* é necessário ativar ele no mesmo projeto criado anteriormente para o *Google Cloud Vision API*, clicando na aba Biblioteca, selecionando a opção *Cloud Translation API* e clicando no botão ativar. A Figura 6 demonstra a tela da API após concluída a ativação.

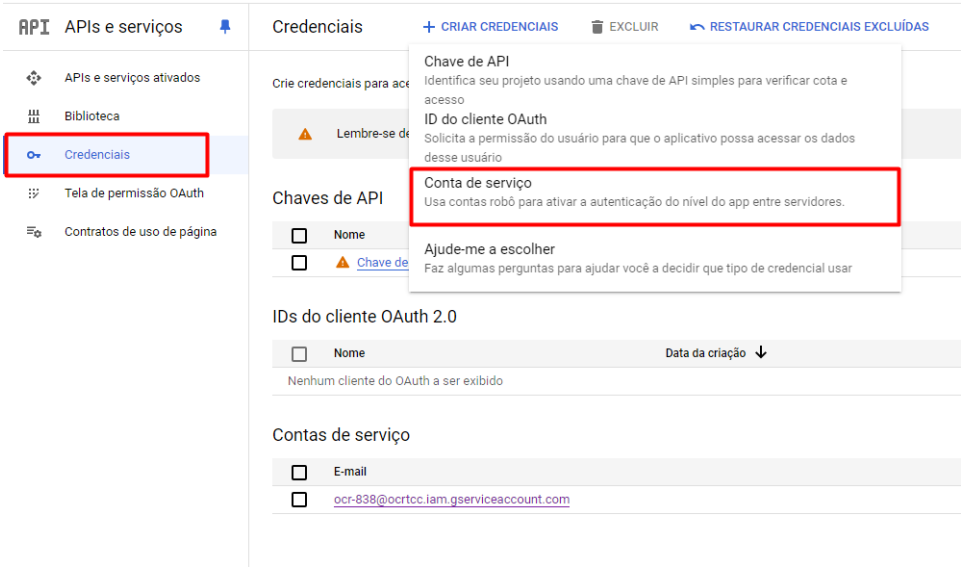
Figura 6 - Ativação do Google Translate.



Fonte: Elaborado pelo autor.

Após adicionada a biblioteca, será necessário criar uma credencial para permitir o acesso da API pelo aplicativo. Na aba credenciais, será selecionado a opção ”CRIAR CREDENCIAIS” e ”Conta de serviç”. A Figura 7 demonstra o passo a passo para a criação da conta de serviço.

Figura 7 - Criação da conta de serviço.



Fonte: Elaborado pelo autor.

Após preenchido os campos será baixado um arquivo com a extensão json, que será utilizado na aplicação para autenticar a conexão com o *Google Translate API* e controlar o número de requisições realizadas.

4 RESULTADOS

Nesta pesquisa o primeiro critério avaliado para a extração dos textos foi o tempo necessário para reconhecer os textos de toda a base de dados. Neste critério o *google cloud vision API* se mostrou superior às outras duas opções, como se pode ver na Tabela 1.

Tabela 1 - Tempo total.

	Pytesseract	Keras OCR	Google Cloud Vision API
Segundos	60	1980	12

Fonte: Elaborado pelo autor.

Em um segundo momento se comparou os textos extraídos com os textos das imagens que contém no próprio site do *The New York Times*. Analisou-se se os textos estavam iguais, onde a tradução foi considerada correta, se os textos mantinham o sentido original contendo apenas erros gramaticais ou então se os textos estavam errados em comparação ao original. Conforme visualizado na Tabela 2.

Tabela 2 - Textos corretamente extraídos.

	Pytesseract	Keras OCR	Google Cloud Vision API
Corretos	1	0	37
Erros Gramaticais	8	4	1
Incorretos	31	36	2

Fonte: Elaborado pelo autor.

A extração de textos utilizando o *pytesseract* não se mostrou eficaz para a aplicação, mesmo realizando o tratamento das imagens os textos não foram extraídos de forma correta nem em imagens com maior definição. Já em imagens com a definição mais baixa só os textos escritos em fontes maiores eram reconhecidos. Também apresentou uma enorme dificuldade em reconhecer caracteres em fontes fora do comum, como é o exemplo da Figura 8.

Figura 8 - Logo do jornal ”The New York Times”.

Fonte: Times (2024).

A biblioteca *Keras OCR* se mostrou menos efetiva para a extração de textos em comparação ao *pytesseract*. Também possuía uma dificuldade em extrair os textos em imagens de menor qualidade.

Figura 9 - Reconhecimento de imagem com o Keras OCR.

Fonte: Elaborado pelo autor.

Como se pode ver na Figura 9 o *Keras OCR* apresentou algumas dificuldades em reconhecer algumas fontes diferentes, como acontece em *”The New York Times”* que foi reconhecido como *”The New lork Times”*. Outra dificuldade apresentada foi o reconhecimento de números, a imagem apresenta o número 100,000, que foi reconhecido como ”ooo00”. Outro ponto é que a biblioteca retornava palavra por palavra e nem sempre na ordem que elas eram exibidas na imagem, dificultando traduções, que necessitam das frases completas para manter o sentido original do texto.

A API do Google Cloud Vision foi superior às outras duas, reconhecendo os caracteres em imagens de baixa qualidade e nos mais variados estilos de escrita. Outro ponto que ele foi superior é que ele retornava os textos inteiros, diferente das outras duas opções que cortavam os textos das imagens, dificultando a tradução.

Vale ressaltar que o *pytesseract* e o *Keras OCR* reconheceram erroneamente as vinte imagens que continham as páginas inteiras, apresentando resultados apenas nas imagens da manchete que continham menos textos.

O módulo de tradução se mostrou efetivo, onde foi mantido o texto original em nomes, como por exemplo a palavra *The New York Times* não era traduzida por ser o nome do jornal, já as palavras que precisavam eram traduzidas corretamente.

5 CONCLUSÃO

Nesta pesquisa buscou-se utilizar recursos de visão computacional, mais especificamente OCR, para extrair e traduzir os textos em imagens históricas com o intuito de auxiliar leitores e pesquisadores. O sistema desenvolvido mostrou-se eficaz diante dos testes apresentados, conectando a extração e a tradução das imagens.

Das três opções de extrações de textos, duas se mostraram ineficazes, e uma delas apresentou os resultados esperados, sendo ela o *Google Cloud Vision API*. Na opção de tradução o *Google Translate API* atendeu todas as expectativas e traduziu corretamente os textos.

É importante ressaltar que este estudo possui limitações, a análise foi realizada em pequena escala e pode não apresentar o mesmo desempenho em cenários de larga escala.

Com base nos conhecimentos adquiridos, bem como nos resultados obtidos, propõe-se para futuros trabalhos: utilizar um maior número de imagens para testes; realizar uma abordagem diferente para o pré-processamento de imagens; editar a imagem via software inserindo o texto traduzido na imagem original, gerando uma nova imagem no idioma selecionado.

REFERÊNCIAS

BANTUPALLI, K.; XIE, Y. **American Sign Language Recognition using Deep Learning and Computer Vision**. [S.l.], 2018. Disponível em: <https://doi.org/10.1109/bigdata.2018.8622141>.

BAZONI, L. S. **Tradução de Textos em Imagens de histórias em Quadrinhos Utilizando Visão computacional E OCR**. Cachoeiro de Itapemirim, 2022. Disponível em: <https://repositorio.ifes.edu.br/handle/123456789/2874>.

CULJAK, I. et al. **A brief introduction to OpenCV**. 2012 Proceedings of the 35th International Convention MIPRO, 2012. 1725-1730 p.

GOMES, D. d. S. **Inteligência Artificial: Conceitos e Aplicações**. [S.l.], 2010. Disponível em: <https://www.professores.uff.br/screspo/wp-content/uploads/sites/127/2017/09/ia_intro.pdf>.

GOMES, F. T.; PARDO, T. A. S. **Estudo e aprimoramento do sistema Apertium de tradução automática entre português e espanhol**. Instituto de Ciências Matemáticas e de Computação (ICMC), USP/São Carlos, 2007.

MENDONçA, F. L. L. d. **Proposta de Arquitetura de um sistema com base em OCR neuronal para Resgate e Indexação de Escritas Paleográficas do sec. XVI AO XIX**. [S.l.], 2008. Disponível em: <https://repositorio.unb.br/handle/10482/1157>.

MUELLER-GASTELL, J.; SENA, M.; TAN, C.-Z. **Multi-digit OCR system for historical records (Computer Vision)**. [s.n.], 2020. Disponível em: <http://cs230.stanford.edu/projects_spring_2020/reports/38792124.pdf>.

OLIVEIRA, C. L. **A importancia da Tradução: Reflexões sobre o Pale do Tradutor.** Vista do a importância da tradução: Reflexões sobre o papel do tradutor, 2019. Disponível em: <https://periodicos.ufac.br/index.php/COMMUNITAS/article/view/1109/pdf>.

TIMES, T. N. Y. **The New York Times**. [S.l.], 2024. Acesso em: 3 jun 2024. Disponível em: <https://www.nytimes.com/search>.