14º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2023

Reconhecimento de cédulas de dinheiro para deficientes visuais utilizando visão computacional

RAISSA R. S. JANUARIO[1], ANDRÉ L. OLIVETE[2]

[1]Graduanda em Bacharelado em Ciência da Computação, IFSP, Câmpus Presidente Epitácio, raissa.santos@aluno.ifsp.edu.br.
[2]Doutor em Ciências Cartográficas, Professor EBTT - Informática, IFSP, Campus Presidente Epitácio, olivete@ifsp.edu.br.

RESUMO: Este projeto tem como objetivo oferecer autonomia e independência em transações financeiras para pessoas com deficiência visual, por meio do desenvolvimento de um aplicativo capaz de identificar cédulas de dinheiro utilizando visão computacional. Sabendo que as cédulas podem perder suas marcas táteis ao longo do tempo, o aplicativo se propõe a suprir essa dificuldade, permitindo que os usuários identifiquem e diferenciem as cédulas de forma rápida e precisa. Para isso foram realizados experimentos com diferentes configurações para uma rede neural convolucional, e então, foi selecionada a configuração com melhor desempenho, a qual obteve uma acurácia de 96,66%.

PALAVRAS-CHAVE: autonomia; transações financeiras; deficiente visual; cédulas; rede neural convolucional; experimentos.

Banknote recognition for visually impaired people using computer vision

ABSTRACT: This project aims to provide autonomy and independence in financial transactions for visually impaired individuals through the development of an application capable of identifying banknotes using computer vision. Considering that banknotes may lose their tactile marks over time, the application aims to overcome this difficulty by allowing users to identify and differentiate banknotes quickly and accurately. For this purpose, experiments were conducted with various configurations for a convolutional neural network, and then, the configuration with the best performance was selected, achieving an accuracy of 96.66%.

KEYWORDS: autonomy; financial transactions; visually impaired; banknotes; convolutional neural network; experiments.

INTRODUÇÃO

A tecnologia está em constante evolução com a criação de novas ferramentas e soluções que transformam a forma como as pessoas se comunicam, trabalham, aprendem e vivem (Santos, 2022). Dessa forma, ela se tornou um recurso extremamente útil para auxiliar pessoas com deficiência, proporcionando maior acessibilidade e inclusão, com a chamada “tecnologia assistiva” que se refere a produtos, equipamentos, dispositivos, recursos, metodologias, estratégias, práticas e serviços que tenham como objetivo promover acessibilidade a pessoas com deficiência ou mobilidade reduzida, visando a autonomia, independência, qualidade de vida e inclusão social (Brasil, 2015).

Segundo dados do último censo do Instituto Brasileiro de Geografia e Estatística (2010), 18,6% da população brasileira possui deficiência visual. E um dos principais problemas enfrentados por essas pessoas, está relacionado às transações financeiras cotidianas, dado que a identificação de cédulas de dinheiro pode ser um obstáculo significativo para a independência e inclusão do deficiente visual.

Conforme apontado por Shimosakai (2010), embora as cédulas possuam marcas de relevo para serem identificadas por meio do tato, essas marcas podem se desgastar com o tempo e o uso constante.

Considerando os fatos supracitados, faz-se importante buscar alternativas que possam suprir as dificuldades enfrentadas por esses indivíduos, tentando mitigar os problemas de acessibilidade enfrentados por eles. Dessa forma, o projeto em questão, propõe uma solução tecnológica que auxilie o deficiente visual a identificar cédulas de dinheiro, contribuindo para uma autonomia em suas atividades financeiras.

MATERIAL E MÉTODOS

O desenvolvimento do projeto consiste em um conjunto de atividades, onde inicialmente foi realizado um levantamento bibliográfico abrangente, com o objetivo de compreender e explorar as tecnologias que seriam utilizadas no projeto. Essa etapa foi fundamental para adquirir um conhecimento aprofundado sobre as técnicas e abordagens existentes, a fim de embasar as escolhas metodológicas feitas ao longo do estudo.

A linguagem de programação Python foi selecionada como a principal ferramenta para implementação do algoritmo, devido à sua ampla adoção na comunidade de aprendizado de máquina e visão computacional. Além disso, foram identificadas diversas bibliotecas especializadas, como OpenCV, NumPy, TensorFlow e Keras, que seriam integradas ao desenvolvimento. Essas bibliotecas fornecem recursos essenciais para o processamento de imagens, manipulação de matrizes, construção e treinamento de redes neurais.

Com o objetivo de testar as ferramentas e alternativas propostas, foi realizada a aquisição das imagens para formar o conjunto de dados. Nessa etapa, foram capturadas imagens de cédulas de duas classes distintas, especificamente as cédulas de R$20,00 e R$50,00. Sendo coletadas 100 imagens de frente e 100 imagens do verso para cada classe, totalizando 400 imagens no conjunto, sendo que para cada uma das imagens foi realizada a rotulação de seu conteúdo.

Essas imagens foram divididas em conjuntos de treinamento e teste. A proporção adotada foi de 20% das imagens destinadas ao conjunto de teste, enquanto 80% foram destinados ao conjunto de treinamento. Durante o processo de treinamento, 20% das imagens de treinamento foram separadas para serem utilizadas como conjunto de validação, permitindo monitorar e ajustar o desempenho do modelo.

Para a identificação das notas foi utilizada uma rede neural convolucional (CNN), que é uma rede neural profunda com uma arquitetura bem adaptada para classificação e reconhecimento de imagens. Dessa forma, foram consideradas diferentes configurações da rede, como a arquitetura, o número de camadas e os parâmetros específicos, como também as técnicas de treinamento, definindo uma função de perda apropriada e o otimizador adequado para ajustar os pesos da rede durante o processo de aprendizado, de forma a obter melhor desempenho. Com o objetivo de evitar o overfitting, que é o ajuste excessivo do modelo aos dados de treinamento, foram utilizadas técnicas de regularização, como a inclusão de camadas de dropout. E com o intuito de selecionar a configuração mais adequada para o projeto, foram realizados alguns experimentos.

O primeiro experimento utilizou um modelo, apresentado na Figura 1, com uma camada convolucional com 6 filtros de tamanho 3x3, ativada pela função de ativação ReLU, seguida de uma camada de MaxPooling 2D, e depois uma camada de flattening foi adicionada para transformar a imagem 2D em uma representação unidimensional para a entrada nas camadas totalmente conectadas, que é a MLP (Multilayer Perceptron) com duas camadas ocultas. A primeira camada oculta contou com 100 neurônios e a segunda com 50 neurônios, ambas ativadas pela função ReLU.

Uma camada de dropout com uma taxa de 0.2 foi colocada para reduzir a chance de overfitting do modelo aos dados de treinamento. E por fim, uma camada de saída com 2 neurônios, ativados pela função softmax, que apresenta a probabilidade de a entrada estar associada a uma determinada classe. A entrada do modelo foi definida com imagens de tamanho 250x250 pixels e um único canal (escala de cinza).

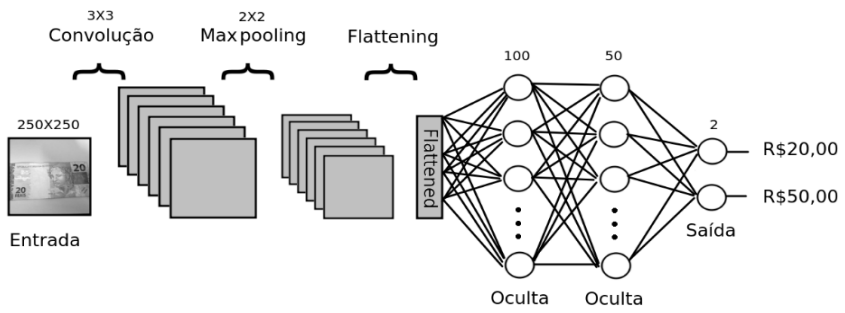

FIGURA 1. Arquitetura da rede neural convolucional no experimento 1.
Fonte: Autores.

No segundo experimento, uma variação na arquitetura da rede neural foi realizada, na qual foram adicionadas uma camada convolucional com 16 filtros de tamanho 3x3 e uma camada de MaxPooling 2D. A escolha da quantidade de filtros nas camadas de convolução e quantidade de camadas ocultas na MLP, foi baseada na arquitetura LeNet-5 de Lecun et al. (1998), que foi o pioneiro das redes neurais convolucionais. Essa arquitetura pode ser visualizada na Figura 2.

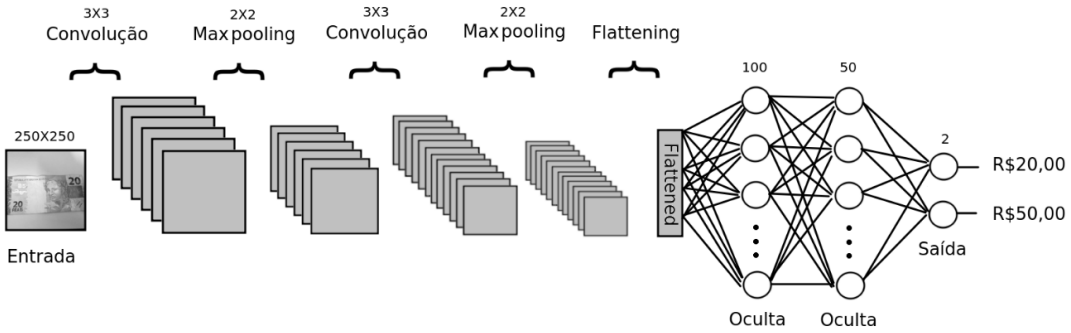

FIGURA 2. Arquitetura da rede neural convolucional no experimento 2
Fonte: Autores.

No terceiro experimento foi aplicada uma técnica de aumento de dados, que consiste no aumento da quantidade e diversidade dos dados do conjunto de dados através de aplicação de pré-processamentos das imagens do conjunto original, onde foram realizadas rotações, espelhamentos verticais e horizontais, desfoque e alterações de escala nas imagens. O conjunto de 400 imagens passou a possuir 2400 imagens. A rede neural convolucional utilizou a mesma arquitetura do primeiro experimento.

No quarto e último experimento, foi realizada uma variação do terceiro experimento, utilizando a mesma técnica de aumento de dados, porém aplicada à rede neural convolucional utilizada no segundo experimento.

Após a análise do desempenho de cada um dos experimentos, por meio das métricas de avaliação, e analisando a capacidade de generalização, foi possível selecionar o que mais se adequa ao projeto para ser implementado no servidor. Com isso, está sendo desenvolvida a aplicação. Sendo implementada em TypeScript, utilizando o *framework* Ionic com Angular, na qual possui como funcionalidades realizar a captura de imagens, bem como upload de imagens, enviando-as ao servidor como requisição, e apresentando o resultado do reconhecimento.

Posteriormente, será desenvolvido um dispositivo físico para captura das imagens, utilizando a placa de prototipação ESP32-CAM que integra uma câmera e o microcontrolador ESP32. Ficando responsável por capturar as imagens das cédulas e as enviar para o aplicativo por meio do módulo de Wi-Fi. Além disso, será realizada a adaptação do modelo para identificar todas as classes de cédula de Real e mais de uma por vez, bem como a não presença delas. E por fim, a hospedagem do servidor em alguma plataforma gratuita, para ser acessado pela aplicação remotamente.

Portanto, dado que o projeto utiliza diversas tecnologias e permeia por vários assuntos dentro da Ciência da Computação, faz-se necessário a utilização de várias ferramentas, linguagens e bibliotecas, que são apresentadas a seguir:

- Computador pessoal Intel Core i5, 512GB de HD e placa de vídeo Intel Corporation HD Graphics 620, para realizar a implementação e treinamento da rede neural convolucional;
- Smartphone com câmera de no mínimo 18 mp, para capturar as imagens que formam a base de dados de treinamento, validação e teste da rede neural;
- Linguagens de programação Python 3.10.6, para o desenvolvimento da rede neural e do algoritmo de identificação no servidor com as bibliotecas: tensorflow 2.12.0, keras 2.12.0, matplotlib 3.7.1, numpy 1.23.4, seaborn 0.12.2, imgaug 0.4.0 e opencv-python 4.7.0.72;
- Linguagem TypeScript 5.1.3 com framework Ionic 7.5.0, para o desenvolvimento do aplicativo para dispositivos móveis capaz de capturar a imagem, fazer a detecção da nota de moeda e informar por voz ao usuário;
- Ambiente de desenvolvimento Arduino IDE e Visual Studio Code para realizar a implementação dos códigos do servidor e da aplicação;
- Aplicativo Jupyter Notebook, para escrita de códigos de implementação, treinamento e avaliação da rede neural;
- ESP32-CAM, para capturar as imagens.

RESULTADOS E DISCUSSÃO

A partir dos experimentos realizados, foi possível analisar o desempenho e capacidade do modelo ao ser treinado de acordo com cada uma das variações. Essa análise permitiu avaliar como ele lida com diferentes tipos de dados, identifica padrões e realiza previsões precisas. Com base nos resultados, foi possível identificar as variações que melhor se adaptam ao problema em questão, proporcionando um melhor desempenho em termos de métricas de avaliação.

No primeiro experimento, apesar do modelo possuir uma arquitetura simples, teve um bom desempenho. Foi possível analisar que ele lidou bem com os dados de teste, obtendo uma acurácia de 96,25% e uma perda de 0.1190, demonstrando que não ocorreu *overfitting*, e então interage bem com dados nunca vistos antes.

Apesar do modelo obtido no primeiro experimento apresentar bons resultados, busca-se desenvolver um algoritmo de identificação ainda mais preciso e confiável. Por esse motivo, foi realizado o experimento 2, porém foi observado uma queda no desempenho do modelo. A acurácia alcançada foi de 93,75%, e a perda registrada foi de 0,1051. Sugerindo a possibilidade de *overfitting*, ou seja, ao adicionar mais camadas convolucionais ao modelo, a rede neural pode ter se especializado demais nos padrões específicos das imagens de treinamento, resultando em uma menor capacidade de generalização para imagens que não foram vistas anteriormente.

A quantidade e variedade de dados de treinamento desempenham um papel fundamental na capacidade do modelo de realizar uma identificação precisa das classes, dado que ele terá mais referências e exemplos para aprender as características distintivas de cada classe. Apesar disso, o experimento 3 obteve desempenho menor do que esperado, sendo que apresentou 94,16% de acurácia e uma perda de 0,2588.

Por fim, aplicando a base de dados em maior quantidade num modelo capaz de capturar características mais abstratas e sutis nas imagens, permitindo uma melhor extração de informações relevantes, foi obtido uma acurácia um pouco superior ao primeiro experimento realizado, sendo de 96,66% e uma perda de 0,1616.

A partir do desempenho dos modelos analisados foi escolhido o modelo e a base de dados do experimento 4, e assim foi desenvolvido o servidor a partir dele, para realizar identificação de imagens enviadas como requisição, e retornar como resposta a classe predita. E com isso, foi desenvolvido uma

aplicação web, que permite capturar imagens e fazer uploads, para então enviar ao servidor requisitando a classe que a cédula da imagem pertence.

Na Figura 3 é apresentada a arquitetura do sistema proposto para a o reconhecimento de cédulas de real, onde a placa de prototipação ESP32-CAM que incorpora uma câmera com resolução SVGA, além de componentes para acesso a rede *wi-fi* e conexão através de *bluetooth*, onde está há um pequeno servidor, que aguarda uma requisição para capturar uma imagem e enviar para o dispositivo móvel.

O aplicativo do dispositivo móvel após receber a imagem, envia para um servidor HTTP Rest, que será responsável pelo processamento do reconhecimento e contagem das cédulas existentes na imagem, devolvendo um texto com a contagem das cédulas para o dispositivo móvel. O dispositivo móvel recebe esse resultado textual representando o valor reconhecido e reproduz para o usuário através do autofalante do dispositivo móvel ou dispositivo *bluetooth* acoplado ao ouvido do usuário.

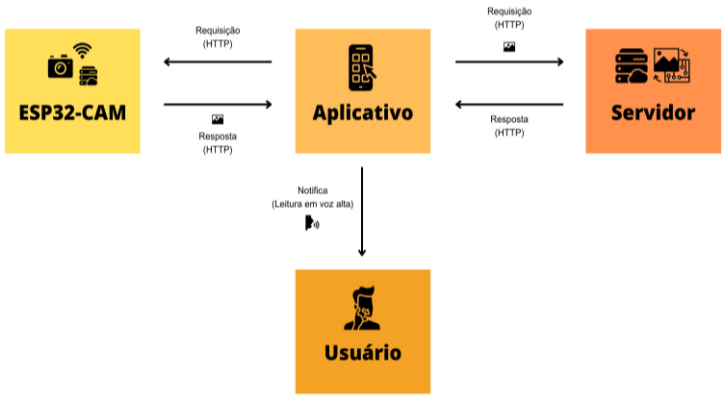

FIGURA 3. Diagrama de componentes do sistema
Fonte: Autores.

CONCLUSÕES

A rede neural convolucional é altamente eficaz para o reconhecimento de imagens, pois possui uma estrutura específica, composta por camadas convolucionais e de pooling, que permite capturar características relevantes das imagens, tornando-a especialmente adequada para lidar com classificação e identificação de padrões visuais.

Apesar disso, é importante destacar que o desempenho do modelo é influenciado por diversos fatores. Alguns desses fatores incluem a configuração dos hiperparâmetros da rede, o tamanho e a diversidade da base de dados utilizada, entre outros. Uma configuração inadequada dos hiperparâmetros, como o número de camadas convolucionais, o tamanho dos filtros e a taxa de *dropout*, pode levar a problemas como o *overfitting*.

Apesar dos resultados satisfatórios obtidos nos experimentos, é importante ressaltar que o objetivo desejado para a aplicação não foi plenamente alcançado. Então, é necessário identificar e analisar os fatores que estão limitando o desempenho do modelo e buscar melhorias para superar essas limitações.

AGRADECIMENTOS

Agradeço ao Instituto Federal de Ciência, Educação e Tecnologia de São Paulo pela oportunidade e todo suporte prestado, e ao meu orientador André Luís Olivete, por me auxiliar em todo o processo, me guiando e incentivando para a realização do projeto.

REFERÊNCIAS

BRASIL. **Lei nº 13.146, de 6 de julho de 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência)**. Diário Oficial da União, Brasília, DF, 7 jul. 2015. Seção 1, p. 1. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm. Acesso em: 6 de Fev. de 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico: Características gerais da população, religião e pessoas com deficiência**. Rio de Janeiro, 2010.
LECUN, Y. et al. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: https://ieeexplore.ieee.org/document/726791. Acesso em: 15 mai. 2023.

LECUN, Y. et al. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: https://ieeexplore.ieee.org/document/726791. Acesso em: 15 mai. 2023.

SANTOS, Aline Cristina dos. **Benefícios da tecnologia para a sociedade**. [S. l.], 6 jun. 2022. Disponível em: https://prensa.li/@alinecristina/beneficios-da-tecnologia-para-a-sociedade/. Acesso em: 17 jun. 2023.

SHIMOSAKAI, Ricardo. **Como os cegos diferenciam as notas de dinheiro?**. [S. l.], 30 out. 2010. Disponível em: https://ricardoshimosakai.com.br/como-os-cegos-diferenciam-as-notas-de-dinheiro/. Acesso em: 25 maio 2023.