Revista Brasileira de Computação Aplicada, Abril, 2025

DOI: 10.5335/rbca.v17i1.15934
Vol. 17, Nº 1, pp. 23–32
Homepage: seer.upf.br/index.php/rbca/index

ARTIGO ORIGINAL

Reconhecimento de cédulas de real usando inteligência artificial para auxiliar pessoas com deficiência visual

Recognition of Brazilian Banknotes Using Artificial Intelligence to Assist Visually Impaired Individuals

Karen de Almeida [1], Almir Olivette Artero [1], Danilo Medeiros Eler [1], Maurício Araújo Dias [1], Danillo Roberto Pereira [1], Francisco Assis da Silva [2]

[1]Universidade Estadual Paulista - UNESP, FCT - Presidente Prudente, SP , [2]Universidade do Oeste Paulista - UNOESTE, Faculdade de Informática de Pres. Prudente, SP

*karen.almeida18@unesp.br; almir.artero@unesp.br; danilo.eler@unesp.br; ma.dias@unesp.br; danillo.pereira@unesp.br; chico@unoeste.br

Recebido: 04/06/2024. **Revisado:** 12/11/2024. **Aceito:** 30/04/2025.

Resumo

O reconhecimento de cédulas de Real ainda é um desafio para pessoas com deficiência visual. Este trabalho propõe uma solução que utiliza uma rede neural convolucional para auxiliar essas pessoas a reconhecerem cédulas de Real, empregando um smartphone com câmera. A implementação resulta em um aplicativo chamado *Banknotes Recognition*, que realiza o reconhecimento de forma rápida, capturando a imagem da cédula com a câmera do próprio smartphone. Os testes realizados com um conjunto de 66 imagens de cédulas de Real demonstraram uma acurácia de 100%, com todas as cédulas sendo classificadas corretamente, mesmo em condições adversas, como notas dobradas, com fundos complexos, sob diferentes condições de iluminação e notas desgastadas.

Palavras-Chave: Acessibilidade, Reconhecimento de Cédulas de Real, Redes Neurais Convolucionais.

Abstract

Recognizing Brazilian Real banknotes remains a challenge for visually impaired individuals. This work presents a proposal that employs a convolutional neural network to assist visually impaired people in recognizing Real banknotes using a smartphone with a camera. The implementation provides an application called *Banknotes Recognition*, which performs the recognition process swiftly by capturing the banknote image with the smartphone's camera. Tests conducted with a set of 66 images of Real banknotes showed an accuracy of 100%, with all banknotes correctly classified even under unfavorable conditions, such as folded notes, complex backgrounds, various lighting conditions, and worn-out notes.

Keywords: Accessibility, Brazilian Banknote Recognition, Convolutional Neural Networks.

1 Introdução

Segundo dados do IBGE (2010), cerca de 18,8% da população brasileira possui algum nível de dificuldade para enxergar e, entre esses, 3,4% apresentam ausência total de visão. Em um país com mais de 200 milhões de habitantes, conclui-se que uma ampla camada da população enfrenta essa deficiência em sua vida diária, incluindo o reconhecimento de cédulas de dinheiro. Para mitigar essas dificuldades, as notas de Real têm sido produzidas com diferentes tamanhos e, adicionalmente, na segunda família do Real, foi incluído o recurso de distinção das cédulas por marcas táteis. No entanto, ainda existem obstáculos em termos de acessibilidade, como, por exemplo, notas desgastadas que perdem o relevo com o tempo.

Almeida and Alves (2019) apresentam uma pesquisa com um grupo de pessoas com deficiência visual para avaliar se as cédulas atuais de Real são realmente acessíveis. O estudo envolveu uma amostra diversificada de indivíduos, incluindo tanto aqueles com perda parcial quanto total da visão. Esta abordagem permitiu uma avaliação abrangente das dificuldades enfrentadas por esse público específico. Como resultado, os autores relatam que, apesar das mudanças implementadas nas cédulas, o processo de reconhecimento de notas continua desafiador. As alterações, como a introdução de diferentes tamanhos de cédulas e a inclusão de marcas táteis na segunda família do Real, não foram suficientes para garantir um reconhecimento eficiente. A pesquisa revelou que as marcas táteis, embora úteis, não proporcionam um reconhecimento perfeito das cédulas, especialmente em condições adversas, como desgaste ou danos às notas.

Diante dessa constatação, a combinação de múltiplas técnicas simultaneamente surge como uma alternativa promissora para resolver o problema. A introdução de novos métodos de reconhecimento pode complementar as marcas táteis, aumentando a precisão e a confiabilidade do processo. Nesse contexto, os avanços recentes em Inteligência Artificial (IA) oferecem uma oportunidade significativa para melhorar a acessibilidade.

Particularmente, os avanços na análise de imagens por meio da IA têm se mostrado extremamente confiáveis. Essas tecnologias possibilitam uma análise detalhada e precisa, que pode ser utilizada de forma complementar para o reconhecimento de cédulas de Real. Esse tipo de análise pode compensar as limitações das marcas táteis, especialmente em cédulas desgastadas ou danificadas.

Com base nessa premissa, este trabalho propõe o uso de uma Rede Neural Convolucional (CNN), uma técnica avançada de Inteligência Artificial que se destaca pela eficiência no reconhecimento de imagens (O'Shea and Nash (2015), Zhang et al. (2021)). As CNNs são particularmente eficazes porque eliminam a necessidade de uma extração manual e complexa de características das imagens. Em vez disso, essas redes são capazes de aprender e identificar automaticamente padrões relevantes, o que as torna ideais para aplicações de reconhecimento de cédulas.

A implementação sugerida envolve a utilização de um aplicativo móvel que emprega a CNN para analisar as imagens das cédulas capturadas pela câmera do smartphone. Esse método não apenas facilita o reconhecimento das cédulas, mas também oferece uma solução prática e acessível para os usuários. A precisão desse sistema foi demonstrada em testes com um conjunto de 66 imagens de cédulas de Real, onde a acurácia alcançada foi de 100%, mesmo em condições adversas.

A combinação de técnicas tradicionais com os avanços da Inteligência Artificial pode revolucionar o reconhecimento de cédulas para pessoas com deficiência visual. Este estudo não apenas confirma as limitações das soluções atuais, mas também abre caminho para o desenvolvimento de métodos mais eficazes e inclusivos, contribuindo significativamente para a acessibilidade financeira.

As demais seções deste trabalho estão organizadas da seguinte forma: na Seção 2, encontra-se a fundamentação teórica das principais técnicas usadas neste trabalho; Na Seção 3, são abordados alguns trabalhos relacionados; A metodologia proposta é discutida na Seção 4; Os experimentos realizados e seus resultados estão detalhados na Seção 5; Por fim, na Seção 6, são apresentadas algumas conclusões e sugestões para trabalhos futuros.

2 Fundamentação Teórica

2.1 Cédulas de Real

Em 1994, o Real se tornou a moeda oficial do Brasil. Posteriormente, em 2010, foi lançada a segunda família do Real, que incorporou elementos antifalsificação mais modernos e novos recursos gráficos. Atualmente, ambas as famílias de notas estão em circulação no país, sendo que a segunda família predomina. Isso ocorre porque o Banco Central recolhe as notas da primeira família à medida que sofrem desgastes (do Brasil, 2023).

A acessibilidade foi um aspecto importante considerado no desenvolvimento da segunda família do Real. Foram introduzidos recursos de diferenciação por tamanho e marcas em relevo para ajudar pessoas com deficiência visual. Em relação ao tamanho, as cédulas foram projetadas para aumentar de comprimento conforme o valor nominal aumenta.

No entanto, esses recursos não resolvem completamente os desafios de acessibilidade. O reconhecimento de uma cédula pelo tamanho pode ser difícil, exigindo um esforço significativo de memorização, especialmente em situações onde é necessário identificar uma cédula de forma isolada. Além disso, as marcas em relevo, embora úteis, tendem a se desgastar com o manuseio contínuo das notas, o que dificulta sua detecção.

Dessa forma, apesar dos avanços incorporados na segunda família do Real, ainda existem obstáculos a serem superados para garantir uma plena acessibilidade. A contínua busca por soluções mais eficazes é essencial para atender melhor às necessidades de todos os cidadãos.

2.2 Redes Neurais Convolucionais (*Convolutional Neural Network - CNN*)

As Redes Neurais Convolucionais (CNNs) são redes neurais artificiais projetadas especificamente para o reconhecimento de padrões em imagens (O'Shea and Nash, 2015). A principal característica das CNNs é a sua capacidade de identificar automaticamente as características (*features*)

relevantes para o aprendizado. Essas redes são estruturas complexas, compostas por várias camadas, e geralmente apresentam um tempo de treinamento bastante elevado. Entre as principais camadas das CNNs estão: camada convolucional, camada de pooling, camada totalmente conectada e outras mais específicas.

A camada convolucional é a construção central de uma CNN. Ela consiste na aplicação de uma coleção de filtros convolucionais (*kernels*) à imagem de entrada, extraindo atributos relevantes e gerando um mapa de características como saída (Alzubaidi et al., 2021); (Barelli, 2018). Esses filtros varrem a imagem, permitindo que a rede identifique padrões como bordas, texturas e outras características importantes.

A camada de *pooling* atua para reduzir o número de parâmetros necessários nas camadas posteriores, com o objetivo de diminuir o custo computacional e a quantidade de parâmetros na rede totalmente conectada que se encontra no bloco final da CNN (Alzubaidi et al., 2021). A operação de pooling realiza uma agregação dos valores do mapa de características, utilizando métodos como agrupamento máximo, mínimo e médio global. O método mais comum é o *max-pooling*, que consiste em manter apenas o elemento de valor máximo dentro de uma região específica do mapa de características, preservando assim os aspectos dominantes.

A camada totalmente conectada, por sua vez, é uma rede neural típica, incluída para realizar a classificação. Normalmente localizada no final da arquitetura da CNN, esta camada possui uma conexão densa, pois cada neurônio está conectado a todos os neurônios da camada anterior (Alzubaidi et al., 2021). Essa densidade de conexões permite que a camada totalmente conectada combine as características extraídas pelas camadas anteriores para produzir a saída final, geralmente uma classificação ou uma previsão.

As CNNs têm revolucionado o campo do processamento de imagens e são amplamente utilizadas em diversas aplicações, como reconhecimento de objetos, análise de imagens médicas, reconhecimento facial e muitas outras áreas onde a identificação de padrões visuais é crucial.

2.3 Modelos de CNN

Ao longo dos anos, diversas arquiteturas foram propostas para as CNNs, com variações significativas para atender a diferentes tipos de tarefas de reconhecimento. Entre essas arquiteturas, destacam-se AlexNet, VGGNet e SqueezeNet, que serão descritas a seguir.

A AlexNet consiste em oito camadas com pesos, incluindo cinco camadas convolucionais e três totalmente conectadas. O modelo começa com uma camada de convolução que utiliza 96 filtros de tamanho 11x11 com stride 4 (Zhang et al., 2021). A seguir, a arquitetura emprega camadas de convolução com filtros menores, de 5x5 e 3x3, respectivamente. Intercaladas entre as camadas convolucionais estão as camadas de *max-pooling*, que reduzem a quantidade de informação processada. Todas as camadas utilizam a função de ativação ReLu, que ajuda a introduzir não linearidades no modelo e acelerar o treinamento.

A VGGNet possui variantes como VGGNet16, com 16 camadas, e VGGNet19, com 19 camadas, ambas de alto custo computacional. A versão original, VGGNet11, é mais enxuta, com um total de 11 camadas. Sua estrutura inclui uma sequência de duas camadas convolucionais usando filtros de 3x3, essenciais para a extração das características primárias da imagem. Estas são seguidas por uma camada de *pooling* de dimensão 2x2 com stride 2. As camadas intermediárias repetem esse padrão de convolução e *pooling*. No final, há três camadas totalmente conectadas (VisoAI, n.d.), todas utilizando a função de ativação ReLu.

A SqueezeNet v1.1 é notável por sua alta precisão combinada com baixa complexidade computacional. Destaca-se por sua contagem reduzida de parâmetros, projetada com foco na eficiência. A arquitetura inclui camadas de convolução, camadas de *pooling* e as inovadoras camadas *fire*. As camadas de convolução extraem mapas de características da imagem de entrada, enquanto as camadas de *pooling* reduzem a resolução da imagem. As camadas *fire* combinam convoluções com filtros de 1x1 e 3x3, diminuindo significativamente o número de parâmetros do modelo. Essa estrutura compacta permite alto desempenho com demanda computacional mínima.

A arquitetura da SqueezeNet v1.1 (veja a Fig. 1) inicia com uma camada de convolução (*conv1*), seguida por oito camadas *fire* (*fire2* a *fire9*). Depois, uma última camada de convolução (*conv10*) é aplicada. A camada de *max-pooling* com passo 2 é inserida entre as camadas *conv1*, *fire4*, *fire8* e *conv10*. Conforme as camadas avançam, o número de filtros aumenta, facilitando a extração do máximo de informações das camadas anteriores.

Essas arquiteturas ilustram a evolução das CNNs para atender a diferentes demandas, balanceando precisão e eficiência computacional.

3 Trabalhos relacionados

Ao longo dos anos, têm surgido diversos estudos voltados para o desenvolvimento de sistemas capazes de reconhecer notas monetárias, especialmente visando contribuir para a autonomia e inclusão de pessoas com deficiência visual. A seguir, são apresentadas algumas dessas abordagens que demonstram a diversidade de técnicas e algoritmos aplicados nessa área, bem como uma discussão dos avanços e lacunas identificados.

Sufri et al. (2017) propôs uma solução para o reconhecimento de notas de Ringgit, a moeda da Malásia. O autor adotou uma abordagem utilizando os algoritmos de classificação *k-Nearest Neighbors* (k-NN) e *Decision Tree Classifier* (DTC). Seu estudo envolveu uma seleção cuidadosa de uma base de dados composta por 168 imagens das cédulas, capturadas em um ambiente controlado. Os resultados obtidos foram notáveis, com uma precisão impressionante de 99,7% para ambos os algoritmos aplicados. Esses resultados reforçam a eficácia dessas técnicas na identificação precisa de notas monetárias, embora em condições de controle.

Zucatelli et al. (2017), por sua vez, desenvolveu uma solução utilizando o algoritmo de *Support Vector Machine* (SVM), uma técnica de aprendizado de máquina supervisionado. Sua abordagem incluiu o treinamento da SVM com 48 imagens de notas de valores entre 2 e 100 reais, abrangendo diversas orientações das cédulas. Apesar de

layer name/type	output size	filter size / stride (if not a fire layer)	depth	s_{1x1} (#1x1 squeeze)	e_{1x1} (#1x1 expand)	e_{3x3} (#3x3 expand)	s_{1x1} sparsity	e_{1x1} sparsity	e_{3x3} sparsity	# bits	#parameter before pruning	#parameter after pruning
input image	224x224x3										-	-
conv1	111x111x96	7x7/2 (x96)	1				100% (7x7)			6bit	14,208	14,208
maxpool1	55x55x96	3x3/2	0									
fire2	55x55x128		2	16	64	64	100%	100%	**33%**	6bit	11,920	5,746
fire3	55x55x128		2	16	64	64	100%	100%	**33%**	6bit	12,432	6,258
fire4	55x55x256		2	32	128	128	100%	100%	**33%**	6bit	45,344	20,646
maxpool4	27x27x256	3x3/2	0									
fire5	27x27x256		2	32	128	128	100%	100%	**33%**	6bit	49,440	24,742
fire6	27x27x384		2	48	192	192	100%	**50%**	**33%**	6bit	104,880	44,700
fire7	27x27x384		2	48	192	192	**50%**	100%	**33%**	6bit	111,024	46,236
fire8	27x27x512		2	64	256	256	100%	**50%**	**33%**	6bit	188,992	77,581
maxpool8	13x12x512	3x3/2	0									
fire9	13x13x512		2	64	256	256	**50%**	100%	**30%**	6bit	197,184	77,581
conv10	13x13x1000	1x1/1 (x1000)	1				**20%** (3x3)			6bit	513,000	103,400
avgpool10	1x1x1000	13x13/1	0									
	activations		parameters				compression info				1,248,424 (total)	**421,098** (total)

Figura 1: Arquitetura da rede SqueezeNet (Tsang, 2019).

alcançar uma acurácia de 88%, os resultados apresentados ressaltam a importância de um ambiente de captura de imagens controlado para garantir a eficácia do sistema.

Duarte (2019) propôs uma abordagem baseada em descritores para encontrar correspondências entre pontos de interesse em imagens. O autor utilizou o *Speeded-Up Robust Features* (SURF), um detector e descritor de características em imagens. Os testes foram realizados em três tipos diferentes de cédulas, com taxas de acurácia que variaram entre 85% e 100%, mesmo sob condições controladas de ambiente. Esses resultados destacam a viabilidade dessa técnica na identificação precisa de notas monetárias.

Aseffa et al. (2022) apresentou uma aplicação para o reconhecimento de notas da Etiópia, utilizando a arquitetura de rede neural convolucional (CNN) MobileNetV2. Seu estudo envolveu o treinamento do modelo por 100 épocas, com um conjunto de dados contendo 30.580 imagens, obtidas utilizando técnicas de aumento de dados. Os experimentos conduzidos revelaram uma precisão de 96,80%. Contudo, a precisão alcançada depende da qualidade das condições de captura das cédulas, exigindo iluminação e posicionamento ideais.

Além disso, um estudo recente trouxe uma revisão sistemática da literatura sobre o reconhecimento de cédulas bancárias falsas, abordando especificamente os desafios e avanços na área Silva et al. (2024). Este trabalho investigou 25 artigos, organizados com base no ano de publicação, na moeda analisada e nas técnicas de visão computacional empregadas. Como destacado em Silva et al. (2024), algumas das abordagens discutidas incluem o uso de redes neurais convolucionais e técnicas híbridas que combinam aprendizado de máquina tradicional e aprendizado profundo. Esse estudo também disponibilizou conjuntos de dados relevantes, contendo imagens de notas falsas e genuínas, que podem contribuir para o avanço dessa linha de pesquisa. A comparação entre os resultados obtidos por diferentes abordagens, como sugerido nesse trabalho, pode oferecer perspectivas valiosas para o desenvolvimento de sistemas mais robustos e acessíveis.

Esses estudos evidenciam o avanço contínuo no desenvolvimento de sistemas de reconhecimento de notas monetárias, ressaltando a importância de abordagens eficazes e acessíveis para garantir a inclusão e a autonomia das pessoas com deficiência visual. Contudo, ainda existem desafios relacionados à generalização dos modelos para diferentes condições de captura e à validação em cenários reais, questões que permanecem como tópicos relevantes para futuras pesquisas.

4 Algoritmo proposto

O método proposto neste trabalho propõe a divisão do processo em três fases distintas, iniciando com a criação de um modelo habilitado para o reconhecimento de cédulas, que utiliza uma Rede Neural Convolucional (*CNN*). Posteriormente, é elaborada uma API com o intuito de simplificar o acesso a esse modelo e, por fim, com a finalidade de tornar o projeto globalmente acessível, é construída uma interface implementada através de um aplicativo.

O método proposto neste trabalho adota uma abordagem dividida em três fases distintas, visando proporcionar um processo eficiente e acessível. Inicialmente, é desenvolvido um modelo capaz de reconhecer cédulas, empregando uma Rede Neural Convolucional. Esta etapa é crucial para a identificação precisa das notas monetárias. Em seguida,

é elaborada uma API (do inglês *Application Programming Interface*) com o objetivo de simplificar o acesso a esse modelo, facilitando sua integração em diferentes sistemas e plataformas. Essa camada de API proporciona uma interface unificada para a utilização do modelo de reconhecimento de cédulas, tornando-o mais acessível e versátil. Por fim, para garantir uma ampla acessibilidade global, é desenvolvida uma interface de usuário implementada por meio de um aplicativo dedicado. Esta interface intuitiva e amigável permite que usuários de diferentes níveis de habilidade e familiaridade com tecnologia possam facilmente utilizar o sistema de reconhecimento de cédulas, promovendo assim a inclusão e a autonomia. Em conjunto, essas três fases compõem uma solução abrangente e integrada para o reconhecimento de cédulas, que visa atender às necessidades de diversos públicos de forma eficaz e acessível. O código fonte est disponível no Github[1].

4.1 Desenvolvimento do modelo de reconhecimento

Na fase inicial do projeto, é empregada uma rede CNN, sendo desenvolvida utilizando a linguagem de programação Python. Esta escolha se deve à vasta disponibilidade de bibliotecas auxiliares que Python oferece, as quais são essenciais para a concepção e implementação da aplicação. Python é conhecido por sua ampla gama de bibliotecas especializadas em aprendizado de máquina e visão computacional, como TensorFlow, PyTorch e Keras, que fornecem ferramentas poderosas para treinar e avaliar modelos de redes neurais convolucionais. Além disso, a linguagem Python é reconhecida por sua simplicidade, flexibilidade e facilidade de uso, tornando-a uma escolha popular entre os desenvolvedores para projetos de inteligência artificial e aprendizado de máquina. Portanto, ao utilizar Python como base para o desenvolvimento da CNN nesta etapa inicial do projeto, é possível aproveitar ao máximo as funcionalidades e recursos disponíveis, facilitando o desenvolvimento e otimizando o desempenho da aplicação de reconhecimento de cédulas.

4.1.1 Definição da arquitetura da CNN

A rede neural utilizada no desenvolvimento deste trabalho é a SqueezeNet v1, que se mostrou mais adequada para lidar com a complexidade do problema em questão. Como mencionado anteriormente, esta rede apresenta uma alta precisão nos resultados e uma baixa complexidade computacional. Isso a torna especialmente recomendada para implementações em dispositivos computacionais com recursos limitados, como smartphones.

A escolha da SqueezeNet v1 para este projeto foi baseada em sua eficiência em termos de desempenho e consumo de recursos. Por ser uma arquitetura compacta e leve, ela pode ser implantada em dispositivos móveis sem comprometer significativamente a velocidade ou a qualidade do reconhecimento de cédulas. Além disso, sua precisão comprovada em problemas similares sugere que é uma escolha sólida para esta aplicação específica.

Portanto, ao optar pela SqueezeNet v1, o projeto visa oferecer uma solução eficaz e acessível para o reconhecimento de cédulas, que pode ser facilmente integrada em dispositivos móveis, proporcionando assim uma experiência de usuário fluida e eficiente.

4.1.2 Treinamento final da CNN

A CNN SqueezeNet v1.1 implementada neste trabalho foi treinada utilizando a técnica de transferência de aprendizado, mais especificamente, o método de ajuste fino. Esse método aproveita um modelo previamente treinado e adapta apenas a camada final (de classificação) para a tarefa específica em questão, de modo a reconhecer as classes do problema em mãos. Para este projeto, a camada final foi configurada para uma saída com sete classes correspondentes às notas de 2, 5, 10, 20, 50, 100 e 200 Reais. Além disso, foram definidas múltiplas taxas de aprendizado, com uma taxa menor para a base do modelo, visando aproveitar o conhecimento prévio adquirido pelo modelo pré-treinado.

Entretanto, para executar esta etapa com sucesso, é necessário realizar um pré-processamento nas imagens, de forma a adequá-las ao formato esperado pela rede neural. Para isso, foi adotada uma etapa de padronização das imagens, que consiste em aplicar transformações específicas para garantir a consistência e qualidade dos dados de entrada.

Com o auxílio da função *transforms*, disponível no pacote *torchvision*, foram aplicadas as seguintes transformações:

1. *Transforms.Resize(256)*: Redimensiona a imagem para uma altura e largura de 256 pixels, garantindo consistência no tamanho das imagens de entrada.
2. *Transforms.CenterCrop(224)*: Realiza um recorte central na imagem para garantir que ela tenha uma dimensão final de 224x224 pixels, o tamanho padrão para muitas redes neurais convolucionais.
3. *Transforms.ToTensor()*: Converte a imagem para o formato de tensor, que é o formato esperado por qualquer rede neural treinada em PyTorch.
4. *Transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])*: Normaliza os valores dos pixels da imagem para uma escala específica, o que ajuda no processo de treinamento da rede neural, facilitando a convergência do algoritmo de otimização.

Essas transformações são cruciais para garantir que as imagens de entrada estejam adequadamente preparadas para o treinamento da rede neural, maximizando assim o desempenho e a eficácia do modelo.

Após a definição das transformações das imagens, foram estabelecidos os parâmetros e funções essenciais para o treinamento da rede neural:

- Taxa de aprendizado: Define a rapidez com que a rede neural aprende. Neste caso, o valor foi definido como 0.0001.
- Número de épocas: Quantidade de vezes que os dados são apresentados à rede durante o treinamento. Foi definido como 20.
- Tamanho do batch: Número de amostras do conjunto de dados usadas pela rede em cada época. Foi estabelecido

[1] https://github.com/karenalmeida18/recognize-notes-api

como 32.
- Penalidade: Controla a complexidade do modelo, penalizando pesos muito altos. O valor definido foi 0.0008.
- Função de perda: Métrica para avaliar o ajuste do modelo, calculando a diferença entre a predição e a saída real dos dados. Foi escolhida a função de entropia cruzada.
- Função de otimização: Tem como objetivo minimizar a perda durante o treinamento. Foi selecionada a função de otimização Adam.

Assim, durante o treinamento, 32 imagens do conjunto de dados foram processadas por vez, ao longo de 20 épocas definidas.

O tempo total necessário para o treinamento foi de aproximadamente quatro horas, representando um tempo consideravelmente reduzido em comparação com o treinamento de outras redes neurais convolucionais. Além disso, o modelo gerado apresentou um tamanho de apenas 2,7 MB após o treinamento, sendo extremamente mais leve em comparação com outras redes convolucionais, que podem chegar a ter tamanhos duzentas vezes maiores.

Esses resultados destacam a eficiência e a compactação do modelo treinado, tornando-o uma escolha viável e eficaz para aplicações em dispositivos com recursos limitados, como smartphones.

4.2 Desenvolvimento da API

A estrutura da API desenvolvida pode ser dividida em duas etapas essenciais para viabilizar a integração do modelo treinado com outros sistemas:

1. Entrada da imagem e envio ao modelo treinado, aplicando o pré-processamento: Nesta etapa, a API recebe a imagem como entrada e aplica as transformações de pré-processamento necessárias para adequá-la ao formato esperado pelo modelo treinado. Isso inclui redimensionamento, recorte e normalização da imagem.
2. Inferência no modelo treinado e atribuição das probabilidades resultantes para cada classe possível: Após o pré-processamento da imagem, a API realiza a inferência no modelo treinado, obtendo as probabilidades associadas a cada classe possível de cédula. Em seguida, é identificado o índice correspondente à classe com a maior probabilidade, determinando assim a classe à qual a imagem pertence.

Com base nesse processo, a API retorna como resposta a classe à qual a imagem pertence, facilitando assim a integração do modelo de reconhecimento de cédulas em outros sistemas.

A API desenvolvida está hospedada em um servidor online chamado Render, uma plataforma de nuvem que oferece um plano gratuito para hospedagem. Essa escolha foi feita visando facilitar o consumo e a integração da API em aplicativos e outros sistemas, garantindo assim uma maior acessibilidade e usabilidade.

4.3 Desenvolvimento do Aplicativo (*Banknotes Recognition*)

O aplicativo Banknotes Recognition foi desenvolvido com o objetivo de oferecer uma experiência de usuário simples e acessível, especialmente para usuários com deficiências visuais. Para isso, foi utilizado o React Native, uma estrutura de aplicativo móvel baseada em JavaScript que permite criar aplicativos nativos para iOS e Android. Além disso, o Expo foi utilizado para simplificar a integração da aplicação com a câmera e áudio dos dispositivos.

A implementação do *Banknotes Recognition* envolve a integração da API criada anteriormente com a interface do aplicativo, com ênfase em recursos de acessibilidade e usabilidade. O aplicativo incorpora as seguintes funcionalidades:

1. Captura de imagens: Os usuários podem capturar imagens das cédulas em tempo real usando a câmera do celular.
2. Integração com API: O aplicativo se conecta à API, que recebe a imagem capturada e retorna informações sobre o valor e a denominação da cédula.
3. Feedback auditivo e textual: O resultado da classificação é exibido visualmente e por áudio, priorizando a usabilidade para usuários com deficiência visual. A biblioteca expo-speech foi utilizada para essa integração via áudio.
4. Recursos de acessibilidade: O código foi estruturado seguindo boas práticas de acessibilidade, com elementos semânticos e considerando questões como contraste e navegação simplificada. Recursos adicionais foram implementados, como a captura de imagens por meio do toque na tela, com instruções fornecidas via áudio narrado.

As telas do aplicativo em funcionamento são mostradas na Fig. 2. A primeira tela é a interface inicial do aplicativo, que orienta o usuário sobre como iniciar a execução da aplicação. Na tela principal, o acesso à câmera do dispositivo é fornecido, com um botão central para capturar a imagem da cédula. Por fim, na terceira tela, é exibida a imagem resultante da captura da cédula, juntamente com a classificação da nota em uma caixa de texto. É importante destacar que todo o texto visual é narrado ao usuário durante a navegação pelas telas, garantindo uma experiência inclusiva e de fácil uso para o público-alvo.

5 Experimentos e resultados

5.1 Elaboração do conjunto de dados

O conjunto de dados utilizado neste artigo foi elaborado exclusivamente para este fim, utilizando imagens capturadas com um dispositivo móvel iPhone XR, equipado com uma câmera de 12 megapixels. As imagens das cédulas foram capturadas em uma variedade de cenários, incluindo cédulas amassadas, rasuradas, com diferentes fundos, dobradas, em superfícies lisas, entre outros. Além disso, foram variadas as posições e iluminação das cédulas para garantir uma ampla diversidade no conjunto de dados e aproximar-se de cenários reais de uso.

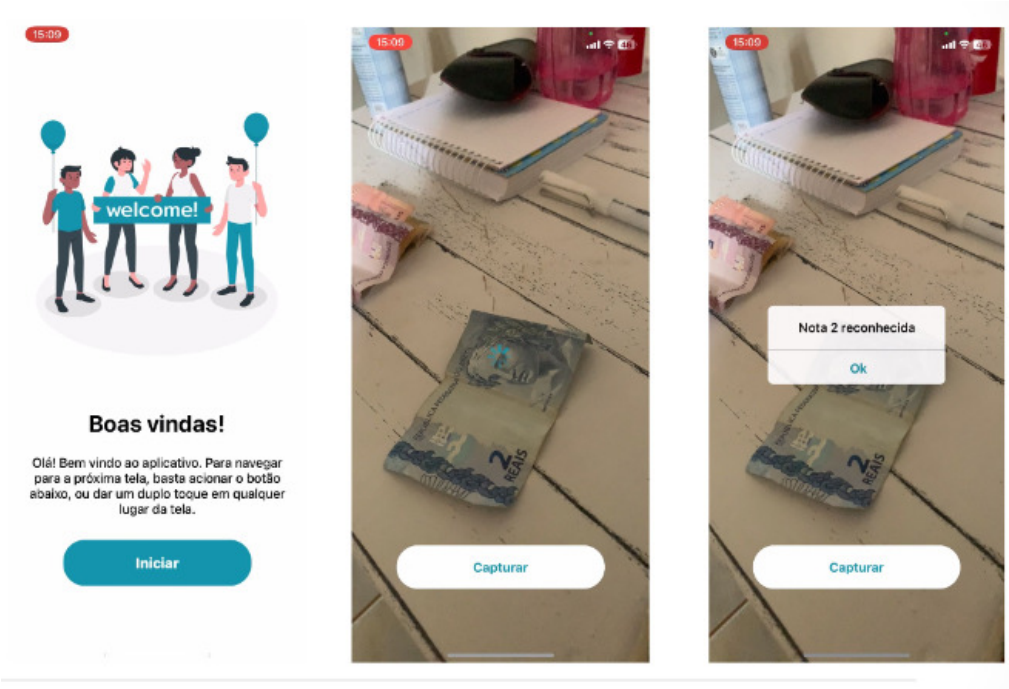


Figura 2: Telas do aplicativo *Banknotes Recognition*

No total, foram capturadas 1.369 imagens, que foram manualmente organizadas em pastas nomeadas de acordo com a classe a que cada cédula pertence. O conjunto de dados foi distribuído em sete classes distintas. Para o treinamento do modelo, 80% dessas imagens foram utilizadas, enquanto os 20% restantes foram reservados para o conjunto de validação. O conjunto de dados está disponível no endereço. [2]

A Tabela 1 apresenta o número total de imagens disponíveis para cada classe de cédula, após a distribuição entre os conjuntos de treinamento e validação.

Tabela 1: Quantidade de imagens obtidas após a separação entre treino e validação.

Classe	Treinamento	Validação
nota-2	164	42
nota-5	160	40
nota-10	209	47
nota-20	146	36
nota-50	177	43
nota-100	153	42
nota-200	87	23

Esse processo de elaboração do conjunto de dados garantiu uma representação diversificada e abrangente das diferentes classes de cédulas, proporcionando ao modelo uma ampla variedade de exemplos para aprendizado e validação. Dado que a quantidade de imagens no conjunto de dados original é considerada insuficiente para um treinamento adequado da rede neural convolucional, neste trabalho, emprega-se a técnica de aumento de dados. Essa técnica consiste na aplicação de diversas transformações nas imagens originais, com o objetivo de gerar um volume maior de imagens, aumentando assim a diversidade do conjunto de dados.

Para automatizar esse processo, foi utilizada a biblioteca auxiliar do Python chamada imgaug. As transformações aplicadas nas imagens para o aumento de dados abrangem uma variedade de variações, tais como translação, rotação em um intervalo de -45 a 45 graus, cisalhamento, ajuste de contraste, adição de ruído gaussiano e transformação de perspectiva.

Ao final do processo de aumento de dados, foram geradas 6.497 imagens para o conjunto de treinamento da rede neural, enquanto 273 imagens foram mantidas para o conjunto de validação. A Tabela 2 apresenta o número total de imagens para cada classe de cédula, após a distribuição entre os conjuntos de treinamento e validação, levando em consideração o conjunto expandido de imagens.

É importante destacar que a técnica de aumento de dados foi aplicada exclusivamente ao conjunto de imagens de treinamento, garantindo assim uma avaliação autêntica para o conjunto de dados de validação.

5.2 Avaliação dos resultados

No processo de avaliação dos resultados, foram selecionadas métricas como acurácia e matriz de confusão, desempenhando papéis cruciais na avaliação da eficácia do modelo. A acurácia é uma métrica amplamente utilizada

[2] https://www.kaggle.com/datasets/karenalmeida340/cdulas-do-real.

Tabela 2: Quantidade de imagens para treinamento após aumento de dados.

Classe	Treinamento	Validação
nota-2	1024	42
nota-5	967	40
nota-10	1160	47
nota-20	877	36
nota-50	1026	43
nota-100	895	42
nota-200	548	23

para medir o desempenho global do modelo. É calculada determinando a proporção entre o número total de previsões corretas e o número total de resultados, fornecendo uma avaliação da capacidade do modelo de fazer previsões precisas. Por outro lado, a matriz de confusão oferece uma representação visual e numérica do desempenho do modelo em relação a classes específicas, sendo relevante para a análise de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos, permitindo uma compreensão mais detalhada da capacidade de classificação do modelo em diferentes cenários.

Para um total de 273 imagens validadas, obteve-se uma acurácia de 99.63%. A Tabela 3 apresenta a matriz de confusão, destacando que a única classificação incorreta ocorreu ao identificar erroneamente uma nota de 20 reais como uma cédula de 50 reais. Esses resultados indicam um alto nível de precisão e desempenho do modelo no reconhecimento das cédulas, com uma taxa de erro mínima.

Tabela 3: Matriz de confusão para os dados de validação.

Notas	2	5	10	20	50	100	200
2	42	0	0	0	0	0	0
5	0	40	0	0	0	0	0
10	0	0	47	0	0	0	0
20	0	0	0	35	1	0	0
50	0	0	0	0	43	0	0
100	0	0	0	0	0	42	0
200	0	0	0	0	0	0	23

Além dos resultados apresentados na Tabela 3, foi conduzido um teste adicional do modelo para garantir sua robustez e capacidade de generalização. Nesse teste, foram adquiridas mais algumas imagens utilizando diferentes dispositivos com diversas câmeras, visando abranger uma ampla gama de características de imagem e qualidade visual. Além disso, um conjunto adicional de imagens foi obtido de fontes externas, como o Google Imagens, com o objetivo de diversificar ainda mais o conjunto de validação extra. Ao todo, foram utilizadas 66 imagens nesse conjunto adicional de validação.

A acurácia final obtida com este conjunto extra foi de 100%, indicando que todas as imagens foram classificadas corretamente pelo modelo. A Fig. 3 apresenta todas as 66 imagens avaliadas nesta validação extra. Observa-se que o aplicativo *Banknotes Recognition* foi capaz de lidar eficientemente com situações desafiadoras, como variações de iluminação, cédulas dobradas, obstruídas pela mão, variações de cores para uma mesma cédula, entre outras dificuldades.

Esses resultados adicionais confirmam a robustez e a capacidade de generalização do modelo, demonstrando sua eficácia em reconhecer cédulas em uma variedade de condições e ambientes, o que reforça a confiança em sua utilização prática.

Os experimentos práticos realizados confirmaram os resultados positivos do modelo, demonstrando sua capacidade de reconhecer notas em situações adversas, sem a necessidade de controle do ambiente. Além disso, uma característica positiva do aplicativo é sua simplicidade de uso, que inclui a captura de cédulas com um toque duplo em qualquer área da tela do smartphone. O tempo de resposta para o reconhecimento da cédula é bastante reduzido, geralmente apenas alguns segundos, o que contribui para uma experiência de usuário eficiente e intuitiva. Essas características tornam o aplicativo acessível e prático para uma ampla gama de usuários, incluindo pessoas com deficiências visuais ou com pouca experiência em tecnologia.

A Tabela 4 apresenta a matriz de confusão para os dados de teste extra, onde é possível observar uma taxa de acertos igual à 100%.

Tabela 4: Matriz de confusão para os dados de teste.

Notas	2	5	10	20	50	100	200
2	9	0	0	0	0	0	0
5	0	10	0	0	0	0	0
10	0	0	12	0	0	0	0
20	0	0	0	13	0	0	0
50	0	0	0	0	11	0	0
100	0	0	0	0	0	9	0
200	0	0	0	0	0	0	2

Na Fig. 4, são apresentados mais dois exemplos que ilustram o desempenho do aplicativo no reconhecimento de cédulas. No primeiro exemplo (a), é notável que a predição foi precisa, mesmo diante da presença de outras cédulas na área periférica da imagem. Isso ressalta a capacidade do aplicativo de identificar corretamente a cédula principal, mesmo em ambientes com múltiplos elementos. Já no segundo exemplo (b), uma situação interessante é observada: uma cédula é identificada com sucesso mesmo quando a imagem capturada é obtida ao apontar a câmera para uma tela de monitor de vídeo que exibe um site da Internet. Esse cenário destaca a adaptabilidade e a robustez do aplicativo, evidenciando sua habilidade em reconhecer cédulas mesmo em situações desafiadoras e inesperadas. Esses exemplos adicionais reforçam a eficácia e a confiabilidade do aplicativo em diversas situações do mundo real, demonstrando sua capacidade de oferecer resultados precisos e consistentes, independentemente das complexidades do ambiente de captura.

Embora o aplicativo não tenha sido testado diretamente com o público-alvo durante esta fase do trabalho, o desenvolvimento foi embasado em boas práticas de acessibilidade (como por exemplo, avisos sonoros), garantindo que as funcionalidades propostas atendam às necessidades identificadas para o reconhecimento de cédulas por

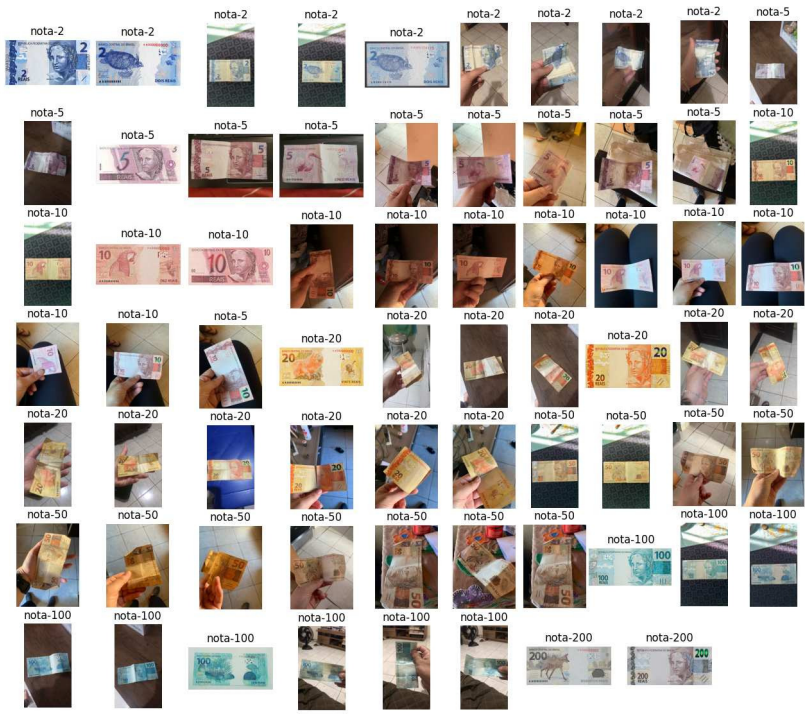
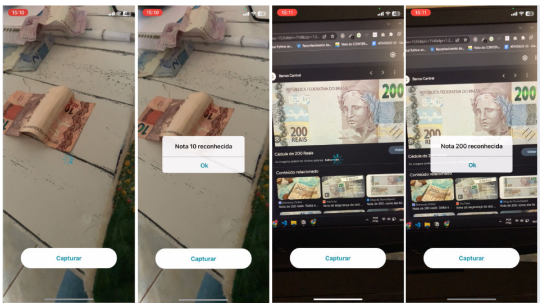


Figura 3: Imagens usadas no teste extra (100% de acurácia).



(a) **(b)**

Figura 4: Capturas do aplicativo em funcionamento reconhecendo as cédulas de 10 e 200 reais.

pessoas com deficiência visual.

6 Conclusões e trabalhos futuros

Diante dos resultados obtidos, com uma acurácia final de 100%, e considerando a facilidade de uso e a rapidez de resposta do aplicativo, a metodologia proposta neste trabalho mostra-se altamente promissora para aplicações em cenários reais. A escolha da rede neural SqueezeNet v1.1 revelou-se crucial, pois ela consegue lidar eficazmente com a complexidade de um problema com sete classes, ao mesmo tempo em que requer recursos computacionais mínimos.

Apesar de não ter sido testado com usuários finais até o momento, os resultados obtidos demonstram que o aplicativo possui potencial para atender às demandas do público-alvo, oferecendo uma solução inclusiva e eficiente para o reconhecimento de cédulas em cenários reais.

É inegável que as técnicas de inteligência artificial têm se mostrado extremamente eficazes na superação de obstáculos enfrentados por pessoas com deficiências em seu dia a dia. A interface do aplicativo, com *feedbacks* sonoros e capturas facilitadas por toques na tela do dispositivo, foram desenvolvidas para que o sistema seja inclusivo e de fácil uso para o público-alvo.

Este trabalho apresenta resultados que se destacam em termos de acurácia e otimização de recursos computacionais, especialmente quando comparados a estudos semelhantes, como o de Sausen and Frozza (2022). Nesse estudo, a arquitetura MobileNet V2 foi utilizada para o reconhecimento de cédulas de dinheiro em Real, alcançando uma precisão média de 95% no *dataset* de teste. É importante ressaltar, entretanto, que os resultados não são diretamente comparáveis, uma vez que cada trabalho utilizou bases de dados distintas, com características e desafios próprios que influenciam tanto a complexidade do problema quanto a avaliação dos modelos.

A escolha da arquitetura SqueezeNet v1.1 no presente trabalho mostrou-se um diferencial essencial. Essa rede conseguiu lidar eficientemente com sete classes diferentes, mantendo o tamanho final do aplicativo em apenas 2,7 MB, o que representa uma solução altamente otimizada em termos de memória e recursos computacionais.

Para trabalhos futuros, está previsto o aprimoramento do aplicativo *Banknotes Recognition*[3] para que possa operar offline, eliminando a necessidade de conexão com um

[3] https://github.com/karenalmeida18/recognize-notes-api

servidor para seu funcionamento. Além disso, serão exploradas estratégias para auxiliar os usuários no posicionamento correto das cédulas diante da câmera, possivelmente utilizando sinais sonoros para orientação. Também será realizada uma etapa de testes com o público-alvo, a fim de validar a usabilidade e eficácia do aplicativo em cenários reais. Essas melhorias têm como objetivo aperfeiçoar ainda mais a experiência do usuário e garantir maior acessibilidade em diferentes contextos e condições de uso.

Referências

Almeida, M. G. and Alves, M. A. F. (2019). Valor invisível: análise da acessibilidade das cédulas da segunda família do real entre pessoas com deficiência visual, *Anais da Mostra de Pesquisa em Ciência e Tecnologia* . Disponível em https://static.even3.com/anais/84522.pdf.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M. and Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions, *Journal of Big Data* **8**: 1–74. https://doi.org/10.1186/s40537-021-00444-8.

Aseffa, D. T., Kalla, H. and Mishra, S. (2022). Ethiopian banknote recognition using convolutional neural network and its prototype development using embedded platform, *Journal of Sensors* **2022**: 1–18. https://doi.org/10.1155/2022/4505089.

Barelli, F. (2018). *Introdução à visão computacional: Uma abordagem prática com Python e OpenCV*, Editora Casa do Código.

do Brasil, B. C. (2023). Cédulas e moedas. Disponpivel em https://www.bcb.gov.br/cedulasemoedas/cedulas.

Duarte, L. E. R. (2019). Reconhecimento de cédulas real utilizando algoritmo surf.

IBGE (2010). Pessoas com deficiência. Disponível em https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/20551-pessoas-com-deficiencia.html.

O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks, *arXiv preprint arXiv:1511.08458* . https://doi.org/10.48550/arXiv.1511.08458.

Sausen, F. S. and Frozza, R. (2022). Aplicativo para auxiliar pessoas com deficiência visual no reconhecimento de cédulas de dinheiro em real com a técnica de redes neurais artificiais, *Revista Brasileira de Computação Aplicada* **14**: 1–16. https://doi.org/10.5335/rbca.v14i3.13058.

Silva, K. W. A., Aires, K. R. T. and Britto Neto, L. S. (2024). Uma revisão sistemática sobre o reconhecimento de cédulas bancárias falsas por meio de visão computacional, *Revista Brasileira de Computação Aplicada* **16**: 45–59. https://doi.org/10.5335/rbca.v16i2.15388.

Sufri, N. A. J., Rahmad, N., As'ari, M. A., Zakaria, N., Jamaludin, M. N., Ismail, L. and Mahmood, N. (2017). Image based ringgit banknote recognition for visually impaired, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* **9**(3-9): 103–111. Disponível em https://jtec.utem.edu.my/jtec/article/view/3133.

Tsang, S.-H. (2019). Revisão: Squeezenet (classificação de imagens). Disponível em https://towardsdatascience.com/review-squeezenet-image-classification-e7414825581a.

VisoAI (n.d.). Vgg very deep convolutional networks (vggnet) – what you need to know. Disponível em https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/.

Zhang, A., Lipton, Z. C., Li, M. and Smola, A. J. (2021). Dive into deep learning 0.17.1 documentation. Disponível em https://pt.d2l.ai.

Zucatelli, F. H. G., Silva, N. D. P. and Suyama, R. (2017). Reconhecimento automático de cédulas do real baseado em máquinas de vetores suporte, *FTT Journal of Engineering and Business* (2). Disponível em https://saijournal.cefsa.org.br/index.php/FTT/article/view/50.