A Systematic Review on Product Recognition for Aiding Visually Impaired People

André Machado, Rodrigo Veras, Kelson Aires and Laurindo Britto Neto

***Abstract*—In recent years, Computer Vision and Machine Learning techniques have been extensively explored in the creation of assistive systems for the visually impaired. One of the most challenging tasks for visually impaired people is object recognition. In this paper, we conducted a systematic review to identify the current state of the art in designing these assistive systems. Due to the huge amount of object categories, we focused on recognizing products, such as those found in grocery stores, pantries and refrigerators. We analyze the techniques used, noting the efficiency and economy of hardware resources such as processing, memory and battery. Thus we verify if they can be used in wearable systems and adapted to existing devices of the Internet of Things (IoT), enabling the proposition of efficient and accessible assistive product recognition systems.**

***Index Terms*—Grocery Product Recognition, Object Recognition, Computer Vision, Wearable System.**

I. Introdução

Segundo a Organização Mundial de Saúde [1], estima-se haver no mundo cerca de 2,2 bilhões de pessoas cegas ou com deficiência visual severa. De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE) [2], no Brasil há 6,6 milhões de pessoas cegas ou com grande dificuldade de enxergar (3,4% da população brasileira), sendo 506,3 mil cegas (0,3% dos brasileiros).

Uma das formas mais importantes de como pessoas cegas ou com baixa visão percebem o mundo é o toque. No entanto, nem todos os objetos são acessíveis para serem sentidos, o que torna difícil perceber o objeto real. Além disso, o processo de aprendizagem para identificar objetos com o toque é muito mais lento do que identificar objetos pela visão. Isso se deve ao fato de que o objeto precisa ser abordado e cuidadosamente sentido, até que uma ideia aproximada possa ser construída no cérebro [3]. Há ainda o problema de existirem objetos diferentes visualmente, porém idênticos ao tato, como pode ser observado na Fig. 1.

Existem diversos campos da área de visão computacional que são explorados na criação de sistemas para o auxílio de pessoas com deficiências visuais, e.g.: detecção de obstáculos [4, 5], reconhecimento de caminho [6], leitura de textos [7, 8], identificação de cédulas monetárias [9], localização e reconhecimento de números de linhas de ônibus [10], assistência para localizar e alcançar objetos [11], reconhecimento de padrões [12], percepção do ambiente em 3D para navegação [13], ajuda a localizar e a atravessar a faixa de pedestres [14, 15], auxílio a encontrar semáforos para pedestre [16], autolocalização em cruzamentos de ruas [17], identificação das cores [18], dentre outros.

(a) (b) (c)

Fig. 1: Objetos análogos ao tato: (a) tubo de cola e protetor solar labial, (b) frasco de colírio e de descongestionante nasal, e (c) uma coleção de DVDs e Blu-rays.

Para ajudar pessoas com deficiências visuais à reconhecer objetos, diversas tecnologias têm sido desenvolvidas utilizando métodos de visão computacional, algumas com pontos fortes em relação às demais. Entretanto, cada um delas possui suas próprias limitações, inclusive na interface com o usuário. Tais tecnologias tiveram sucesso limitado ao reconhecer objetos, devido às condições descontroladas nos ambientes, como: grandes variações na iluminação, brilho, fundo, orientação do objeto em relação à câmera, oclusões parciais, tamanho e distância dos objetos etc. Segundo Jafri et al. [19], o reconhecimento de objetos no auxílio de pessoas com deficiências visuais é um campo que ainda está no princípio.

Jafri et al. [19] categorizaram as abordagens em visão computacional para o reconhecimento de objetos no auxílio a pessoas com deficiências visuais em dois grupos: (1) Baseadas em Etiquetas (BEs) e (2) Não Baseadas em Etiquetas (NBEs). As abordagens BEs se limitam a identificação de objetos previamente rotulados. Alguns exemplos de sistemas dessa categoria são: Badge3D [20], fornece o reconhecimento de objetos e detecção de obstáculos; Trinetra [21], auxilia cegos na identificação de produtos em uma mercearia; Al-Khalifa et al. [22], propõe a adição de *QR codes* a objetos.

As abordagens NBEs utilizam informações físicas do objeto, como a sua forma, tamanho e cor, para determinar a sua identidade. Jafri et al. [19] subdividiram essa categoria em

Esta pesquisa foi realizada com apoio da Fundação de Amparo à Pesquisa do Estado do Piauí (FAPEPI), por meio dos editais FAPEPI/CAPES Nº 005/2018 e FAPEPI/MCT/CNPq Nº 007/2018 (PPP).

A. L. Machado é mestrando do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Piauí (PPgCC/UFPI), Teresina, Brasil (e-mail: 4ndre.machado@gmail.com).

R. M. S. Veras é professor adjunto do Deparatamento de Computação da Universidade Federal do Piauí (DC/UFPI), Teresina, Brasil (e-mail: rveras@ufpi.edu.br).

K. R. T. Aires é professor associado do Departamento de Computação da Universidade Federal do Piauí (DC/UFPI), Teresina, Brasil (e-mail: kelson@ufpi.edu.br).

L. S. Britto Neto é professor adjunto do Departamento de Computação da Universidade Federal do Piauí (DC/UFPI), Teresina, Brasil (e-mail: laurindoneto@ufpi.edu.br).

mais duas: (1) Modelagem 3D e (2) Modelagem 2D.

Na Modelagem 3D é possível a utilização de visão estéreo para construir modelos 3D de objetos, de modo semelhante ao que ocorre com os olhos humanos [23]. Por exemplo, Hub e Hartter [24] desenvolveram um sistema de modelagem 3D para reconhecer objetos e também rastrear objetos móveis. Entretanto, essas abordagens aumentam o custo, pois exigem mais de uma câmera.

Na Modelagem 2D, geralmente, é utilizada uma única câmera para capturar dados de imagem. Isso viabiliza o seu uso em dispositivos *wearables* ("vestíveis"), tornando-a mais viável para auxiliar pessoas com deficiência visual. O reconhecimento é então realizado com base em várias características extraídas desses dados. Alguns exemplos de sistemas dessa subcategoria são: Schauerte et al. [25] usaram uma *webcam* anexada ao pulso do usuário para detectar objetos; o sistema DORA [26] faz uso da cor e das bordas detectadas para reconhecimento de objetos, usando Redes Neurais Artificiais [27]; Kumar e Ganesan [28] criaram um reconhecedor para objeto, imagens ou faces; o sistema criado por Fuangkaew e Patanukhom [29] busca imagens na internet automaticamente, para reconhecer objetos não cadastrados no sistema; Drishti [30] é um sistema que reconhece face, objeto e texto; Balasuriya et al. [31] encontravam objetos em ambientes interno e externo, empregando *Regions with CNN feature* (R-CNN) [32] e *Recurrent Neural Network* (RNN) [33]; Kacorri et al. [34] reconheceram objetos usando um pequeno número de exemplos; o sistema de Sosa-García e Odone [35] associa um objeto com uma determinada marca, um modelo ou um tipo; Intelligent eye [36] é um sistema que realiza o reconhecimento de objetos e de notas bancárias, além de detecção de luz e de cores, utilizando CNNdroid [37].

Como existe uma grande variedade de classes ou categorias distintas de objetos que as pessoas com deficiência visual possuem necessidade de reconhecer, foi necessário definir que tipo de objetos será trabalhado nesta pesquisa. A classe escolhida foi a de produtos que, geralmente, são encontrados em prateleiras de mercearias, como também em prateleiras de supermercados, geladeiras ou despensas. Um sistema para reconhecer tal categoria de objetos seria bastante útil para uma pessoa com deficiência visual, visto que ele poderia verificar, sem ajuda de terceiros, se na geladeira de sua casa ou em sua despensa, por exemplo, há uma caixa de leite ou de suco.

Nos últimos anos, tem-se visto o advento dos dispositivos *wearable* [38, 39], ou "tecnologia vestível", i.e., qualquer dispositivo de computação que seja pequeno o bastante para ser usado ou transportado junto ao corpo, incorporado a acessórios ou utilizados em peças de roupa, tais como óculos, relógio inteligente (*smartwatch*), telefone inteligente (*smartphone*), dispositivos portáteis, entre outros. Eles vêm ganhando popularidade com o crescimento da chamada Internet das Coisas (*Internet of Things* – IoT) [40], um novo paradigma que, rapidamente, ganha mais espaço no cenário moderno das telecomunicações. A ideia básica desse conceito é a presença pervasiva de uma variedade de coisas ou objetos, como etiquetas RFID, sensores, atuadores, telefones celulares etc. – que, por meio de esquemas de endereçamento único, são capazes de interagir e cooperar uns com os outros para atingir objetivos comuns [41].

De acordo com Britto Neto et al. [42], sistemas que prestam auxílio a usuários com deficiência visual, no reconhecimento do ambiente circundante, são complexos devido a limitações tecnológicas do projeto, isto é, hardware e algoritmos para o reconhecimento de objeto, bem como em razão de aspectos sociotécnicos, isto é, a interação entre o usuário, a tecnologia e todo o contexto de uso. Com base nisso, soluções mais robustas devem considerar incrementar ao máximo os seguintes fatores: dispositivo portátil; sistema *wearable*; baixo custo; execução em tempo real; baixo consumo de memória, poder de processamento e bateria; alta taxa de acertos; robustez em diversos tipos de ambientes; êxito em condições de ambiente descontroladas; *feedback* apropriado a pessoas com deficiências visuais; facilidade de manuseio.

O objetivo deste trabalho é, por meio de uma revisão sistemática da literatura, analisar os estudos mais relevantes no reconhecimento de produtos de mercearia, identificando o estado da arte, bem como as lacunas existentes. Deve ser viável aplicar essas abordagens em tecnologias *wearable* ao auxílio de pessoas com deficiências visuais. Serve, portanto, como base para a proposição de novas abordagens voltadas ao auxílio de pessoas com deficiências visuais no reconhecimento de itens mercantis. Este trabalho está inserido no contexto de um projeto maior, atualmente financiado pela parceria entre a FAPEPI e o CNPq, que visa desenvolver novas abordagens em visão computacional para auxiliar pessoas com deficiência visual a realizar tarefas do seu cotidiano.

Este artigo está organizado da seguinte forma: a Seção II descreve os passos executados na realização da revisão sistemática; a Seção III relata a revisão da literatura, destacando pontos relevantes dos trabalhos selecionados; a Seção IV apresenta e discorre sobre as bases de dados utilizadas por esses estudos; a Seção V discute importantes pontos deste trabalho; e, finalmente, as conclusões e perspectivas para trabalhos futuros são fornecidas pela Seção VI.

II. Metodologia da Revisão de Literatura

Neste trabalho foi utilizada a ferramenta StArt (*State of the Art through Systematic Review*) [43] para realizar uma revisão sistemática da literatura baseada na metodologia PRISMA [44], procurando responder à seguinte questão da pesquisa: "Quais abordagens em reconhecimento de objetos, em específico produtos comumente encontrados em prateleiras de mercearia, para auxílio de pessoas com deficiência visual, mais se aproximam do estado da arte e são facilmente adaptáveis a sistemas *wearable*?"

As seguintes etapas foram seguidas:

1) **Planejamento**: definição dos objetivos, questões de pesquisa, palavras-chave, bases de busca bibliográfica e critérios de seleção;
2) **Execução**: busca nas bases bibliográficas escolhidas;
3) **Seleção**: leitura das partes principais do trabalho, respondendo aos critérios de inclusão e exclusão;
4) **Extração**: leitura completa dos trabalhos selecionados, respondendo às questões de pesquisa e reaplicando os critérios de inclusão e exclusão;

5) **Sumarização**: criação da tabela com o resumo da revisão e discussão dos resultados.

A. Definição das Questões de Pesquisa

Com base na questão de pesquisa inicial, foram elaboradas outras perguntas, que foram aplicadas a cada um dos artigos selecionados em uma etapa posterior.

1) Qual a abordagem proposta?
2) Quais as bases de dados utilizadas?
3) Quais experimentos foram realizados?
   a) Qual pré-processamento foi realizado na base de dados?
   b) Como a base de dados foi dividida?
   c) Quais metodologias e métricas de comparação foram usadas?
   d) Quais abordagens foram comparadas?
4) Quais as abordagens com melhores resultados nas bases de dados utilizadas?

Com base nesses questionamentos sobre a literatura, pretendeu-se identificar as abordagens que geraram melhores resultados, seus campos de atuação, seus pontos fortes e fracos. Desse modo, este trabalho possibilita uma análise que forneça, para pesquisadores iniciantes nesta área de pesquisa, um ponto de partida para a proposição de novas abordagens em reconhecimento de objetos, dentro do escopo de aplicações voltadas para pessoas com deficiência visual, que alcance melhores taxas de acurácia e economize recursos do hardware *wearable*, como processamento, memória e bateria.

Apesar de se relatar um pouco sobre a fase de detecção realizada por alguns dos trabalhos estudados, focou-se na etapa do reconhecimento dos objetos. Levou-se em conta as técnicas utilizadas, os descritores e classificadores comparados, a metodologia experimental e quais os melhores resultados obtidos nas bases de dados utilizadas.

B. Seleção das Bases Bibliográficas

As bases bibliográficas IEEE Xplore Digital Library (IEEE Xplore), ACM Digital Library (ACM) e Elsevier's Scopus (Scopus) foram selecionadas pelos seus grandes acervos, pela relevância que possuem em ciência da computação, por terem a funcionalidade de exportar as referências dos trabalhos no formato *.bibtex*, e por serem acessíveis pelo Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PP-CAPES).

A biblioteca ACM é híbrida, pois, além dos estudos primários, também indexa trabalhos oriundos de outras bases. Já biblioteca Scopus consiste em um motor de busca capaz de retornar artigos indexados por diversas bases bibliográficas, em que várias delas são ou totalmente ou parcialmente acessíveis pelo PP-CAPES. Por exemplo, a base bibliográfica Springer teve diversos artigos indexados pela Scopus. Contudo, nem todos os artigos da Springer eram acessíveis pelo PP-CAPES. As bases ACM e IEEE Xplore também tiveram estudos indexados pela Scopus, resultando em alguns resultados duplicados, a maioria detectada automaticamente pelo StArt.

C. Processo de Busca

O processo de busca foi refinado de modo a identificar um número relevante de artigos dentro da abordagem escolhida, além de filtrar aqueles que fogem ao tema da pesquisa. Considerou-se, primeiramente, as abordagens em visão computacional que se aproximam do estado da arte em reconhecimento de objetos, viáveis ao auxílio para pessoas com deficiências visuais, tendo como ponto secundário a abordagem ser aplicável a dispositivos *wearable*. Devido à grande amplitude desse campo, restringiu-se o espaço de busca para produtos usualmente encontrados em prateleiras de mercearias, supermercados, geladeiras ou despensas.

Na definição das *strings* de busca, foram realizados vários testes empíricos, com várias palavras, até que elas retornassem uma boa quantidade de artigos relevantes. Os melhores resultados foram obtidos com as palavras-chave "*grocery recognition*" e "*computer vision*". Assim, exigiu-se que as *strings* de busca "*grocery*" e "*recognition*" estivessem no título, no resumo ou nas palavras-chave. Conforme pode ser visto na Tabela I, cada base bibliográfica teve a sua respectiva *string* de busca.

TABELA I
Bases bibliográficas e respectivas *strings* de busca

Base de busca	*String* de busca
IEEE Xplore	*grocery recognition*
ACM	acmdlTitle:(*grocery recognition*) AND recordAbstract:(+*grocery* +*recognition*)
Scopus	(TITLE-ABS-KEY (*grocery* AND *recognition*)) AND ("*computer vision*") AND (LIMIT-TO (SUBJAREA, "COMP"))

A última busca foi realizada no mês de dezembro do ano 2019, então um número maior de trabalhos poderá ser encontrado caso a mesma busca seja refeita, visto que as bases estão periodicamente incrementando artigos a seus acervos. Na Fig. 2, é exibida a contribuição das bases de busca bibliográfica para os estudos encontrados, com base nas *strings* de busca utilizadas.

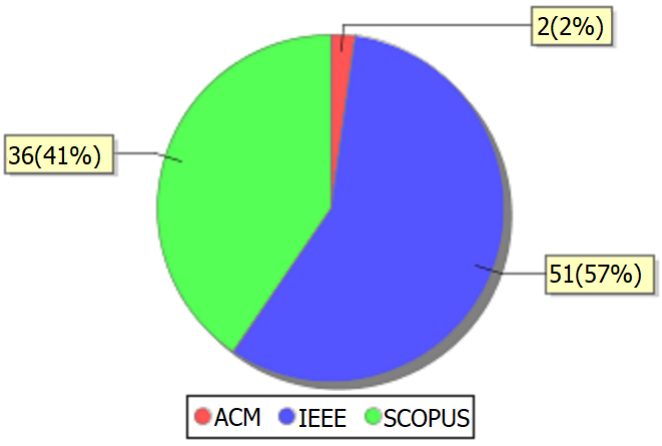


Fig. 2: Quantidade de estudos encontrados em cada base de busca bibliográfica observada.

D. Triagem de Artigos

Os critérios para seleção (inclusão e exclusão) para a Triagem de Artigos foram definidos em conformidade com os objetivos e questões de pesquisa.

1) Critérios de Inclusão:

- Ser escrito em inglês *AND*;
- Ser acessível pelo sistema Periódicos da Capes *AND*;
- Tratar de reconhecimento de produtos de mercearia.

2) Critérios de Exclusão:

- Ser baseado em etiquetas *OR*;
- Utilizar visão estereoscópica *OR*;
- Concentrar-se em Reconhecimento Ótico de Caracteres (OCR) *OR*;
- Focar na detecção dos objetos em imagens de prateleira.

3) Processo de Seleção dos Estudos: O processo de seleção de estudos ocorreu por meio da leitura do título e do resumo de todos os trabalhos, de modo a identificar a adequação aos critérios de seleção. Houve também uma leitura parcial de alguns trabalhos, visando responder aos critérios de seleção.

A base ACM retornou apenas dois artigos na busca inicial. A base IEEE Xplore apanhou 51 estudos, o maior número entre as três. A base Scopus teve 36 trabalhos encontrados inicialmente, porém vários deles também foram obtidos nas outras bases. Do total de 89 estudos, foram encontradas e removidas 19 duplicações, resultando em 70 trabalhos, dos quais se rejeitou 43 por não satisfazerem a todos os critérios de inclusão. Portanto foram aceitos 27 artigos para a fase de extração. O total de artigos aceitos, rejeitados e duplicados pode ser visto na Fig. 3.

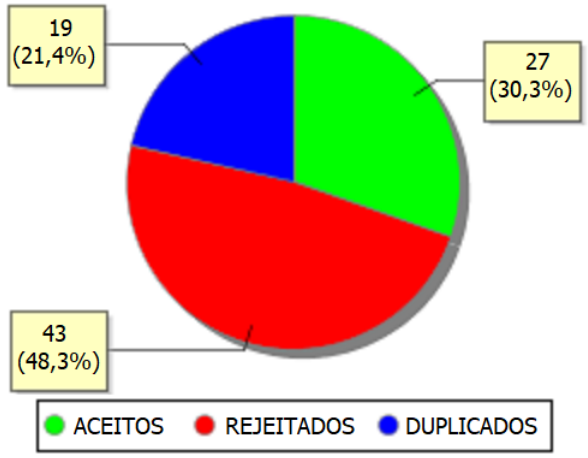


Fig. 3: Quantidade de estudos aceitos, rejeitados e duplicados na fase de seleção.

E. Extração de Dados e Sumarização

Durante a etapa de extração, continuou-se a aplicar os critérios de inclusão e exclusão aos artigos aceitos anteriormente. Também foram analisadas as referências desses trabalhos, como uma estratégia para a inclusão de mais trabalhos relevantes, resultando na adição de mais um trabalho.

Após uma leitura completa, a base IEEE Xplore alcançou sete estudos aceitos: [45–51]. A base ACM teve um trabalho aceito: [52]. A base Scopus obteve dois artigos aceitos: [53, 54], mais o artigo [55], que não foi encontrado em nenhuma das bases de busca bibliográficas, mas foi alcançado pela análise de referências de outro artigo da Scopus. Com isso, um total de 11 trabalhos atenderam a todos os critérios de inclusão e não incorreram em nenhum dos critérios de exclusão.

Finalmente, na etapa de sumarização, resumiu-se as principais informações dos trabalhos aceitos, respondendo qual foi a abordagem proposta, em quais bases de dados realizaram os experimentos, quais comparações em reconhecimento de objetos foram feitas e quais abordagens obtiveram os melhores resultados nas bases de dados utilizadas. Na Fig. 4, são apresentados quantos trabalhos foram aceitos e quantos foram rejeitados nessa fase, considerando também o artigo adicionado pela análise de referências dos estudos.

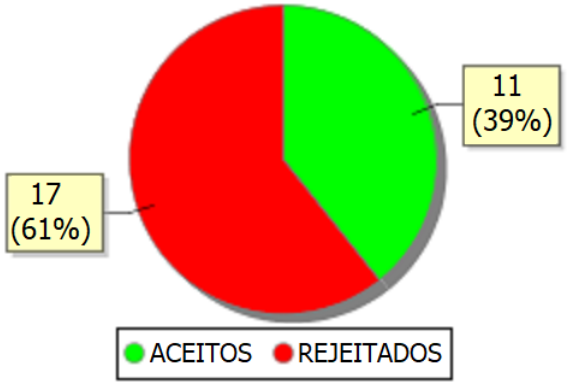


Fig. 4: Quantidade de estudos aceitos e rejeitados na fase de extração.

III. Análise dos Resultados

A seguir estão relacionados os trabalhos selecionados nesta revisão da literatura. As primeiras pesquisas encontradas envolvem a utilização de descritores de características, tais como *Scale-Invariant Feature Transform* (SIFT) [56] e *Speeded Up Robust Features* (SURF) [57]. Em seguida, foi incrementado aos trabalhos o uso de *Bag-of-Words* (BoW) [58], além de *Support Vector Machine* (SVM) [59] para classificação. Os sistemas mais recentes dessa revisão literária apresentaram o uso de redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs) [60]. As CNNs são, portanto, o atual estado da arte em reconhecimento de objetos. A seguir estão relacionados os trabalhos selecionados nesta revisão da literatura, divididos quanto à abordagem ser principalmente baseada em descritores de características ou em CNNs.

A. Abordagens Baseadas em Descritores de Características

Em Merler et al. [45] foi proposto o uso do algoritmo SIFT combinado a uma abordagem baseada em *K-Nearest Neighbors* (K-NN) [61]. O artigo introduz a base de dados GroZi-120 (veja a Seção IV para mais detalhes sobre as bases de dados). Houve uma comparação entre três descritores: SIFT; Boosted Haar-like Features – mais conhecido pelo nome dos autores, Viola-Jones [62]; e *Color Histogram* (CH) [63], testado com os espaços de cores YCbCr, HSV e Lab – o

terceiro espaço de cor obteve o melhor resultado, apenas com os canais ab. Foram comparados quatro classificadores. A primeira abordagem foi baseada em K-NN. A segunda abordagem utilizou CH com três diferentes métricas para cálculo de distância: Distância Euclidiana (ED), χ^2 (Qui-quadrado) e Interseção de Histograma (*Histogram Intersection* - HI) [64] usando *L1 Distance* (Distância Manhattan), obtendo uma melhor performance com HI. A terceira abordagem obteve o melhor resultado ao utilizar o descritor SIFT, computando um BoW às amostras de treinamento de cada produto. A quarta abordagem usou Viola-Jones, realizando o treinamento com o algoritmo Adaboost (ADA) [65]. A média das curvas ROC de todos os produtos foi a métrica utilizada no reconhecimento de objetos. Baseando-se no resultado da média das curvas ROC de todos os produtos, a abordagem com SIFT foi a melhor, devido à sua invariância à escala, à rotação e por outros fatores. A comparação do histograma de cores mostrou um desempenho muito bom, o que se justifica, já que os produtos de mercearia tendem a ser coloridos para atrair a atenção e se distinguir dos concorrentes. O método de Viola-Jones, no entanto, teve um desempenho apenas ligeiramente melhor que o aleatório.

Winlock et al. [46] desenvolveram o Shelf-Scanner, um programa que combina descritores do algoritmo SURF com CH para gerar novos descritores, reduzindo a dimensionalidade deles por meio de Análise de Componentes Principais (*Principal Component Analysis* – PCA) [66]. O CH usou o sistema de cores Lab, descartando o canal L de luminância, devido à grande variação de luminosidade entre as imagens de treinamento e teste. O classificador multi-classe Naive Bayes inspirado por NIMBLE [67] foi usado para realizar o reconhecimento de produtos de supermercado. Ainda, o trabalho utilizou a ideia de construir mosaicos, progressivamente, para distinguir as novas regiões a serem processadas daquelas que já apareceram nos quadros anteriores. O sistema recebe itens que estão em uma lista de compras e tenta encontrá-los. O sistema usa a base de dados GroZi-120. Melhorias foram propostas para que o sistema funcione em tempo real.

Rivera-Rubio et al. [47] submeteram três abordagens: SIFT + K-Means [68] + SVM; SIFT + *Locality-constrained Linear Coding* (LLC) [69]; e SIFT + PCA + *Fisher Vector Encoding* (FV) [70] + SVM. No algoritmo de agrupamento K-means, usaram duas diferentes quantidades de palavras visuais. Ainda, utilizaram dois valores distintos para a densidade de amostras no método FV. Os autores introduziram a base de dados SHORT-30, que posteriormente foi atualizada e passou a se chamar SHORT-100. Foram comparadas as acurácias mínima, média e máxima entre as abordagens propostas, realizando testes em um conjunto de imagens capturadas por *smartphones* e outros em um conjunto de imagens extraídas de vídeos. No primeiro conjunto, a melhor acurácia média foi de 77,51%, da abordagem SIFT + K-Means + SVM, enquanto a maior acurácia foi de 97,64%, alcançada pela abordagem SIFT + LLC. No segundo conjunto, a melhor acurácia média e a maior acurácia foram alcançadas pela abordagem SIFT + LLC, resultando, respectivamente, em 69,41% e 91,88%.

Varol e Salih [53] consideraram o problema do reconhecimento de uma mercadoria específica em imagens de prateleiras de diversos estabelecimentos. Trabalhou-se especificamente com embalagens de cigarro, montando o banco de dados Grocery, que pode ser utilizado tanto para detecção quanto ao reconhecimento dos produtos já recortados. Em primeiro lugar, os objetos de interesse foram detectados utilizando um classificador em cascata (Viola-Jones), treinado com características de Histograma de Gradientes Orientados (*Histogram of Oriented Gradients* - HOG) [71] para identificar as embalagens de cigarro. A região detectada é formada pela imagem do logotipo e da advertência da embalagem. Assim, foi utilizado para classificação apenas 40% da imagem do topo da embalagem, onde se encontra o logotipo, removendo a parte de advertência que é comum entre as marcas de cigarro. Posteriormente, a imagem do logotipo foi representada combinando informações de forma e cor, por meio da abordagem BoW. As informações de forma foram descritas pelo SIFT, enquanto as cores pelo modelo de cores HSV, devido sua melhor invariância à iluminação. Com a ajuda do K-means foram criados vocabulários visuais de forma, agrupando os vetores de característica SIFT, e de cores, agrupando os valores tridimensionais do HSV. Com isso, para cada imagem, o descritor BoW foi formado pela concatenação dos histogramas normalizados dos vocabulários de forma e cor. Finalmente, foi utilizado o classificador SVM no reconhecimento da marca do produto. A acurácia foi medida usando apenas SIFT (85,9%), apenas HSV (60,5%) e com ambos (92,3%).

Em Advani et al. [48], o protótipo Third-Eye foi apresentado para auxiliar pessoas com deficiências visuais a realizar compras de supermercado, por meio da combinação de aceleradores de hardwares, algoritmos de visão computacional, com uma câmera cada, um par de óculos e uma luva *wearable*. A primeira tarefa do sistema é identificar a categoria de produtos (*coarse classification*) pela câmera dos óculos, utilizando SVM com *Gist Of The Scene* [72, 73]. O usuário repete esse processo até chegar ao corredor desejado. O próximo passo é delimitar o conteúdo importante da imagem. Isso é feito segmentando a imagem para encontrar Regiões de Interesse (*Regions of Interest* - RoI), por meio de *Template Matching* [74], gerando sementes por meio da técnica de Atenção pela Maximização da Informação (*Attention by Information Maximization* - AIM) [75], e recorrendo ao SURF para o reconhecimento do produto (*fine-grained classification*). Passa-se a usar a luva para guiar o usuário a pegar o item, por meio de sua câmera mais quatro motores de vibração. Usou-se ainda uma *Shallow* CNN – isto é, uma CNN rasa, geralmente com apenas uma camada oculta – combinada com descritores HOG, a fim de identificar outras pessoas no supermercado e evitar colisões com elas. Foi utilizada uma Unidade de Processamento Gráfico (*Graphics Processing Unit* - GPU) para diminuir a latência. Por fim, os pesquisadores construíram um grafo para ajudar no reconhecimento dos itens que normalmente são encontrados juntos no supermercado, em que os nós do grafo correspondem aos produtos contendo a quantidade de vezes que foram vistos, enquanto cada aresta do grafo armazena a quantidade de vezes que dois produtos conectados foram vistos juntos.

B. Abordagens Baseadas em CNNs

Jund et al. [55] fizeram uso de Transferência de Aprendizagem (*Transfer Learning*) [76], que consiste em reutilizar

uma CNN pré-treinada em uma outra base de dados, tendo escolhido a arquitetura CaffeNet [77]. Foi então realizado o Ajuste Fino (ou *Fine Tuning*) sobre a CNN (*Fine-tuned CNN* - FT-CNN), que consiste em definir as camadas treináveis e substituir as últimas camadas por novas, de classificação e de saída. Foi alcançada uma acurácia média de 78,9% na base de dados Freiburg Groceries, criada pelos próprios autores.

Franco et al. [54] apresentaram uma abordagem para a detecção e reconhecimento de produtos em prateleiras de supermercados. No estágio do reconhecimento, eles exploraram duas abordagens. Na primeira, eles utilizaram o descritor SURF, o método de agrupamento K-means e o modelo BoW. Na abordagem seguinte, eles utilizaram a arquitetura AlexNet [78] – uma CNN pré-treinada em mais de um milhão de imagens do ILSVRC12, subconjunto da ImageNet [79] – usando ED. Durante o processo realizado pelo SURF, a imagem foi dividida em regiões para que o histograma de ocorrências fosse calculado em cada uma e não de maneira global, concatenando a imagem completa em seguida. A base de dados GroZi-120 foi tomada para os testes. Os autores compararam os descritores BoW (com SURF e K-means) e CNN. No reconhecimento, eles compararam apenas as abordagens BoW e CNN, cada uma com e sem pré-seleção baseada em cor. CH, SIFT e ADA foram comparados apenas pela abordagem de detecção. A métrica de comparação de reconhecimento foi, para cada produto, o cálculo da curva ROC, calculando a Área sob a curva ROC (*Area Under ROC* - AUROC) para identificar a acurácia global. Os valores AUROC calculados mostraram que o desempenho das duas abordagens é muito semelhante.

Geng et al. [52] elaboraram um *framework* que detecta e reconhece produtos mercantis, fazendo uso de aprendizado *One-Shot* [80], que requer apenas um exemplo de treinamento de cada classe, possuindo ainda a capacidade de adicionar novos produtos sem a necessidade de retreinar todas as classes. Eles executaram testes com e sem *Attention Map* (ATmap) [81], que é empregado para identificar regiões discriminativas, tendo bons resultados com os algoritmos BRISK [82] e SIFT. Escolheram o algoritmo SIFT como descritor de características, por ele apresentar melhor equilíbrio entre precisão e *recall*. Para reconhecer as instâncias detectadas, recorreram a um classificador baseado em CNN muito profunda, o VGG-16 [83], que foi pré-treinado e ajustado pela biblioteca Keras [84]. Os testes foram feitos na base GroZi-120, além de outras bases de dados, que tem imagens de prateleiras no conjunto de testes. Os autores compararam os descritores SIFT, OpponentSIFT [85], C-SIFT [85], RGB-SIFT [85], BRISK, SURF e sem ATmap. Também compararam os classificadores ResNet-50 [86] e VGG-16. Quatro abordagens foram comparadas: [87], [54], [88] e VGG-16 com ATmap usando SIFT, proposta pelos autores. Os autores aplicaram os mesmos protocolos de avaliação dos trabalhos relacionados. As abordagens sem ATmap, com SIFT e com BRISK foram testadas e comparadas com o estado da arte das bases de dados, observando os atributos precisão, *recall* e *mean average precision* (mAP). Em cada comparação com o estado da arte das bases de dados, algumas dessas abordagens propostas obtiveram o melhor resultado em ao menos um dos parâmetros observados.

Kanuri et al. [49] implementaram um aplicativo que reconhece produtos mercantis e *commodities*. Dispuseram-se do modelo Inception v3 [89]. Baseando-se ainda em aspectos desse modelo, os autores usaram uma arquitetura formada por uma rede de classificadores, em que um classificador mestre recorre a um cálculo de pontuação para redirecionar uma dada imagem aos três mais bem pontuados dentre dez classificadores independentes, todos implementações do Inception. O classificador mestre foi treinado em 4.000 épocas e os demais em 500. A base de dados utilizada foi criada pelos próprios autores.

Klasson et al. [50] utilizaram três diferentes CNNs pré-treinadas: AlexNet, VGG-16 e DenseNet-169 [90]. Elas serviram como descritores de características, sendo combinadas a um classificador SVM. Foram extraídas diferentes camadas das CNNs como vetores de características, com e sem a aplicação de *Fine-tuning* sobre elas. Também foram combinadas com uma outra técnica de aprendizado de máquina, chamada *Variational Autoencoders* (VAEs) [91], além de *Canonical Correlation Analysis* (CCA) [92], porém não melhoraram os resultados nos testes. Ademais, foi proposta uma outra abordagem, utilizando apenas FT-CNNs como descritores e classificadores. Por fim, os autores criaram a base de dados Grocery Store, que permite testes de classificação para reconhecer produtos específicos (*fine-grained classification*) e também de categorias de produto (*coarse-grained classification*). Os testes com produtos específicos foram melhores utilizando a FT-CNN DenseNet-169 com o classificador SVM, obtendo uma acurácia de 85,0%. Já os testes para classificar categorias de produtos tiveram melhor resultado usando a CNN DenseNet-169 sem *Fine-Tuning* com o classificador SVM, alcançando 85,2% de acurácia.

Pintado et al. [51] manipularam uma CNN em um protótipo de Raspberry Pi 3 para construir um par de óculos especializado em reconhecer itens da seção de produtos em uma mercearia, o qual captura uma imagem usando uma câmera acoplada, identifica o produto que se encontra na imagem e reporta o seu preço ao usuário. Para o treino da CNN foram empregadas as bibliotecas Tensorflow [93] e Keras, ambas de código aberto. Durante o pré-processamento, redimensionaram as imagens para 100x100 e as converteram para a escala de cinza. Desse modo, o treinamento visou encontrar nas imagens, características importantes não afetadas pela cor. A rede foi treinada em 10 épocas com mini-lotes (*mini-batchs*) de tamanho 100. Para o desenvolvimento da CNN ao dispositivo, foi utilizada a biblioteca de visão computacional OpenCV, que é de código aberto e foi desenvolvida com um forte foco em aplicações de tempo real [94]. Os autores alcançaram altas taxas de acurácia (99,35%). No entanto, não foi um aumento tão significativo em relação aos trabalhos com que foi comparado: 97,5% [95] e 98,43% [96]. O sistema não foi capaz de funcionar em tempo real.

Para facilitar a comparação entre os trabalhos vistos nesta revisão, criou-se a Tabela II, contendo, para cada um deles, a citação, a abordagem, as bases de dados utilizadas e as comparações realizadas para o reconhecimento de objetos.

TABELA II
Resumo da Revisão Sistemática.

Trabalho	Abordagens propostas	Base de dados	Abordagens comparadas
Merler et al. (2007) [45]	SIFT + K-NN	GroZi-120	1) CH + ED 2) CH + χ^2 3) CH + Lab 4) SIFT 5) Viola-Jones
Winlock et al. (2010) [46]	SURF + CH + PCA + Naive Bayes	GroZi-120	Sem comparações com outras abordagens
Rivera-Rubio et al. (2014) [47]	1) SIFT + K-Means + SVM 2) SIFT + LLC 3) SIFT + FV + PCA + SVM	SHORT-100	Sem comparações com outras abordagens
Varol e Salih (2015) [53]	SIFT + HSV + K-Means + BoW + SVM	Grocery	**Descritores:** 1) SIFT (forma) 2) HSV (cor) 3) Ambos
Jund et al. (2016) [55]	FT-CNN (CaffeNet)	Freiburg Groceries	Sem comparações com outras abordagens
Advani et al. (2017) [48]	*Coarse classification*: Descritor Gist + SVM *Fine-grained classification*: SURF	Sem testes em bases de dados	Sem comparações com outras abordagens
Franco et al. (2017) [54]	1) SURF + HI + K-Means + BoW 2) CNN + ED	GroZi-120	**Descritores:** 1) SURF + K-Means + BoW 2) CNN
Geng et al. (2018) [52]	SIFT + VGG-16	GroZi-120	**Descritores:** 1) Sem ATmap 2) SIFT 3) OpponentSIFT 4) C-SIFT 5) RGB-SIFT 6) BRISK 7) SURF
Kanuri et al. (2018) [49]	FT-CNN (Inception v3) + rede de classificadores	Base própria	**Classificadores:** 1) ResNet-50 2) VGG-16 1) Vanilla CNN 2) FT-CNN (Inception v3) 3) Abordagem proposta
Klasson et al. (2019) [50]	1) CNN + SVM 2) FT-CNNs	Grocery Store	**CNNs ou Descritores:** 1) AlexNet 2) VGG-16 3) DenseNet-169 **Classificadores:** 1) SVM 2) VAE + SVM 3) VAE + SVM-ft 4) VAE-CCA + SVM 5) VAE-CCA + SVM-ft
Pintado et al. (2019) [51]	Raspberry Pi 3 + CNN	Usou imagens do site Sprouts	1) [95] 2) [96] 3) Abordagem proposta

IV. Bases de Dados

Nesta seção são descritas as bases de dados utilizadas pelos estudos da revisão sistemática (Seção III), usadas na identificação de produtos comumente encontrados em mercearias. Algumas das bases foram elaboradas para o reconhecimento de produtos, distinguindo produtos de uma mesma categoria, enquanto outras foram feitas para o reconhecimento de categorias de produtos, ou ainda para ambos.

Por conta das restrições do escopo deste artigo, que focou na etapa de reconhecimento, não foram adicionadas bases cujas imagens de teste se limitavam a prateleiras com diversas classes de produtos, visto que são usadas por trabalhos de detecção e não somente de reconhecimento. Todas as bases aceitas possuem imagens de teste em que apenas uma classe aparece em cada imagem. Com isso, este trabalho considerou o cenário de reconhecer um produto próximo ou na mão do usuário, ou ainda o cenário de reconhecer um produto que já foi detectado, independente de como foi realizada a detecção.

A Tabela III apresenta um resumo com as principais informações dessas bases de dados. Na coluna **Base de dados**, além da referência, são encontrados o nome da base e o ano de criação. Pela coluna **Obtenção dos dados** é possível verificar a forma de aquisição das imagens de treino e de teste. A coluna **Classes** informa a quantidade de gêneros que serão classificados, além de indicar se a classificação é categórica (*coarse classification*) ou se cada classe se refere a apenas um produto (*fine-grained classification*). Na coluna **Divisão dos dados** é visto a forma que os trabalhos dividiram os dados da base para realizar testes de acurácia. A coluna **Amostras** mostra a quantidade de imagens em cada conjunto das bases. A coluna **Amostras/Classe** exibe a taxa de amostras por classe. Por fim, a coluna **Máscara** informa quais dos conjuntos (treinamento, teste ou ambos) fornecem máscaras para o recorte do produto na imagem, retirando o fundo, ou se já possuem apenas o produto recortado nas imagens.

TABELA III
Bases de dados usadas no reconhecimento de produtos mercantis.

Base de dados	Obtenção dos dados	Classes	Divisão dos dados	Amostras	Amostras/Classe	Máscara
GroZi-120 (2007) [45]	Treino: imagens da web Teste: *frames* de vídeos Vídeos: *smartphone*	120 (*fine-grained*)	*hold out*	Treino: 676 Teste: 11.194 Vídeos: 29	Treino: 5,6 Teste: 93,3	Treino
SHORT-100 (2014) [47]	Treino: Nikon D7100 SLR Teste: 30 diferentes *smartphones*	30 (*fine-grained*)	*k-fold* ($K = 5$)	*Treino* 30 classes: 1.080 Sem classe: 2.520 *Teste* ST-SG: 2.797 VF-SG: 92.293 ST-BF: 1.225 VF-BF: 39.121	*Treino* 30 classes: 36 *Teste* ST-SG: 93,2 VF-SG: 3.076,4 ST-BF: 40,8 VF-BF: 1.304,0	Treino
Grocery (2015) [53]	Quatro câmeras	10 (*fine-grained*)	*hold out*	Treino: 3.701 Sem classe: 10.440 Teste: 2.744	Treino: 370,1 Teste: 274,4	Ambos
Freiburg Groceries (2016) [55]	*Smartphone*	25 (*coarse-grained*)	*k-fold* ($K = 5$)	4.947	197,9	Nenhum
Grocery Store (2019) [50]	*Smartphone*	81 (*fine-grained*) 43 (*coarse-grained*)	*hold out*	Treino: 2.640 Teste: 2.485	*Fine-grained:* Treino: 32,6 Teste: 30,7 *Coarse-grained*: Treino: 62,9 Teste: 59,2	Nenhum

Finalmente, a Tabela IV contém um resumo com os melhores resultados de acurácia obtidos pelos autores das bases apresentadas. Excepcionalmente, nas bases SHORT-100 e Freiburg Groceries foram relatados resultados de acurácia média, devido aos autores utilizarem nessas bases o método de validação cruzada *k-fold*. Apesar da base GroZi-120 ter sido amplamente utilizada pelos trabalhos encontrados, esses estudos não apresentaram a acurácia de reconhecimento geral, como feito pelos autores das demais bases públicas.

A. GroZi-120

Dos autores Merler et al. [45], o banco de dados GroZi-120 foi desenvolvido em 2007 e possui 120 produtos de mercearia, de diferentes categorias. Eles variam em cor, tamanho, opacidade, forma e rigidez. As imagens foram obtidas em duas diferentes representações na base de dados: uma capturada *in vitro*, para treinamento, e a outra capturada *in situ*, para testes.

Os autores capturaram as imagens *in vitro* a partir da web, mais especificamente em lojas de supermercado como o Froogle ("http://www.froogle.com", atualmente denominado de Google Shopping), usando um script para rastrear (*crawl*) automaticamente a web procurando as imagens dos produtos usando o Código Universal de Produtos (Universal Product Code – UCP) [45]. Elas possuem condições ideais de iluminação e perspectiva. Já as imagens *in situ* foram adquiridas por meio de vídeos registrados em lojas reais e, portanto, possuem limitações comuns de cenários reais, como escala, cor, rotação, distorções de perspectiva, iluminação, borramentos, sombras e oclusões. Elas foram selecionadas a cada cinco quadros da aparição do produto nos vídeos.

As imagens *in vitro* possuem uma média de aproximadamente 5,6 amostras por classe, com um total de 676 imagens para os 120 itens. Elas possuem máscaras para recortar o produto, com o objetivo de retirar o fundo da imagem. As imagens *in situ* possuem uma média aproximada de 93,3 amostras por classe, com um total de 11.194 imagens.

B. SHORT-100

Criada por Rivera-Rubio et al. [47] em 2014, a base era chamada de SHORT-30, pois possuía imagens de 30 produtos nos conjuntos de treinamento e de teste. Porém, após sua atualização, somente o conjunto de treinamento foi atualizado para 100 produtos. Cada produto do conjunto de treino possui 36 imagens, obtidas pela câmera Nikon D7100 SLR em condições ideais de estúdio, capturadas com rotação horizontal do produto, permitindo que os produtos sejam vistos por diversos ângulos. Com isso, são 1.080 imagens de treino nas 30 classes, mais 2.520 imagens dos 70 produtos restantes (sem classe), totalizando 3.600 imagens. A base fornece máscaras para o recorte do produto em cada imagem do treinamento.

O conjunto de teste consiste em imagens adquiridas por meio de 30 *smartphones* distintos. Durante a captura, enquanto uma das mãos segurava o *smartphone*, a outra segurava o produto, e no fundo da imagem também podem aparecer outras coisas do ambiente. Esse conjunto é dividido em *Sighted* (SG) – imagens tiradas por pessoas com visão – e *Blindfolded* (BF) – imagens capturadas por pessoas com os olhos vendados. Cada uma dessas divisões do conjunto de teste ainda se subdivide em imagens fixas (*Still Images* - ST) e imagens retiradas de *frames* de vídeo (*Video Frames* - VF). A quantidade de imagens em cada um desses subconjuntos é: ST-SG: 2.797; VF-SG: 92.293; ST-BF: 1.225; VF-BF: 39.121. No total, o conjunto de teste possui 135.436 imagens. Apesar da divisão dos dados em treino e teste, apenas para os grupos de teste ST-SG e VF-SG, os autores calcularam a acurácia média, de cada um desses grupos, a partir de uma validação cruzada *k-fold* com $K = 5$, fornecendo um código fonte com a partição dos *folds* para a utilização imediata da base de dados SHORT-100. Alcançaram 77,51% de acurácia média no conjunto ST-SG e 69,41% no conjunto VF-SG. Os autores não relataram os desvios padrões obtidos a partir dos cálculos dessas acurácias médias.

TABELA IV
Resumo com os melhores resultados nas bases de dados apresentadas.

Base de dados	Melhor resultado	Método utilizado
SHORT-100 (2014) [47]	Conjunto ST: I: Acurácia média: 77,51% Conjunto VF: II: Acurácia média: 69,41%	Conjunto ST: I: SIFT + K-Means + SVM Conjunto VF: II: SIFT + LLC
Grocery (2015) [53]	Acurácia: 92,3%	SIFT + HSV
Freiburg Groceries (2016) [55]	Acurácia média: 78,9±0,5%	FT-CNN (CaffeNet)
Grocery Store (2019) [50]	Acurácia: 85,0% (*fine-grained*) Acurácia: 85,2% (*coarse-grained*)	FT-CNN (DenseNet-169) + SVM

C. Grocery

A base Grocery foi criada em 2015 por Varol e Salih [53]. Dentre as bases de dados aceitas neste trabalho, a Grocery é a única cujas classes consistem somente em um produto específico (embalagens de cigarro), variando apenas entre diferentes marcas. Apesar da base possuir imagens de prateleiras de produtos, para testes de detecção, ela também possui o recorte de cada um desses produtos (i.e., o produto já detectado), viabilizando testes que se limitam ao reconhecimento, sem se preocupar com a abordagem de detecção a ser utilizada.

As imagens da base foram tiradas em torno de 40 lojas, a partir de quatro câmeras diferentes (iPhone5S, iPhone4, Nikon Coolpix S3 e Sony Cyber-shot DSC-W300), variando a distância e o ângulo da câmera em relação à prateleira. As classes consistem em dez marcas de cigarro. A base possui 13.184 imagens de produtos no conjunto de treinamento, que foram extraídas de 354 imagens de prateleiras, em que o fundo foi removido. Restaram, ainda, 10.440 produtos não rotulados em nenhuma classe, que podem ser usados como classe negativa. O conjunto de treinamento tem 3.701 imagens de produtos, enquanto o de teste possui 2.744 imagens.

A base fornece ainda as imagens com o recorte apenas da marca para todos os produtos rotulados. Foi nesses conjuntos que os autores realizaram testes de reconhecimento, alcançando uma acurácia de 92,3%.

D. Freiburg Groceries

A base foi criada em 2016 por Jund et al. [55] e contém 4.947 imagens de produtos capturadas por *smartphones*, em supermercados da Alemanha. A Freiburg Groceries é dividida em 25 classes, que não correspondem a produtos específicos, mas a categorias de produtos (*coarse classification*). As classes são desbalanceadas, cada uma possuindo de 97 a 370 imagens. Não existe separação em conjuntos de treinamento e de teste, portanto é recomendável que os testes sejam feitos da mesma forma que os autores da base os executaram, por meio do método de validação cruzada *k-fold*, com $K = 5$. As imagens de cada classe são distribuídas em cinco *folds* com o mesmo tamanho. Os autores também informaram quais imagens foram alocadas para cada *fold* nos seus testes. Eles alcançaram uma acurácia média de 78,9%, com desvio padrão de 0,5%.

E. Grocery Store

Elaborado por Klasson et al. [50] em 2019, o repositório *Grocery Store* contém 5.125 imagens naturais de 81 classes de frutas, legumes e produtos em caixas (suco, leite e iogurte), para classificação *fine-grained*, capturadas por uma câmera de *smartphone* em diferentes mercearias. As imagens são também divididas em 43 categorias, em que, por exemplo, os produtos *Royal Gala* e *Granny Smith* pertencem à mesma categoria *Apple*, viabilizando a classificação *coarse-grained*. Para cada classe são fornecidos ainda um ícone e uma descrição do produto. A base possui os conjuntos de treinamento e de teste definidos. Seus autores alcançaram 85,0% de acertos no conjunto *fine-grained* e 85,2% no conjunto *coarse-grained*.

V. Discussão

Por meio de análises sobre os trabalhos aceitos, é possível observar que a partir de 2016 os estudos se concentraram mais em CNNs, tendo três tipos básicos de abordagens: CNNs na descrição e na classificação [49–51, 54, 55]; somente como descritores [50]; e somente como classificadores [52]. Nessa e em diversas outras áreas de visão computacional, as abordagens com CNNs lideram o estado da arte.

Apesar dos ótimos resultados relatados pelos trabalhos que utilizaram CNNs no reconhecimento multi-classe, é necessário incrementar técnicas recentes a alguns deles, que melhorem o treinamento das redes, como a utilização de Aumento de Dados [97], a adição de camadas de *Drop Out* [98] e também de camadas de *Batch Normalization* [99] em redes mais antigas, como a VGG-16, que, além de melhorar os resultados, reduz bastante o tempo necessário ao treinamento da rede. Diversas taxas de aprendizado podem ser testadas com diferentes algoritmos de otimização, como SGD [100], Adam [101] e Adadelta [102]. Também podem ser utilizadas diversas outras CNNs pré-treinadas, ou ainda diversos outros descritores ou classificadores combinados à CNN. É possível ainda explorar outras técnicas dos campos de aprendizado de máquina e de visão computacional, combinando-as com as utilizadas pelos autores.

Todavia, há duas bases de dados que possuem uma grande diferença na aquisição das imagens dos conjuntos de treinamento e de teste, o que pode prejudicar os resultados de alguns métodos, especialmente a capacidade de generalização de CNNs: as bases GroZi-120 e SHORT-100. Uma justificativa dos autores dessas bases terem feito essa separação é que, por conta de as imagens de treino terem maior qualidade, mais de suas características serão encontradas pelos descritores. Isso pode ser vantajoso para técnicas mais antigas, no entanto, para CNNs, a rede poderá tender a se especializar demais

nas características das imagens de treinamento, provocando um super-ajuste ao conjunto de treino, fenômeno conhecido como *overfitting* [103]. Em trabalhos futuros, portanto, essas duas bases poderão exigir maiores melhorias nas abordagens com CNNs.

O cenário das bases de dados simula diferentes hipóteses de aplicações para pessoas com deficiências visuais. Por exemplo, se em um dado sistema for o usuário quem captura as imagens de treinamento e de teste, fazendo isso nos mesmos ambientes e com a mesma câmera, o uso de CNNs provavelmente terá ótimos resultados. Por outro lado, em um sistema em que as imagens de treinamento são baixadas da internet ao invés de serem capturadas pelo usuário, provavelmente irá ocorrer uma grande diferença entre as imagens de treino e de teste, o que pode diminuir os acertos de métodos com CNN. Diante disso, a forma que um sistema realiza o treinamento é importante na decisão da abordagem a ser usada.

Além de observar a taxa de acurácia dos métodos, também é preciso verificar se funcionam em tempo real em sistemas *wearable*, minimizando o uso do processador, economizando memória e bateria. Há, por exemplo, CNNs pré-treinadas que, apesar de suas altas taxas de acerto, dispõem de um número muito grande de camadas, nós e parâmetros, o que faz o treinamento demorar dias, até mesmo em computadores com GPU. Em contrapartida, existem estudos com propostas de CNNs menores com boas taxas de acerto, voltadas a dispositivos *wearable*, como em [104–106].

Contudo, é importante considerar que o custo do treinamento de CNNs depende, dentre outros fatores, do número de amostras e de épocas, da arquitetura da rede utilizada e do *hardware* que irá realizar o processamento. Por isso, algumas abordagens recomendam a utilização de GPUs para realizar o treinamento da rede [13, 48, 107], comunicando-se via redes sem fio com os aparelhos *wearable* do usuário. O volume e a qualidade dos dados também são muito importantes ao treinamento da CNN [108], todavia, também podem aumentar a carga sobre o processador. Muitas vezes, reduzir um pouco a resolução das imagens otimiza o treinamento sem uma perda significante nos resultados.

Um bom sistema deve, primeiramente, ser capaz de realizar o reconhecimento e a localização de produtos, buscando maximizar os acertos em diversos ambientes. E, como características secundárias, deseja-se combinar ideias que garantam uma interface adequada ao usuário, de acordo com as suas limitações, aumentando assim a acessibilidade e a usabilidade do sistema.

VI. Conclusão

Neste trabalho foi realizada uma revisão detalhada sobre o reconhecimento de produtos de mercearia, identificando aspectos relevantes sobre o tema, principalmente com relação aos algoritmos e às bases de dados utilizadas. Isso possibilita uma melhor compreensão do assunto e ajuda na proposição de novas abordagens.

Este estudo reuniu os mais recentes trabalhos de importantes bases de busca bibliográfica em ciência da computação. Ele poderá ser usado como base para se propor novas abordagens em reconhecimento de itens mercantis. As técnicas de reconhecimento também podem ser combinadas com técnicas de detecção, construindo um sistema completo, para detectar e reconhecer produtos de mercearia.

Referências

[1] WHO, *Blindness and vision impairment*, [online] Disponível em: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment [Acesso em: Abr. 2020], 2019.

[2] IBGE, "Releitura dos dados de pessoas com deficiência no censo demográfico 2010 à luz das recomendações do grupo de washington", IBGE, Rio de Janeiro, Nota tecnica 01/2018, 2018.

[3] J. E. Jan, E. P. Scott e R. D. Freeman, *Visual impairment in children and adolescents*. New York : Grune & Stratton, 1977.

[4] R. Tapu, B. Mocanu, A. Bursuc e T. Zaharia, "A smartphone-based obstacle detection and classification system for assisting visually impaired people", em *The IEEE ICCV Workshops*, 2013.

[5] A. Rodríguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán e A. Cela, "Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback", *Sensors*, vol. 12, n.º 12, pp. 17 476–17 496, 2012.

[6] J. Coughlan e R. Manduchi, "Color targets: Fiducials to help visually impaired people find their way by camera phone", *EURASIP Journal on Image and Video Processing*, vol. 2007, n.º 1, p. 096 357, 2007.

[7] M. George, D. Mircic, G. Sörös, C. Floerkemeier e F. Mattern, "Fine-grained product class recognition for assisted shopping", em *IEEE ICCVW*, 2015, pp. 546–554.

[8] P. G. Bhat, D. K. Rout, B. N. Subudhi e T. Veerakumar, "Vision sensory substitution to aid the blind in reading and object recognition", em *ICIIP*, 2017, pp. 1–6.

[9] L. P. Sousa, R. M. S. Veras, L. H. S. Vogado e L. S. Britto Neto, "Metodologia de identificação de cédulas monetárias para deficientes visuais", *RSC*, vol. 8, n.º 1, 2018.

[10] C. Guida, D. Comanducci e C. Colombo, "Automatic bus line number localization and recognition on mobile phones—a computer vision aid for the visually impaired", em *ICIAP*, G. Maino e G. L. Foresti, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 323–332.

[11] K. Thakoor, N. Mante, C. Zhang, C. Siagian, J. Weiland, L. Itti e G. Medioni, "A system for assisting the visually impaired in localization and grasp of desired objects", em *ECCV 2014*, L. Agapito, M. M. Bronstein e C. Rother, eds., Cham: Springer International Publishing, 2015, pp. 643–657.

[12] X. Yang, S. Yuan e Y. Tian, "Assistive clothing pattern recognition for visually impaired people", *IEEE THMS*, vol. 44, n.º 2, pp. 234–243, 2014.

[13] V. Pradeep, G. Medioni e J. Weiland, "Robot vision for the visually impaired", em *IEEE CVPR*, 2010, pp. 15–22.

[14] V. N. Murali e J. M. Coughlan, "Smartphone-based crosswalk detection and localization for visually impaired pedestrians", em *ICMEW*, 2013, pp. 1–7.

[15] D. Ahmetovic, C. Bernareggi, A. Gerino e S. Mascetti, "Zebrarecognizer: Efficient and precise localization of pedestrian crossings", em *ICPR*, 2014, pp. 2566–2571.

[16] J. Roters, X. Jiang e K. Rothaus, "Recognition of traffic lights in live video streams on mobile devices", *IEEE TCSVT*, vol. 21, n.º 10, pp. 1497–1511, 2011.

[17] G. Fusco, H. Shen e J. M. Coughlan, "Self-localization at street intersections", em *CRV*, 2014, pp. 40–47.

[18] M. Samara, J. AlSadah, M. Driche e Y. Osais, "A color recognition system for the visually impaired people", em *ICETAS*, 2017, pp. 1–5.

[19] R. Jafri, S. A. Ali, H. R. Arabnia e S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: A survey", *The Visual Computer*, vol. 30, n.º 11, pp. 1197–1222, 2014.

[20] G. Iannizzotto, C. Costanzo, P. Lanzafame e F. La Rosa, "Badge3d for visually impaired", em *IEEE CVPR*, 2005, pp. 29–29.

[21] P. E. Lanigan, A. M. Paulos, A. W. Williams, D. Rossi e P. Narasimhan, "Trinetra: Assistive technologies for grocery shopping for the blind", em *IEEE ISWC*, 2006, pp. 147–148.

[22] H. S. Al-Khalifa, "Utilizing qr code and mobile phones for blinds and visually impaired people", em *ICCHP*, K. Miesenberger, J. Klaus, W. Zagler e A. Karshmer, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1065–1069.

[23] C. Mendes e D. Wolf, "Desvio de obstáculos utilizando um método estéreo semi-global", em *Encontro Nacional de Inteligência Artificial*, 2010, pp. 788–799.

[24] A. Hub e T. Hartter, "Interactive localization and recognition of objects for the blind", em *21st Annual International Technology and Persons with Disabilities Conference*, California State University, 2006.

[25] B. Schauerte, M. Martinez, A. Constantinescu e R. Stiefelhagen, "An assistive vision system for the blind that helps find lost things", em *ICCHP*, K. Miesenberger, A. Karshmer, P. Penaz e W. Zagler, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 566–572.

[26] W. Fink, M. Tarbell, J. Weiland e M. Humayun, *Dora: Digital object recognition audio — assistant for the visually impaired.* 2004.

[27] J. K. Basu, D. Bhattacharyya e T.-H. Kim, "Use of artificial neural network in pattern recognition", em *IJSEA*, vol. 4, 2010.

[28] A. L. Kumar e R. Ganesan, "Improved navigation for visually challenged with high authentication using a modified sift algorithm", em *ICGCCEE*, 2014, pp. 1–5.

[29] S. Fuangkaew e K. Patanukhom, "Stereo image based object localization framework for visually impaired people using edge orientation histogram and co-occurrence matrices", em *INISCom*, 2015, pp. 113–121.

[30] A. Kapur e S. Kapur, "Drishti: An ultra-low cost visual-aural assistive technology for the visually impaired", em *i-CREATe*, Kaki Bukit TechPark II,, Singapore: Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, 2015, 14:1–14:4.

[31] B. K. Balasuriya, N. P. Lokuhettiarachchi, A. R. M. D. N. Ranasinghe, K. D. C. Shiwantha e C. Jayawardena, "Learning platform for visually impaired children through artificial intelligence and computer vision", em *SKIMA*, 2017, pp. 1–7.

[32] G. Gkioxari, R. Girshick e J. Malik, "Contextual action recognition with r*cnn", em *IEEE ICCV*, 2015, pp. 1080–1088.

[33] A. Graves, A. Mohamed e G. Hinton, "Speech recognition with deep recurrent neural networks", em *IEEE ICASSP*, 2013, pp. 6645–6649.

[34] H. Kacorri, K. M. Kitani, J. P. Bigham e C. Asakawa, "People with visual impairment training personal object recognizers: Feasibility and challenges", em *ACM CHI*, New York, NY, USA: ACM, 2017, pp. 5839–5849.

[35] J. Sosa-García e F. Odone, ""Hands on"visual recognition for visually impaired users", *ACM TACCESS*, vol. 10, n.º 3, 8:1–8:30, 2017.

[36] M. Awad, J. E. Haddad, E. Khneisser, T. Mahmoud, E. Yaacoub e M. Malli, "Intelligent eye: A mobile application for assisting blind people", em *IEEE MENACOMM*, 2018, pp. 1–6.

[37] S. S. Latifi Oskouei, H. Golestani, M. Hashemi e S. Ghiasi, "Cnndroid: Gpu-accelerated execution of trained deep convolutional neural networks on android", em *ACM MM*, Amsterdam, The Netherlands, 2016, pp. 1201–1205.

[38] Cambridge Dictionary, *Wearable*, [online] Disponível em: http://dictionary.cambridge.org/dictionary/english/wearable [Acesso em: 11 Dez. 2018].

[39] Dictionary.com, *Wearable computer*, [online] Disponível em: https://www.dictionary.com/browse/wearable-computer [Acesso em: 11 Dez. 2018].

[40] F. Xia, L. T. Yang, L. Wang e A. Vinel, "Internet of things", *International Journal of Communication Systems*, vol. 25, n.º 9, p. 1101, 2012.

[41] L. Atzori, A. Iera e G. Morabito, "The internet of things: A survey", *Computer Networks*, vol. 54, n.º 15, pp. 2787 –2805, 2010.

[42] L. S. Britto Neto, V. R. M. L. Maike, F. L. Koch, M. C. C. Baranauskas, A. de R. Rocha e S. K. Goldenstein, "A wearable face recognition system built into a smartwatch and the visually impaired user", em *ICEIS*, INSTICC, SciTePress, 2015, pp. 5–12.

[43] E. Hernandes, A. Zamboni, S. Fabbri e A. D. Thommazo, "Using gqm and tam to evaluate start-a tool that supports systematic review", *CLEI Electronic Journal*, vol. 15, n.º 1, pp. 3–3, 2012.

[44] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman e T. P. Group, "Preferred reporting items for systematic reviews and meta-analyses: The prisma statement", *PLOS Medicine*, vol. 6, n.º 7, pp. 1–6, 2009.

[45] M. Merler, C. Galleguillos e S. Belongie, "Recognizing groceries in situ using in vitro training data", em *IEEE CVPR*, 2007, pp. 1–8.

[46] T. Winlock, E. Christiansen e S. Belongie, "Toward real-time grocery detection for the visually impaired", em *IEEE CVPR*, 2010, pp. 49–56.

[47] J. Rivera-Rubio, S. Idrees, I. Alexiou, L. Hadjilucas e A. A. Bharath, "Small hand-held object recognition test (short)", em *IEEE WACV*, 2014, pp. 524–531.

[48] S. Advani, P. Zientara, N. Shukla, I. Okafor, K. Irick, J. Sampson, S. Datta e V. Narayanan, "A multitask grocery assist system for the visually impaired: Smart glasses, gloves, and shopping carts provide auditory and tactile feedback", *IEEE CEM*, vol. 6, n.º 1, pp. 73–81, 2017.

[49] S. N. Kanuri, S. P. Navali, S. R. Ranganath e N. V. Pujari, "Multi neural network model for product recognition and labelling", em *ICACCI*, 2018, pp. 1837–1842.

[50] M. Klasson, C. Zhang e H. Kjellström, "A hierarchical grocery store image dataset with visual and semantic labels", em *IEEE WACV)*, 2019, pp. 491–500.

[51] D. Pintado, V. Sanchez, E. Adarve, M. Mata, Z. Gogebakan, B. Cabuk, C. Chiu, J. Zhan, L. Gewali e P. Oh, "Deep learning based shopping assistant for the visually impaired", em *IEEE ICCE*, 2019, pp. 1–6.

[52] W. Geng, F. Han, J. Lin, L. Zhu, J. Bai, S. Wang, L. He, Q. Xiao e Z. Lai, "Fine-grained grocery product recognition by one-shot learning", em *ACM MM*, New York, NY, USA: ACM, 2018, pp. 1706–1714.

[53] G. Varol e R. S. Kuzu, "Toward retail product recognition on grocery shelves", em *ICGIP 2014*, Y. Wang, X. Jiang e D. Zhang, eds., International Society for Optics e Photonics, vol. 9443, SPIE, 2015, pp. 46 –52.

[54] A. Franco, D. Maltoni e S. Papi, "Grocery product detection and recognition", *Expert Syst. Appl.*, vol. 81, n.º C, pp. 163–176, 2017.

[55] P. Jund, N. Abdo, A. Eitel e W. Burgard, *The freiburg groceries dataset*, 2016. arXiv: 1611.05799.

[56] D. G. Lowe, "Object recognition from local scale-invariant features", em *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[57] H. Bay, T. Tuytelaars e L. Van Gool, "SURF: Speeded up robust features", em *ECCV*, A. Leonardis, H. Bischof e A. Pinz, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[58] Y. Zhang, R. Jin e Z.-H. Zhou, "Understanding bag-of-words model: A statistical framework", *IJMLC*, vol. 1, pp. 43–52, 2010.

[59] J. A. Suykens e J. Vandewalle, "Least squares support vector machine classifiers", *Neural processing letters*, vol. 9, n.º 3, pp. 293–300, 1999.

[60] S. B. Lo, S. A. Lou, M. T. Freedman, M. V. Chien e S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection", *IEEE T-MI*, vol. 14, n.º 4, pp. 711–718, 1995.

[61] Y. Gao, B. Zheng, G. Chen, W. Lee, K. C. K. Lee e Q. Li, "Visible reverse k-nearest neighbor query processing in spatial databases", *IEEE TKDE*, vol. 21, n.º 9, pp. 1314–1327, 2009.

[62] P. Viola, M. Jones et al., "Rapid object detection using a boosted cascade of simple features", *CVPR*, vol. 1, pp. 511–518, 2001.

[63] J. Morovic, J. Shaw e P.-L. Sun, "A fast, non-iterative and exact histogram matching algorithm", *Pattern Recognition Letters*, vol. 23, n.º 1-3, pp. 127–135, 2002.

[64] M. J. Swain e D. H. Ballard, "Color indexing", *IJCV*, vol. 7, n.º 1, pp. 11–32, 1991.

[65] Y. Freund e R. E. Schapire, "Experiments with a new boosting algorithm", em *ICML*, 1996, pp. 148–156.

[66] J. Shlens, *A tutorial on principal component analysis*, 2014. arXiv: 1404.1100.

[67] L. Barrington, T. K. Marks, J. Hui-wen Hsiao e G. W. Cottrell, "NIMBLE: A kernel density model of saccade-based visual memory", *Journal of Vision*, vol. 8, n.º 14, pp. 17–17, 2008.

[68] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman e A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation", *IEEE TPAMI*, vol. 24, n.º 7, pp. 881–892, 2002.

[69] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang e Y. Gong, "Locality-constrained linear coding for image classification", em *IEEE CVPR*, Citeseer, 2010, pp. 3360–3367.

[70] M. A. Fischler e R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Commun. ACM*, vol. 24, n.º 6, pp. 381–395, 1981.

[71] N. Dalal e B. Triggs, "Histograms of Oriented Gradients for Human Detection", em *International Conference on Computer Vision & Pattern Recognition (CVPR '05)*, C. Schmid, S. Soatto e C. Tomasi, eds., vol. 1, San Diego, United States: IEEE Computer Society, 2005, pp. 886–893.

[72] A. Oliva e A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope", *Int. J. Comput. Vision*, vol. 42, n.º 3, pp. 145–175, 2001.

[73] O. Murthy e M. Hanmandlu, "A study on the effect of outliers in devanagari character recognition", em *IJCA*, vol. 32, 2011.

[74] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley Publishing, 2009.

[75] N. D. B. Bruce e J. K. Tsotsos, "An information theoretic model of saliency and visual search", em *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*, L. Paletta e E. Rome, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 171–183.

[76] S. J. Pan e Q. Yang, "A survey on transfer learning", *IEEE TKDE*, vol. 22, n.º 10, pp. 1345–1359, 2010.

[77] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama e T. Darrell, "Caffe: Convolutional architecture for fast feature embedding", *arXiv:1408.5093*, 2014.

[78] A. Krizhevsky, I. Sutskever e G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Commun. ACM*, vol. 60, n.º 6, pp. 84–90, 2017.

[79] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg e L. Fei-Fei, "Imagenet large scale visual recognition challenge", *IJCV*, vol. 115, n.º 3, pp. 211–252, 2015.

[80] L. Fei-Fei, R. Fergus e P. Perona, "One-shot learning of object categories", *IEEE TPAMI*, vol. 28, n.º 4, pp. 594–611, 2006.

[81] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng e S. Li, "Salient object detection: A discriminative regional feature integration approach", em *IEEE CVPR*, 2013, pp. 2083–2090.

[82] S. Leutenegger, M. Chli e R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints", em *ICCV*, 2011, pp. 2548–2555.

[83] K. Simonyan e A. Zisserman, "Very deep convolutional networks for large-scale image recognition", em *ICLR*, 2015.

[84] F. Chollet et al., *Keras: The Python Deep Learning library*, Astrophysics Source Code Library, 2018.

[85] K. van de Sande, T. Gevers e C. Snoek, "Evaluating color descriptors for object and scene recognition", *IEEE TPAMI*, vol. 32, n.º 9, pp. 1582–1596, 2010.

[86] K. He, X. Zhang, S. Ren e J. Sun, "Deep residual learning for image recognition", em *IEEE CVPR*, 2016, pp. 770–778.

[87] M. George e C. Floerkemeier, "Recognizing products: A per-exemplar multi-label image classification approach", em *ECCV*, D. Fleet, T. Pajdla, B. Schiele e T. Tuytelaars, eds., Cham: Springer International Publishing, 2014, pp. 440–455.

[88] L. Karlinsky, J. Shtok, Y. Tzur e A. Tzadok, "Fine-grained recognition of thousands of object categories with single-example training", em *IEEE CVPR*, Los Alamitos, CA, USA: IEEE Computer Society, 2017, pp. 965–974.

[89] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens e Z. Wojna, "Rethinking the inception architecture for computer vision", em *IEEE CVPR*, 2016, pp. 2818–2826.

[90] G. Huang, Z. Liu, L. van der Maaten e K. Q. Weinberger, "Densely connected convolutional networks", em *IEEE CVPR*, 2017.

[91] C. Doersch, *Tutorial on variational autoencoders*, 2016. arXiv: 1606.05908.

[92] G. Andrew, R. Arora, J. Bilmes e K. Livescu, "Deep canonical correlation analysis", em *ICML*, 2013, pp. 1247–1255.

[93] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., "Tensorflow: A system for large-scale machine learning", em *USENIX OSDI*, Savannah, GA: USENIX Association, 2016, pp. 265–283.

[94] G. Bradski e A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. "O'Reilly Media, Inc.", 2008.

[95] S. M. R. Bagwan e L. J. Sankpal, "Visualpal: A mobile app for object recognition for the visually impaired", em *IC4*, 2015, pp. 1–6.

[96] R. Kumar e S. Meher, "A novel method for visually impaired using object recognition", em *ICCSP*, 2015, pp. 0772–0776.

[97] L. Perez e J. Wang, "The effectiveness of data augmentation in image classification using deep learning", *arXiv:1712.04621*, 2017.

[98] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever e R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *JMLR*, vol. 15, n.º 1, 1929–1958, 2014.

[99] S. Ioffe e C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015. arXiv: 1502.03167.

[100] S. Ruder, "An overview of gradient descent optimization algorithms", *arXiv:1609.04747*, 2016.

[101] D. P. Kingma e J. Ba, "Adam: A method for stochastic optimization", *arXiv:1412.6980*, 2014.

[102] M. D. Zeiler, "Adadelta: An adaptive learning rate method", *arXiv:1212.5701*, 2012.

[103] D. M. Hawkins, "The problem of overfitting", *J. Chem. Inf. Comput. Sci.*, vol. 44, n.º 1, pp. 1–12, 2004.

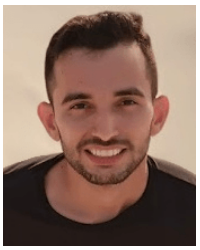
[104] S. Li, D. Liu, C. Xiang, J. Liu, Y. Ling, T. Liao e L. Liang, "Fitcnn: A cloud-assisted lightweight convolutional neural network framework for mobile devices", em *IEEE RTCSA*, 2017, pp. 1–6.

[105] K. Yanai, R. Tanno e K. Okamoto, "Efficient mobile implementation of a cnn-based object recognition system", em *ACM MM*, New York, NY, USA: Association for Computing Machinery, 2016, 362–366.

[106] L. Tobías, A. Ducournau, F. Rousseau, G. Mercier e R. Fablet, "Convolutional neural networks for object recognition on mobile devices: A case study", em *ICPR*, 2016, pp. 3530–3535.

[107] F. Hu, Z. Zhu e J. Zhang, "Mobile panoramic vision for assisting the blind via indexing and localization", em *ECCV 2014*, L. Agapito, M. M. Bronstein e C. Rother, eds., Cham: Springer International Publishing, 2015, pp. 600–614.

[108] C. Zhang, P. Patras e H. Haddadi, "Deep learning in mobile and wireless networking: A survey", *IEEE COMST*, vol. 21, n.º 3, pp. 2224–2287, 2019.



André de Lima Machado received his Bachelor's degree in Computer Science from Federal University of Piauí (2014). He is currently a Master Degree student in Computer Science. His current research's interest are applications in Computer Vision and Neural Networks.

Rodrigo de Melo Souza Veras received the B.S. (2005) degree in Computer Science at Federal University of Piauí and M.Sc.(2007) degree in Computer Science at Federal University of Ceará. He obtained his Ph.D (2014) degree in Teleinformatics Engineering at Federal University of Ceará. Currently, he is Professor at Federal University of Piauí.

Kelson Romulo Teixeira Aires received the B.S. (1999) degree in Electric Engineering, M.Sc. (2001) and Ph.D (2009) degrees in Computer Engineering at Federal University of Rio Grande do Norte. Currently, he is Professor at Federal University of Piauí.

Laurindo de Sousa Britto Neto received the B.Sc. degree in Information Systems from the FATEPI (2004), the M.Sc. degree in Systems and Computing from the UFRN (2007) and the Ph.D. degree in Computer Science from the Unicamp (2016). He is a Professor at the Department of Computing, UFPI. His interests lie in computer vision, image processing, computer graphics and human-computer interaction.