

FUNDAMENTOS DA VISÃO COMPUTACIONAL: ARCABOUÇO TEÓRICO DO RECONHECIMENTO ARTIFICIAL DE IMAGENS E VÍDEOS

FUNDAMENTALS OF COMPUTER VISION: THEORETICAL FRAMEWORK OF ARTIFICIAL RECOGNITION OF IMAGES AND VIDEOS

Bruno Vicente Nunes de Oliveira 1
Filipe Torres de Melo 2

Resumo: Este trabalho visa compreender os aspectos da visão computacional sob o prisma teórico. Na contemporaneidade, tal tecnologia é aplicada em diversos segmentos, desde o agronegócio até os serviços relacionados à saúde. Ademais, em consonância às recentes pesquisas, a utilização de mecanismos de visão computacional, em nível mundial, cresce consideravelmente. Tendo em vista essa particularidade, este artigo tem por objetivo central delinear os fundamentos da visão computacional por meio de arcabouço teórico acerca do reconhecimento artificial de imagens e vídeos. Para tanto, classifica-se o presente estudo em pesquisa básica. No que concerne ao processo de coleta de informações, incorporou-se a realização de uma pesquisa bibliográfica por intermédio de revisão de literatura. Nessa seara, destaca-se que foram considerados, especialmente, artigos científicos e livros oriundos das bases de dados Web of Science e Scopus. Em termos de resultados, contribuiu-se com um arcabouço conceitual sistematizado da tecnologia ora discutida.

Palavras-chave: Visão Computacional. Inteligência Artificial. Tecnologia. Inovação. Imagens e Vídeos.

Abstract: This work aims to understand aspects of computer vision from a theoretical perspective. Nowadays, this technology is applied in several segments, from agribusiness to health-related services. Furthermore, in line with recent research, the use of computer vision mechanisms worldwide is growing considerably. Bearing this particularity in mind, this article's central objective is to outline the foundations of computer vision through a theoretical framework regarding the artificial recognition of images and videos. Therefore, the present study is classified as basic research. Regarding the information collection process, bibliographical research was carried out through a literature review. In this area, it is noteworthy that scientific articles and books from the Web of Science and Scopus databases were especially considered. In terms of results, a systematized conceptual framework of the technology discussed was contributed.

Keywords: Computer Vision. Artificial Intelligence. Technology. Innovation. Images and Videos

-
- 1 Mestre em Propriedade Intelectual e Transferência de Tecnologia para a Inovação, pós-graduado em Gestão Pública e Bacharel em Ciências Econômicas pela Universidade Federal de Alagoas – UFAL. Bacharel em Direito pela Universidade Estadual de Alagoas – UNEAL. Lattes: <http://lattes.cnpq.br/1811131159155077>. ORCID: <https://orcid.org/0000-0001-7047-0938>. E-mail: brunonunes.oli@gmail.com
 - 2 Mestrando em Ciência da Informação pela UFAL. Bolsista da Fundação de Amparo à Pesquisa do Estado de Alagoas – FAPEAL. Graduado em Direito pela Universidade Estadual de Alagoas e em Letras – Língua Portuguesa pela Universidade Estácio de Sá. Especialização em Revisão de Textos pela PUC – Minas e Especialização em Docência e Gestão do Ensino Superior pela Universidade Estácio de Sá. Lattes: <http://lattes.cnpq.br/1992326536730280>. ORCID: <https://orcid.org/0000-0002-8991-4284>. E-mail: filipe_torresml@hotmail.com

Introdução

O ser humano possui cinco sentidos essenciais para sua existência, visão, audição, olfato, paladar e tato. Todavia, a visão é, na concepção de Davies (2012), sem dúvidas, o principal elemento receptor do homem, a qual fornece grande parte das informações cotidianas e necessárias para as tomadas de decisões.

À vista disso, em consonância ao pensamento de Nixon e Aguado (2020, p. 4), a visão humana é um sistema sofisticado, que evoluiu ao longo de milhares de anos, tendo por objetivo a defesa e a sobrevivência. Destarte, a função precípua do olho é “formar uma imagem, em que requer a habilidade de focar, seletivamente, em objetos de interesse. Isto é conseguido pelos músculos ciliares que seguram a lente ocular”.

Nesse ínterim, muitas das imagens captadas se tornam redundantes e são comprimidas pelas várias camadas do córtex visual, “de modo que os centros superiores do cérebro precisam interpretar, abstratamente, apenas uma pequena fração dos dados”. Todavia, a quantidade de informações recebidas diariamente por meio do olho humano é superior a todo o restante de dados que se obtêm dos outros sentidos (Davies, 2012, p. 1).

Outra particularidade da visão humana é, de acordo com Davies (2012, p. 2), a facilidade com que a interpretação das imagens é realizada. Nessa perspectiva, vê-se uma cena como ela é, por exemplo, árvores em uma paisagem, não sendo necessárias deduções e interpretações óbvias. Além disso, “as respostas são efetivamente imediatas e normalmente estão disponíveis em um décimo de segundo”.

Desse modo, o olho possibilita que os humanos possam perceber e compreender o mundo ao seu redor, enquanto a visão computacional tem por desafio duplicar o efeito da visão humana ao identificar e entender uma imagem captada por um dispositivo eletrônico. Sonka, Hlaváč, Boyle (2014, p. 39) explicitam na obra *“Image Processing, Analysis, and Machine Vision”* que em outros livros, além deste, “se debruçariam longamente sobre essa frase e o significado da palavra ‘duplicado’, se a visão computacional está simulando ou imitando sistemas humanos, isso é um território filosófico”.

Nessa acepção, a visão computacional pode ser conceituada como um sistema de computador que treina máquinas para captar e analisar imagens do mundo real, por meio do reconhecimento de padrões (Palsule; Mishra, 2020). Sob a ótica de Savekar e Kumar (2021), a visão computacional se concentra em capturar, entender e interpretar informações visuais importantes contidas em dados de imagem e vídeo, cujo fito é orientar tomadas de decisões.

À luz disso, esta tecnologia, segundo Russell e Norving (2013), foi estabelecida em 1963 no *Massachusetts Institute of Technology* – MIT, por intermédio da tese do cientista da computação Lawrence Roberts, considerada uma das primeiras publicações no campo que se tem conhecimento, a qual introduziu ideias-chave acerca do tema.

Isto posto, ainda de acordo com Davies (2012), a necessidade de um sistema de processamento robusto, em paralelo às complexas variáveis de cenários, aclara que a visão computacional deve ser compreendida, de fato, como um dos desafios mais difíceis de resolução na atualidade. Portanto, permitir que computadores possam ver e entender não é algo simplório (Sonka; Hlaváč; Boyle, 2014).

Apesar disso, intuitivamente, a visão humana e dos computadores aparentam ter a mesma função. “O objetivo de ambos os sistemas é interpretar dados espaciais, dados que são indexados por mais de uma dimensão”. Nessa lógica, embora ambas as visões sejam funcionalmente semelhantes, não pode haver a ilusão, neste momento, que um sistema de visão computacional presente, de forma exata, a função do olho humano (Nixon; Aguado, 2020, p. 3-4).

Apesar disso, a visão computacional se configura como uma das áreas mais promissoras de pesquisa e investimento em inteligência artificial e ciência da computação, haja vista, por exemplo, o aumento da demanda por sistemas de visão computacional em aplicações automotivas e a alta demanda por inspeção e automação de qualidade (Savekar; Kumar, 2021).

Sob essa perspectiva, prevê-se que o mercado de sistemas de visão computacional para ambientes de manufatura pode alcançar, até o ano de 2028, US\$ 14,18 bilhões (Venturus, 2022). Outrossim, “o tamanho do mercado global de visão computacional foi avaliado em US\$ 11,3 bilhões

em 2020 e deve expandir a uma taxa de crescimento anual composta de 7,3% de 2020 a 2028” (Palsule; Mishra, 2020, p, 18).

À face do exposto, delimitar o conceito de visão computacional é, ao menos num primeiro momento, uma tarefa mais concludente, tendo em vista que a compreensão sobre a visão humana é um conceito concreto, diferentemente da IA que recorre a pensamentos filosóficos para sua concepção.

Tendo em vista essa particularidade, este artigo tem por objetivo central delinear os fundamentos da visão computacional, por meio de arcabouço teórico acerca do reconhecimento artificial de imagens e vídeos. Para tanto, classifica-se o presente estudo em pesquisa básica. No que concerne ao processo de coleta de informações, incorporou-se a realização de uma pesquisa bibliográfica por intermédio de revisão sistemática. Nessa seara, destaca-se que foram considerados artigos científicos e livros oriundos das bases de dados *Web of Science* e *Scopus*.

Conceito e síntese histórica da visão computacional

Em consonância ao pensamento de Shanmugamani, (2018, p. 6), a visão computacional “é a ciência de entender ou decifrar imagens e vídeos” por meio de complexos algoritmos. Para Holm *et al.* (2020, p. 2), a visão computacional é o “campo da ciência da computação que se concentra na quantificação do conteúdo de informação visual de imagens digitais”.

Ademais, na concepção de Khemasuwan *et al.*, (2020, p. 6), a visão computacional é uma subcategoria do *deep learning*, a qual identifica objetos a partir de pixels de imagens brutas. Nesse sentido, os especialistas, primeiramente, rotulam cada imagem, representando o que é compreendido como verdade, então, a visão computacional se “concentra no reconhecimento de imagem e vídeo que lida com tarefas atribuídas, como classificação, detecção e interpretação de objetos, a fim de categorizar resultados predefinidos”.

Nesse contexto, o “objetivo principal da visão computacional é fazer com que as máquinas vejam o mundo da mesma forma que os humanos” (BHATT *et al.*, 2020, p. 1-2). Dessa forma, as tecnologias de inteligência artificial, que constituem a visão computacional, permitem que a máquina compreenda o mundo por meio de suas próprias experiências (Pereira; Moreschi, 2020, p. 2). À luz disso, a visão computacional como uma área de estudo é, conforme Forsyth e Ponce (2012, p. xvii), uma fronteira intelectual. “Como qualquer fronteira, é excitante e desorganizada, e muitas vezes, não há autoridade confiável para apelar”.

Posto isto, tece-se uma síntese histórica acerca do tema. Cabe ressaltar que a história da visão computacional possui uma linha escassa dos fatos, diferentemente de como ocorre com a inteligência artificial, em que há um conjunto mais expressivo de informações. O fato é que os progressos históricos da visão computacional estão atrelados à história da IA.

No entanto, a visão computacional foi, possivelmente, idealizada na década de 1960, por meio da tese de Lawrence Gilman Roberts (1963), intitulada “*Machine Perception of Three-Dimensional Solids*”, enquanto estudante do *Massachusetts Institute of Technology*. “Foi uma das primeiras publicações no campo, introduzindo ideias-chave como detecção de aresta e correspondência baseada em modelo” (Russell; Norving, 2013, p. 1110).

Em sua tese, Roberts (1963, p. 2) analisa que:

A fotografia é uma projeção em perspectiva de um conjunto de objetos que podem ser construídos a partir da transformação de modelos tridimensionais conhecidos, e que os objetos são suportados por outros objetos visíveis ou por um plano de fundo. Essas suposições permitem que um computador obtenha uma descrição tridimensional razoável das informações de borda em uma fotografia por meio de um processo matemático topológico.

Em suma, Roberts (1963) descreveu, em sua tese, um programa de computador, o qual podia processar uma fotografia em um desenho de linha e, por conseguinte, a transformação em uma representação tridimensional, que exibia uma estrutura com todas as linhas ocultas removidas. À

vista disso, tinham-se três processos principais, os quais foram desenvolvidos e programados por intermédio de seu estudo:

O processo de entrada produz um desenho de linha a partir de uma fotografia. Em seguida, o programa de construção 3D produz uma lista de objetos tridimensional a partir do desenho de linha. Quando isso é concluído, o programa de exibição 3-D pode produzir uma projeção bidimensional dos objetos de qualquer ponto de vista. Destes processos, o programa de entrada é o mais restritivo, enquanto o programa 2-D para 3-D e o programa 3-D para 2-D são capazes de lidar com quase todos os objetos de superfície planar (Roberts, 1963, p. 8-9).

Apesar do pioneirismo de Roberts na década de 1960, Russell e Norving (2013, p. 1110) colocam que o progresso foi lento nas décadas de 1960, 1970 e 1980, haja vista o baixo nível de recursos computacionais e, principalmente, de armazenamento. Já as décadas seguintes, em conformidade com Sonka, Hlaváč e Boyle (2014, p. 27), foram marcadas por um amadurecimento do campo e um “crescimento significativo de aplicações ativas; sensoriamento remoto, diagnóstico técnico, orientação de veículos autônomos, imagens biomédicas”.

Em 2005, Sebe et al. (2005, p. 15) identificaram, naquele momento, que:

A capacidade de raciocinar e a capacidade de aprender são as duas principais capacidades associadas a esses sistemas. Nos últimos anos, avanços teóricos e práticos estão sendo feitos no campo da visão computacional e reconhecimento de padrões por meio de novas técnicas e processos de aprendizagem, representação e adaptação. Provavelmente é justo afirmar, no entanto, que o aprendizado representa a próxima fronteira desafiadora para a visão computacional.

Em 2012, Forsyth e Ponce (2012, p. 19) analisavam, na conjunta obra “*Computer Vision - A Modern Approach*”, que a visão computacional estava em um ponto extraordinário em seu desenvolvimento. “Esse florescimento foi impulsionado por várias tendências, dentre elas, o fato de computadores e sistemas de imagem terem se tornado mais acessíveis”.

Avante, nos últimos 10 anos, o campo da visão computacional se transformou drasticamente. No início dos anos 2000, “a opinião comum dos especialistas era de que levariam décadas para que os computadores ganhassem uma compreensão de alto nível das imagens digitais” (Smits; Wevers, 2021, p. 2). No entanto, já em 2015, o jornal britânico *The Guardian* anunciou que os “computadores são melhores que humanos em reconhecer e classificar imagens”, no tocante às categorias predefinidas.

Outrossim, as câmeras digitais criadas no início da década de 1970 e o desenvolvimento de tecnologias e algoritmos complexos em IA, a partir do final do século XX, “proporcionaram a criação de modernos softwares de visão computacional”. Desse modo, há, na contemporaneidade, uma “demanda significativa para reconhecer objetos e imagens automaticamente” (Bhatt et al., 2020, p. 1).

Nesse cenário, contemporaneamente, existem muitos sistemas em uso, câmeras “inspecionam peças mecânicas para verificar o tamanho, alimentos são inspecionados quanto à qualidade e imagens usadas em astronomia se beneficiam de técnicas de visão computacional” (Nixon; Aguado, 2020, p. 1).

Isso ocorre graças ao crescente número de novos algoritmos de *machine learning* e *deep learning*, “acessíveis por meio de bibliotecas de *software* livre e de código aberto, juntamente à disponibilidade de grandes conjuntos de dados e ao aumento do uso de unidades de processamento gráfico – GPU” (Hosseini et al., 2021, p. 286).

Nessa lógica, este ramo da IA “está preenchendo a lacuna entre os talentos humanos e da máquina em um ritmo vertiginoso”. Muitos acadêmicos, cientistas, empresas e entusiastas estão trabalhando em vários elementos tecnológicos no domínio da visão computacional (Bhatt et al., 2020, p. 1-2).

Isso tudo significa que as tecnologias da visão computacional têm estado em um cenário intenso de mudanças, marcadas por avanços e conquistas quase exponenciais, e “testemunhada pela qualidade, e muitas vezes, pela quantidade de contribuições notáveis para o campo publicadas em importantes conferências e revistas como ICCV e IEEE” (Davies, 2012, p. 21).

Diante disso, as revistas e as conferências acadêmicas sobre o tema não poderiam deixar de serem explicitadas. Segundo Russell e Norving (2013), as principais revistas de visão computacional são *Transactions on Pattern Analysis and Machine Intelligence* – IEEE, fundada em 1979 e a revista *Intelligence and International Journal of Computer Vision* – IJCV, constituída em 1987. Já as conferências, incluem-se a *International Conference on Computer Vision* – ICCV (1987), *Computer Vision and Pattern Recognition* – CVPR (1983) e *European Conference on Computer Vision* – ECCV (1990).

Desse modo, enquanto há uma reflexão do contínuo desenvolvimento, cresce o número de cursos e cadeiras acadêmicas em processamento de imagens digitais, bem como visão computacional oferecidos em Instituições de Ensino de todo o mundo. Todavia, “o assunto sofre, no entanto, com a escassez de textos ‘completos’ no sentido de serem acessíveis ao principiante, de uso para os instruídos e atualizados” (Sonka; Hlaváč; Boyle, 2014, p. 27). Bem, tal fato, aparentemente, ainda persiste no ano de 2023.

Sistema de visão computacional

A imersão no tema de visão computacional não é uma tarefa tão simples, principalmente quando a ideia é uma abordagem descritiva de sua origem e de suas funcionalidades. Os livros e artigos científicos disponíveis quase sempre tratam de modelos matemáticos, descrição de algoritmos complexos, bem como processamento de imagens e vídeos por meio de tecnologias preexistentes, a exemplo de técnicas de utilização de câmeras digitais inteligentes, técnicas de reconhecimento de cenários e objetos, mecanismos *machine learning* e *deep learning*, além de programações computacionais que utilizam redes neurais artificiais e redes neurais convolucionais, ou do termo inglês *convolutional neural network*.

Além do mais, boa parte desses documentos é de língua inglesa, o que gera para estudantes e pesquisadores brasileiros um pouco mais de afincamento para a concretude do entendimento acerca das particularidades dos sistemas de visão computacional. No entanto, a concepção geral é a de que a visão computacional está interconectada transversalmente aos sistemas de *machine learning*, *deep learning* e redes neurais convolucionais. Nesse ínterim, a missão dos próximos parágrafos é compor uma trilha cronológica desses mecanismos vinculados à visão computacional.

Machine learning e deep learning

A capacidade de aprender é essencial para um comportamento inteligente. Isso inclui atividades como “memorizar, observar e explorar situações para aprender fatos, melhorar habilidades motoras e cognitivas por meio de novas práticas e organização do conhecimento em representações apropriadas” (Faceli *et al.*, 2011, p. 2-3).

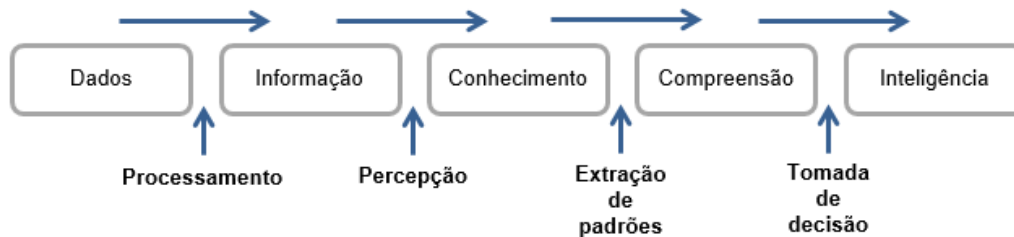
À luz disso, sob a perspectiva computacional, Cozman e Neri, (2021, p. 28) descrevem que, sem dúvida, as técnicas relacionadas à *machine learning* foram as que mais obtiveram atenção nos últimos anos. Na década de 1990, o “aprendizado de máquina era um dos possíveis focos de atenção dentro de IA. Todavia, hoje é uma fração dominante da área e, em certos aspectos, mais popular que a própria IA”.

Sendo assim, o *machine learning* lida com a compreensão da inteligência, ao passo que desenvolve algoritmos que aprendem com dados disponíveis, assim o computador adquire experiência e melhora seu comportamento de aprendizado ao longo do tempo (Holzinger, 2018).

Outrossim, o reconhecimento de padrões por intermédio do *machine learning* é, talvez, a tecnologia mais popular de IA, pois, “com base nos modelos de aprendizado, realizam-se previsões em dados desconhecidos e imperceptíveis ao ser humano” (Prateek, 2017, p. 31). Nessa concepção,

na Figura abaixo, esquematiza-se, segundo Prateek (2017), como os dados brutos são convertidos em sabedoria por meio de vários níveis de processamento de *machine learning*.

Figura 1. Níveis de processamento de IA em *machine learning*.



Fonte: Prateek (2017).

Sob essa perspectiva, o *machine learning* “é quase inteiramente dominado por técnicas que extraem padrões de grandes bases de dados”. Nesse sentido, há uma concentração expressiva em sistemas estatísticos, fundamentado em mecanismos inspirados na biologia (Cozman; Neri, 2021, p. 28). Destarte, “a classificação de dados do *machine learning* resolve o problema de identificar a categoria à qual um novo ponto de dados pertence” (Prateek, 2017, p. 31).

Diante disso, a construção de modelo de classificação, com base em *machine learning*:

Dá-se com base no conjunto de dados de treinamento contendo pontos de dados e os rótulos correspondentes. Por exemplo, para verificar se a imagem fornecida contém o rosto de uma pessoa ou não, é necessário um conjunto de dados de treinamento contendo classes correspondentes a essas duas classes: face e sem face. Em seguida, treina-se o modelo com base nas amostras de treinamento (Prateek, 2017, p. 31).

Nesse sentido, essa tecnologia é usada em robótica, reconhecimento de fala, previsões do mercado financeiro, bem como no reconhecimento de imagem e em sistemas de visão computacional (Prateek, 2017). No entanto, em consonância ao pensamento de Holzinger (2018, p. 1), os grandes desafios dessa tecnologia “estão na criação de sentido, na compreensão do contexto e na tomada de decisões diante de incertezas”.

Sob essa ótica, uma das principais restrições é que os programas de aprendizado são limitados ao conjunto de dados disponibilizados. “Se o conjunto de dados for pequeno, os modelos de aprendizado também serão limitados” (Prateek, 2017, p. 31).

Nesse contexto, o *machine learning*, especialmente impulsionado pelo *deep learning*, revolucionou a IA, bem como ultrapassou os humanos a alcançarem melhores desempenhos em algumas áreas, do reconhecimento visual de objetos até jogos complexos. Diante disso, o *deep learning* processa, de modo eficaz, “a enorme quantidade de dados disponíveis atualmente e possui grande poder e velocidade computacionais das máquinas modernas” (Costa *et al.*, 2021, p. 61).

A aplicação dessa tecnologia, que é uma subárea do *machine learning*, tornou-se realidade a partir dos anos 2000, com os avanços tecnológicos exponenciais das unidades de processamento gráfico, as quais possibilitaram a construção de “requisitos computacionais complexos de multiplicação de múltiplas matrizes com processamento paralelo” (Mutasa; Sun; Ha, 2021, p. 1).

Diante disso, o *deep learning* reveste-se de significância para a IA, haja vista a estruturação de novos métodos, como as redes neurais artificiais – RNA e redes neurais *convolucionais*, as quais revolucionaram a prática nas áreas de visão computacional e processamento de linguagem natural – PLN, além de praticamente todos os demais campos de aplicação de modelagem quantitativa (Cozman; Neri, 2021). Mutasa, Sun e Ha (2021, p. 1) ressaltam que:

Os algoritmos de *deep learning* extraem e combinam automaticamente representações de nível superior de uma

entrada para executar a tarefa necessária. Há entusiasmo em torno da capacidade do *deep learning* de descobrir conexões de dados, anteriormente desconhecidas, e realizar tarefas mais complexas.

Sendo assim, o uso de *deep learning* em visão computacional pode ser ordenado em diferentes mecanismos: “classificação, detecção, segmentação e processamento, tanto em imagens quanto em vídeos”. Dessa forma, ao utilizar técnicas desse tipo de sistema, a visão computacional se apropria das propriedades da visão humana em um computador (Shanmugamani, 2018, p. 21).

Destarte, o desenvolvimento recente das abordagens de *deep learning* provou ser útil na “detecção e classificação de objetos a partir de imagens, em particular, por intermédio do uso de variações de redes neurais *convolucionais*” (Sheaves *et al.*, 2020, p. 6). Nesse contexto, a supracitada rede será apresentada no próximo tópico.

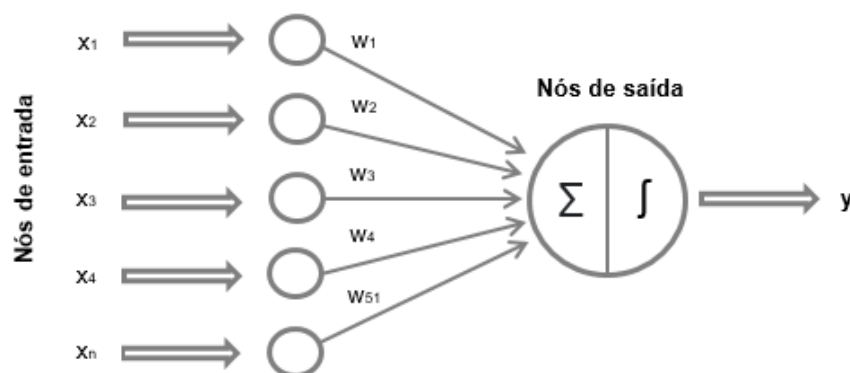
Redes neurais artificiais e redes neurais *convolucionais*

As redes neurais artificiais, por meio do *machine learning* e *deep learning*, simulam o sistema nervoso humano, “tratando as unidades computacionais em um modelo de *machine learning* semelhante aos neurônios humanos” (Aggarwal, 2018, p. 7). No nível mais básico, “uma rede neural artificial é uma função que inclui unidades, que também podem ser chamadas de neurônios” (Taulli, 2019, p. 73).

As redes neurais artificiais contêm arranjos em nós, semelhantes a corpos celulares humanos, as quais se articulam com outros nós por meio de conexões programadas, que são ponderadas com base em sua capacidade de fornecer um resultado almejado (Choi *et al.*, 2019, p. 7).

A grande contribuição da RNA “é criar inteligência artificial por meio de aprendizado, cuja arquitetura computacional simule o sistema nervoso humano” (Aggarwal, 2018, p. 7). Nessa conjectura, a rede neural mais simples é chamada de “*perceptron*”. Esta rede neural contém uma única camada de entrada e um único nó de saída, conforme a Figura 2.

Figura 2. Arquitetura básica de um *perceptron*



Fonte: Aggarwal (2018).

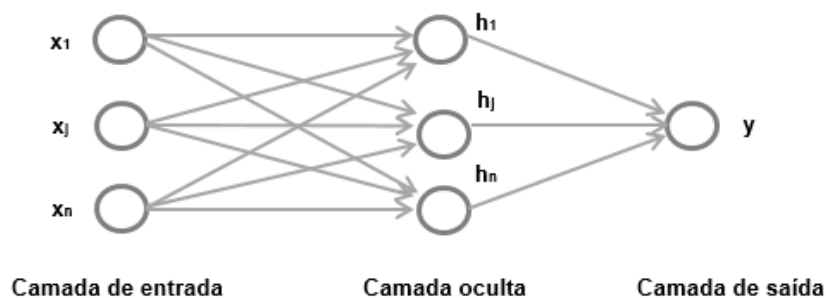
A instância de treinamento é dada pela fórmula , em que cada $[x_1...x_n]$ contém n variáveis de recurso e contém o valor observado da variável de classe binária. A camada de entrada contém n nós que transmitem as n características $[x_1...x_n]$ com arestas de peso $[w_1 ...w_n]$ para um nó de saída. A camada de entrada não executa nenhum cálculo por si só. Todavia, a função linear $=$ é calculada no nó de saída. Posteriormente, o sinal desse valor real é usado para prever a variável dependente de (Aggarwal, 2018).

Em outras palavras, o *perceptron* recebe várias entradas de dados e realiza uma soma ponderada, por meio de pesos, para produzir uma saída. Nessa lógica, a soma passa por uma

função degrau unitário, neste caso, para um problema de classificação binária. Sendo assim, um *perceptron* só aprende funções simples ao compreender os pesos a partir de exemplos, esse processo é chamado de treinamento (Shanmugamani, 2018).

Por conseguinte, têm-se, também, os *perceptrons* multicamadas, os quais são conectados para formar camadas ou unidades ocultas. “As unidades ocultas formam a base não linear que mapeia as camadas de entrada para as camadas de saída em um espaço de menor dimensão”. Os valores são calculados pela adição ponderada das entradas com vieses (Shanmugamani, 2018, p. 11).

Figura 3. Diagrama de um *perceptron* multicamada



Fonte: Shanmugamani (2018).

Desse modo, várias entradas $[x_1...x_n]$ são passadas por uma camada oculta de *perceptrons* $[h_1...h_n]$ e somadas à saída. Nessa vertente, cada camada oculta pode ter uma das funções de ativação, as quais decidem se um *perceptron* deve ser ativado ou não. Além do mais, o número de camadas ocultas e *perceptrons* podem ser escolhidos com base no problema (Shanmugamani, 2018).

Posto isto, em visão computacional, as redes neurais artificiais foram desenvolvidas em analogia à funcionalidade da visão humana, as quais executam combinações, de modo progressivo, entradas de neurônios ou *perceptrons*, que reconhecem desde características simples, como linhas ou cores no campo visual, até representações mais complexas (Mutasa; Sun; Ha, 2021).

Nos últimos anos, “os métodos baseados em *deep learning* tornaram as técnicas mais poderosas que detectam automaticamente os recursos da imagem como um mapa de recursos” (Joshi; Patel, 2020, p. 6). Diante disso, a rede neural *convolucional* tornou-se uma das redes neurais artificiais mais representativas no campo da aprendizagem profunda (Li *et al.*, 2021, p. 1).

A diferença entre redes neurais comuns e redes neurais *convolucionais* está no tipo de camadas que são usadas e os dados são tratados nas entradas. A RNC assume que as entradas são imagens, o que lhes permite extrair propriedades específicas. Isso torna as redes neurais *convolucionais* muito mais eficientes em lidar com técnicas de processamento de imagens (Prateek, 2017, p. 400).

Por certo, as redes neurais profundas, segundo Khemasuwan *et al.* (2020, p. 6):

São comparáveis à do padrão de conectividade dos neurônios no cérebro humano, seguindo um modelo hierárquico que cria uma estrutura semelhante a um funil para fornecer uma camada totalmente conectada, em que todos os neurônios estão interligados. De interesse histórico, a precisão de visão computacional de última geração foi comparada, favoravelmente, com a precisão humana em tarefas de classificação de objetos.

Outrossim, Bhatt *et al.* (2020, p. 1) ratifica que a RNC beneficiou a comunidade de visão computacional ao fornecer resultados de ponta em “processamento de vídeo, reconhecimento de objetos, classificação e segmentação de imagens, processamento de linguagem natural, reconhecimento de fala e muitos outros campos”. Diante desses fatos, na atualidade, as redes

neurais *convolucionais* executam tarefas de reconhecimento de imagens, quase tão bem quanto os humanos (Correia *et al.*, 2020, p. 2).

Ademais, enquanto os métodos baseados em machine learning exigem dados rotulados, a rede neural convolucional pode executar dados não rotulados (Joshi; Patel, 2020, p. 2). Além disso, a introdução de grandes quantidades de dados e hardware prontamente disponíveis abriram novos rumos para projetos que utilizam RNC. “Vários conceitos inspiradores para o progresso da RNC são investigados, incluindo funções alternativas de ativação, regularização e otimização de parâmetros” (Bhatt *et al.*, 2020, p 1).

Em termos mais técnicos, as redes neurais convolucionais são robustas em comparação a uma rede neural regular. Diante disso, utilizar processamento de imagens em uma simples rede, provocaria um overfitting, haja vista o expressivo quantitativo de perceptrons necessários para a concretização das tarefas (Shanmugamani, 2018).

Esse tipo de procedimento requer um grande número de perceptrons, pois uma imagem é considerada volumosa, diante dos fatores de dimensão, largura e profundidade. Sendo assim, as redes neurais convolucionais possuem pesos, vieses e saídas por meio de uma ativação não linear. Cada uma das camadas transforma o volume de entrada em um volume menor de saída (Shanmugamani, 2018).

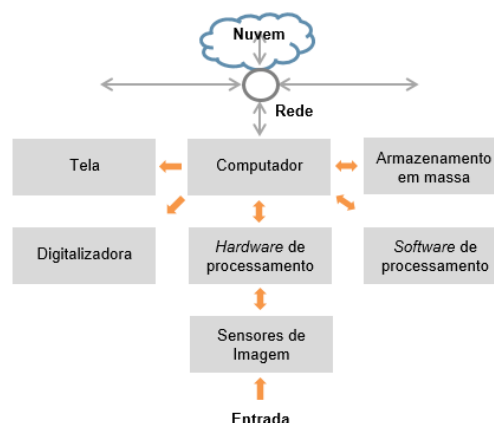
Diante disso, Li *et al.* (2021, p. 1) ratifica que as redes neurais convolucionais, em conjunto à visão computacional, permitiu o que era considerado impossível nos últimos séculos, “como reconhecimento facial, veículos autônomos, supermercados de autoatendimento e tratamentos médicos inteligentes”. Para tanto, atrelado à RNC, os sistemas de visão computacional necessitam de um grupo de componentes e etapas de processamento de imagens, os quais são descritos na seção subsequente.

Componentes e processamento de imagens em sistemas de visão computacional

Como visto nos tópicos anteriores, para que ocorra a concretude das técnicas de um sistema de visão computacional, é necessário um conjunto de tecnologias digitais que se inter-relacionam, como os mecanismos de *deep learning* e as redes neurais *convolucionais*. Outrossim, conforme Nixon e Aguado (2020), um sistema de visão computacional básico requer, ao menos, uma câmera digital e sua interface com *softwares* de processamento de imagem.

Todavia, para além disso, Gonzalez e Woods (2018) expõem que a tendência continua em direção à miniaturização e combinação de pequenos computadores de uso geral com *hardware* e *software* de processamento de imagem especializado. Nesse ínterim, os autores estabelecem os componentes básicos que compõem um sistema típico de uso geral usado para processamento de imagens digitais, os quais estão descritos na Figura 4.

Figura 4. Componentes de um sistema de processamento de imagens

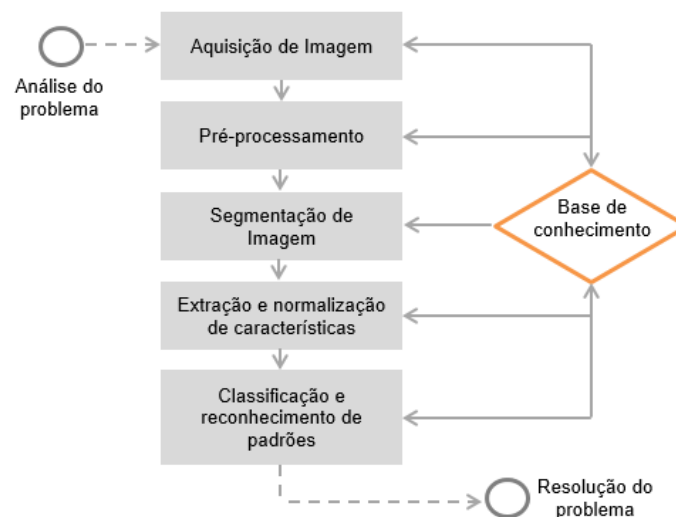


Fonte: Gonzalez e Woods (2018).

Segundo Gonzalez e Woods (2018), há, ao menos, nove componentes básicos para processamento de imagens. São eles: sensores físicos de imagens, os quais convertem as entradas em dados digitais; dispositivos de digitalização, impressoras a laser, câmeras de filme, dispositivos sensíveis ao calor, unidades de jato de tinta e unidades digitais, como discos ópticos e CD-ROM; *hardwares* de processamento de imagem especializado, cujas funções permitem rápidas transferências de dados; *software* de processamento de imagem, que consiste em módulos especializados que realizam tarefas específicas; computador, que pode variar de um PC a um supercomputador; armazenamento em massa, o qual envolve mecanismos de compactação de imagens; telas para exibição das imagens; rede e comunicação em nuvem, quase padrão em qualquer sistema de computador em uso hoje.

Posto isto, estão atrelados a esses *softwares* e *hardwares* alguns requisitos mínimos para que o computador reconheça e interprete determinada imagem. Isso ocorre por intermédio de etapas típicas de processamento de imagens em sistemas de visão computacional, conforme é explicitado na 3ª (2009) e 4ª (2018) edição da obra *Digital Image Processing* de Gonzalez e Woods e simplificado por Almeida (2018), como exemplo abaixo.

Figura 5. Etapas típicas no processamento de imagens em visão computacional



Fonte: Gonzalez e Woods (2018); Almeida (2018).

Conforme Almeida (2018), a estrutura de um sistema de visão computacional contempla etapas sequenciais, mediante a descrição de um problema visual, em que todas estão interligadas a uma base de conhecimento dos resultados obtidos. Desse modo, têm-se as etapas:

- I. Aquisição de imagem: é a primeira etapa de um processamento de imagens de um sistema de visão computacional. Segundo Gonzalez e Woods (2018, p. 41), “a aquisição pode ser tão simples quanto receber uma imagem que já está em formato digital”;
- II. Pré-processamento: envolve elementos que assegurem a qualidade da imagem, tais como redução de ruídos, ajuste de luz e contraste; processamento morfológico e compactação de imagem (Gonzalez; Woods, 2018);
- III. Segmentação de imagem: divide uma imagem em suas partes ou objetos constituintes. De modo geral, a segmentação é uma das tarefas mais complexas no processamento digital de imagens. “Um procedimento de segmentação robusto leva o processo a uma solução bem-sucedida para problemas de imagem que exigem que os objetos sejam identificados individualmente” (Gonzalez; Woods, 2018, p. 42-43);
- IV. Extração e normalização de características: utiliza o resultado da segmentação. Geralmente, por meio desse resultado, obtêm-se dados brutos de pixel, permitindo a entrada para os algoritmos de classificação de padrões (Gonzalez; Woods, 2018);
- V. Classificação e reconhecimento de padrões: é o processo que atribui um “rótulo (por

exemplo, “veículo”) a um objeto com base em seus descritores de recursos”, utilizando métodos de *deep learning* e redes neurais convolucionais (Gonzalez; Woods, 2018, p. 43).

Na Figura 6, Nixon e Aguado (2020) explicitam essa última etapa, a classificação e reconhecimento de padrões. No exemplo, após o transcorrer das quatro primeiras etapas de um sistema de visão computacional, o *software* reconhece pessoas por gênero nas ruas de uma cidade qualquer.

Figura 6. Classificação de pessoas por gênero com utilização de *deep learning*



Fonte: Nixon; Aguado (2020).

Esse é o resultado final de um sistema de visão computacional. É a partir da classificação e reconhecimento de padrões que os computadores tomam decisões, com base em outros sistemas de inteligência artificial. Todavia, conforme Nixon e Aguado (2020, p. 574), o percorrer “é bastante longo e complicado, e requer muito poder computacional”. Outrossim, o processamento e a classificação das imagens da Figura 6 são obtidos em tempo real, por meio de *deep learning*, utilizando-se técnicas de redes neurais convolucionais, aplicadas a um sensor de vídeo, após o sistema ter sido treinado (Nixon; Aguado, 2020).

Aplicação da visão computacional

A visão computacional tem muitas aplicações, incluindo “direção autônoma, inspeção industrial e realidade aumentada” (Shanmugamani, 2018, p. 6). Essa tecnologia tem sido muito utilizada nos campos de “agricultura, vigilância, saúde, entre outros campos da ciência e indústria” (Pangal *et al.*, 2021, p. 1).

Smits e Wevers (2021) ratificam o rápido desenvolvimento, nos últimos anos, da aplicabilidade da visão computacional:

Cientistas, governos e profissionais da indústria têm empregado modelos de visão computacional para tomar decisões de alto risco na sociedade. Eles são usados para uma ampla gama de tarefas, como inspeção automática na fabricação, controle de processos industriais, diagnóstico médico, policiamento, vigilância, navegação (carros autônomos), organização de dados visuais (em mídias sociais), marketing personalizado e uma infinidade de aplicações militares (Smits; Wevers, 2021, p. 2).

Por conseguinte, não é objetivo dos parágrafos subsequentes esgotarem a discussão da crescente multidisciplinaridade da visão computacional, mas exemplificar, por meio de algumas tecnologias existentes, suas reais aplicações. Sendo assim, suas funcionalidades e seus mecanismos são transversais e, desse modo, em consonância aos autores Shanmugamani, (2018), Pangal *et al.* (2021) e Smits e Wevers (2021), abrangem ou podem abranger todos os segmentos da sociedade e da economia. Nessa acepção, delineiam-se, na sequência, quatros diferentes usos de visão computacional.

Exposto isto, explicita-se, em primeiro lugar, talvez a mais conhecida e popular tecnologia de visão computacional, o reconhecimento facial e seus mecanismos. Ngoc *et al.* (2020, p. 473) descreve, em seu trabalho intitulado “*Automatic Facial Expression Recognition System Using Convolutional Neural Networks*”, o reconhecimento de expressão facial – REF, em que o principal objetivo é “classificar uma determinada imagem facial em uma das sete emoções humanas: raiva, desgosto, medo, alegria, tristeza, surpresa e neutro”.

Em termos concretos, Simão, Fragoso e Roberto (2020) expressam, no estudo “Reconhecimento Facial e o Setor Privado” do IDEC, que a empresa ViaQuatro, concessionária da Linha 4 Amarela do metrô de São Paulo, instalou, nas plataformas de embarque e desembarque, por ela administradas, um sistema de “Portas Interativas Digitais”. A tecnologia consiste, basicamente:

Em câmeras, instaladas junto a painéis de publicidade, que identificam a presença de rostos humanos e reconhecem as emoções das pessoas diante de determinado anúncio publicitário, além de permitir a coleta de dados sobre o número de pessoas que transitam na plataforma e suas características de gênero, etnia e faixa etária (Simão; Fragoso; Roberto, 2020, p. 22).

Outra tecnologia de visão computacional é o reconhecimento de imagens, materializadas, dentre outras ferramentas de mercado, pelo *Google Cloud Vision*, que é, segundo Pereira e Moreschi, (2020), o resultado de muito trabalho humano e de *machine learning*. O mecanismo é baseado na recente expansão de inovações do Google, o qual utiliza algoritmos especializados e exaustivos dados para o treinamento de suas máquinas inteligentes, cujo intuito é permitir que os computadores enxerguem e compreendam o mundo.

A visão computacional também tem sido crescentemente aplicada na automação de diagnósticos de saúde. Na contemporaneidade, já é possível contar com diagnósticos automáticos acertados e precisos e, algumas vezes, até mais precisos que aqueles dados pelos profissionais da área de saúde. A indústria *iFlytek* desenvolveu um robô que conseguiu passar no exame nacional para licenciamento de médicos da China. O “Médico Assistente robô” é capaz de apontar as queixas dos pacientes, avaliar as imagens de tomografia computadorizada e fazer o diagnóstico inicial (Ludemir, 2021).

Por último e mais desafiante, são os sistemas de visão computacional aplicados em veículos autônomos. Conforme Guedes e Machado (2020, p. 53), esse tipo de veículo se depara com cenários complexos e próximos do realmente enfrentado por humanos, como a “percepção do ambiente, o reconhecimento de pedestres e a predição de suas intenções, a condução do veículo em condições meteorológicas adversas”, como chuvas, raios solares, dentre outros.

Em 2009, o Google assumiu a liderança em pesquisas relacionadas a veículos autônomos, mas não foi a única empresa a investir em projetos desse cunho. Em 2013, a Mercedes Benz apresentou o *S500 Intelligent Drive*, em um “passeio autônomo de 103 km na histórica rota *Bertha Benz de Mannheim a Pforzheim* na Alemanha” (Janai *et al.*, 2020, p. 14).

O veículo Mercedes S500 foi equipado, segundo Janai *et al.* (2020, p. 14):

Com *hardware* de sensor próximo à produção. A detecção de objetos e a análise do espaço livre foram realizadas usando radar e visão estéreo. A visão monocular foi usada para detecção de semáforos e classificação de objetos. Dois algoritmos de visão complementares, baseados em recursos de ponto e baseados em marcação de pista, permitiram a

localização com precisão de centímetros em relação a mapas HD anotados manualmente. Enquanto se concentra em uma única rota, o esforço demonstrou que a condução autônoma em ambientes complexos do centro da cidade com base em *hardware* próximo à produção e mapas HD é viável.

É certo que sistemas de visão computacional aplicados em carros autônomos possuem um nível maior de complexidade, diante das n variáveis que os cercam, todavia, encontrar esses veículos nas ruas se tornará uma realidade próxima, ao menos é o que demonstram os projetos veiculados em mídias das empresas Google, Uber, Tesla, Mercedes-Benz, dentre outras.

Conforme o portal eletrônico da NVIDIA (2020), líder global em computação acelerada, há um planejamento de criar, em mútua cooperação à Mercedes-Benz, um revolucionário sistema de computação em veículo e infraestrutura de computação de IA. A pretensão é que, a partir de 2024, esse sistema seja inserido em toda a frota de veículos da empresa, permitindo-lhes funções de condução automatizadas atualizáveis.

Diante disso, a visão computacional, enquanto um ramo da inteligência artificial, também possui variadas frentes de pesquisa e diferentes níveis de complexidade de aplicação. Isso é refletido por intermédio dos mecanismos descritos em tópicos anteriores e nos exemplos expostos nesta seção. Tudo indica que é só o início de uma nova era da humanidade.

Considerações finais

Como visto, no decorrer deste trabalho, a visão computacional está se tornando uma tecnologia cotidiana. O crescimento de sua utilização por diversos meios ocorre de maneira exponencial. Além do mais, na atualidade, há, no mercado, uma razoável lista de *softwares* que permitem o desenvolvimento de mecanismos de processamento de imagens, com base em exaustivos treinamentos de máquinas por meio de técnicas de *machine learning*, *deep learning*, bem como redes neurais artificiais e redes neurais convolucionais.

Apesar disso, a compreensão sobre a temática não é algo relativamente simplório e, quase sempre, o domínio da origem e das funcionalidades da visão computacional ocorre por parte de um seleto grupo de técnicos e profissionais altamente especializados. Outrossim, em consonância ao relato deste trabalho, os livros e artigos científicos disponíveis, quase sempre, tratam de modelos matemáticos, descrição de algoritmos complexos, bem como processamento de imagens e vídeos.

Além do mais, boa parte desses documentos é de língua inglesa, o que gera para estudantes e pesquisadores brasileiros um pouco mais de afinco para a concretude do entendimento acerca das particularidades dos sistemas de visão computacional. Tendo isso em vista, buscou-se, assim, mitigar a baixa compreensão de tal tecnologia, em que os esforços foram direcionados ao estudo dos fundamentos da visão computacional por meio de arcabouço teórico acerca do reconhecimento artificial de imagens e vídeos.

Referências

ADAMIAN, N; *et al.* An Open-Source Computer Vision Tool for Automated Vocal Fold Tracking From Videoendoscopy. *Laryngoscope*, v. 0, p. 1–7, 2020. DOI: 10.1002/lary.28669.

AGGARWAL, Charu C. **Neural Networks and Deep Learning**. NY, USA: Springer, 2018.

ALMEIDA, Dyego de Oliveira. **Métodos de Visão Computacional Aplicados a Extração de Características de Ambientes Urbanos em Imagens de Satélite de Baixa Resolução**. 2018. Dissertação (Mestrado em Direito) - Programa de Pós-Graduação do Instituto de Informática, Universidade Federal de Goiás, Goiânia, 2018.

BHATT, Dulari; *et al.* CNN Variants for Computer Vision: History, Architecture, Application, Challenges

and Future Scope. **Electronics**, v.10, n. 20, p. 1-28, 2021. DOI: 10.3390/electronics10202470.

CERTI. **Visão computacional: conheça as aplicações e a importância para a transformação digital**. CERTI, 2020. Disponível em: <https://certi.org.br/blog/visao-computacional/>. Acesso em: 20 jul. 2022.

CHOI, Rene Y.; *et al.* Introduction to Machine Learning, Neural Networks, and Deep Learning. **TVST**, Special Issue, v. 9, n. 2, 2020. DOI: 10.1167/tvst.9.2.14.

CORREIA, David L. P.; *et al.* Leveraging Artificial Intelligence for Large-Scale Plant Phenology Studies From Noisy Time-Lapse Images. **IEEE Access**, v. 8, 2020.

COSTA, Lucas; *et al.* Determining leaf stomatal properties in citrus trees utilizing machine vision and artificial intelligence. **Precision Agriculture**, v. 22, 2021. DOI: 10.1007/s11119-020-09771-x.

COZMAN, Fabio G.; NERI, Hugo. O que, afinal, é Inteligência Artificial? *In*: COZMAN, Fabio G.; PLONSKI, Guilherme Ary; NERI, Hugo (Org.). **Inteligência artificial: avanços e tendências**. São Paulo: Instituto de Estudos Avançados, 2021.

DAVIES, E. R. **Computer and Machine Vision: Theory, Algorithms, Practicalities**. 4. ed. Waltham, USA: Elsevier/AP, 2012.

FORSYTH, David A.; PONCE, Jean. **Computer Vision: a Modern Approach**. 2. ed. Boston, USA: Pearson, 2012.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. 4. ed. New York, USA: Pearson, 2018.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. 3. ed. New York, USA: Pearson, 2009.

GUEDES, Marcelo Santiago; MACHADO, Henrique Felix de Souza. **Veículos Autônomos Inteligentes e a Responsabilidade Civil nos Acidentes de Trânsito no Brasil: Desafios regulatórios e propostas de solução e regulação**. v. 2. Brasília: ESMPU, 2020.

HOLM, Elizabeth A.; *et al.* **Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis**. Metallurgical And Materials Transactions A, 2020. DOI: 10.1007/s11661-020-06008-4.

HOLZINGER, Andreas. Introduction to MACHine Learning & Knowledge Extraction (MAKE). **Mach. Learn. Knowl. Extr.** v. 1, n. 1, 2018. DOI:10.3390/make101000.

HOSSEINI, Kasra; *et al.* Maps of a Nation? The Digitized Ordnance Survey for New Historical Research. **Journal of Victorian Culture**, v. 26, n. 2, p. 284–299, 2021. DOI: 10.1093/jvcult/vcab009.

JANAI, Joel; *et al.* Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. **Foundations and Trends in Computer Graphics and Vision**. v. 12, n. 1–3, p. 1–308, 2021. DOI: 10.1561/06000000079.

JOSHI, Khushbu; PATEL, Manish I. Recent advances in local feature detector and descriptor: a literature survey. **International Journal of Multimedia Information Retrieval**, 2020. DOI: 10.1007/s13735-020-00200-3.

KHEMASUWAN, Danai; *et al.* Artificial intelligence in pulmonary medicine: computer vision,

predictive model and COVID-19. **Eur Respir Rev** v. 29, 2020. DOI: 10.1183/16000617.0181-2020.

LI, Yu; *et al.* Crash report data analysis for creating scenario-wise, spatio-temporal attention guidance to support computer vision-based perception of fatal crash risks. **Accident Analysis and Prevention**, v. 151, 2021. DOI: 10.1016/j.aap.2020.105962.

LUDEMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, 2021.

MUTASA, Simukayi; SUN, Shawn; HA, Richard. Understanding artificial intelligence based radiology studies: CNN architecture. **Clinical Imaging**. v. 80, p. 72–76, 2021. DOI: 10.1016/j.clinimag.2021.06.033.

NGOC, Hung; *et al.* Automatic Facial Expression Recognition System Using Convolutional Neural Networks. In: VAN TOI, V. *et al.* (eds.), **7th International Conference on the Development of Biomedical Engineering in Vietnam (BME7)**, IFMBE Proceedings. DOI: 10.1007/978-981-13-5859-3_82.

NIXON, Mark S.; AGUADO, Alberto S. **Feature Extraction and Image Processing for Computer Vision**. 4. ed. London, UK: Academic Press/ Elsevier, 2020.

NVIDIA. **Mercedes-Benz and NVIDIA to Build Software-Defined Computing Architecture for Automated Driving Across Future Fleet**. NVIDIA, 2020. Disponível em: <https://nvidianews.nvidia.com/news/mercedes-benz-and-nvidiato-build-software-defined-computing-architecture-for-automated-driving-across-future-fleet>. Acesso em: 20 jul. 2022.

PALSULE, Siddhesh; MISHRA, Sandeep. **Computer Vision Market Analysis and Segment Forecasts to 2028**. Grand View Research, 2020.

PANGAL, Dhiraj J.; *et al.* A Guide to Annotation of Neurosurgical Intraoperative Video for Machine Learning Analysis and Computer Vision. **World Neurosurgery**, 2021. DOI: 10.1016/j.wneu.2021.03.022.

PEREIRA, Gabriel; MORESCHI, Bruno. Artificial intelligence and institutional critique 2.0: unexpected ways of seeing with computer vision. **AI & Society**, 2020. DOI: 10.1007/s00146-020-01059-y.

PEREIRA, Gabriel; MORESCHI, Bruno. Artificial intelligence and institutional critique 2.0: unexpected ways of seeing with computer vision. **AI & Society**, 2020. DOI: 10.1007/s00146-020-01059-y.

PRATEEK, Joshi. **Artificial Intelligence with Python: Build real-world Artificial Intelligence applications with Python to intelligently interact with the world around you**. Mumbai, India: Packt, 2017.

ROBERTS, Lawrence Gilman. **Machine Percepton of Three-dimensional Solids**. 1961. (Tese de doutorado), Massachusetts Institute of Technology, Cambridge, 1961.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Tradução: SIMILLE, Regina Célia. Rio de Janeiro: Elsevier, 2013.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. 4. ed. Hoboken: Pearson, 2021.

SAVEKAR, Avinash; KUMAR, Vineet. **AI in Computer Vision Market: global opportunity analysis and industry forecast, 2021-2030**. Portland: Allied Market Research, 2021.

SEBE, N.; *et al.* **Machine Learning in Computer Vision**. Dordrecht, Netherlands: Springer, 2005.

SHANMUGAMANI, Rajalingappaa. **Deep Learning for Computer Vision**: Expert techniques to train advanced neural networks using Tensor Flow and Keras. Mumbai, India: Packt, 2018.

SHEAVES, Marcus; *et al.* Optimizing video sampling for juvenile fish surveys: Using deep learning and evaluation of assumptions to produce critical fisheries parameters. **Fish and Fisheries**, p. 1–18, 2020. DOI: 10.1111/faf.12501.

SIMÃO, Bárbara; FRAGOSO, Nathalie; ROBERTO, Enrico. **Reconhecimento Facial e o Setor Privado: guia para a adoção de boas práticas**. São Paulo: InternetLab/IDEC, 2020.

SMITS, Thomas; WEVERS, Melvin. The agency of computer vision models as optical instruments. **Visual Communication**, v. 0, n. 0, p. 1–21, 2021. DOI: 10.1177/1470357221992097.

SONKA, Milan; HLAVÁČ, Vaclav; BOYLE, Roger. **Image Processing, Analysis, and Machine Vision**. 4. ed. Stamford, USA: Cengage Learning, 2014.

TAULLI, Tom. **Artificial Intelligence Basics**: a non-technical introduction. Monrovia, USA: Apress, 2019.

VENTURUS. **Visão Computacional**. Venturus, 2022. Disponível em: <https://www.venturus.org.br/visao-computacional>. Acesso em: 20 jul. 2022.

Recebido em 8 de abril de 2023
Aceito em 15 de setembro de 2023