

Visão Computacional: reconhecimento de padrões em imagens

Matheus Kenzo Tuda Roriz Silva¹, Gilberto Fernandes Junior¹

¹Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

matheus.kenzo.tuda@uel.br, gilfernandes@uel.br

Abstract. *Pattern recognition in images is an important area of computer vision, with applications ranging from security to healthcare. Convolutional neural networks have gained prominence for their ability to extract relevant features and recognize complex patterns in large datasets. This project aims to explore the effectiveness of Convolutional neural networks in image pattern recognition, highlighting their advantages over traditional methods. The expected results of this study will contribute to the advancement of computer vision by demonstrating the accuracy and robustness of Convolutional neural networks.*

Resumo. *O reconhecimento de padrões em imagens é uma área importante da visão computacional, com aplicações que vão desde a segurança até a saúde. As redes neurais convolucionais têm se destacado por sua capacidade de extrair características relevantes e reconhecer padrões complexos em grandes volumes de dados. Este projeto busca explorar a eficácia das redes neurais convolucionais no reconhecimento de padrões em imagens, destacando suas vantagens em relação a métodos tradicionais. Espera-se que os resultados deste estudo contribuam para o avanço da visão computacional, evidenciando a precisão e a robustez das redes neurais convolucionais.*

1. Introdução

”Podem as máquinas pensar?” essa é a questão levantada por Alan Turing em seu artigo seminal *Computing Machinery and Intelligence*, no qual ele vislumbra a possibilidade da construção de máquinas capazes de pensar e aprender da mesma forma que os seres humanos. Turing propôs um teste, conhecido como o ”Teste de Turing”, como um experimento mental para avaliar se um computador poderia simular inteligência humana. Se um computador, ao responder perguntas escritas, não pudesse ser distinguido de um ser humano, ele seria considerado inteligente[22].

Para que um computador tenha sucesso nesse teste, ele precisaria possuir capacidades como processamento de linguagem natural para se comunicar de forma eficaz, representação de conhecimento para armazenar informações, raciocínio automatizado para responder a perguntas e gerar novas conclusões, e aprendizado de máquina para adaptar-se a novas situações e identificar padrões.

Embora Turing tenha considerado desnecessário simular fisicamente um ser humano para demonstrar a inteligência de uma máquina, pesquisadores posteriores concluíram que, para um ”Teste de Turing Total”, seria necessário que a máquina interagisse com o mundo real. Isso implica a necessidade de tecnologias como visão computacional e reconhecimento de fala para perceber e interpretar o ambiente, além de robótica para manipular objetos e movimentar-se no mundo físico[17].

A visão computacional é a ciência responsável por simular a capacidade de visão de uma máquina, permitindo que ela perceba e interprete o mundo ao seu redor. Isso é feito através da extração de informações a partir de imagens capturadas por câmeras de vídeo, sensores, scanners e outros dispositivos. As informações extraídas dessas imagens permitem à máquina reconhecer, classificar e manipular objetos, além de fazer inferências sobre o conteúdo da imagem[3].

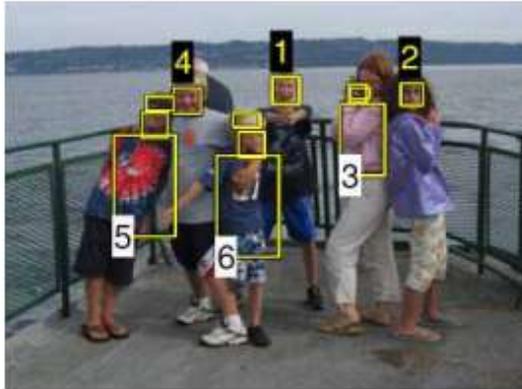


Figura 1. O sistema de visão do ser humano não tem problemas para interpretar as variações sutis de translucidez e sombreado desta fotografia, além de ter uma noção dos objetos que compõe seu fundo.

O ser humano é capaz de perceber um mundo de objetos tridimensionais com aparente facilidade. Por exemplo, ao olhar para um vaso de flores na mesa em uma mesa, podemos discernir o formato e a translucidez de cada pétala, através dos sutis padrões de luz e sombra que brincam em sua superfície, e segmentar cada flor do fundo da cena da figura 1. Da mesma forma, ao observar um retrato de grupo de pessoas, é relativamente fácil contar e identificar todas as pessoas na foto e até mesmo inferir suas emoções a partir das expressões faciais da figura 2a.

Pesquisadores em visão computacional têm buscado desenvolver técnicas matemáticas para recuperar a forma tridimensional e a aparência de objetos a partir de imagens. Atualmente, técnicas avançadas permitem a construção precisa de modelos 3D de ambientes a partir de múltiplas fotografias parcialmente sobrepostas figura 2c. Além disso, a correspondência estéreo possibilita a criação de modelos de superfície 3D densos e precisos demonstrados na figura 2d. A segmentação de pessoas e objetos em imagens tem alcançado sucesso moderado demonstrados na figura 2a. No entanto, apesar desses avanços significativos, o sonho de um computador que possa interpretar uma imagem com a mesma riqueza de detalhes e contexto causal que uma criança de dois anos continua sendo um desafio complexo, tanto por não ser apenas um processo visual, mas de todo um contexto do ambiente e assimilação a partir dos conhecimentos do indivíduo.[21]

A importância da visão computacional se reflete em suas vastas aplicações em diversas áreas, tais como física, biologia, indústria, segurança, entre outras, contribuindo para avanços significativos na ciência e na tecnologia. Na medicina, por exemplo, os Sistemas Computacionais de Apoio ao Diagnóstico (CAD) são essenciais. Como o olho humano é limitado em relação à percepção de alguns padrões em imagens e na detecção de imagens com ruídos ou contraste muito baixo, o uso de recursos computacionais pode



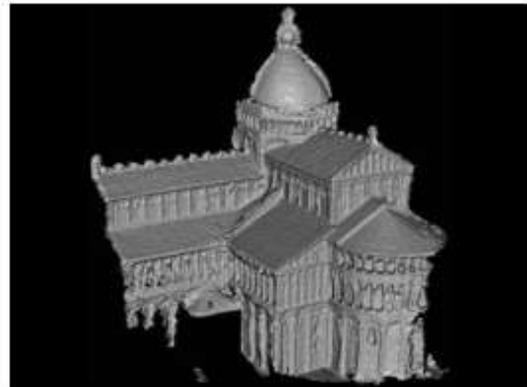
(a)



(b)



(c)



(d)

Figura 2. Alguns exemplos de algoritmos e aplicações de visão computacional. (a) Algoritmos para detecção de rosto, juntamente com algoritmos de detecção de roupas e cabelos baseados em cores, permitindo localizar e reconhecer os indivíduos nesta imagem (Sivic, Zitnick, and Szeliski 2006) © 2006 Springer[19]. (b) A segmentação de instâncias de objetos pode delinear cada pessoa e objeto em uma cena complexa (He, Gkioxari et al. 2017) © 2017 IEEE[9]. (c) Estrutura de algoritmos de movimento pode reconstruir um modelo de ponto 3D esparsa de uma grande cena complexa a partir de centenas de fotografias parcialmente sobrepostas (Snavely, Seitz, and Szeliski 2006) © 2006 ACM[20]. (d) Algoritmos de correspondência estéreo podem construir um modelo 3D detalhado de uma fachada de edifício a partir de centenas de fotografias com exposições diferentes tiradas da internet (Goesele, Snavely et al. 2007) © 2007 IEEE[6].

ajudar neste reconhecimento. Isso serve, por exemplo, como um auxiliar no diagnóstico de algumas doenças, evitando um diagnóstico impreciso ou equivocado. Dessa forma, seria uma espécie de segunda opinião para o médico, com o intuito de confirmar uma suspeita ou abrir novos horizontes à detecção de certa doença. Tais sistemas incluem o diagnóstico de fraturas cranianas, lesões pulmonares[2] e mamografia. Nestes sistemas, a visão computacional é utilizada para detectar anomalias, caracterizando-as e classificando-as, baseando-se num vasto conjunto de imagens e informações. Outro fato que contribui para a utilização da visão computacional é que a análise de uma radiografia torna-se mais precisa quando utilizam-se dois radiologistas para a detecção de anomalias. Dessa forma, poderia ser utilizado um sistema automatizado no lugar de um dos radiologistas, fazendo uma junção do resultado para um diagnóstico mais preciso.

Além disso, a visão computacional tem uma forte utilização na indústria, onde pode ser aplicada desde o controle de qualidade de determinado produto fabricado até em áreas mais específicas, como a caracterização e classificação de minério de ferro. Essa tecnologia não só melhora a eficiência e a precisão dos processos industriais, mas também reduz o desperdício e aumenta a produtividade.[4]

Dessa maneira, o objetivo deste trabalho de conclusão de curso é realizar uma análise comparativa das redes neurais convolucionais (CNNs) no reconhecimento de padrões em imagens. O foco será avaliar métricas de desempenho como precisão, *recall*, *F1-score* e *mean Average Precision* (mAP), comparando os resultados obtidos com outras abordagens de reconhecimento de padrões presentes na literatura. A proposta é elaborar um modelo eficaz de CNN, demonstrar suas vantagens e limitações em relação a outros métodos, e identificar as melhores situações de uso para cada tipo de técnica de reconhecimento.

Este documento está organizado da seguinte forma: A Seção 2 apresenta os conceitos, métodos, técnicas e revisão do estado da arte necessários para a elaboração do modelo proposto. A Seção 3 apresenta o objetivo a ser alcançado uma vez que o desenvolvimento proposto esteja concluído. A Seção 4 descreve como os objetivos serão atingidos fazendo uso da fundamentação teórico-metodológica e a revisão do estado da arte. A Seção 5 apresenta o cronograma de execução das atividades citadas na seção anterior. Finalmente, na seção 6 serão descritas as contribuições e resultados do trabalho.

2. Fundamentação Teórico-Metodológica e Estado da Arte

2.1. Processamento de imagem

O campo do processamento de imagens digitais refere-se ao tratamento de imagens digitais através de um computador[7]. Uma imagem é definida como uma função bidimensional, $a(x, y)$, sendo x e y coordenadas espaciais, e a amplitude a da imagem em qualquer par de coordenadas (x, y) sendo a intensidade ou nível de cinza naquele ponto na imagem, quando todos os valores de x , y , e de intensidade de a são finitos e discretos. A imagem pode conter subimagens, também chamadas de regiões de interesse (ROIs) ou apenas regiões, considerando o fato de que imagens frequentemente contêm coleções de objetos, cada um dos quais pode ser a base para uma região. Em alguns sistemas de processamento de imagens, temos a possibilidade de aplicamos operações específicas em uma região selecionada. Assim, por exemplo, uma parte da imagem pode ser processada para

suprimir desfoque de movimento, enquanto outra parte pode ser processada para melhorar a reprodução das cores.

Uma imagem digital $a[m, n]$ descrita em um espaço discreto 2D é derivada de uma imagem analógica $a(x, y)$ em um espaço 2D contínuo através de um processo de digitalização. A imagem contínua 2D $a(x, y)$ é dividida em n linhas e m colunas. A interseção de uma linha e uma coluna é chamada de *pixel*[24], mas também podemos chamar de elemento da imagem ou *pels*[7]. O valor atribuído as coordenadas inteiras $[m, n]$ com $m = 0, 1, 2, \dots, M-1$ e $n = 0, 1, 2, \dots, N-1$ são $a[m, n]$. Vale ressaltar que $a(x, y)$ é na verdade uma função que teria muitas variáveis incluindo parametros como profundidade (z), cor (l) e tempo (t), mas consideraremos um caso de uma imagem 2D, estática e monocromática, para fins de exemplo.[24].

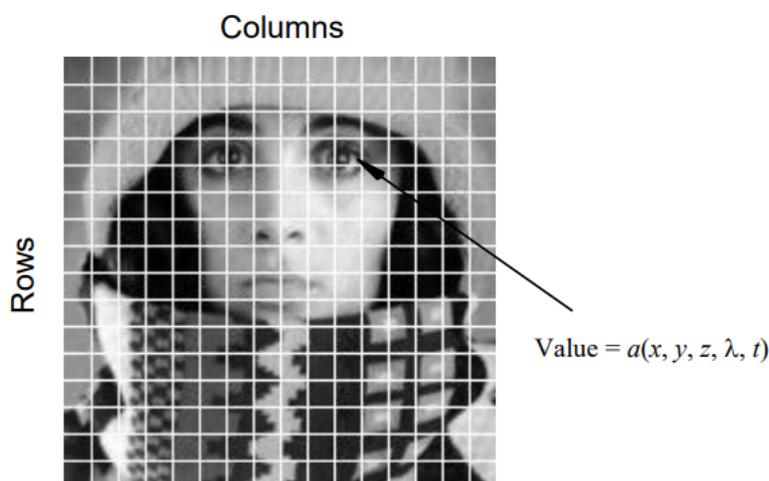


Figura 3. Imagem contínua digitalizada. O pixel apontado é referente as coordenadas $[m=10, n=3]$ tem 110 como valor de brilho[24].

A imagem da figura 3 é um exemplo de uma imagem subdividida em $n = 16$ linhas e $m = 16$ colunas. O valor atribuído a cada pixel corresponde ao brilho médio do pixel, arredondado para o inteiro mais próximo. O processo de representar a amplitude do sinal 2D em uma determinada coordenada como um valor inteiro com L diferentes níveis de cinza é geralmente conhecido como quantização de amplitude ou simplesmente quantização[24].

Entre nossos 5 sentidos, podemos destacar a visão como a mais avançada, logo não é de se surpreender que as imagens desempenhem o papel mais importante na percepção humana. Porém, diferente do ser humano, que estão limitados à faixa visual do espectro eletromagnético (EM), as máquinas podem cobrir quase todo o espectrograma EM, variando de raios gama a onda de rádio. Elas tem a possibilidade de operar em imagens geradas por fontes que os humanos não estão acostumados a associar com imagens. Isso inclui ultrassom, microscopia eletrônica e imagens geradas por computador. Assim, o processamento digital de imagens abrange um campo amplo e vasto de aplicações[7].

Historicamente, uma das primeiras aplicações de processamento de imagem remonta ao início deste século, quando se buscavam maneiras de melhorar a qualidade de impressão de imagens digitalizadas transmitidas pelo sistema Bartlane de transmissão de

imagens por cabo submarino entre as cidades de Londres e Nova York. Os primeiros sistemas Bartlane, no começo da década de 1920, codificavam uma imagem em cinco níveis de intensidade distintos.

O grande impulso da área veio cerca de três décadas depois, com o advento dos primeiros computadores digitais de grande porte e o início do programa espacial dos Estados Unidos. O uso de técnicas computacionais de aprimoramento de imagens começou no Jet Propulsion Laboratory na cidade de Pasadena localizada no estado americano da Califórnia em 1964, quando imagens da lua transmitidas por uma sonda Ranger foram processadas por computador para corrigir vários tipos de distorção inerentes à câmera de TV acoplada à sonda. Essas técnicas serviram de base para métodos aprimorados de realce e restauração de imagens em outros programas espaciais subsequentes, como por exemplo, as expedições tripuladas da série Apollo.

De 1964 até os dias de atuais, a área de processamento de imagens tem mostrado um crescimento significativo, com aplicações que permeiam quase todos os campos da atividade humana. Como na medicina, com o uso de imagens no diagnóstico médico tornou-se rotineiro, e os avanços em processamento de imagens têm permitido tanto o desenvolvimento de novos equipamentos quanto a facilidade de interpretação de imagens produzidas por equipamentos mais antigos, como os de raio X. Em Biologia, a capacidade de processar automaticamente imagens obtidas de microscópios, por exemplo, contando o número de células de um certo tipo presentes em uma imagem, facilita enormemente a execução de tarefas laboratoriais com alta precisão e repetibilidade. O processamento e a interpretação automática de imagens captadas por satélites auxiliam trabalhos em Geografia, Sensoriamento Remoto, Geoprocessamento e Meteorologia, entre outras áreas. Técnicas de restauração de imagens ajudam arqueólogos a recuperar fotos borradas de artefatos raros, já destruídos[14].

2.2. Aprendizado de máquina

O aprendizado de máquina é um campo da inteligência artificial que visa desenvolver técnicas computacionais para aprendizado automático e construção de sistemas capazes de adquirir conhecimento de forma autônoma. Um sistema de aprendizado é um programa de computador que toma decisões com base em experiências acumuladas, derivadas da solução bem-sucedida de problemas anteriores. Diversos sistemas de aprendizado de máquina possuem características específicas e compartilhadas que permitem sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizada[15].

2.2.1. Redes neurais

As redes neurais são uma tentativa de simular os sistemas nervosos biológicos, consistindo em uma rede de elementos interconectados, semelhantes aos neurônios, que produzem um padrão de saída em resposta a um padrão de entrada, inspirando-se no funcionamento do cérebro humano[16].

Uma rede neural artificial é um sistema distribuído e paralelo que utiliza um mecanismo distinto da inteligência artificial tradicional e das tecnologias convencionais de processamento de informações. Este mecanismo permite que ela supere as limitações da

IA baseada em lógica ao lidar com intuição e dados não estruturados. Além disso, as redes neurais artificiais possuem vantagens como adaptação, organização e capacidade de aprendizagem em tempo real.

Nos últimos 10 anos, houve grandes progressos em relação a pesquisa sobre redes neurais artificiais. Conquistas significativas foram alcançadas em áreas como reconhecimento de padrões, robótica inteligente, controle automático, previsão e estimativa, biologia, medicina, economia, entre outras. Essas redes têm conseguido resolver problemas práticos que os computadores modernos não conseguem solucionar, demonstrando uma inteligência superior em diversas aplicações[23].

As redes neurais são compostas por unidades de processamento simples, chamadas de neurônios artificiais, organizadas em camadas. As camadas são conectadas por pesos ajustáveis que são aprendidos durante o treinamento da rede. O treinamento envolve o uso de um algoritmo de retropropagação, que ajusta os pesos com base no erro entre a saída prevista e a saída esperada. Esse processo permite que a rede aprenda a reconhecer padrões complexos e generalize a partir de exemplos de treinamento[8].

Além das redes neurais artificiais (ANNs), existem outros tipos especializados de redes neurais, como redes neurais convolucionais (CNNs), usadas principalmente para processamento de imagens, e redes neurais recorrentes (RNNs), adequadas para dados sequenciais. Cada tipo de rede neural é projetado para lidar com diferentes tipos de problemas e dados, ampliando as possibilidades de aplicação dessas tecnologias[1].

2.2.2. Aprendizado profundo

O aprendizado de máquina é um dos campos da computação moderna, o aprendizado profundo, também conhecido como *deep learning*, é um subconjunto desse campo[18], representando um avanço significativo no campo da inteligência artificial, permitindo que modelos computacionais com múltiplas camadas de processamento aprendam representações complexas e abstratas dos dados. Essa abordagem tem revolucionado várias áreas, incluindo o reconhecimento de fala e de objetos visuais, detecção de objetos, descoberta de medicamentos e genômica, alcançando resultados que antes eram considerados inatingíveis.

Uma das principais forças do *deep learning* é a sua capacidade de descobrir padrões intrincados em grandes volumes de dados. Utilizando o algoritmo de retropropagação, esses modelos ajustam seus parâmetros internos para melhorar a precisão a cada camada, refinando continuamente a representação dos dados.

Redes neurais convolucionais profundas (CNNs) são particularmente eficazes em tarefas de processamento de imagens, vídeos, fala e áudio, capturando detalhes sutis e nuances que métodos anteriores não conseguiam. Por outro lado, redes neurais recorrentes (RNNs) são especializadas em lidar com dados sequenciais, como texto e fala, permitindo a análise de contextos temporais e sequências de eventos.[11]

2.3. Redes Neurais Convolucionais (CNNs)

Uma rede neural convolucional, um dos métodos que o aprendizado profundo engloba, utiliza uma imagem de entrada e atribui pesos aprendíveis a vários objetos na imagem,

distinguindo um do outro. A CNN é usada para construir a maioria dos algoritmos de visão computacional, ela em comparação com outros métodos, demanda menos pré-processamento. Logo a CNN é o algoritmo mais eficaz para termos a compreensão do material de imagem[10]. Vale ressaltar também, que demonstrou classificação, reconhecimento, segmentação e recuperação de imagens excepcionais.[13]

Por conta da realização da CNN, ela acabou despertou o interesse de pessoas fora da área acadêmica. Empresas como a Microsoft, Google, ATT, NEC e Facebook estão entre as empresas envolvidas no desenvolvimento e avanço da arquitetura CNN [5]. Além disso, elas possuem grupos de pesquisa ativos que estão investigando novos designs de CNN. No momento, modelos profundos baseados em CNN estão sendo usados pela maioria dos pioneiros em competições de processamento de imagem e visão computacional. Por conta disso, existem várias variantes do design básico da CNN[12].

3. Objetivos

Os objetivos deste trabalho incluem o aprofundamento do estudo das redes neurais convolucionais (CNNs), entendendo melhor suas arquiteturas, princípios de funcionamento e aplicações no campo da visão computacional, a aplicação de um modelo de CNN em um problema específico de classificação de objetos, utilizando um conjunto de dados relevante e avaliar a eficácia do modelo desenvolvido, comparando seu desempenho em termos de precisão, *recall* e outras métricas com outras abordagens de reconhecimento de padrões em imagens. Além disso, pretende-se identificar as vantagens e limitações das CNNs em relação a métodos tradicionais, proporcionando uma análise crítica sobre a aplicabilidade dessas redes neurais em diferentes contextos. O trabalho também irá comparar os resultados obtidos com CNNs com outros modelos de aprendizado presentes na literatura, que serão levantados durante o processo de fundamentação teórica, em prol de fornecer uma perspectiva mais abrangente sobre o desempenho relativo das diferentes técnicas de reconhecimento de padrões em imagens.

4. Procedimentos metodológicos/Métodos e técnicas

Primeiramente, será feita a revisão bibliográfica visando aprofundar o conhecimento sobre a arquitetura de redes neurais convolucionais (CNNs) e compreender melhor os diversos modelos já existentes, buscando identificar as principais características, vantagens e limitações de cada abordagem.

Após esse levantamento, será escolhido um conjunto de dados adequado para a tarefa de reconhecimento de padrões em imagens. Este conjunto de dados passará por um processo de pré-processamento, que incluirá técnicas de normalização, padronização e aumento de dados, a fim de melhorar a qualidade e enriquecer a diversidade dos exemplos para treinamento.

Em seguida, será implementado um modelo de CNN utilizando uma arquitetura apropriada, baseando-se nas estudadas durante a revisão bibliográfica. Este modelo será treinado com nosso conjunto de dados pré-processado, com ajustes nos hiperparâmetros e utilização de técnicas de regularização para evitar *overfitting*.

Ao término do treinamento, o modelo será avaliado utilizando métricas de desempenho tais como precisão, *recall*, *F1-score* e *mean Average Precision* (mAP). Os resultados obtidos serão comparados com outras abordagens de reconhecimento de padrões em

imagens, permitindo a realização de uma análise crítica da eficácia e eficiência do modelo que foi desenvolvido.

Além disso, essa comparação incluirá a análise de outros modelos de aprendizado de máquina, destacando suas vantagens e limitações em relação ao modelo de CNN implementado. Com base nos resultados obtidos, serão feitas correções e ajustes no modelo, seguidos pela escrita e revisão do Trabalho de Conclusão de Curso, que incluirá todas as etapas realizadas, os resultados obtidos, as comparações feitas e as conclusões finais.

5. Cronograma de Execução

Atividades:

1. Revisão bibliográfica;
2. Seleção do conjunto de dados;
3. Pré-processamento de dados;
4. Treinamento da rede neural;
5. Construção do modelo;
6. Avaliação e comparação com outras abordagens;
7. Escrita do Trabalho de Conclusão de Curso;

Tabela 1. Cronograma de Execução

	jul	ago	set	out	nov	dec	jan	fev
Atividade 1	•	•	•					
Atividade 2	•	•						
Atividade 3		•	•					
Atividade 4		•	•	•				
Atividade 5			•	•	•			
Atividade 6				•	•	•		
Atividade 7				•	•	•	•	•

6. Contribuições e/ou Resultados esperados

Com a conclusão do trabalho, espera-se demonstrar a eficácia do *deep learning* no reconhecimento de padrões em imagens, utilizando técnicas de visão computacional. Através da implementação e avaliação de redes neurais convolucionais (CNNs) em um conjunto de dados específico, pretende-se evidenciar a precisão e a robustez dessas abordagens na identificação e classificação de diferentes padrões visuais. Além disso, a comparação com outras técnicas de reconhecimento de padrões permitirá destacar as vantagens e possíveis limitações do uso de *deep learning* nesta área. Os resultados esperados incluem a obtenção de altas taxas de acurácia e a validação da aplicabilidade dessas técnicas em cenários reais, contribuindo para o avanço do conhecimento e das práticas na área de visão computacional.

7. Espaço para assinaturas

Londrina, 29 de julho de 2024.

Referências

- [1] Charu C Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer, 2018.
- [2] Paulo Eduardo Ambrósio, Paulo M. de Azevedo-Marques, and Clóvis Simão Trad. Técnicas de inteligência artificial aplicadas ao desenvolvimento de sistema de apoio ao diagnóstico de lesões pulmonares. In *Proceedings of the Conference on Centro de Ciências das Imagens e Física Médica (CCIFM)*, Faculdade de Medicina de Ribeirão Preto (FMRP), Universidade de São Paulo (USP), Brasil, ano. Centro Universitário Barão de Mauá (CeUBM), Brasil.
- [3] Dana H. Ballard. *Computer Vision*. Prentice-Hall, 1982.
- [4] Danilo de Milano and Luciano Barrozo Honorato. Visão computacional. *Faculdade de Tecnologia, Universidade Estadual de Campinas*, 2010.
- [5] Li Deng, Dong Yu, et al. Deep learning: methods and applications. *Foundations and trends® in signal processing*, 7(3–4):197–387, 2014.
- [6] M. Goesele, N. Snavely, R. Szeliski, and E. H. Adelson. Multi-view stereo for community photo collections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2007.
- [7] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Pearson, Upper Saddle River, NJ, 4 edition, 2018.
- [8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [9] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017.
- [10] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53:5455–5516, 2020.
- [11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [12] Yann LeCun, Koray Kavukcuoglu, and Clément Faret. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems*, pages 253–256. IEEE, 2010.
- [13] Xiaolong Liu, Zhidong Deng, and Yuhan Yang. Recent progress in semantic image segmentation. *Artificial Intelligence Review*, 52:1089–1106, 2019.
- [14] Ogê Marques Filho and Hugo Vieira Neto. *Processamento digital de imagens*. Brasport, 1999.
- [15] Maria Carolina Monard and José Augusto Baranauskas. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32, 2003.
- [16] Phil Picton and Phil Picton. *What is a neural network?* Springer, 1994.

- [17] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition, 2021.
- [18] Pramila P. Shinde and Seema Shah. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6, 2018.
- [19] J. Sivic, A. Zitnick, and R. Szeliski. Efficient visual search of video archives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Springer, 2006.
- [20] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *Proceedings of the ACM SIGGRAPH Conference*. ACM, 2006.
- [21] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2nd edition, 2020.
- [22] A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.
- [23] Yu-chen Wu and Jun-wen Feng. Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656, 2018.
- [24] Ian T Young, Jan J Gerbrands, and Lucas J Van Vliet. *Fundamentals of image processing*, volume 841. Delft University of Technology Delft, 1998.