

# INTELIGÊNCIA ARTIFICIAL PARA IDENTIFICAÇÃO DE LIVROS PELA CAPA E ANOTAÇÃO VIRTUAL EM LIVROS FÍSICOS: UMA ABORDAGEM PARA AUXILIAR OS ESTUDOS ACADÊMICOS

Vitor Jacinto Frassetto<sup>1</sup>, Marlon de Matos de Oliveira<sup>2</sup>

**Resumo:** Um dos desafios atuais no ambiente acadêmico é a integração das vantagens do livro físico com as possibilidades oferecidas pelas tecnologias digitais. Nesse sentido, esta pesquisa propõe o uso de inteligência artificial para o reconhecimento automático de livros a partir de imagens de suas capas, integrando funcionalidades como anotações virtuais e a catalogação das informações dos livros identificados. Para isso, foi proposto um protótipo buscando superar as limitações do uso exclusivo de livros físicos, como a impossibilidade de realizar anotações diretamente nas páginas, as dificuldades na organização, catalogação e recuperação de informações bibliográficas, além da limitação no acesso prático e centralizado às anotações realizadas. A metodologia adotada envolve a aplicação de técnicas de visão computacional e aprendizado de máquina, com foco na extração de características visuais por meio de redes neurais convolucionais, para identificar com precisão o livro consultado. Além disso, foi implementado um sistema interativo que permite ao usuário criar anotações virtuais sobre imagens, promovendo uma interação mais dinâmica e personalizada com o material acadêmico. O desenvolvimento deste protótipo demonstrou resultados promissores, evidenciando sua eficácia no reconhecimento automático de livros a partir de imagens de capa, alcançando níveis de acurácia superiores a 99%, mesmo quando submetido a imagens capturadas.

**Palavras-chave:** reconhecimento de livros; visão computacional; aprendizado de máquina; anotações virtuais; inteligência artificial

---

<sup>1</sup>Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), vitorfrassetto.vf@unesc.net

<sup>2</sup>Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), marlon.oliveira@unesc.net

**ABSTRACT:** One of the current challenges in the academic environment is integrating the advantages of physical books with the possibilities offered by digital technologies. In this context, this research proposes the use of artificial intelligence for the automatic recognition of books based on images of their covers, integrating functionalities such as virtual annotations and the cataloging of the identified book information. To this end, a prototype was developed to overcome the limitations of using only physical books, such as the inability to make notes directly on the pages, difficulties in organizing, cataloging, and retrieving bibliographic information, and the lack of practical and centralized access to annotations. The adopted methodology involves the application of computer vision and machine learning techniques, focusing on the extraction of visual features through convolutional neural networks to accurately identify the consulted book. Additionally, an interactive system was implemented to allow users to create virtual annotations on images, promoting a more dynamic and personalized interaction with academic material. The development of this prototype demonstrated promising results, highlighting its effectiveness in the automatic recognition of books from cover images, achieving accuracy levels above 99%, even when tested with captured images

**Keywords:** book recognition; computer vision; machine learning; virtual annotations; artificial intelligence.

## 1 INTRODUÇÃO

Os livros, tanto didáticos quanto literários, são, sem dúvida, ferramentas valiosas para promover o desenvolvimento educacional e o hábito da leitura. Geralmente, representam o primeiro material de leitura ao qual os alunos têm acesso adequado, contribuindo para o cultivo do hábito de ler e facilitando o processo de aprendizagem (Godoi; Silva, 2016). Além de seu papel fundamental no desenvolvimento das habilidades de escrita e leitura, o livro físico permite uma interação única com o conteúdo, ao fazer anotações e destacar partes importantes nas páginas e textos, o que favorece um entendimento mais profundo e a memorização (nVersos Editora, 2022).

A possibilidade de fazer esse tipo de anotação nos livros físicos pode ajudar a superar uma barreira significativa para os estudantes e favorecer sua compreensão e memorização do conteúdo e a interação com o texto, já que a disponibilidade limitada de exemplares em uma biblioteca

pode levar a situações em que os alunos não são permitidos a fazer anotações na própria página do livro. Além disso, enquanto as informações digitais podem ser facilmente buscadas e referenciadas online, as referências bibliográficas contidas apenas em livros físicos podem não estar facilmente acessíveis, o que representa uma dificuldade significativa para os estudantes no momento de consultá-las ou referenciá-las após a leitura do livro. Diante desse cenário, a adoção de recursos digitais que facilitem o acesso a essas informações e permitam uma interação mais ativa com o material torna-se cada vez mais necessária.

Nesse contexto, a integração de tecnologias digitais surge como uma alternativa promissora para complementar a aprendizagem tradicional, sem a necessidade de substituí-la completamente (Silva, 2013). Um exemplo disso é o *blended learning*, que consiste na prática de combinar recursos digitais e métodos tradicionais para enriquecer a experiência educacional (Tayebinik; Puteh, 2013). Uma das principais formas pelas quais as tecnologias digitais melhoram o progresso dos estudantes é por meio da personalização do aprendizado, mediante o uso de ferramentas educacionais e plataformas de aprendizagem. Essas ferramentas permitem que o conteúdo seja ajustado às necessidades, habilidades e ritmos individuais, favorecendo uma maior compreensão, retenção do conhecimento e melhoria na qualidade do ensino (Dutra et al., 2024).

Nesse sentido, o reconhecimento de padrões surge como um elemento central, sendo definido por Dores (2018) como a busca por estruturas em diferentes tipos de dados, tanto quantitativos quanto qualitativos. Esses dados podem incluir informações numéricas, textuais ou visuais, como registros médicos, fotos aéreas e características demográficas. Os seres humanos, por sua vez, possuem habilidades naturais para essa tarefa, associando formas e sons de maneira intuitiva (Krug et al., 2008). A informatização dessa capacidade, por meio da visão computacional, visa automatizar atividades repetitivas e suscetíveis a falhas.

A identificação automática de objetos é uma aplicação prática do reconhecimento de padrões, embora enfrente desafios, especialmente em cenários complexos que envolvem variações de iluminação, posicionamento, ângulo, escala, textura, sombras, deformações, oclusões e outras características. Para lidar com esses obstáculos, integra-se a visão computacional, que busca replicar a capacidade humana de reconhecer padrões visuais e interpretar contextos (Oliveira; Melo, 2023).

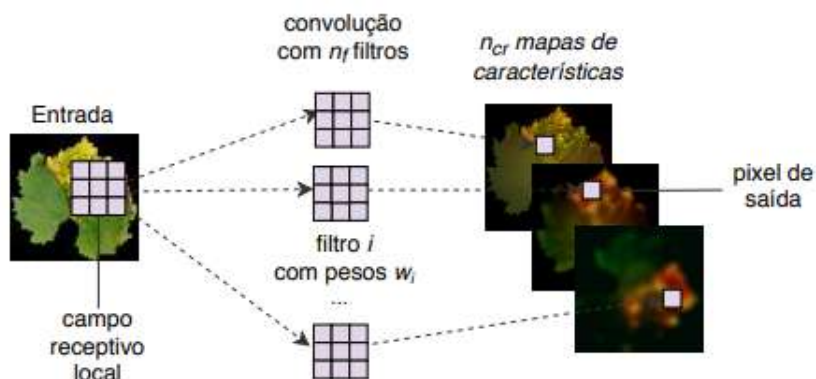
Dentre os métodos de identificação automática de livros, o uso

da imagem da capa destaca-se por sua acessibilidade, estabilidade visual entre exemplares da mesma edição e, principalmente, por não demandar custos adicionais com dispositivos externos ou à dependência de terceiros. Diferentemente dos códigos QR, que exigem manutenção contínua para garantir que os links vinculados permaneçam ativos e atualizado (Paul; Nairkar, 2024), ou das etiquetas RFID, cujo custo pode se tornar elevado em implementações de larga escala — conforme apontado por Shi, Tang e Lu (2021) —, a padronização das capas publicadas comercialmente permite seu uso como identificador visual confiável em sistemas de reconhecimento por visão computacional. Nesse processo, o uso de inteligência artificial e o processamento de imagens são fundamentais (Graciano, 2007).

Complementando esse cenário, o aprendizado de máquina se destaca como técnica capaz de gerar conhecimento a partir de dados. Ele constrói hipóteses com base em exemplos anteriores (Ludermir, 2021). Um de seus principais métodos, o aprendizado indutivo, divide-se em duas categorias: supervisionado e não supervisionado. (Monard; Baranauskas, 2003).

Entre as arquiteturas mais eficazes nesse campo estão as Redes Neurais Convolucionais (CNNs), amplamente utilizadas na visão computacional devido à sua alta performance na identificação de padrões visuais (Utsch, 2018; Silva, 2004). O funcionamento dessas redes começa com a aplicação de camadas convolucionais que extraem características relevantes das imagens (Rezende, 2019), conforme ilustrado na Figura 1.

Figura 1 - Operação de Convolução.

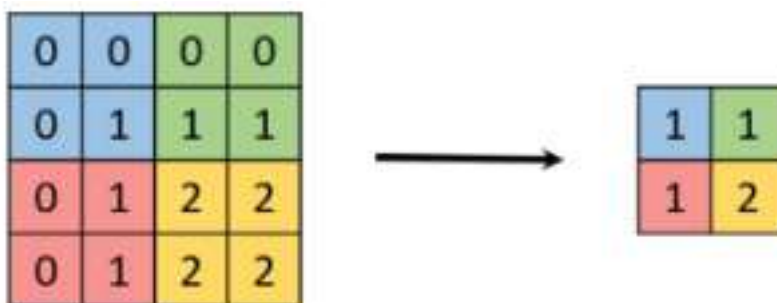


Fonte: Rezende (2019).

Seguido por camadas de pooling que reduzem a dimensionalidade e o custo computacional (Imamura et al., 2021), conforme Figura 2

abaixo.

Figura 2 - Max-pooling.



Fonte: Rezende (2019).

A etapa final envolve camadas totalmente conectadas que realizam assim tarefas de raciocínio complexo, como, por exemplo, a classificação de objetos em uma imagem (Utsch, 2018).

Com isso, a aplicação de inteligência artificial na identificação de livros físicos junto com anotações virtuais para aprimoramento do estudo acadêmico representa uma abordagem promissora para melhorar a gestão de conhecimento e a experiência de aprendizado dos estudantes, já que algumas das vantagens apresentadas do livro digital em relação aos livros físicos são a capacidade de destaque de texto e anotações do leitor (Mesquita; Conde, 2008), além da maior facilidade na catalogação das referências bibliográficas.

Dentro dessa proposta, a combinação de tecnologias de reconhecimento de imagem e anotação virtual oferece uma solução abrangente para facilitar o acesso, a organização e a interação com o material acadêmico. A relevância dessa abordagem é respaldada por vários estudos no campo da educação e tecnologia, como, por exemplo, o trabalho de Yufei et al. (2020) que destaca a crescente demanda por soluções tecnológicas que possam simplificar e personalizar a experiência de estudo, especialmente em um contexto acadêmico.

Portanto, esta pesquisa tem como objetivo geral desenvolver um protótipo baseado em inteligência artificial para reconhecer livros pela capa e integrar um sistema de anotações virtuais em livros físicos.

Os objetivos específicos desta pesquisa consistem em analisar os benefícios da integração de tecnologias digitais ao processo de estudo; desenvolver um protótipo de IA capaz de reconhecer o livro por meio de uma imagem da capa; implementar um sistema de catalogação dos livros identificados, que permita criar anotações, como fazer destaques sobre as

imagens enviadas e inserir comentários de forma intuitiva; avaliar o protótipo, mensurando a acurácia do reconhecimento.

Para estruturar este trabalho, a pesquisa foi dividida em cinco seções. Na primeira seção, apresenta-se a introdução, a segunda seção reúne os trabalhos correlatos, a terceira seção descreve os materiais e métodos utilizados no desenvolvimento do protótipo, a quarta seção expõe a discussão e os resultados obtidos, o quinto capítulo apresenta as conclusões e sugestões para trabalhos futuros.

## **2 TRABALHOS CORRELATOS**

Verne et al. (2022) propuseram uma abordagem baseada em redes neurais convolucionais (CNNs) para detecção e reconhecimento de plantas de pequeno porte em ambientes controlados. Utilizando uma base de dados composta por 1.050 imagens de 10 espécies diferentes, capturadas com fundo uniforme e sob variadas condições de iluminação, o sistema foi treinado após pré-processamento e expansão dos dados, alcançando 92,19% de acurácia na validação. O tempo médio de processamento por imagem foi de 375 milissegundos. O estudo demonstrou a viabilidade da técnica e sugeriu melhorias futuras, como o uso de arquiteturas mais avançadas, aumento do conjunto de dados e refinamento de parâmetros para elevar a precisão e reduzir o custo computacional.

Yang e Shen (2024) propuseram um sistema de reconhecimento automático de capas de livros utilizando técnicas de visão computacional e OCR, com uma interface desenvolvida em MATLAB. O sistema identifica regiões de texto nas capas por meio do algoritmo MSER e aplica filtragem morfológica para remover ruídos, permitindo a extração precisa do conteúdo textual. Foram avaliados dois mecanismos de OCR — o embutido no MATLAB e o Tesseract — sendo este último o mais eficaz, com 78,2% de acurácia na extração de texto e 86% de correspondência em buscas online. O estudo demonstrou a eficiência da solução e indicou seu potencial para ser expandido a dispositivos móveis e a outros tipos de mídia, como capas de CDs e jornais.

Hu, Yan e Xie (2024) desenvolveram um sistema automatizado para reconhecimento de livros em bibliotecas utilizando redes neurais convolucionais (CNNs), que combina técnicas de segmentação, correção de distorções e extração de características visuais para identificar páginas de livros mesmo sob condições adversas, como variações de perspectiva e iluminação. Utilizando uma arquitetura baseada no modelo VGG-F, o sistema

processa imagens de forma eficiente e precisa, alcançando taxas de acerto superiores a 99% na recuperação das cinco páginas mais semelhantes e operando com média de 103,42 ms por imagem, o que o torna adequado para bibliotecas de pequeno e médio porte.

### 3 MATERIAIS E MÉTODOS

Esta seção apresenta os recursos utilizados e os procedimentos adotados para o desenvolvimento do protótipo de reconhecimento de capas de livros com anotações virtuais. A solução foi construída com base em técnicas de visão computacional e aprendizado de máquina, englobando etapas como pré-processamento de imagens, extração de características com redes neurais e uma interface interativa para visualização e anotação.

O primeiro passo consiste no treinamento de uma base de dados disponibilizada por Finsterwald (2024), na plataforma *Kaggle*, composta por 5.700 imagens, acompanhadas de seus respectivos metadados. Para cada imagem, são aplicadas etapas de pré-processamento e técnicas de *data augmentation*, com o objetivo de ampliar a diversidade e a quantidade de amostras disponíveis para o treinamento. Em seguida, realiza-se o processo de treinamento utilizando algoritmos de inteligência artificial sobre essa base de dados expandida, especificamente com a aplicação da arquitetura VGG16 para extração de características visuais e de um classificador *Multilayer Perceptron* (MLP) para a etapa de classificação, visando aumentar a precisão e a robustez do modelo no momento do reconhecimento.

O fluxo do sistema inicia-se com o envio da imagem pelo usuário, a qual passa por um processo de remoção do fundo e ajuste de contraste, com o intuito de melhorar a visibilidade e a qualidade da imagem, frequentemente comprometidas devido à captura em ambientes não controlados. Em seguida, a imagem é processada pela rede neural VGG16, responsável pela extração de características relevantes que serão comparadas com aquelas previamente armazenadas no modelo treinado. O classificador treinado realiza a identificação da imagem enviada pelo usuário com base nos vetores de características extraídos. Por fim, o sistema retorna como resultado as três classes ou imagens mais prováveis, possibilitando ao usuário identificar rapidamente o livro correspondente e acessar as funcionalidades de catalogação e anotações virtuais associadas.

Após o reconhecimento, o usuário pode verificar os livros já reconhecidos e suas informações, assim como realizar anotações diretamente sobre imagens enviadas pelo usuário, integrando o físico com o virtual e

auxiliando na aprendizagem e armazenagem das informações. A Figura 3 ilustra este processo por meio de um fluxograma.

Figura 3 - Fluxograma.



Fonte: Elaborado pelo autor.

### 3.1 BASE DE DADOS

A base de dados empregada foi obtida na plataforma *Kaggle*, no conjunto denominado *Book Cover Dataset*, disponibilizado por Finsterwald (2024), e foi escolhida uma base composta por 5700 imagens de capas de livros acompanhadas de seus metadados, como título, autor, Amazon Index (ASIN), nome do arquivo, URL da imagem, ID categoria, categoria, armazenados em um arquivo CSV, como mostra o exemplo da Tabela 1.

Tabela 1 - Exemplo de dados da base utilizada no protótipo.

| ASIN       | Título                              | Autor                         | Categoria                    |
|------------|-------------------------------------|-------------------------------|------------------------------|
| 044310073X | Oral and Maxillofacial Surgery...   | –                             | Medical Books                |
| 1438005687 | Barron's GRE, 21st Edition          | Sharon W. Green M.A.          | Test Preparation             |
| 0060750715 | George Balanchine: The Ballet Maker | Robert Gottlieb               | Biographies & Memoirs        |
| 1580237959 | A Partner in Holiness, Vol. 2       | Rabbi Jonathan P. Slater DMin | Religion & Spirituality      |
| 0135137829 | Construction Scheduling, 2nd Ed.    | Jay S. Newitt                 | Arts & Photography           |
| 0312556411 | Literature and Its Writers          | Ann Charters                  | Literature & Fiction         |
| 0393339157 | Straight on Till Morning            | Mary S. Lovell                | Engineering & Transportation |
| 0521456924 | Diagrammatica                       | Martinus Veltman              | Science & Math               |
| 0898699223 | Book of Common Prayer 1979          | –                             | Christian Books & Bibles     |

Fonte: Elaborado pelo autor.

Para viabilizar a integração com o protótipo, os arquivos de imagem foram baixados e organizados em um único diretório, mantendo a correspondência exata entre o nome de cada arquivo e o identificador respectivo no CSV (Filename). Como havia apenas uma imagem por livro na base de dados, foram aplicadas técnicas de *data augmentation* com o objetivo de ampliar o número de amostras e melhorar a capacidade de generalização do modelo. Para cada imagem original e sua versão equalizada (com ajuste de contraste), foram geradas quatro variações por meio de transformações que simulam condições reais de captura, como mudanças na iluminação, no ângulo de visão e na qualidade da imagem.

Para viabilizar a integração com o protótipo, os arquivos de imagem foram baixados e organizados em um único diretório, mantendo a correspondência exata entre o nome de cada arquivo e o identificador respectivo no CSV (*Filename*). Como tinha apenas uma imagem de cada livro na base de dados, para ampliar esse número e também melhorar a precisão na hora do reconhecimento, foram aplicadas técnicas de *data augmentation* com o objetivo de ampliar o número de amostras e melhorar a capacidade de generalização do modelo. Para cada imagem original e sua versão equalizada (com ajuste de contraste), foram geradas quatro variações por meio de transformações que simulam condições reais de captura, como alterações na iluminação, posicionamento e qualidade da imagem.

Inicialmente, foram aplicadas transformações geométricas às imagens, como translações e escalas aleatórias de até 5%, além de rotações de até 10 graus, simulando variações no posicionamento e ângulo de captura das capas dos livros. Também foram realizadas distorções de perspectiva entre 2% e 5% e ajustes aleatórios nos parâmetros de brilho, contraste, saturação e matiz, com o objetivo de reproduzir diferentes condições de iluminação e configurações de dispositivos. Para simular a degradação da qualidade das imagens, aplicou-se desfoque gaussiano. Além disso, foram inseridas até três regiões aleatórias apagadas (*Coarse Dropout*), com tamanhos entre  $10 \times 10$  e  $20 \times 20$  pixels, representando oclusões típicas, como sombras, dedos e objetos. Por fim, todas as imagens foram redimensionadas para  $224 \times 224$  pixels, garantindo uniformidade e compatibilidade com os modelos de redes neurais convolucionais.

### 3.2 DESENVOLVIMENTO DO PROTÓTIPO

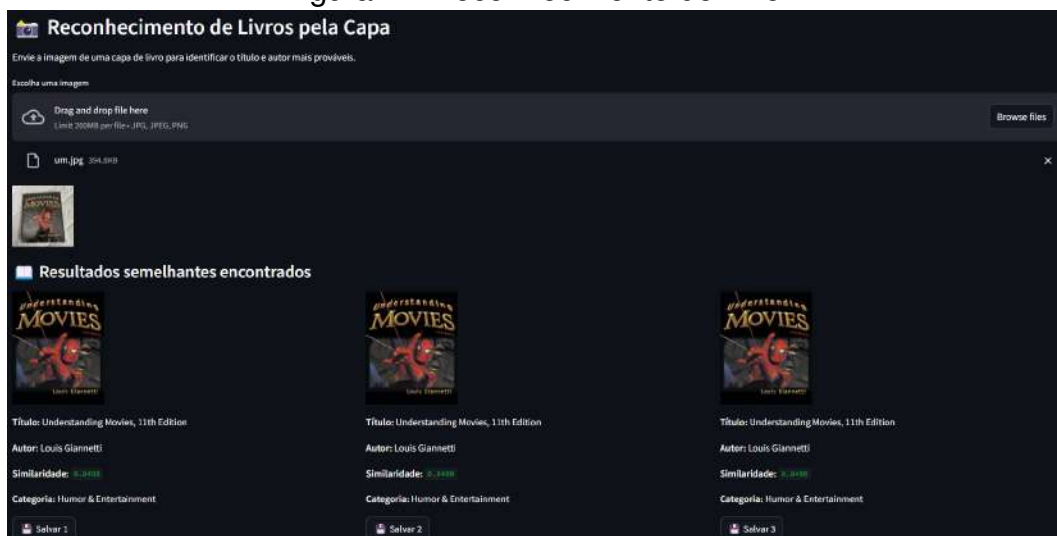
O desenvolvimento do protótipo foi realizado em Python 3.12.10, integrando bibliotecas especializadas em visão computacional, *deep learning*, manipulação de dados e processamento de imagens. Para as tarefas de visão computacional e pré-processamento de imagens, foram utilizadas as bibliotecas *OpenCV* e *Pillow* (PIL). O *TensorFlow* e o *Keras* foram empregados para a construção e o treinamento das redes neurais convolucionais e do classificador *Multilayer Perceptron* (MLP). A manipulação e análise de dados foram realizadas com o apoio do *NumPy* e do *Pandas*. Esta combinação permitiu a implementação eficiente de cada etapa do protótipo, desde o pré-processamento das imagens até a interação com o usuário final por meio de uma interface.

A interface visual foi desenvolvida com a biblioteca *Streamlit*, que

é uma ferramenta de código aberto voltada à construção de aplicações web interativas. Com o uso dessa tecnologia, foi possível criar uma interface simples e funcional, permitindo que os usuários carregassem imagens de capas de livros, visualizassem os resultados do reconhecimento e interagissem com as funcionalidades de anotação virtual. Para as anotações virtuais, foi integrado o componente *streamlit-drawable-canvas*, que no protótipo, foi usado para adicionar uma área interativa de desenho diretamente sobre uma imagem enviada pelo usuário. Essa funcionalidade permite que o usuário destaque áreas relevantes, simulando o comportamento de marcações em livros físicos.

A abordagem proposta é estruturada em três módulos principais que atuam de forma integrada. O primeiro módulo, dedicado ao reconhecimento de capas, é responsável pelo recebimento da imagem da capa do livro fornecida pelo usuário e, a partir desta, executar as etapas de pré-processamento, extração de características e inferência do resultado utilizando um modelo de visão computacional treinado para identificação dos livros, conforme ilustrado na Figura 4.

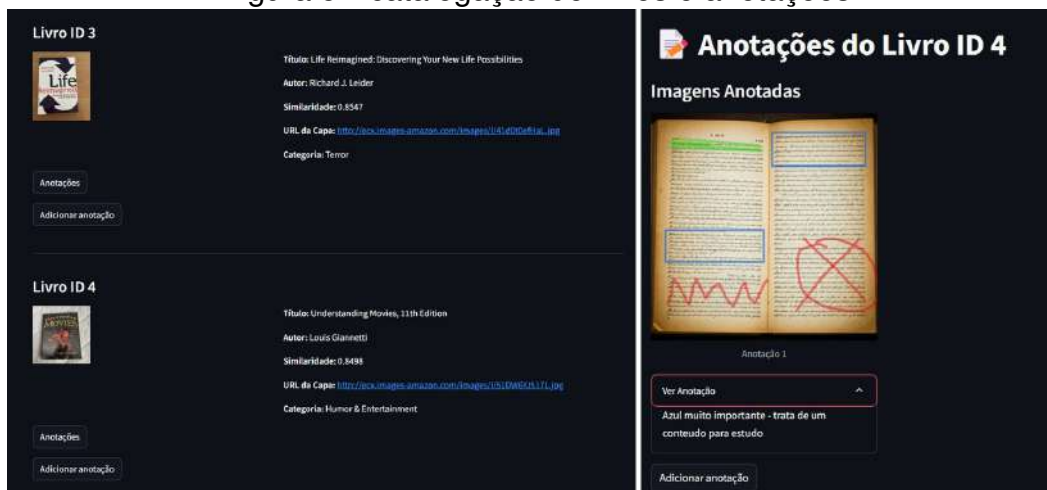
Figura 4 - Reconhecimento de livro.



Fonte: Elaborado pelo autor.

O segundo módulo, denominado catálogo, apresentado na Figura 5, é responsável pelo gerenciamento dos livros reconhecidos, organizando e armazenando seus metadados de forma estruturada. Além disso, esse módulo permite que o usuário localize o conteúdo e as anotações realizadas, além de permitir a criação de novas anotações associadas aos livros catalogados.

Figura 5 - catalogação de livros e anotações



Fonte: Elaborado pelo autor.

Por fim, o terceiro módulo, correspondente ao sistema de criação de anotações virtuais, consiste em uma interface gráfica interativa que permite ao usuário realizar anotações diretamente sobre a imagem da capa enviada, que pode corresponder a uma página do livro catalogado. Essa funcionalidade inclui a possibilidade de realizar desenhos livres, adicionar destaques visuais, além de um campo de texto para a inserção de informações complementares, conforme ilustrado na Figura 6.

Figura 6 - Anotações virtuais



Fonte: Elaborado pelo autor.

### 3.3 ANOTAÇÕES VIRTUAIS

As anotações virtuais representam uma funcionalidade complementar ao reconhecimento de capas, voltada ao enriquecimento do uso da ferramenta como auxílio nos estudos acadêmicos em uma biblioteca universitária e organização pessoal. Após a identificação de uma capa, o sis-

tema disponibiliza uma interface interativa por meio da biblioteca *streamlit-drawable-canvas*, permitindo que o usuário desenhe livremente sobre a imagem.

Estão disponíveis ferramentas como lápis, formas geométricas (círculos, retângulos, linhas e movimentação dessas formas geométricas), além de opções para escolha de cores, ajuste de opacidade e definição da espessura do traço. O componente também possibilita a exclusão de elementos desenhados, permitindo ao usuário modificar ou refazer partes específicas da anotação conforme necessário. Além disso, o sistema também conta com uma caixa de texto que permite ao usuário inserir resumos ou informações relacionadas à anotação feita sobre o livro específico.

Essas anotações são salvas e armazenadas como uma imagem .png em uma pasta do projeto e seu caminho é salvo em um arquivo csv com a referência também ao livro onde foi feita a anotação, assim podendo criar várias anotações para o mesmo livro. Essa abordagem visa simular, em ambiente digital, o hábito comum de destacar ou anotar pontos relevantes em imagens de livros físicos, adaptando-se a práticas modernas de estudo e consulta.

### 3.4 TREINAMENTO E RECONHECIMENTO

Os processos de treinamento e reconhecimento apresentam etapas semelhantes, embora possuam particularidades específicas.

Para o processamento de imagens, tanto na etapa de treinamento quanto na de inferência (reconhecimento), utilizou-se a biblioteca OpenCV para a leitura e equalização das cores. Essa etapa envolve ajustes na saturação, no contraste e na nitidez das imagens, proporcionando uma aproximação mais fiel às cores e características reais das imagens reais. Esses ajustes são fundamentais para uniformizar a qualidade visual das imagens e otimizar a extração de características relevantes pelos modelos de reconhecimento.

Adicionalmente, somente na etapa de reconhecimento, foi empregado o pacote *backgroundremover* para realizar a segmentação e a remoção automática do fundo das imagens. Complementarmente, aplicou-se uma técnica baseada na detecção de quadriláteros, com o objetivo de corrigir eventuais inclinações nas imagens de capa, bem como auxiliar no ajuste adequado da perspectiva e na melhoria da segmentação do fundo. Esses procedimentos são essenciais para garantir a consistência visual entre as imagens processadas e minimizar ruídos que possam interferir no desem-

penho do modelo.

Além disso, especificamente na etapa de treinamento, utilizou-se a ferramenta *Albumentations* para a definição de um procedimento composto por técnicas de *data augmentation*. Nesse procedimento, geraram-se variantes a partir das imagens originais sem a aplicação de remoção automática de fundo, uma vez que, para fins de treinamento, as capas encontravam-se em condições adequadas e o pacote *backgroundremover* frequentemente removia elementos internos importantes da capa, e também geraram-se variantes a partir das versões tratadas por meio da equalização de contraste. Essas transformações aumentaram a diversidade do conjunto de dados, melhorando a capacidade do modelo de generalizar e reconhecer imagens com variações visuais.

A extração de características foi implementada utilizando rede neural *VGG16* sem as camadas densas finais e com pooling médio, carregada com pesos pré-treinados no *ImageNet* por meio do *TensorFlow/Keras*. O uso desses pesos pré-treinados no *ImageNet* é fundamental, pois permite que a rede já possua filtros capazes de identificar padrões visuais genéricos, como bordas, texturas e formas, que são úteis para diversas tarefas de visão computacional. Cada imagem de entrada foi redimensionada para 224×224 pixels, convertida em array *NumPy* e pré-processada para ajustar as médias e desvios do *ImageNet*, garantindo normalização adequada dos valores de pixel. A saída da *VGG16* consistiu em vetores de 512 dimensões, obtidos a partir de uma camada de pooling que sintetiza informações espaciais.

Na fase de inferência, os vetores de 512 dimensões gerados pela *VGG16* foram apresentados a um classificador *Multilayer Perceptron* (MLP) implementado em *Keras*. A arquitetura do MLP foi composta por uma primeira camada totalmente conectada com 512 neurônios e ativação *ReLU*, responsável por modelar relações não lineares entre as características extraídas, seguida de *BatchNormalization*, para estabilizar e acelerar o treinamento, além de um *Dropout* com taxa de 40%, para reduzir o risco de overfitting.

Na sequência, adicionou-se uma segunda camada totalmente conectada com 256 neurônios e ativação *ReLU*, para refinar ainda mais a representação das características, também seguida por *BatchNormalization* e *Dropout* de 30%, reforçando a regularização do modelo.

Por fim, a camada de saída foi configurada com um número de unidades igual ao número de classes, empregando ativação *Softmax* para

produzir uma distribuição de probabilidades. O MLP foi treinado com o otimizador *Adam* (taxa de aprendizado de  $1 \times 10^{-4}$ ) e função de perda *sparse\_categorical\_crossentropy*, ajustando os pesos da rede neural, com base no erro calculado pela função de perda, por retropropagação. Durante a inferência, cada vetor de 512 dimensões fornecido pela *VGG16* é processado pelo MLP, que retorna as probabilidades de associação a cada classe. Esse fluxo garante um reconhecimento preciso e com baixo tempo de resposta.

### 3.5 TESTES E VALIDAÇÕES

Após o treinamento do modelo de classificação, foi realizada uma etapa de validação para acompanhar o desempenho ao longo das 100 épocas aplicadas inicialmente. Para isso, utilizou-se um parâmetro de divisão interna que separou automaticamente 20% dos dados do conjunto de treino exclusivamente para validação. A cada época, todo o conjunto de treino era dividido em 1424 lotes de 32 imagens para cálculo das métricas de treinamento, enquanto os 20% reservados para validação eram divididos em 284 lotes de 32 imagens para medir a acurácia e a função de perda em dados não vistos. Esse procedimento permitiu acompanhar a evolução das métricas *training accuracy*, *training loss*, *validation accuracy* e *validation loss*, ajudando também na prevenção de *overfitting* por meio do callback *EarlyStopping*, que interrompeu o treinamento assim que a performance na validação deixou de melhorar.

Com o término do treinamento e validação, foi realizada a avaliação final do modelo utilizando o conjunto de teste, composto por 20% dos dados originais que não foram utilizados em nenhuma etapa anterior. Essa avaliação retornou métricas importantes como a perda (loss) e a acurácia (accuracy) do modelo. Essa etapa é essencial para medir a capacidade do modelo de generalizar, ou seja, de manter um bom desempenho quando aplicado a novos dados, diferentes dos usados durante o processo de aprendizado.

Dessa forma, o processo de validação e teste garantiu uma avaliação completa e confiável do modelo treinado. A validação interna acompanhou o desempenho ao longo do treinamento, evitando problemas como o sobreajuste, enquanto a avaliação final com o conjunto de teste forneceu uma métrica objetiva e imparcial do desempenho geral do modelo. Com isso, foi possível assegurar que o modelo estivesse devidamente treinado e validado antes de ser aplicado a situações reais.

Adicionalmente, foram selecionadas 30 imagens reais de capas, capturadas em ambientes não controlados, que passaram pelo mesmo fluxo de pré-processamento (remoção de fundo e equalização de contraste) para o reconhecimento. Esse conjunto complementar permitiu avaliar a robustez prática do sistema diante de variações típicas de uso cotidiano.

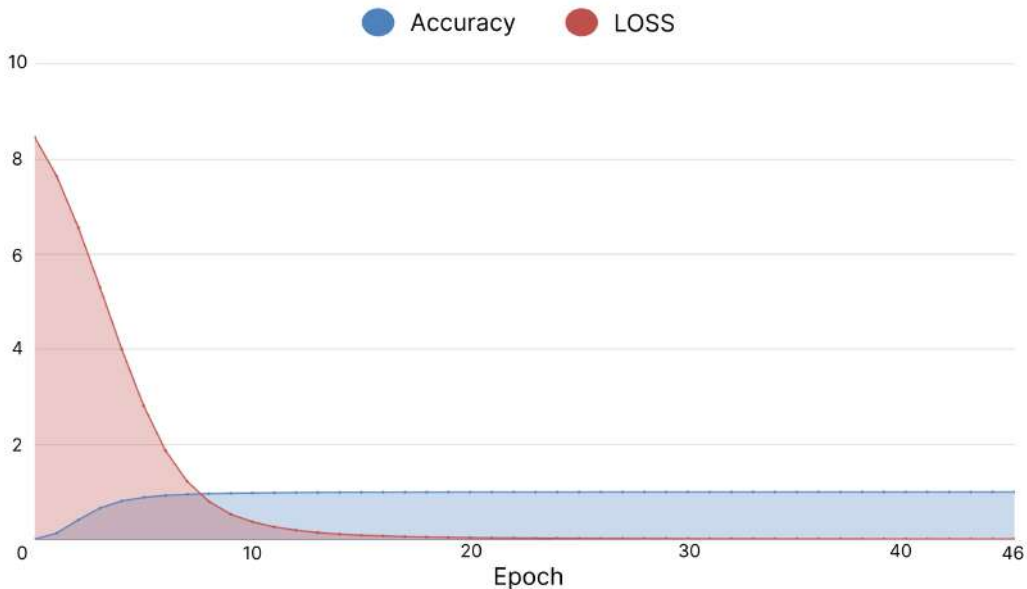
#### 4 DISCUSSÃO E RESULTADOS

A avaliação do desempenho foi realizada utilizando métricas padrão, como acurácia e função de perda (*loss*), tanto no conjunto de validação durante o treinamento quanto no conjunto de teste final, reservado exclusivamente para a avaliação final do modelo.

Durante o treinamento, a acurácia no conjunto de validação apresentou uma evolução rápida e consistente, demonstrando a eficácia do modelo na aprendizagem das características visuais das capas de livros. O processo teve início com uma acurácia extremamente baixa, de aproximadamente 0,17% na primeira época, mas rapidamente progrediu para 64% já na quarta época. A melhoria foi ainda mais expressiva, atingindo 99% de acurácia na décima quinta época. Embora o treinamento estivesse inicialmente configurado para 100 épocas, a utilização do mecanismo de *EarlyStopping* interrompeu o processo de forma antecipada na 46<sup>a</sup> época com acurácia de 99,9%, já que manteve estabilidade de 99% de acurácia desde a 15<sup>a</sup>. Esse resultado evidencia não apenas a eficiência do modelo, mas também a adequação dos parâmetros e estratégias de regularização adotadas, garantindo um equilíbrio entre desempenho e prevenção de sobreajuste, conforme ilustrado na Figura 7.

De maneira igualmente significativa, a função de perda acompanhou essa trajetória de aprimoramento com uma redução acentuada nos valores ao longo das épocas. Inicialmente, a função de perda situava-se em 8,5921 na primeira época, diminuindo para 5,5694 na quarta época, até atingir um patamar de estabilidade em torno de 0,0162 a partir da 30<sup>a</sup> época. Esse valor permaneceu praticamente constante até a 46<sup>a</sup> época, quando o treinamento foi interrompido, culminando em uma perda mínima de 0,0077.

Figura 7 - Acurácia e função de perda no conjunto de validação.



Fonte: Elaborado pelo autor.

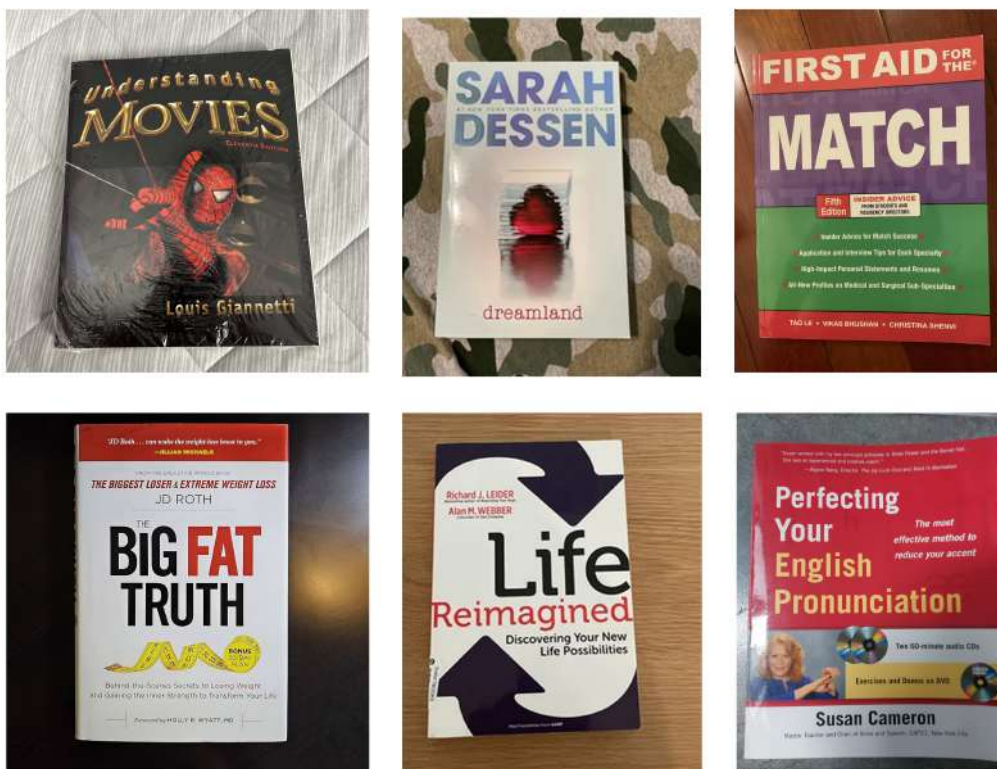
Para a avaliação final, realizada sobre o conjunto de teste, o modelo alcançou uma acurácia de aproximadamente 98,01%, com uma função de perda de 0,1004. Esses resultados demonstram uma excelente capacidade de generalização, uma vez que o modelo manteve um desempenho elevado mesmo diante de imagens que não haviam sido apresentadas anteriormente durante o treinamento. O fato de o modelo ter conseguido identificar corretamente padrões em dados completamente novos evidencia a efetividade do processo de aprendizado e a robustez da arquitetura utilizada.

Além dos resultados obtidos com a validação interna de desempenho, o modelo foi submetido a um teste manual composto por 30 imagens de capas de livros, exemplos na Figura 8, representando cenários variados de qualidade de imagem, presença de reflexos e diferentes padrões visuais. Dentre essas, o modelo acertou corretamente a classificação de 27 livros, resultando em uma taxa de acerto de 90%. Conseqüentemente, houve 3 erros (10% de falhas), todos associados a problemas específicos de pré-processamento das imagens, como falhas no recorte automático e dificuldades na equalização do contraste em capas brilhantes.

Em relação às probabilidades associadas aos acertos, a média das probabilidades preditas foi 88,33%, indicando que o modelo não apenas acertou a maior parte das predições, mas o fez com um alto nível de confiança. Dentre os acertos, 17 ocorreram com probabilidades superiores a 99%, evidenciando uma classificação altamente segura em grande parte

dos casos. Apenas três acertos ocorreram com probabilidades inferiores a 70%, sugerindo que, embora a predição tenha sido correta, nesses casos o modelo demonstrou incerteza considerável.

Figura 8 - Exemplo em situações reais testadas manualmente.



Fonte: Elaborado pelo autor.

Apesar dos resultados promissores, este estudo apresenta algumas limitações importantes. Primeiramente, a diversidade da base de dados pode ter impactado o desempenho do modelo, especialmente no que se refere ao aprendizado das classes. Como cada livro possuía apenas um exemplar de capa disponível, a ampliação da variabilidade do conjunto foi realizada exclusivamente por meio de técnicas de *data augmentation*, o que, embora tenha aumentado a quantidade de exemplos, pode não ter sido suficiente para representar toda a complexidade e diversidade visual que ocorre em contextos reais.

Além disso, outra limitação refere-se aos desafios enfrentados no pré-processamento das imagens. Reflexos de luz, causados pelas diferenças nos materiais utilizados em algumas capas, como acabamentos mais brilhantes ou texturizados, dificultaram a segmentação e o tratamento adequado da imagem, como mostrado na Figura 9.

Figura 9 - Exemplo de problemas no recorte da capa.

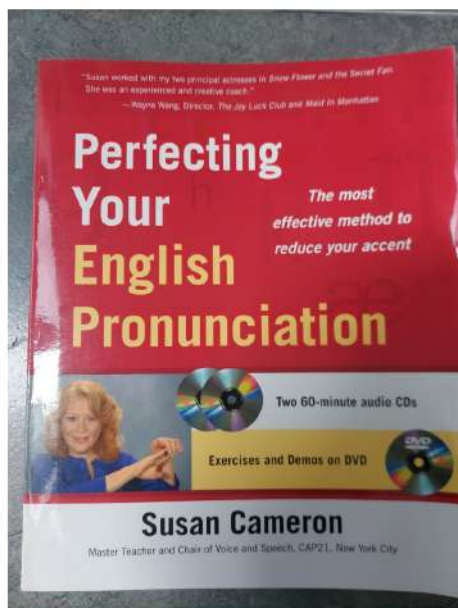


Imagem Original

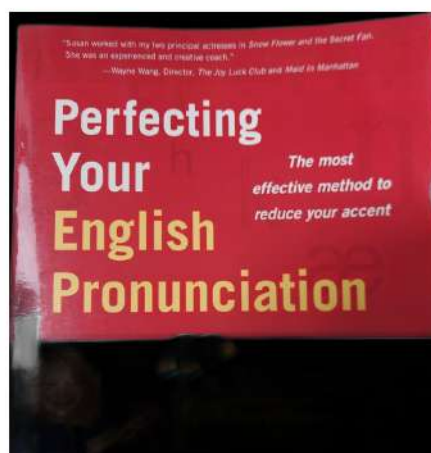


Imagem Recortada

Fonte: Elaborado pelo autor.

Outro fator que comprometeu o desempenho foi a presença de imagens enviadas com fundos de cor idêntica ou muito semelhante à da capa, o que prejudicou a eficácia do recorte automático, resultando em separações imprecisas e, conseqüentemente, afetando negativamente a qualidade da extração de características e a acurácia do reconhecimento.

Embora não tenham sido encontrados trabalhos com exatamente o mesmo propósito de reconhecer livros com base na identificação da capa em um banco de dados, há similaridades com o estudo de Hu, Yan e Xie (2024), que propuseram um sistema automatizado para recuperação de páginas de livros utilizando redes neurais convolucionais (CNNs). Ambos os trabalhos compartilham o uso de técnicas de visão computacional e arquiteturas baseadas na VGG, com foco na extração de características visuais para permitir a identificação precisa de materiais bibliográficos.

Em termos de desempenho, Hu, Yan e Xie relataram uma taxa de recuperação de páginas de 93,18% no teste inicial, atingindo até 99,31% quando consideravam as cinco principais configurações. Já o modelo proposto, que combina extração de características via VGG16, obteve 98.01% de acurácia no conjunto de teste para reconhecimento de capas de livros.

## 5 CONCLUSÃO

O presente trabalho teve como objetivo desenvolver um protótipo baseado em inteligência artificial para reconhecer livros pela capa e integrar um sistema de anotações virtuais em livros físicos. Para isso, foram utilizadas técnicas de visão computacional e redes neurais convolucionais para a extração de características visuais, combinadas com um classificador MLP. Esse objetivo foi plenamente atingido por meio da implementação de um pipeline que inclui etapas de pré-processamento das imagens, extração de vetores de características com a arquitetura VGG16 e posterior classificação realizada pelo modelo treinado, especificamente para distinguir entre as diferentes capas de livros, além da implementação de um módulo intuitivo para anotações digitais sobre as imagens.

Os principais resultados demonstraram que o modelo foi capaz de atingir uma acurácia de 98,01% no conjunto de teste, além de apresentar desempenho consistente em avaliações práticas, com 88,33% de similaridade média e 90% de acerto no reconhecimento de livros em um conjunto adicional de 30 imagens reais. Esses dados reforçam a viabilidade do sistema para aplicações em reconhecimento automatizado de livros, inclusive em condições não controladas.

No que diz respeito ao sistema de anotações virtuais, este foi desenvolvido para permitir que o usuário crie destaques diretamente sobre as imagens para cada livros reconhecidos, além de permitir escrever comentários, simulando interações semelhantes às feitas em livros físicos. Essa funcionalidade contribui para a personalização do aprendizado e facilita a organização do conteúdo estudado, especialmente em contextos em que a escrita direta nos exemplares físicos não é permitida, como em bibliotecas. Além disso, o recurso promove a sistematização do conhecimento e potencializa o uso de materiais digitais como ferramentas complementares no processo educacional, ampliando o acesso e a interatividade com o conteúdo.

Como sugestões para trabalhos futuros, destaca-se a possibilidade de expandir a base de dados com múltiplas imagens reais de cada capa, visando aprimorar a generalização do modelo e reduzir a dependência exclusiva de técnicas de *data augmentation*. Além disso, podem ser exploradas arquiteturas mais recentes de redes neurais, assim como a incorporação de módulos adicionais de segmentação e correção geométrica, a fim de aumentar ainda mais a precisão do reconhecimento em cenários

mais desafiadores. Por fim, a adaptação do protótipo para uso em dispositivos móveis, permitindo maior portabilidade e acessibilidade aos usuários em diferentes contextos de uso, como bibliotecas, salas de aula ou ambientes de estudo pessoais, promovendo ainda mais a usabilidade e aplicabilidade do sistema no ambiente acadêmico.

## REFERÊNCIAS

- DORES, A. F. das. **Reconhecimento de Padrões com Inteligência Artificial**. Trabalho de Conclusão de Curso de Graduação em Ciência da Computação — Faculdade Anhanguera de São José, São José dos Campos, 2018.
- DUTRA, I. T. L. et al. **O Impacto das Tecnologias Digitais no Processo de Ensino-Aprendizagem**. [s.n.], 2024. v. 28. Acesso em: 19 maio 2024. Disponível em: <<https://doi.org/10.5281/zenodo.10982842>>.
- FINSTERWALD, M. **Book Cover Dataset**. 2024. Acesso em: 18 mar. 2025. Disponível em: <<https://www.kaggle.com/datasets/mexwell/book-cover-dataset>>.
- GODOI, M. M. de; SILVA, J. B. de M. **A Importância do Livro Didático no Ensino de História**. Anápolis, GO, 2016. 180–185 p. Acesso em: 19 maio 2024. Disponível em: <<http://www.seminariodeestagio.ccseh.ueg.br/>>.
- GRACIANO, A. B. V. Dissertação de Mestrado. **Rastreamento de objetos baseado em reconhecimento estrutural de padrões**. São Paulo: [s.n.], 2007. Acesso em: 18 out. 2024. Disponível em: <<https://doi.org/10.11606/D.45.2007.tde-05112009-204609>>.
- HU, J.; YAN, Y.; XIE, Z. **Automatic Recognition Technology of Library Books Based on Convolutional Neural Network Model**. [S.l.]: HighTech and Innovation, 2024. v. 5. 200–212 p. Acesso em: 18 out. 2024.
- IMAMURA, M. E. et al. **Detecção e Reconhecimento de Placas de Licenciamento Veicular em Tempo Real usando CNN**. [s.n.], 2021. v. 13. 89–99 p. Acesso em: 18 out. 2024. ISSN 2178-8332. Disponível em: <<http://journal.unoeste.br/index.php/ce/index>>.
- KRUG, et al. **ANÁLISE E RECONHECIMENTO DE PADRÕES USANDO PROCESSAMENTO DE IMAGENS E INTELIGÊNCIA ARTIFICIAL**. [s.n.], 2008. Acesso em: 19 out. 2024. ISSN 2713-4323. Disponível em: <<http://www.periodicos.ulbra.br/index.php/ic/article/view/1649>>.
- LUDERMIR, T. B. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências**. [S.l.]: Instituto de Estudos Avançados da Universidade de São Paulo, 2021. v. 35. 85–94 p. ISSN 0103-4014.

MESQUITA, I. C. A.; CONDE, M. G. **A evolução gráfica do livro e o surgimento dos e-books**. São Luís, MA, Brasil, 2008. Trabalho apresentado no GT – Jornalismo e Editoração, do Iniciacom.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos Sobre Aprendizado de Máquina**. 1. ed. Barueri-SP: Manole Ltda, 2003. 89–114 p. ISBN 85-204-168.

nVersos Editora. **A importância dos livros físicos**. 2022. Acesso em: 19 maio 2024. Disponível em: <<https://www.nversoseditora.com/post/a-importancia-dos-livros-fisicos>>.

OLIVEIRA, B. V. N. de; MELO, F. T. de. **Fundamentos da Visão Computacional: Arcabouço Teórico do Reconhecimento Artificial de Imagens e Vídeos**. [s.n.], 2023. v. 10. 313–324 p. Acesso em: 20 out. 2024. Disponível em: <<https://revista.unitins.br/index.php/humanidadeseinovacao/article/view/9078>>.

PAUL, M.; NAIKAR, S. **Innovative Use of QR Codes in Academic Libraries: Benefits and Challenges**. [S.l.: s.n.], 2024. v. 11. 565-572 p. Acesso em: 30 jun. 2024.

REZENDE, V. C. **Metodologia para a Classificação Automática de Doenças em Plantas Utilizando Redes Neurais Convolucionais**. Dissertação (Dissertação de Mestrado) — Universidade Federal do Pará, Belém, PA, 2019. Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Tecnologia.

SHI, X.; TANG, K.; LU, H. **Smart library book sorting application with intelligence computer vision technology**. Emerald Publishing Limited, 2021. v. 39. 220–232 p. Acesso em: 30 jun. 2024. Disponível em: <<https://doi.org/10.1108/LHT-10-2019-0211>>.

SILVA, G. C. **Tecnologia, educação e tecnocentrismo: as contribuições de Álvaro Vieira Pinto**. [S.l.]: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2013. v. 94. 839–857 p. ISSN 2176-6681.

SILVA, J. D. S. da. **Uso de Redes Neurais em Visão Computacional e Processamento de Imagens**. São José dos Campos, SP, 2004. Relatório de estágio SPE/RHAE, Processo nº 260.048/95-0, realizado no Departamento de Engenharia Nuclear, Universidade do Tennessee, Knoxville, EUA.

TAYEBINIK, M.; PUTEH, M. **Blended Learning or E-learning?** [s.n.], 2013. 1–10 p. Acesso em: 19 maio 2024. Disponível em: <<https://arxiv.org/abs/1306.4085v1>>.

UTSCH, K. G. **Uso de Redes Neurais Convolucionais para Classificação de Imagens Digitais de Lesões de Pele**. Projeto de Graduação — Universidade Federal do Espírito Santo, Vitória, ES, janeiro 2018. Departamento de Engenharia Elétrica, Centro Tecnológico.

VERNE, T. S. et al. **DETECÇÃO E RECONHECIMENTO DE PLANTAS DE PEQUENO PORTE UTILIZANDO APRENDIZAGEM DE MÁQUINA.** [S.l.: s.n.], 2022. v. 14. 36–45 p.

YANG, L.; SHEN, X. **Book Cover Recognition.** [s.n.], 2024. Acesso em: 16 out. 2024. Disponível em: <[https://web.stanford.edu/class/ee368/Project\\_Autumn\\_1617/Reports/report\\_yang\\_shen.pdf](https://web.stanford.edu/class/ee368/Project_Autumn_1617/Reports/report_yang_shen.pdf)>.

YUFEI, L. et al. **Review of the Application of Artificial Intelligence in Education.** [S.l.]: International Journal of Innovation Creativity and Change, 2020. 548-562 p. ISSN 2201-1323.