

Artificial Intelligence for Book Cover Recognition and Virtual Annotation on Physical Books: An Approach to Support Academic Study

Vitor Jacinto Frassetto¹, André Faria Ruaro¹, Leila Laís Gonçalves¹, Marlon Oliveira¹

¹Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc)
Av. Universitária, 1105 – Universitário – Criciúma – SC – Brazil

{vitorfrassetto.vf, andre.ruaro, llg, marlon.oliveira}@unesc.net

Abstract. *This research presents an AI-based prototype for automatic book recognition through cover images, integrating virtual annotation and digital cataloging features to support academic study. It addresses limitations of physical books, such as the inability to annotate pages and difficulties in organizing and retrieving references. The method uses computer vision and machine learning, extracting visual features via convolutional neural networks for accurate identification. An interactive system was also developed for creating virtual annotations on images, enhancing user interaction. The prototype showed promising results, with over 99% accuracy even under uncontrolled image conditions.*

Resumo. *Esta pesquisa apresenta um protótipo de inteligência artificial para reconhecimento automático de livros por imagens de capas, com funcionalidades de anotação virtual e catalogação digital para apoio aos estudos acadêmicos. A proposta supera limitações dos livros físicos, como a falta de anotações diretas e dificuldades na organização e recuperação de referências. A metodologia aplica visão computacional e aprendizado de máquina, utilizando redes neurais convolucionais para identificação precisa. Um sistema interativo permite criar anotações virtuais nas imagens, ampliando a interação com o conteúdo. O protótipo obteve resultados promissores, com acurácia superior a 99%, mesmo em imagens capturadas em condições não controladas.*

1. Introdução

O livro didático é, indiscutivelmente, uma ferramenta pedagógica fundamental, frequentemente servindo como o primeiro contato formal dos alunos com o material de leitura e fomentando o desenvolvimento de hábitos de estudo e aprendizado contínuo [de Godoi and de Moraes Silva 2016]. Além de seu papel na aquisição de habilidades de escrita e leitura, o formato físico do livro permite interações importantes, como a realização de anotações e destaque. Essas práticas ativas não apenas promovem uma compreensão mais profunda, mas também auxiliam significativamente na memorização do conteúdo [nVersos Editora 2022]. O valor tátil e a experiência de leitura imersiva que o livro físico proporciona permanecem centrais no processo educacional.

Apesar das vantagens inerentes ao formato físico, a praticidade do livro pode ser limitada em um cenário educacional cada vez mais dinâmico. A posse transitória ou compartilhada de exemplares dificulta o registro de anotações personalizadas e permanentes. Além disso, em um contexto onde a pesquisa e a organização do conhecimento se tornam

predominantemente digitais, anotações físicas perdem em acessibilidade e capacidade de busca em comparação com materiais digitais, o que pode dificultar a revisão e o aprofundamento do aprendizado. Essa desconexão entre o material físico e as conveniências do ambiente digital impõe desafios à gestão e recuperação de referências.

Nesse contexto, a integração de tecnologias digitais surge como um complemento estratégico à aprendizagem tradicional. Modelos como o *blended learning*, que combinam recursos convencionais com plataformas digitais, demonstram o potencial de enriquecer a experiência educacional [Silva 2013, Tayebinik and Puteh 2013]. Essas ferramentas permitem adaptar o ensino aos ritmos e necessidades individuais dos estudantes, aumentando a compreensão e a retenção do conhecimento [Dutra et al. 2024]. A chave reside em encontrar soluções que harmonizem o valor pedagógico do livro físico com a eficiência e as funcionalidades do universo digital.

Para alcançar essa integração e superar as limitações do suporte impresso sem abrir mão de seus benefícios, métodos avançados de Inteligência Artificial (IA) apresentam-se como uma abordagem promissora. Especificamente, o reconhecimento de padrões e a visão computacional buscam automatizar a identificação e a interpretação de dados visuais, tarefa que, embora intuitiva para humanos, exige técnicas robustas para agilizar processos e reduzir erros em sistemas automatizados. Nesse domínio, o aprendizado de máquina se destaca por sua capacidade de gerar conhecimento diretamente a partir de dados, e as Redes Neurais Convolucionais (CNNs) são amplamente empregadas devido à sua alta performance na identificação e classificação de padrões visuais em imagens, extraíndo características relevantes para tarefas complexas de raciocínio [de Oliveira and de Melo 2023, Graciano 2007].

No cerne das soluções de visão computacional, o aprendizado de máquina se destaca pela sua capacidade de gerar conhecimento diretamente a partir de dados, construindo modelos preditivos baseados em exemplos [Ludermir 2021]. Suas abordagens, como o aprendizado indutivo (supervisionado e não supervisionado) [Monard and Baranauskas 2003], são essenciais para sistemas que precisam aprender a identificar objetos visuais. Entre as arquiteturas mais eficazes para este propósito, as Redes Neurais Convolucionais (CNNs) são amplamente empregadas devido à sua alta performance na identificação e classificação de padrões visuais em imagens [Lucieri et al. 2020, da Silva 2004]. As CNNs funcionam através de camadas convolucionais que extraem características relevantes das imagens, culminando em camadas totalmente conectadas que realizam tarefas de raciocínio complexo, como a classificação precisa de objetos [Lucieri et al. 2020].

Assim, a aplicação da IA para identificar livros físicos por meio de suas capas e possibilitar a criação de anotações virtuais sobre eles representa uma abordagem inovadora. Essa estratégia não só aprimora a gestão do conhecimento, mas também enriquece a experiência de aprendizado, ao capitalizar as vantagens do mundo digital, como o destaque de texto e a facilidade de catalogação, sem desconsiderar o valor do material impresso [Mesquita and Conde 2008]. A combinação do reconhecimento de imagem com a anotação virtual facilita o acesso, a organização e a interação com o material acadêmico, respondendo a uma demanda crescente por soluções tecnológicas que personalizam e simplificam o estudo contemporâneo [Yufei et al. 2020].

Diante deste cenário, esta pesquisa tem como objetivo geral desenvolver um protótipo baseado em inteligência artificial capaz de reconhecer livros pela capa e integrar um sistema de anotações virtuais diretamente em livros físicos. Para tal, os objetivos específicos incluem: (1) analisar os benefícios da integração de tecnologias digitais ao processo de estudo; (2) desenvolver um protótipo de IA apto a reconhecer o livro por meio de uma imagem da capa; (3) implementar um sistema de catalogação dos livros identificados, que permita a criação de anotações, como destaque e comentários intuitivos sobre as imagens enviadas; e (4) avaliar o protótipo, mensurando a acurácia do reconhecimento.

Para uma organização clara, este trabalho está estruturado em cinco seções. A primeira seção corresponde a esta introdução. A segunda seção detalha os trabalhos correlatos à presente pesquisa. A terceira seção descreve os materiais e métodos empregados no desenvolvimento do protótipo. A quarta seção apresenta a discussão e os resultados obtidos. Por fim, o quinto capítulo compila as conclusões do estudo e oferece 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 and 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 et al. 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.

[Iwana et al. 2016] propuseram um sistema de classificação de gêneros literários a partir de imagens de capas de livros, utilizando redes neurais convolucionais profundas. O modelo, baseado na arquitetura AlexNet pré-treinada no ImageNet, foi adaptado para

classificar 30 gêneros distintos. O conjunto de dados criado inclui 137.788 capas com metadados, sendo desafiador pela ambiguidade visual entre gêneros. A abordagem de *transfer learning* obteve 24,7% de acurácia Top-1, superando a LeNet adaptada (13,5%). A análise qualitativa mostrou que o modelo aprendeu associações visuais relevantes, como cores, objetos e estilos tipográficos específicos, embora muitas capas apresentassem características ambíguas. O estudo reforça o potencial de redes profundas para tarefas complexas de classificação visual e sugere futuras pesquisas em classificação multirrótulo, ajustes de hiperparâmetros e uso combinado de texto e imagem.

3. Materiais e Métodos

Esta seção detalha os recursos e procedimentos utilizados no desenvolvimento do protótipo de reconhecimento de capas de livros com anotações virtuais, empregando técnicas de visão computacional e aprendizado de máquina. A solução engloba desde a aquisição e o pré-processamento de imagens até a extração de características com redes neurais, a construção do classificador e a implementação de uma interface interativa para visualização e anotação.

O processo inicia-se com o treinamento de uma base de dados composta por 5.700 imagens e metadados, disponibilizada por [Finsterwald 2024] na plataforma *Kaggle*. As imagens passam por pré-processamento e *data augmentation* para ampliar a diversidade das amostras. Em seguida, aplica-se a arquitetura VGG16 para extração de características visuais e um classificador *Multilayer Perceptron* (MLP) para a classificação, visando maior precisão e robustez no reconhecimento.

O fluxo do sistema inicia-se com o envio da imagem pelo usuário, que passa por remoção de fundo e ajuste de contraste para melhorar sua qualidade. A imagem é então processada pela VGG16 para extração de características, que são comparadas com o modelo treinado. O classificador identifica a imagem com base nesses vetores, retornando as três classes mais prováveis com mais de 50% de probabilidade, permitindo ao usuário identificar o livro e acessar as funcionalidades de catalogação e anotações virtuais.

Após o reconhecimento, o usuário pode acessar informações dos livros identificados e realizar anotações sobre as imagens enviadas, integrando o conteúdo físico ao virtual. A Figura 1 ilustra esse processo por meio de um fluxograma.

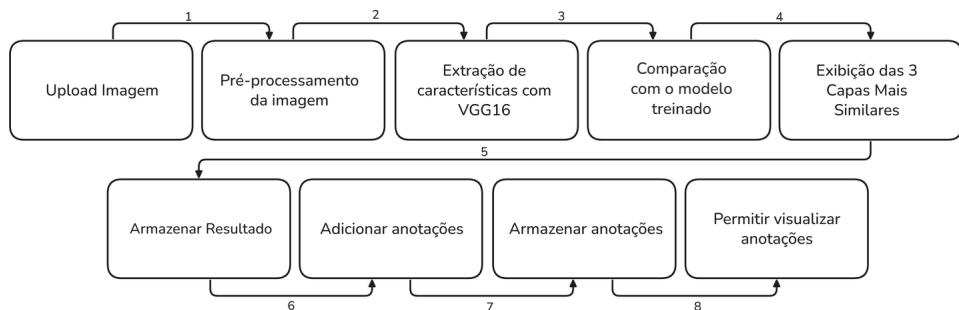


Figura 1. Fluxograma das etapas de desenvolvimento do trabalho.

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.

Para viabilizar a integração com o protótipo, as imagens foram organizadas em um diretório, mantendo a correspondência com os identificadores no CSV (*Filename*). Como havia uma imagem por livro, aplicaram-se técnicas de *data augmentation* para ampliar a variabilidade e melhorar a generalização do modelo. As transformações simulam condições reais de captura, incluindo translações e escalas de até 5%, rotações de até 10°, distorções de perspectiva de 2–5%, e ajustes aleatórios de brilho, contraste, saturação e matiz. Para simular degradações, aplicou-se desfoque gaussiano e inseriram-se até três regiões apagadas (*Coarse Dropout*) de 10×10 a 20×20 pixels, representando oclusões como sombras ou dedos. Todas as imagens foram redimensionadas para 224×224 pixels, garantindo uniformidade para as redes neurais convolucionais.

O protótipo foi desenvolvido em Python 3.12.10, integrando bibliotecas especializadas em visão computacional, *deep learning*, manipulação e processamento de dados. Para pré-processamento das imagens, usaram-se *OpenCV* e *Pillow*; para construção e treinamento das redes convolucionais e do classificador *Multilayer Perceptron* (MLP), *TensorFlow* e *Keras*; e para manipulação e análise, *NumPy* e *Pandas*.

A interface visual foi desenvolvida com a biblioteca *Streamlit*, permitindo a criação de uma interface web simples e interativa, permitindo o envio de imagens, a visualização dos resultados e a realização de anotações virtuais. Para as anotações, foi integrado o componente *streamlit-drawable-canvas*, que adiciona uma área interativa de desenho sobre a imagem enviada, permitindo destacar regiões relevantes, simulando marcações em livros físicos.

A abordagem proposta é estruturada em três módulos principais que atuam de forma integrada. O primeiro módulo, voltado ao reconhecimento de capas, recebe a imagem enviada pelo usuário e executa o pré-processamento, extração de características e inferência com um modelo treinado para identificar o livro, conforme ilustrado na Figura 2.

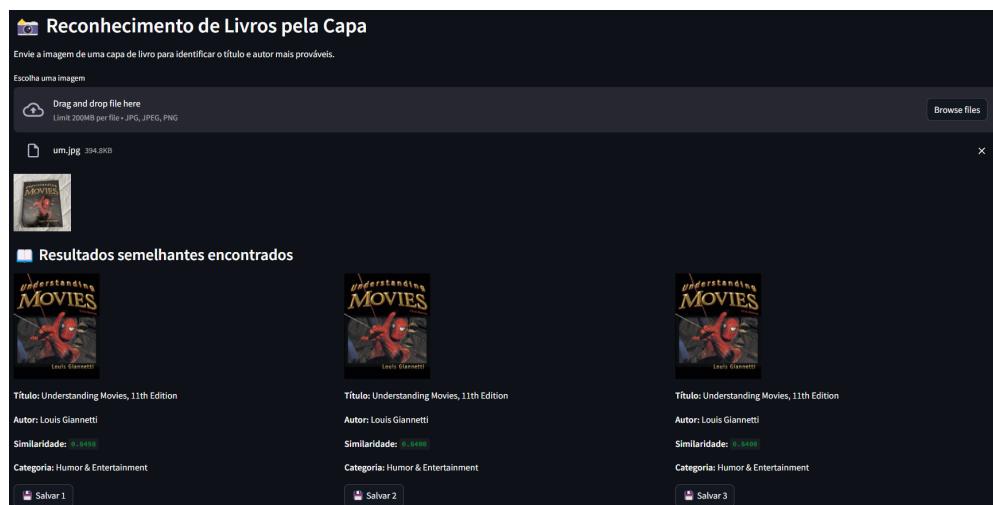


Figura 2. Reconhecimento de livro.

O segundo módulo, denominado catálogo, gerencia os livros reconhecidos, organizando seus metadados. Também permite ao usuário visualizar anotações e criar novas anotações associadas aos livros catalogados.

Por fim, o terceiro módulo, ilustrado na Figura 3, corresponde ao sistema de anotações virtuais. Ele oferece uma interface interativa para que o usuário faça anotações diretamente sobre a imagem enviada, permitindo desenhos livres e destaque visuais além de um campo para inserção de textos complementares.



Figura 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 sistema 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.

Para o processamento de imagens, tanto na etapa de treinamento quanto na de reconhecimento, utilizou-se a biblioteca OpenCV para leitura e equalização de cores, ajustando saturação, contraste e nitidez. Esses ajustes uniformizam a qualidade visual e

otimizam a extração de características relevantes pelos modelos de reconhecimento.

Adicionalmente, somente na etapa de reconhecimento, empregou-se o pacote *backgroundremover* para segmentar e remover o fundo, seguido de detecção de quadriláteros para corrigir inclinações e ajustar a perspectiva. Esses procedimentos garantem consistência visual e reduzem ruídos que possam prejudicar o desempenho do modelo.

Além disso, especificamente na etapa de treinamento, utilizou-se a ferramenta *Albumentations* para aplicar técnicas de *data augmentation*, gerando variantes das imagens originais sem remoção de fundo, já 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 com equalização de contraste. Essas transformações aumentaram a diversidade do conjunto de dados, melhorando a capacidade de generalização do modelo.

A extração de características foi implementada utilizando a rede neural *VGG16*, composta por 13 camadas convolucionais e 3 camadas totalmente conectadas, organizadas em blocos com filtros 3×3 , *ReLU* como função de ativação e camadas de *max pooling* 2×2 para redução progressiva da dimensão espacial. Utilizou-se a versão 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 é fundamental, pois a rede já possui filtros capazes de identificar padrões visuais genéricos, como bordas, texturas e formas. 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* global 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 1×10^{-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.

Após o treinamento do modelo de classificação, realizou-se uma etapa de validação para acompanhar o desempenho ao longo das 100 épocas iniciais. Utilizou-

se uma divisão interna que separou automaticamente 20% dos dados de treino para validação. A cada época, 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 acurácia e perda em dados não vistos. Esse procedimento permitiu monitorar as métricas *training accuracy*, *training loss*, *validation accuracy* e *validation loss*, além de recall e F1-score, prevenindo *overfitting* com o callback *EarlyStopping*, que interrompeu o treinamento ao detectar estagnação na validação.

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.

Os códigos fontes deste trabalho encontram-se no GitHub¹

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, observou-se uma rápida evolução da acurácia de validação e uma redução consistente da função de perda, refletindo a eficácia do modelo na aprendizagem das características visuais das capas. A acurácia subiu de 0,17% na primeira época para 64% na quarta e estabilizou em 99% a partir da 15^a, levando ao encerramento antecipado na 46^a época via *EarlyStopping*. Paralelamente, a perda caiu de 8,59 para 0,0077, mantendo-se estável nas últimas épocas. Esses resultados, ilustrados na Figura 4, demonstram bom desempenho e prevenção de sobreajuste.

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

¹<https://github.com/vitorfrassetto/TCC-FINAL>

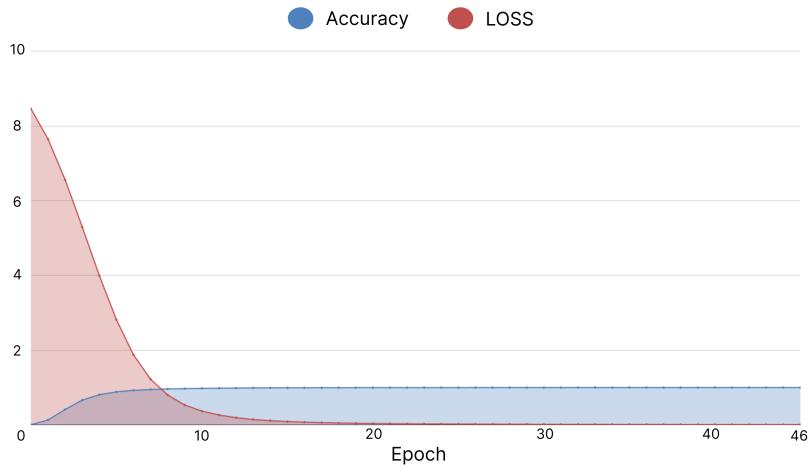


Figura 4. Acurácia e função de perda no conjunto de validação.

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 da acurácia e da perda, foram calculadas métricas adicionais para avaliar o desempenho do modelo. O *recall* ponderado foi 0,9783, indicando boa identificação das classes relevantes, e o *F1-score* ponderado atingiu 0,9767, mostrando equilíbrio entre precisão e sensibilidade. A eficiência também foi avaliada, com tempo médio de processamento por imagem de 67,76 ms (mínimo 53,10 ms, máximo 283,51 ms), indicando resposta adequada para aplicações práticas e possível integração em sistemas quase em tempo real.

Além da validação interna, o modelo foi testado manualmente com 30 imagens reais de capas de livros (Figura 5), que apresentavam variações de qualidade, reflexos e padrões visuais. O modelo classificou corretamente 27 capas, alcançando 90% de acerto. Os 3 erros (10%) ocorreram devido a problemas no pré-processamento, 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 previsões, mas o fez com um alto nível de confiança. Dentre os acertos, 17 tiveram probabilidades acima de 99%, evidenciando classificações seguras. Apenas três acertos apresentaram probabilidades inferiores a 70%, sugerindo incerteza considerável nesses casos, apesar da correção da previsão.

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-



Figura 5. Exemplo em situações reais testadas manualmente.

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 6.



Imagen Original

Imagen Recortada

Figura 6. Exemplo de problemas no recorte da capa.

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, consequentemente, afetando negativamente a qualidade da extração de características e a acurácia do reconhecimento.

Embora não existam trabalhos exatamente iguais que reconheçam livros pela capa em um banco de dados, há similaridades com [Hu et al. 2024], que propuseram um sistema automatizado para recuperação de páginas de livros usando redes neurais convolucionais (CNNs). Ambos utilizam visão computacional e arquiteturas baseadas na VGG, focando na extração de características visuais para identificar materiais bibliográficos.

Hu, Yan e Xie relataram taxa de recuperação de 93,18% no teste inicial, chegando a 99,31% considerando as cinco melhores configurações. Já o modelo proposto, que combina extração via VGG16, alcançou 98,01% de acurácia no reconhecimento de capas de livros.

5. Conclusão

O presente trabalho teve como objetivo desenvolver um sistema automatizado para o reconhecimento de livros a partir de imagens de suas capas, utilizando 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, posteriormente, classificação realizada por meio do modelo treinado, especificamente para distinguir entre as diferentes classes de capas de livros.

Os principais resultados demonstraram que o modelo foi capaz de atingir uma acurácia de 98,01% no conjunto de teste, além de apresentar um desempenho consistente em avaliações práticas, com 88,33% de similaridade média e 90% de acerto nos livros reconhecidos, em um conjunto adicional de 30 imagens reais. Esses resultados indicam a viabilidade do sistema para o reconhecimento automatizado de livros, mesmo quando exposto a imagens obtidas em condições variadas. Além disso, destaca-se seu potencial para auxiliar estudos por meio de sistemas de catalogação de livros e anotações virtuais em imagens. Assim, apoia as práticas educacionais com um ambiente propício para o enriquecimento da experiência de estudo, promovendo a personalização do aprendizado, a sistematização do conhecimento e facilitando a gestão de informações em ambientes acadêmicos.

Para trabalhos futuros, sugere-se expandir o dataset com capas de livros brasileiros de diferentes editoras e analisar o desempenho do modelo frente a distintos tipos de capas (coloridas, monocromáticas, com mais texto ou imagens). A incorporação de OCR pode permitir a extração direta de informações da capa. Também é recomendada a investigação dos hiperparâmetros, melhorias no pré-processamento de imagens e a comparação com ferramentas como o Google Lens. Outras possibilidades incluem a integração com sistemas de bibliotecas, funcionalidades colaborativas e o desenvolvimento de um aplicativo móvel.

Referências

- da Silva, J. D. S. (2004). Uso de redes neurais em visão computacional e processamento de imagens. Relatório de atividades, Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, SP. Relatório de estágio SPE/RHAE, Processo nº 260.048/95-0, realizado no Departamento de Engenharia Nuclear, Universidade do Tennessee, Knoxville, EUA.
- de Oliveira, B. V. N. and de Melo, F. T. (2023). *Fundamentos da Visão Computacional: Arcabouço Teórico do Reconhecimento Artificial de Imagens e Vídeos*, volume 10. Acesso em: 20 out. 2024.

- Dutra, I. T. L., de Oliveira, F. C., Fachin, E. A. G., da Silva, R. L., de Paula Filho, P. N., Domingues, K. M., and Guimarães, U. A. (2024). *O Impacto das Tecnologias Digitais no Processo de Ensino-Aprendizagem*, volume 28. Acesso em: 19 maio 2024.
- Finsterwald, M. (2024). Book cover dataset. Acesso em: 18 mar. 2025.
- Graciano, A. B. V. (2007). *Rastreamento de objetos baseado em reconhecimento estrutural de padrões*. São Paulo. Acesso em: 18 out. 2024.
- Hu, J., Yan, Y., and Xie, Z. (2024). *Automatic Recognition Technology of Library Books Based on Convolutional Neural Network Model*, volume 5. HighTech and Innovation. Acesso em: 18 out. 2024.
- Iwana, B. K., Rizvi, S. T. R., Ahmed, S., Dengel, A., and Uchida, S. (2016). Judging a book by its cover. *arXiv preprint arXiv:1610.09204*.
- Lucieri, A., Sabir, H., Siddiqui, S. A., Rizvi, S. T. R., Iwana, B. K., Uchida, S., Dengel, A., and Ahmed, S. (2020). Benchmarking deep learning models for classification of book covers. *SN computer science*, 1(3):139.
- Ludermir, T. B. (2021). *Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências*, volume 35. Instituto de Estudos Avançados da Universidade de São Paulo.
- Mesquita, I. C. A. and Conde, M. G. (2008). *A evolução gráfica do livro e o surgimento dos e-books*. São Luís, MA, Brasil. Trabalho apresentado no GT – Jornalismo e Editoração, do Iniciacom.
- Monard, M. C. and Baranauskas, J. A. (2003). *Conceitos Sobre Aprendizado de Máquina*. Manole Ltda, Barueri-SP, 1 edition.
- nVersos Editora (2022). A importância dos livros físicos. Acesso em: 19 maio 2024.
- Silva, G. C. (2013). *Tecnologia, educação e tecnocentrismo: as contribuições de Álvaro Vieira Pinto*, volume 94. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- Tayebinik, M. and Puteh, M. (2013). *Blended Learning or E-learning?* Acesso em: 19 maio 2024.
- Verne, T. S., da Silva, F. A., de Almeida, L. L., Pereira, D. R., and Artero, A. O. (2022). *Detecção e Reconhecimento de Plantas de Pequeno Porte Utilizando Aprendizagem de Máquina*, volume 14.
- Yang, L. and Shen, X. (2024). *Book Cover Recognition*. Acesso em: 16 out. 2024.
- Yufei, L., Saleh, S., Jiahui, H., and Abdullah, S. (2020). *Review of the Application of Artificial Intelligence in Education*. International Journal of Innovation Creativity and Change.