

Classificação de imagens usando Redes Neurais Convolucionais SIN393

1st Arthur Roberto de Paula Neto

Matrícula: 8079

UFV-CRP

Rio Paranaíba, Brasil

arthur.r.neto@ufv.br

I. INTRODUÇÃO

A classificação de imagens é uma tarefa fundamental no campo da visão computacional, com aplicações que vão desde o diagnóstico médico até sistemas de segurança baseados em reconhecimento facial. O objetivo principal dessa tarefa é categorizar imagens em classes predefinidas, utilizando informações extraídas dos dados visuais para identificar padrões e características distintas. Com o avanço das técnicas de aprendizado profundo, as Redes Neurais Convolucionais (CNNs) se destacaram como uma das abordagens mais eficazes para resolver problemas relacionados à classificação de imagens, devido à sua capacidade de capturar características espaciais e hierárquicas de forma automática.

Este trabalho tem como objetivo desenvolver e avaliar um modelo de classificação de imagens utilizando Redes Neurais Convolucionais, aplicado ao Oxford-IIIT Pet Dataset, um conjunto de dados amplamente utilizado em tarefas de visão computacional. Este dataset apresenta desafios significativos, como a variabilidade nos fundos das imagens e a alta similaridade visual entre algumas classes, o que torna a tarefa de classificação ainda mais complexa.

As CNNs são particularmente úteis neste contexto devido à sua arquitetura especializada, que combina camadas convolucionais para extração de características, operações de pooling para redução da dimensionalidade e camadas densas para a previsão final das classes. Além disso, a utilização de técnicas como aumento de dados (data augmentation) e ajuste dinâmico da taxa de aprendizado contribui para melhorar a capacidade do modelo de generalizar para novos dados.

Neste estudo, foram exploradas etapas fundamentais, como o pré-processamento das imagens, o treinamento e a validação do modelo, bem como a análise de métricas relevantes, incluindo precisão, recall, F1-score e acurácia geral. Gráficos como as curvas de perda e acurácia ao longo das épocas fornecem insights sobre o progresso do treinamento, enquanto a matriz de confusão auxilia na identificação de padrões de erros recorrentes.

Portanto, este trabalho não apenas busca implementar um modelo eficaz para a classificação de imagens, mas também contribui para o entendimento das limitações e potencialidades das Redes Neurais Convolucionais em cenários complexos de

visão computacional. Acredita-se que os resultados obtidos possam servir como base para futuras melhorias e aplicações práticas em diversas áreas que demandam classificação de imagens de alta precisão.

II. REVISÃO BIBLIOGRÁFICA

O estudo de Silva reforça a eficácia das CNNs em tarefas de classificação de imagens, especialmente no domínio do reconhecimento facial. Ele ilustra como a arquitetura convolucional é capaz de lidar com dados visuais complexos e extrair características discriminantes. Essa contribuição se alinha aos avanços na área, destacando o papel das CNNs na solução de problemas computacionais desafiadores, como a classificação de imagens em diferentes domínios.

O trabalho também evidencia desafios relacionados ao uso de CNNs, como:

- A necessidade de bases de dados grandes e bem rotuladas para o treinamento eficaz.
- O alto custo computacional, que demanda hardware especializado, como GPUs.

Em suma, o estudo oferece uma base sólida para compreender o uso de CNNs em tarefas de classificação de imagens, sendo um exemplo relevante de aplicação prática e resultados bem-sucedidos. [1]

O trabalho de Luiz Carlos Marques Junior, intitulado "Classificação de Plantas Daninhas em Banco de Imagens Utilizando Redes Neurais Convolucionais", apresenta uma aplicação prática de CNNs no contexto da agricultura de precisão. Ele destaca o uso de aprendizado profundo para identificar e classificar cinco espécies de plantas daninhas que apresentam resistência a herbicidas. Essa abordagem visa mitigar problemas econômicos e ambientais associados ao uso excessivo de herbicidas.

As Redes Neurais Convolucionais são uma escolha natural para problemas de classificação de imagens devido à sua capacidade de aprender automaticamente características relevantes de dados visuais. No trabalho, as CNNs foram utilizadas para diferenciar espécies específicas de plantas daninhas, que frequentemente apresentam padrões visuais semelhantes ao ambiente ou às culturas ao seu redor. Isso torna a tarefa de classificação desafiadora, mas ideal para o uso de CNNs.

O estudo contribui para a aplicação de CNNs na agricultura de precisão, destacando sua utilidade na identificação e manejo de plantas daninhas resistentes. A utilização de aprendizado profundo pode reduzir o uso excessivo de herbicidas, diminuindo custos para os agricultores e impactos ambientais. [2]

O trabalho de Souza, tem como objetivo principal avaliar o desempenho de diferentes arquiteturas de Redes Neurais Convolucionais (ResNet50, InceptionV3 e InceptionResNetV2) na tarefa de reconhecimento de cenas. Essa tarefa, que é uma forma específica de classificação de imagens, consiste em categorizar imagens em diferentes classes, como "praia", "deserto" ou "floresta". O estudo destaca o uso de pesos pré-treinados na base ImageNet como técnica de transferência de aprendizado para melhorar os resultados.

As CNNs têm se consolidado como a principal abordagem para classificação de imagens devido à sua capacidade de extrair automaticamente características relevantes de dados visuais. O trabalho explora isso ao aplicar CNNs na base de dados Places, que contém milhões de imagens rotuladas em diversas classes semânticas.

- Transferência de Aprendizado: O uso de pesos pré-treinados em modelos, como os disponibilizados pela ImageNet, permite que as CNNs comecem o treinamento a partir de uma base já ajustada, o que reduz a necessidade de grandes conjuntos de dados específicos para o problema em questão.

O trabalho de Souza, é particularmente relevante para a classificação de imagens em domínios onde as classes possuem alta similaridade visual, como cenas de montanhas e geleiras. A análise de métricas, como precisão, revocação e pontuação F1, mostra que as classes mais semelhantes são as que apresentam maior dificuldade para as CNNs.

Benefícios das CNNs:

- Capacidade de generalizar padrões em grandes conjuntos de dados.
- Redução da necessidade de intervenção manual para extração de características.

Embora o foco do trabalho seja no reconhecimento de cenas, as técnicas e arquiteturas avaliadas são amplamente aplicáveis a outras tarefas de classificação de imagens, como:

- Identificação de objetos em imagens médicas.
- Detecção de sinais de trânsito em sistemas de direção autônoma.
- Categorização de produtos em sistemas de e-commerce. [3]

O estudo "Uso de Aprendizagem de Máquina e Redes Neurais Convolucionais Profundas para a Classificação de Áreas Queimadas em Imagens de Alta Resolução Espacial" de Petronio Diego Silva de Oliveira aborda o uso de Redes Neurais Convolucionais (CNNs) e algoritmos de aprendizado de máquina para identificar e classificar áreas queimadas em imagens de alta resolução espacial obtidas por satélites. Ele destaca a relevância do sensoriamento remoto para o monitoramento de incêndios florestais e sua aplicação na gestão ambiental, especialmente no bioma Cerrado. O trabalho

reflete uma tendência crescente na utilização de aprendizado profundo em análise de imagens, com foco em problemas ambientais.

As CNNs são apresentadas como uma solução eficaz para tarefas de classificação de imagens devido à sua capacidade de aprender características hierárquicas e relevantes diretamente dos dados. No estudo, elas foram comparadas com algoritmos tradicionais de aprendizado de máquina, como K-Nearest Neighbors (KNN), Random Forest (RF) e Support Vector Machine (SVM), destacando-se por seu desempenho superior.

O estudo demonstra a viabilidade de usar CNNs para classificação de imagens em alta resolução espacial, como as fornecidas pelos satélites PlanetScope. A capacidade dessas redes de lidar com grandes volumes de dados e extrair características discriminantes é particularmente relevante para a gestão de recursos naturais e o monitoramento ambiental.

- Impacto Prático:

- Redução de erros em classificações ambientais.
- Agilidade na identificação de áreas críticas, como zonas queimadas, permitindo intervenções rápidas.

O trabalho contribui significativamente para o campo da classificação de imagens, mostrando que as CNNs são ferramentas robustas e precisas para aplicações de sensoriamento remoto. Destaca a importância de arquiteturas convolucionais especializadas para preservar informações espaciais, essenciais em imagens ambientais.

O autor sugere que futuras pesquisas podem explorar novas arquiteturas de aprendizado profundo e combinar técnicas de aprendizado supervisionado e não supervisionado para melhorar ainda mais os resultados. [4]

III. MATERIAL E MÉTODOS

O Oxford-IIIT Pet Dataset, amplamente utilizado em tarefas de classificação e segmentação de imagens, é um recurso essencial para treinar e avaliar modelos nessas áreas. Ele contém imagens de gatos e cães de várias raças, com anotações que possibilitam a classificação da raça do animal e a segmentação de suas formas, fornecendo máscaras de contorno específicas.

Esse conjunto de dados é ideal para diversas tarefas, incluindo a classificação de raças, onde o objetivo é prever a raça específica do animal, e a segmentação de imagens, que visa identificar e separar o animal do fundo. Ele pode ser usado tanto para classificações binárias, como distinguir entre gatos e cães, quanto para classificações multiclasse, abrangendo todas as raças.

O dataset é composto por aproximadamente 7.349 imagens, cada uma contendo um único animal em diferentes poses, condições de iluminação e variações no fundo. As anotações, disponíveis no arquivo annotations/list.txt, incluem o nome do arquivo, a classe do animal (1 para gatos e 2 para cães), a divisão entre dados de treino e teste, e as máscaras para segmentação, armazenadas no diretório annotations/trimaps/. As imagens estão no formato JPEG, enquanto as máscaras de segmentação estão em PNG.

Ao todo, o conjunto apresenta 37 raças, sendo 12 de gatos e 25 de cães. Cada classe está bem representada, com cerca de 200 imagens por raça. Isso torna o dataset valioso para várias aplicações práticas, como sistemas de reconhecimento de raças, ferramentas para abrigos de animais, análises na área de saúde veterinária, e uso educacional, como em cursos que ensinam redes neurais convolucionais.

Entretanto, trabalhar com o Oxford-IIIT Pet Dataset apresenta alguns desafios, como a variabilidade nos fundos das imagens, que podem ser bastante complexos e dificultar o treinamento de modelos mais simples, além de contornos pouco nítidos dos animais em algumas imagens, o que pode complicar as tarefas de segmentação.

Em termos de aprendizado profundo, o dataset é amplamente empregado no treinamento de redes convolucionais (CNNs). Essas redes processam imagens de entrada no formato [3, 224, 224] (três canais de cor: RGB, com altura e largura de 224 pixels) e as transformam em previsões de classe. A arquitetura de uma CNN pode ser dividida em dois blocos principais: o extrator de características e o classificador. O primeiro, formado por camadas convolucionais e de pooling, identifica características visuais relevantes da imagem, como bordas, texturas e formas. O segundo, composto por camadas totalmente conectadas, utiliza essas características para prever a classe da imagem.

As camadas convolucionais aprendem diferentes níveis de abstração, desde bordas simples até formas mais complexas. As operações de pooling reduzem o tamanho dos mapas de características, tornando a rede mais robusta a variações nas imagens, como rotações ou translações. Já as camadas totalmente conectadas transformam as características aprendidas em uma probabilidade para cada classe. A saída final da rede prevê a classe da imagem, com uma função de ativação, como a softmax, convertendo os valores em probabilidades.

Esse fluxo de processamento demonstra como o Oxford-IIIT Pet Dataset, aliado a arquiteturas de CNNs, é um recurso poderoso para abordar desafios de classificação e segmentação, ao mesmo tempo em que oferece uma base sólida para o desenvolvimento de sistemas de visão computacional robustos e eficazes.

IV. RESULTADOS E DISCUSSÃO

Os resultados obtidos durante o treinamento e avaliação do modelo são apresentados nas figuras e métricas descritas a seguir. O gráfico de perda (Figura 1) demonstra que a perda no conjunto de treinamento diminui de forma consistente ao longo das épocas, indicando que o modelo está aprendendo os padrões dos dados. No entanto, a perda de validação, apesar de uma redução inicial, apresenta uma estabilização nas últimas épocas. Esse comportamento pode indicar que o modelo está atingindo o limite de sua capacidade de generalização para novos dados.

A acurácia (Figura 2) apresenta uma tendência crescente tanto para o conjunto de treinamento quanto para o de validação. Curiosamente, a acurácia de validação permanece consistentemente superior à de treinamento, o que sugere que

os dados de validação podem possuir características mais simples em comparação ao conjunto de treinamento. A acurácia de treinamento se estabiliza em um nível inferior, indicando que o modelo não apresenta sinais de sobreajuste evidente.

O gráfico de taxa de aprendizado (Figura 3) revela que essa métrica foi ajustada dinamicamente ao longo do treinamento, com uma redução progressiva. Esse ajuste contribui para um refinamento no aprendizado nas últimas épocas, evitando alterações significativas nos pesos do modelo à medida que o treinamento avança.

O relatório de classificação (Tabela 1) revela discrepâncias significativas no desempenho do modelo entre as classes. Métricas como precisão, recall e F1-score variam consideravelmente, com algumas classes apresentando valores abaixo de 0,2, indicando dificuldade do modelo em diferenciá-las. Em contrapartida, classes como "Egyptian Mau" demonstram desempenho superior, possivelmente devido a características intrínsecas dos dados dessas classes que facilitam sua identificação. A matriz de confusão (Figura 4) complementa essa análise, evidenciando padrões de confusão entre classes específicas, possivelmente relacionadas a similaridades visuais.

A acurácia geral do modelo no conjunto de validação foi de 36,7 por cento, o que indica um desempenho limitado no contexto de um problema de classificação de múltiplas classes. Esse resultado pode ser atribuído a fatores como o desequilíbrio ou a insuficiência de dados em algumas classes, a complexidade inerente ao problema, bem como à arquitetura do modelo ou à configuração de seus hiperparâmetros.

Esses resultados sugerem a necessidade de melhorias, como a ampliação e o balanceamento do conjunto de dados, a aplicação de técnicas de aumento de dados e a experimentação com arquiteturas mais complexas. Além disso, ajustes nos hiperparâmetros, como a taxa de aprendizado e o número de épocas, podem contribuir para um desempenho mais robusto em futuras iterações.

V. CONCLUSÃO

Neste trabalho, desenvolvemos e avaliamos um modelo baseado em Redes Neurais Convolucionais (CNNs) para a tarefa de classificação de imagens do Oxford-IIIT Pet Dataset. Embora os resultados demonstrem que o modelo foi capaz de aprender os padrões dos dados de treinamento e apresentar um desempenho consistente no conjunto de validação, os índices gerais de acurácia, como os 36,7

Os gráficos de perda e acurácia destacaram o progresso do treinamento, com melhorias graduais ao longo das épocas, mas também apontaram para a estabilização do aprendizado e possíveis desafios no modelo para lidar com a complexidade do problema. Além disso, a análise do relatório de classificação e da matriz de confusão revelou que o desempenho varia amplamente entre as classes, com algumas apresentando métricas de precisão, recall e F1-score insatisfatórias. Esse comportamento pode ser atribuído a fatores como a alta similaridade visual entre classes, a variabilidade das imagens ou o desequilíbrio no conjunto de dados.

Para superar essas limitações, é fundamental explorar estratégias de aprimoramento, como o balanceamento das classes, o aumento do volume de dados, a aplicação de técnicas de aumento de dados e o ajuste fino dos hiperparâmetros. Experimentos com arquiteturas de redes mais complexas também podem elevar o desempenho do modelo, especialmente em tarefas que envolvem grande quantidade de classes.

Em suma, os resultados obtidos ilustram tanto o potencial quanto os desafios na aplicação de CNNs em problemas de classificação multiclasse, apontando caminhos para futuras pesquisas e melhorias no desempenho do modelo. Este estudo reforça a importância de arquiteturas bem ajustadas e conjuntos de dados robustos para alcançar avanços significativos na área de visão computacional.

REFERENCES

- [1] R. E. V. d. Silva, “Um estudo comparativo entre redes neurais convolucionais para a classificação de imagens,” 2018.
- [2] L. C. Marques Junior, “Classificação de plantas daninhas em banco de imagens utilizando redes neurais convolucionais,” 2019.
- [3] V. Souza, L. Silva, A. Santos, and L. Araújo, “Análise comparativa de redes neurais convolucionais no reconhecimento de cenas,” *Anais do Computer on the Beach*, vol. 11, pp. 419–426, 2020.
- [4] P. D. S. d. Oliveira, “Uso de aprendizagem de máquina e redes neurais convolucionais profundas para a classificação de áreas queimadas em imagens de alta resolução espacial,” 2020.