

Estudo de modelos de redes neurais convolucionais para a classificação de vias urbanas limpas e sujas

1st Eliana Maria Silva de França

Departamento de Estatística e Informática (DEINFO)

Universidade Federal Rural de Pernambuco - UFRPE

Recife, Brazil

eliana.franca@ufrpe.br

Abstract—The article deals with the application of Convolutional Neural Network (CNN) models for trash detection in images of urban roads. The automatic identification of littered urban areas aims to help local authorities manage and clean public spaces. The research uses a database containing a limited number of images divided into two categories: roads with litter and roads without litter. Due to the limited database, the transfer learning technique is used, training only the last layer of the CNN, the fully connected one. The aim is to use pre-trained CNN models to properly classify the images into the two categories, in order to evaluate and compare the performance of the models used.

keywords—Convolutional Neural Networks, Learning Transfer, Urban Cleaning

Resumo—O artigo trata da aplicação de modelos de Redes Neurais Convolucionais (CNN) para detecção de lixo em imagens de vias urbanas. A identificação automática de áreas urbanas cheias de lixo visa ajudar as autoridades locais a gerir e limpar os espaços públicos. A pesquisa utiliza um banco de dados contendo um número limitado de imagens divididas em duas categorias: estradas com lixo e estradas sem lixo. Devido ao banco de dados limitado, é utilizada a técnica de aprendizagem por transferência, treinando apenas a última camada da CNN, a totalmente conectada. O objetivo é utilizar modelos CNN pré-treinados para classificar adequadamente as imagens nas duas categorias, a fim de avaliar e comparar o desempenho dos modelos utilizados.

Palavras chaves—Redes Neurais Convolucionais, Transferência de Aprendizagem, Limpeza Urbana

I. INTRODUÇÃO

Esta seção tem como objetivo introduzir a temática e motivação deste estudo, os quais serão descritos e explicados nas subseções a seguir.

Os tópicos a serem abordados são:

- A. Contexto;
- B. Problema e Motivação;
- C. Objetivos;
- D. Impacto e Resultados;
- E. Contribuições.

A. Contexto

A visão computacional é um campo da inteligência artificial que capacita os computadores a interpretar e compreender o mundo visual. Por meio de algoritmos e análise de imagens, essa tecnologia é capaz de identificar padrões visuais e reconhecer objetos em fotografias ou vídeos.

Os governos distritais tem investido significativamente em recursos e estratégias direcionadas para garantir a higiene e o cuidado ambiental em seus espaços urbanos. Por meio de programas e ações específicas.

Os governos tem potencial para aprimorar significativamente a gestão de limpeza urbana por meio de investimentos em tecnologias emergentes, como a visão computacional. A implementação dessas ferramentas pode ser uma estratégia inovadora para agilizar a identificação de áreas com acúmulo de resíduos e direcionar os recursos de maneira mais eficiente.

B. Problema e Motivação

Aplicação de modelos de Redes Neurais Convolucionais para a detecção de lixo em imagens de ruas urbanas

Ao considerar a viabilidade de investimentos na visão computacional, especificamente Redes Neurais Convolucionais (CNNs), na detecção de lixo em imagens capturadas das ruas dos distritos, a administração municipal poderá não apenas acelerar os processos de limpeza, mas também reduzir custos operacionais a longo prazo, melhorando a qualidade de vida dos cidadãos e promovendo a higiene ambiental da cidade.

C. Objetivos

Usar modelos pré-treinados de CNN para classificar imagens de ruas contendo lixo ou não corretamente, para avaliar e comparar o desempenho de cinco modelos clássicos.

D. Impacto e Resultados

Este trabalho pretende fornecer os seguintes impactos para a sociedade:

- Influenciar positivamente nas tomadas de decisões de gestores públicos;
- Apoiar no embasamento para aplicação de políticas públicas, através de projeções nas características estudadas neste estudo;
- Possibilitar uma alternativa no apoio à tomada de decisão, na manutenção da qualidade de vida dos cidadãos;
- Promover a higiene ambiental da cidade.

E. Contribuições

Este trabalho pretende auxiliar a comunidade científica através das seguintes contribuições:

- Fornecer uma análise de vários algoritmos CNNs para a detecção de vias públicas com ou sem lixo.
- Demonstrar como o aprendizado de máquina pode ajudar a combater problemas que afligem a sociedade, através de bases de dados.
- Propagar o estudo de inteligência artificial no idioma brasileiro.

F. Organização do Trabalho

Este trabalho está organizado em 6 seções. Na seção II é feita uma abordagem teórica dos termos, algoritmos e técnicas utilizadas nos experimentos. Na seção III são apresentados os trabalhos correlatos, demonstrando as semelhanças e divergências com este estudo. Na seção IV será explanado de maneira aprofundada a abordagem da proposta. Na seção V é explicado a metodologia seguida no trabalho, quais experimentos foram realizados, apresentado os métodos, ferramentas e tratamentos, assim como os resultados obtidos. Na seção VI serão apresentadas as conclusões deste trabalho.

II. REFERENCIAL TEÓRICO

Nesta seção serão explicados, brevemente, os termos técnicos, tecnologias e procedimentos utilizados neste estudo, assim como serão apresentados os algoritmos e técnicas que foram utilizadas para os experimentos e avaliação dos resultados obtidos.

1) *Inteligência Artificial*: É a área da computação que estuda o aprendizado de máquinas artificiais. Ao contrário dos seres humanos, os computadores não são providos de criatividade e capacidade analítica, mas em contrapartida, possuem um alto poder de processamento de informações, o que humanos não possuem. Essa área da ciência da computação tem como objetivo unir as capacidades humanas com a velocidade de processamento dos computadores. O aprendizado de máquina pode ser dividido em três tipos: Aprendizado de máquina supervisionado, Aprendizado de máquina não supervisionado e Aprendizado de máquina semi-supervisionado. Neste estudo vamos abordar o Aprendizado de máquina supervisionado.

2) *Aprendizado de máquina supervisionado*: Um método de aprendizado de máquina supervisionado é um método que precisa que um ser humano ensine para a máquina como ele deve determinar uma dada instância, a partir de uma base de dados de treino. Podemos encontrar dois tipos de aprendizado de máquina supervisionado - classificação e regressão. A classificação ocorre quando as instâncias possuem um atributo que contém rótulos para cada instância de uma base de dados, por exemplo, determinar uma pessoa alta ou magra de acordo com suas características (atributos), já a regressão envolve cálculo de funções para determinar o valor e uma variável dependente a partir de variáveis independentes.

3) *Redes Neurais Convolucionais*: São um tipo de arquitetura de rede neural amplamente empregada em tarefas de visão computacional e processamento de imagens. Inspiradas no funcionamento do córtex visual humano, as CNNs são compostas por camadas convolucionais que realizam a extração de

características por meio de operações de convolução, seguidas por camadas de *pooling* para redução dimensional e por fim uma camada final com neurônios totalmente conectadas para a classificação.

No âmbito da visão computacional, as cinco redes neurais convolucionais - *ResNet18*, *VGG16*, *DenseNet121*, *MobileNet2* e *AlexNet* - representam marcos significativos no desenvolvimento de arquiteturas capazes de lidar com a complexidade das imagens urbanas e sua aplicação na limpeza de vias públicas.

A *ResNet18*, derivada da família *ResNet*, revolucionou a capacidade de treinar redes profundas ao introduzir conexões residuais. Essa abordagem permitiu uma melhor propagação de gradientes, facilitando o treinamento e a eficiência na identificação de resíduos urbanos.

A *VGG16*, reconhecida por sua estrutura simples e profunda, foi uma das primeiras a demonstrar o potencial das CNNs na extração de características de imagens complexas. No entanto, sua profundidade pode gerar desafios computacionais e de *overfitting* em conjuntos de dados menores.

Por sua vez, a *DenseNet121*, com sua estrutura de conexões densas entre camadas, otimizou a propagação de informações ao longo da rede, lidando de maneira eficaz com o desafio do desaparecimento de gradientes e mostrando-se eficiente na identificação de resíduos urbanos.

A *MobileNetV2* se destaca pela sua eficiência, sendo especialmente projetada para implementação em dispositivos com recursos limitados. Sua leveza não compromete substancialmente a precisão, tornando-a uma escolha viável para a detecção de resíduos em ambientes urbanos.

Por fim, a *AlexNet*, embora mais antiga em termos de desenvolvimento, foi pioneira ao destacar o potencial das CNNs na visão computacional. Apesar de suas limitações em comparação com arquiteturas mais recentes, seu impacto histórico é inegável.

O estudo dessas cinco arquiteturas não apenas ressalta a evolução contínua das CNNs na análise de imagens urbanas, mas também demonstra o papel vital dessas redes na identificação e classificação de resíduos, contribuindo diretamente para soluções inovadoras na limpeza e manutenção de ambientes públicos.

4) *Transferência de Aprendizado*: É um paradigma em aprendizado de máquina no qual o conhecimento adquirido em um modelo treinado posteriormente é aplicado para melhorar o desempenho em outro modelo relacionado. Esse conceito é especialmente relevante em redes neurais convolucionais (CNNs), onde modelos pré-treinados em grandes conjuntos de dados, como *ImageNet*, são adaptados para novos conjuntos de dados ou tarefas específicas. Ao reutilizar os pesos e as representações aprendidas em camadas anteriores do modelo pré-treinado, a transferência de aprendizagem permite a extração de características relevantes e a melhoria do desempenho mesmo em conjuntos de dados menores. Essa abordagem tem sido fundamental para acelerar o desenvolvimento de modelos eficazes em diversas aplicações, como reconhecimento

de imagens, processamento de linguagem natural e outras áreas de aprendizado de máquina.

III. TRABALHOS RELACIONADOS

O artigo *YOLO TrashNet: Garbage Detection in Video Streams* aborda a detecção em tempo real de resíduos abandonados por meio da análise de fluxos de vídeo em um contexto de gestão de resíduos em cidades inteligentes. Emprega uma versão otimizada da rede *YOLOv3* para identificar e reconhecer lixo, visando melhorar a eficiência na gestão de resíduos urbanos. Descreve a criação de um conjunto de dados específico para treinar a rede, incluindo a coleta, pré-processamento e anotação de imagens. Detalha o ambiente e parâmetros de treinamento utilizados, além de avaliar os resultados obtidos, incluindo métricas de precisão e revocação em diferentes classes de lixo. Conclui delineando áreas para futuras melhorias, como expansão do *dataset* e inclusão de novas classes de objetos. Comparativamente, destaca-se pela abordagem de detecção de resíduos em tempo real em contraste com a avaliação da limpeza das ruas, foco do nosso trabalho que possui poucas imagens e por isto usa transferência de aprendizado.

O artigo *Urban Street Cleanliness Assessment Using Mobile Edge Computing and Deep Learning*, aborda a relevância da limpeza de ruas em cidades inteligentes e os desafios das estratégias manuais na detecção de lixo urbano. Propõe-se uma solução inovadora que emprega computação de borda móvel e aprendizado profundo para avaliar a limpeza das vias urbanas. Utilizando câmeras em veículos para capturar imagens das ruas, o processo ocorre em servidores de borda e, posteriormente, na nuvem central. O modelo *Faster R-CNN* foi empregado para identificar categorias e quantidades de lixo, resultando em um mapa visual da limpeza das ruas. As contribuições incluem um novo *framework* de computação de borda, o uso do *Faster R-CNN* e a disponibilização de um conjunto de dados público de lixo urbano. Diferentemente deste estudo, nosso estudo visa comparar cinco tipos de modelos diferentes para avaliar o desempenho de cada um com a mesma base de dados.

IV. ABORDAGEM DA PROPOSTA

APLICAÇÃO DE MODELOS DE REDES NEURAIS CONVOLUCIONAIS PARA A DETECÇÃO DE LIXO EM IMAGENS DE RUAS URBANAS

A identificação manual de áreas urbanas com lixo é um desafio para as autoridades locais. A falta de métodos automatizados dificulta a gestão eficiente e a limpeza de espaços públicos.

Usando técnicas de visão computacional, propomos analisar cinco modelos de CNNs para classificar ruas com ou sem lixo para viabilizar o saneamento público com mais eficácia.

A. Conjunto de dados

Os dados deste estudo foram obtidos do *dataset Clean/Littered Road Classification* disponibilizado pela plataforma *Kaggle*. São 237 imagens, divididos em duas classes: Imagens de ruas urbanas limpas (*clean*) e imagens de ruas

com lixo (*dirty*). Por se tratar de uma base de dados pequena, o uso de modelos pré-treinados é essencial para este estudo.

Uma amostragem das imagens antes do treinamento podem ser verificadas na figura 1.



Figura 1. Amostragem das imagens antes do treinamento. Fonte da imagem: a autora.

V. EXPERIMENTOS E RESULTADOS

Na seção anterior, vimos a abordagem da proposta e a descrição dos dados utilizados. Nesta seção iremos apresentar a metodologia utilizada para este estudo e os resultados obtidos da metodologia aplicada.

A. Ambiente Experimental

Nesta subseção será apresentado o ferramental técnico utilizado neste estudo.

1) *Linguagem de Programação e Bibliotecas Utilizadas*: Todas as etapas que exigiram codificação, da análise dos dados à avaliação dos modelos treinados foram executadas utilizando a linguagem de programação Python, versão 3. As bibliotecas e suas aplicações no experimento estão descritas na tabela I.

Tabela I
BIBLIOTECAS PYTHON UTILIZADAS NO EXPERIMENTO. FONTE: A AUTORA.

Biblioteca	Aplicação
PyTorch	Implementações de aprendizado de máquina, CNN e Transferência de Aprendizado
Matplotlib	Plotagem de gráficos

2) *Ambiente de Desenvolvimento*: O ambiente de desenvolvimento utilizado para este estudo foi o Google Colab, devido à facilidade de utilização, direta no navegador, já possuir muitas bibliotecas pré-instaladas e acesso a mais recursos computacionais, tais como mais memória RAM, disco, e poder utilizar GPUs e TPUs mais rápidas. Para que o treinamento possa ser mais rápido foi utilizado T4 GPU disponibilizado no Google Colab.

B. Pré-Processamento de Dados

Os dados foram divididos em duas partições, uma de validação e outra de treinamento. A distribuição dos dados pode ser verificada na tabela II, assim como a taxa utilizada para determinar o tamanho da partição de validação.

Tabela II
DISTRIBUIÇÃO DOS DADOS. FONTE: A AUTORA.

Clean			
Total	Treinamento	Validação	Taxa
113	74	38	33%
Dirty			
Total	Treinamento	Validação	Taxa
124	83	41	33%

C. Treinamento dos modelos

Por se tratar de um estudo que utiliza modelos pré treinados, foi necessário fazer implementações para aproveitar o aprendizado de detectação de objetos já contido no modelos. Portanto foi feita o treinamento da ultima camada, a totalmente conectada. Os treinamentos foram realizados em 25 épocas para aproveitar da robustez dos modelos.

D. Predições dos modelos

Nesta subseção serão apresentados exemplos de imagens preditas por cada modelo após o treinamento.

As imagens com o título "*predicted:clean*" são referentes as imagens preditas como ruas sem lixo e as imagens com o título "*predicted:dirty*" são as imagens preditas como imagens de ruas com lixo.

As predições da *ResNet18* estão na figura 2, as da rede *VGG16* na figura 3, assim como as predições das redes *DenseNet121*, *MobileNetv2*, *AlexNet*, respectivamente nas figuras 4, 5 e 6.



Figura 2. Predições da rede *ResNet18*. Fonte da imagem: a autora.



Figura 3. Predições da rede *VGG16*. Fonte da imagem: a autora.



Figura 4. Predições da rede *DenseNet121*. Fonte da imagem: a autora.



Figura 5. Predições da rede *MobileNetv2*. Fonte da imagem: a autora.



Figura 6. Predições da rede *AlexNet*. Fonte da imagem: a autora.

Observa-se que todas as imagens mostradas foram classificadas corretamente, e que mesmo as imagens com outros tipos de objetos, tais como automóveis, pessoas, canteiros, sinalização de vias, efeitos de luz, etc não interferiram na classificação correta das imagens.

E. Resultados e Discussão

Todos os modelos obtiveram acurácias altas e semelhantes, no entanto, a *ResNet18* e a *VGG16* demonstraram a maior acurácia, sendo que a *ResNet18* concluiu o treinamento com um minuto a menos em relação a *VGG16*, 75% da duração do treinamento da *VGG16*, portanto este modelo foi o que obteve o maior desempenho neste estudo.

A acurácia obtida de cada modelo, assim como o tempo de execução do treinamento em GPU, podem ser verificados na tabela III.

Tabela III
RESULTADOS DOS 5 MODELOS UTILIZADOS. FONTE: A AUTORA.

Modelo	Acurácia	Tempo de Treinamento
ResNet18	0.987500	3m 3s
<i>VGG16</i>	0.987500	4m 1s
<i>DenseNet121</i>	0.962500	3m 53s
<i>MobileNetv2</i>	0.975000	3m 14s
<i>AlexNet</i>	0.975000	2m 59s

VI. CONCLUSÕES

Este estudo demonstrou a eficácia da transferência de aprendizado utilizando cinco Redes Neurais Convolucionais (CNNs) clássicas previamente treinadas para facilitar a tarefa de identificação de vias públicas com ou sem lixo. A aplicação desses modelos pré-treinados permitiu a rápida adaptação das redes escolhidas para o problema proposto obtendo ótimos resultados, o que possibilita avanços significativos na automação e otimização da limpeza de vias urbanas. Como trabalhos futuros poderemos implementar um sistema automatizado nas vias das cidades e utilizar fotos dos locais nos quais há a possibilidade de implementação do sistema.

REFERÊNCIAS

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.
- [2] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [4] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [5] M. Sandler, A. Howard, M. Zhu, and others, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [6] P. Zhang, Q. Zhao, J. Gao, W. Li, and J. Lu, "Urban Street Cleanliness Assessment Using Mobile Edge Computing and Deep Learning," *IEEE Access*, vol. 7, Institute of Electrical and Electronics Engineers (IEEE), pp. 63550–63563, 2019. doi: 10.1109/access.2019.2914270.
- [7] B. D. Carolis, F. Ladogana, and N. Macchiarulo, "YOLO TrashNet: Garbage Detection in Video Streams," 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE, May 2020. doi: 10.1109/eais48028.2020.9122693.
- [8] Karim, Faizal and Rajbangshi, Krishnav. (2022, December). Clean/Littered Road Classification, Version 2. Retrieved October 10, 2023 from <https://www.kaggle.com/datasets/faizalkarim/cleandirty-road-classification/data>.