Indústria 5.0: Oportunidades e Desafios para Arquitetura e Construção

13º Simpósio Brasileiro de Gestão e Economia da Construção e 4º Simpósio Brasileiro de Tecnologia da Informação e Comunicação na Construção

ARACAJU-SE | 08 a 10 de Novembro

[1]APRENDIZAGEM PROFUNDA PARA ANÁLISE DA OCUPAÇÃO DO ESPAÇO PÚBLICO AVALIANDO A PRESENÇA DE MODAIS DE TRÂNSITO

Deep Learning Applied to Public Occupation Analysis Through Transit Mode Presence Evaluation

Sofia Puppin Rontani
Unicamp | Limeira, São Paulo | s163474@dac.unicamp.br

Caroline Kehl
Unicamp | Campinas, São Paulo | c216094@dac.unicamp.br

Regina Ruschel
Unicamp | Campinas, São Paulo | ruschel@fec.unicamp.br

Eloisa Dezen-Kempter
Unicamp | Limeira, São Paulo | eloisak@unicamp.br

RESUMO

No campo da Arquitetura, Engenharia e Construção (AEC), o *Machine Learning* (ML) tem sido usado em aplicações como design generativo, análises de desempenho e reconhecimento de imagens, entre outros. Neste contexto, o objetivo deste trabalho foi verificar a aplicação de *Deep Learning* (DL) por meio de *Convolutional Neural Networks* (CNNs) para *Object Detection* (Detecção de Objetos). O objetivo do código de DL neste trabalho foi quantificar a presença de diferentes modais de trânsito em determinada área urbana para subsidiar a análise de ocupação do espaço público. O código Detectron/Mask R-CNN de detecção de instâncias em imagens associado ao *dataset* COCO foi capaz de identificar padrões de ocupação do espaço urbano coerentes com o contexto existente na rua de uso misto, comercial e residencial, com lazer noturno. O padrão dá destaque aos modais de trânsito, carro e pedestres. Foram listadas considerações, que podem auxiliar futuros trabalhos acerca da detecção de instâncias em imagens de câmeras urbanas com o código e *dataset* empregados.

Palavras-chave: Aprendizado profundo; Detecção de Objetos; Redes Neurais Convulsionais; Máscara-CNN; Análise urbana.

ABSTRACT

In the Architecture, Engineering and Construction (AEC) field, Machine Learning (ML) has been used in applications such as generative design, performance analysis and recognition of images, among others. In this context, this work presents an application of Deep Learning (DL) through Convolutional Neural Networks (CNNs) for Object Detection. The objective of DL in this work is to quantify the presence of different modes of traffic in a given urban area, to support the analysis of public space occupancy. The Detectron/Mask R-CNN code for image instance detection, combined with the COCO dataset, was able to identify patterns of urban space occupancy consistent with the existing context of a mixed-use street, with commercial and residential areas and nighttime leisure activities. The pattern highlights modes of transportation such as cars and pedestrians. Several considerations were listed, which can assist future work on instance detection in urban camera images using the employed code and dataset.

Keywords: Deep Learning; Object Detection; Convolutional Neural Networks; Mask-CNN; Urban analysis.

1 INTRODUÇÃO

Entender como a mobilidade se apresenta em determinada área urbana é elemento chave para possibilitar o diagnóstico de um local. Para isso, deve ser feita a avaliação do comportamento urbano das instâncias que convivem no meio urbano, tais como automóveis, pedestres e bicicletas. Geralmente, nos estudos urbanos, essa análise ocorre através de observação sistemática (RÚDIO, 2003), em campo, e de maneira manual, pela observação e coleta de dados. Esse tipo de trabalho é moroso e não possui precisão e, por isso, automatizar esse processo seria de grande importância para os estudos urbanos.

A Inteligência Artificial (IA) é um ramo da ciência da computação que busca construir mecanismos que simulem a capacidade cognitiva do ser humano. O ML não é um requisito indissociável da IA, entretanto, está inserido neste universo (ALPAYDIN, 2016). O *Deep Learning* (DL), por sua vez, é um subconjunto de um campo do ML que consiste em aprender automaticamente, utilizando arquiteturas compostas por

[1]RONTANI, S. P. *et al*. Aprendizagem profunda para análise da ocupação do espaço público avaliando a presença de modais de trânsito. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 4., 2023, Aracaju. **Anais [...].** Porto Alegre: ANTAC, 2023.

múltiplas camadas de processamento não linear, cuja estrutura encontra-se baseada na estrutura de um neurônio. Segundo Arnold *et al*. (2016), treinar arquiteturas profundas é uma tarefa difícil e os métodos clássicos, que se mostraram eficazes quando aplicados a arquiteturas rasas, não são tão eficientes quando adaptados a arquiteturas profundas. Especificamente, em um esquema de aprendizado profundo, cada camada é tratada separadamente e treinada sucessivamente: uma vez que as camadas anteriores foram treinadas, uma nova camada é treinada a partir da codificação dos dados de entrada pelas camadas anteriores (ARNOLD *et al*., 2016).

Convolutional Neural Network (CNN), ou Rede Neural Convolucional, é um tipo de rede neural capaz de interpretar imagens como dados de entrada (inputs), atribuindo pesos aos diversos aspectos e instâncias (objetos) contidas nela e, então, ser capaz de diferenciar uma da outra. Conforme Goodfellow, Bengio e Courville (2016), a estrutura de uma CNN é análoga ao padrão de conectividade de neurônios no cérebro humano. Nas CNNs, cada neurônio responde a estímulos apenas em uma região restrita (filtragem). Esse tipo de rede é usado principalmente em reconhecimento de imagens e processamento de vídeos, que consistem em problemas computacionais de classificação. Enquanto nos métodos anteriores os filtros são feitos manualmente, com treinamento supervisionado, nas CNNs o aprendizado ocorre de forma não-supervisionada (GOODFELLOW; BENGIO; COURVILLE, 2016). São exemplos do uso de redes neurais e DL a classificação de imagens no aplicativo Google *Street View* (GOODFELLOW; BENGIO; COURVILLE, 2016) e no reconhecimento de padrões aplicados à plataforma Airbnb (HALDAR *et al*., 2019).

A visão computacional acontece por rastreamento de padrões. Os algoritmos de ML consideram uma imagem como uma matriz de pixels e automatizam as tarefas de monitoramento, inspeção e vigilância (KHAN; AL-HABSI, 2020). Um dos principais objetivos da visão computacional é a compreensão de imagens complexas – incluindo reconhecer os objetos presentes e determinar os atributos desses objetos e da cena, caracterizar as relações entre objetos e fornecer uma descrição semântica da cena (LIN *et al*., 2014). Devido aos avanços recentes nas tecnologias digitais e à disponibilidade de dados confiáveis, o DL demonstrou sua capacidade e eficácia na resolução de problemas de aprendizagem complexos. Em particular, as CNNs têm demonstrado sua eficácia em aplicações de detecção e reconhecimento de imagens (SHAWAHNA, 2019; CARRARA, 2019).

Isso posto, verifica-se que há mais de duas décadas, códigos vêm sendo desenvolvidos para reconhecimento automático, segmentação e classificação de imagens nas mais diversas áreas de aplicação (áreas médica, agricultura, governamental, militar entre outras). Entretanto, ainda é considerada uma tarefa complexa para os profissionais que não têm conhecimento específico em ciências da computação.

O estudo apresentado é fruto de um exercício de aprendizagem realizado durante uma disciplina de pós-graduação de introdução à inteligência artificial[1].Nesse sentido, a pergunta que tentamos responder com este trabalho é: de que maneira tecnologias de IA poderiam ser implementadas para resolução de problemas do campo da arquitetura e urbanismo, neste caso, automatização do processo de levantamento de campo ao que se refere à contagem e classificação de diferentes modais, sem a necessidade de equipe técnica especialista em ciência da informação e computação? Com intuito de aprimorar atividades de Arquitetura e Urbanismo apoiado às tecnologias da Ciência da Computação, o objetivo deste trabalho é demonstrar a aplicação ML para detectar e quantificar instâncias de modais de trânsito em imagens de câmeras urbanas visando a análise da ocupação do espaço urbano.

2 SEGMENTAÇÃO SEMÂNTICA DE IMAGENS

Recentemente, abordagens baseadas em DL foram aplicadas para interpretar cenas urbanas de forma precisa. Lateef e Ruichek (2019) fizeram um levantamento desses métodos, primeiro categorizando-os em dez classes diferentes de acordo com os conceitos das arquiteturas e, depois, fornecendo uma visão geral dos conjuntos de dados disponíveis publicamente onde foram avaliados. No entanto, as redes neurais profundas requerem um número substancial de amostras de treinamento que são de difícil obtenção. Coletar imagens processadas em nível de pixel é um processo demorado. Por isso, a utilização de dados sintéticos está se tornando predominante. Entretanto, a simples aplicação de modelos treinados em dados sintéticos leva a uma queda dramática de desempenho em imagens reais.

A classificação semântica ainda apresenta alguns desafios fundamentais, como traçar limites de edifícios. No artigo de Kang *et al*. (2018), é proposta uma estrutura para classificar os usos de edificações. O método proposto baseado-se em CNNs que classificam as estruturas da fachada a partir de imagens do ponto de vista do observador, como as do Google StreetView, somadas a imagens de sensoriamento remoto, que

geralmente mostram apenas a cobertura das edificações (KANG *et al*., 2018). No artigo de Zhao *et al*. (2019), os autores apresentam uma nova estratégia para determinação de elementos semânticos e compreensão da cena urbana em imagens de alta resolução. As precisões de classificação do conjunto de dados chegaram a 91% no nível do objeto e 88% no nível da cena (ZHAO *et al*., 2019). Por outro lado, no trabalho de Gao *et al*. (2020), é proposto um método de aprendizagem não-supervisionado que supera outros métodos de última geração.

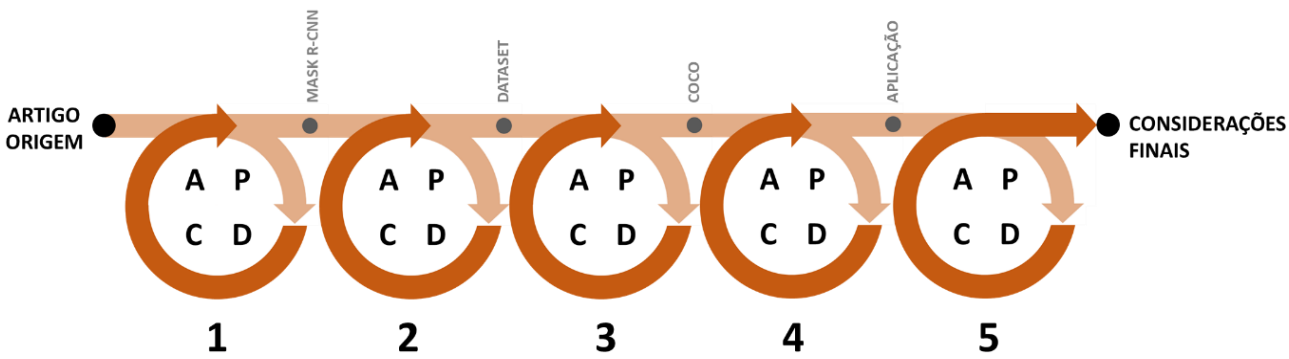
No contexto das análises de vídeo, Lou *et al*. (2019) propõem um novo método de detecção e rastreamento de veículos em movimento com base em vídeo a fim de fazer uso das câmeras de vigilância existentes e melhorar a capacidade de gerenciamento do tráfego urbano. No estudo, é usado o algoritmo Mask R-CNN para detectar contornos de veículos em ambiente de tráfego urbano complexo. Os resultados experimentais mostram que o método proposto pode atingir 95% de acurácia média com velocidade de 2,86 fps, e que ele pode processar efetivamente diferentes condições climáticas e de tráfego (LOU *et al*, 2019).

Segundo Lin *et al*. (2014), os sistemas de reconhecimento funcionam bem em visualizações icônicas. Entretanto, ainda há muitos problemas para reconhecer objetos de outra forma, como por exemplo localizados em posições distintas, no fundo de uma imagem, parcialmente obstruídos ou em meio à desordem, o que ocorre na composição das cenas cotidianas reais. A identidade de muitos objetos só pode ser resolvida usando o contexto, devido ao tamanho pequeno ou à aparência ambígua na imagem (LIN *et al*., 2014). Para impulsionar a pesquisa em raciocínio contextual, imagens que retratam cenas em vez de objetos isolados, são necessárias.

3 MÉTODO

Este artigo se caracteriza como uma pesquisa qualitativa e exploratória, visando a aplicação de ML para resolução de um problema na área de Arquitetura e Urbanismo. Os métodos e técnicas descritos a seguir seguem a ordem cronológica dos acontecimentos. Para tanto, realizou a investigação por meio de PDCA (ANDRADE, 2003), desenvolvendo 5 ciclos (Figura 1).

Figura 1: Ciclos PDCA do desenvolvimento deste trabalho



Fonte: As autoras.

O primeiro ciclo resultou na seleção de um código de detecção de instâncias em imagens a partir do estudo seminal Ji *et al* (2019). Este ciclo resultou na identificação do código Mask R-CNN (HE et al., 2017). O segundo ciclo não será apresentado neste artigo, mas representou a tentativa de criar e treinar um *dataset*. No terceiro ciclo, dada a dificuldade vivenciada no ciclo anterior, optou-se pela utilização de *dataset* existente e adequadamente treinado para a identificação de objetos em cenas complexas. A solução encontrada foi explorar uma aplicação possível com o conjunto de dados original do Detectron, o COCO *Dataset*. No quarto, estudou-se os objetos que estavam contidos no *dataset* COCO, verificou-se a possibilidade de utilizar as próprias instâncias treinadas no modelo para detecção e quantificação de instâncias modais de trânsito em imagens urbanas, tais como carros, bicicletas, pedestres, ônibus etc. Foi definido que as imagens utilizadas seriam advindas de câmeras urbanas *open source*. Finalmente, no quinto ciclo foi realizada a aplicação do código no *dataset* e conferidas visualmente as detecções de modais extraídas.

4 RESULTADOS

Neste item, estão descritos os passos executados nesta aplicação.

4.1 Detectron/Mask R-CNN

Detectron[2] *Mask* R-CNN[3] foi o código utilizado neste estudo sendo desenvolvido com *Python* em Google *Collab*. Para a aplicação foi utilizado equipamento portátil com Processador Intel® Core™ i7-9750H CPU 2.60GHz 2.59 GHz 9a geração, com 16GB de RAM e placa de vídeo Nvidia GeForce RTX.

4.2 Caracterização do COCO *dataset*

Conforme Lin *et al*. (2014) apresentam, o d*ataset* COCO[4] tem 2.500.000 exemplos rotulados em 328.000 imagens. No *dataset* COCO, para cada categoria encontrada, as instâncias individuais são rotuladas, verificadas e segmentadas. Algumas categorias têm grande número de instâncias (parede: 20.213, janela: 16.080, cadeira: 7.971) enquanto a maioria tem um número relativamente modesto de instâncias (barco: 349, avião: 179, luminária de chão: 276). Dentro do conjunto de dados, cada categoria de objeto tem um número significativo de instâncias. O *dataset* apresenta 91 categorias de objetos comuns.

Dada a ambiguidade inerente da rotulagem, cada um desses estágios tem inúmeras vantagens e desvantagens (LIN *et al*., 2014). Segundo os autores, os conjuntos de dados relacionados ao reconhecimento de objetos podem ser divididos em três grupos: aqueles que abordam principalmente (I) a classificação de objetos nas imagens, (II) a detecção de objetos e (III) a rotulagem semântica de cenas. Dentro do conjunto de dados COCO, os autores garantem que cada categoria de objeto tenha um número significativo de instâncias (LIN *et al*., 2014).

4.3 CONJUNTO DE IMAGENS APLICADAS

As imagens utilizadas neste trabalho foram extraídas de câmeras urbanas *open source*, localizadas no município de São Paulo, um projeto da Prefeitura Municipal de São Paulo, denominado City Câmeras[5]. Trata-se de um projeto de ação integrada e tecnologia com intuito de aumentar a segurança urbana. Essas câmeras eram transmitidas 24 horas por dia, em uma plataforma aberta para a população. Segundo informações do projeto, o principal diferencial do programa é o uso de câmeras de segurança residenciais e pontos comerciais, que já se encontram distribuídas por São Paulo, além das câmeras dos órgãos públicos.

Foram definidos três grupos de amostras, descritos a seguir:

- **Grupo1:** Imagens de Comparação entre **horas de um dia inteiro**– total da amostragem: 24
- **Grupo 2:** Imagens de Comparação entre **dias da semana no mesmo horário durante o DIA** às 12:00, de 29/06/2021 a 05/07/2021 – total da amostragem: 7
- **Grupo 3:** Imagens de Comparação entre **dias da semana no mesmo horário durante a NOITE** às 22:00, de 29/06/2021 a 05/07/2021 – total da amostragem: 7

Foram selecionadas imagens de uma mesma câmera, localizada no Bairro do Itaim, na Rua Leopoldo Couto de Magalhães Junior, na altura da Rua Clodomiro Amazonas. Dentre as quatorze câmeras disponíveis no momento desta pesquisa, esta posição foi escolhida por se tratar de uma imagem com boa visibilidade do leito carroçável e dos passeios. As demais câmeras apresentavam obstruções e pouca visibilidade, ora do passeio, ora do leito carroçável.

4.4 DETECÇÃO DAS INSTÂNCIAS DE MODAIS DE TRÂNSITO

Apresentam-se a seguir os quadros extraídos do código (Figuras 2, 3 e 4 e Tabelas 1, 2 e 3), em formato *.csv*, que foram editadas no *Excel*. As imagens analisadas automaticamente pelo código também foram conferidas manualmente e, dessa forma, foram identificados os falsos negativos e falsos positivos. Há um número significativo de instâncias que o código não reconheceu nas imagens.

A partir da análise de todas as imagens do experimento, foi possível inferir:

- As instâncias que mais aparecem fazem parte das categorias pessoa (*person*) e carro (*car*);
- Existem categorias que aparecem na detecção que não dizem respeito a instâncias urbanas;

Cada categoria foi separada nos quadros em três colunas, sendo a primeira “D” de Detecção, que representa as instâncias detectadas automaticamente em cada imagem. A segunda coluna foi identificada por “FP”, falso positivo, e a terceira por “FN”, falso negativo, que são instâncias que foram detectadas manualmente. As cores que aparecem nas colunas representam o aumento da frequência. Quanto mais escura a cor, maior a frequência. Na tabulação das ocorrências de detecção de, quando se apresenta para uma instância somente da categoria D significa que a detecção foi precisa, não tendo apresentado falsos positivos ou negativos.

Figura 2: Comparação Grupo 1

DIA INTEIRO - 1 julho 2021 - Quinta-feira																																									
Horário	person			bike				car				motor cycle			truck			boat			fire hydrant			parking meter			bird			dog			cow			backpack			suitcase			chair			potted plant			bus*
	D	FP	FN	D				D	FP	FN		D	FP	FN	D			D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	D	FP	FN	FN
0:00	0			0	0			0			0	0			0	1	1	0			0			0			0			0			0			0			0			0			
1:00	0			0	0			0			0	0			0	0		0			0			0			0			0			0			0			0			0			
2:00	0			0	0			0			0	0			0	0		1	1		0			0			0			1	1		0			0			0			
3:00	0			0	0			0			0	0			0	0		0			0			0			0			0			0			0			0			
4:00	0			0	0			0			0	0			0	1	1	0			0			0			0			0			0			0			0			
5:00	0			0	0			0			0	0			0	0		0			0			0			0			0			0			0			0			
6:00	0			0	0			0			0	0			0	0		0			0			0			0			0			0			0			0			
7:00	1			0	0			0			0	1	1		0	0		0			0			0			0			0			0			0			0			
8:00	2			0	3			0			0	1	1		0	0		0			0			0			0			0			0			0			0			
9:00	2		2	0	5			0			1	0			0	0		0			0			0			0			0			0			0			0			
10:00	0		1	0	5		2	1			0	1	1		0	0		0			0			0			0			0			0		1	0			0			
11:00	2			0	7	2	3	1			0	0			0	0		0			0			0			0			0			0			0			0			1	1	
12:00	0			0	2		1	0		1	0	1	1		0	0		0			0			0			0			0			0			0			0			1	1	
13:00	6			0	8	1	2	0			0	0			0	0		0			0			0			0			0			0			0			0			0		
14:00	4		1	0	4		3	1			0	1	1		0	0		0			0			0			0			0			0			0			0			0		
15:00	1			1	4	1	4	0			0	1	1		0	0		0			0			0			1	1		0			1			0			1	1		0		
16:00	1	1		0	6	1	4	0		1	1	0			0	0		0			0			0			1	1		0			0	1	1	0			0			0		
17:00	4			0	5		1	3	2		0	0			0	0		0			0			0			1	1		0			0			0			0			1	1	
18:00	1			0	3		5	0			0	0			1	0		0			0			0			0			1	1		0			0			0			0		
19:00	5			0	2		1	0			0	0			0	0		0			0			0			0			0			0			0			0			0		
20:00	1			0	8	3	1	0			0	0			0	0		0			0			0			0			0			0			0			0			0		1
21:00	2	1		0	3	1	2	0		2	0	0			0	0		1	1		0			0			0			0			0			0			0			0		
22:00	2			0	2		3	0			0	0			0	0		0			0			0			0			0			0			1	1		0			0		
23:00	0		1	0	1		1	0			0	0			0	0		0			0			0			0			0			0			0			0			1	1	
23:59	0			0	1			0			0	0			0	0		0			0			0			0			0			0			0			0			0		
Total	34	2	5	1	69	9	33	6	2	4	2	6	6		1	2	2	2	2		3	3		1	1		2	2		2			1	1		1	1		4	4		1

*A instância "bus" não foi detectada em nenhum dos horários.

Fonte: As autoras.

Tabela 1: Relação entre quantidades de instâncias detectadas e reais com os falsos positivos e falsos negativos para o conjunto do Grupo 1

DIA INTEIRO - 1 julho 2021 - Quinta-feira		
TIPO	**QNT**	**%**
DETECÇÃO (D)	135	-
FALSOS POSITIVOS (FP)	35	26%
FALSOS NEGATIVOS (FN)	45	31%
REAL = D - FP + FN	**145**	-

Fonte: As autoras.

No grupo 1 (Figura 2 e tabela 1), foram detectadas pelo código 135 instâncias, quando na realidade deveriam ter sido detectadas 145. Os falsos positivos representam 26% dos detectados automaticamente (D) dentro desta amostra, e os falsos negativos representam 31% da contagem real. Não houveram falsos negativos e positivos das categorias “carro” (*car*) e “pessoa” (*person*) das 00:00 às 8:00.

Figura 3: Comparação Grupo 2

SEMANA 12:00																
dia	person		bicycle	car			motor cyde		truck	boat		dog	backpack		potted plant	
	D	FN	D	D	FP	FN	D	FN	D	D	FP	D	D	FN	D	FP
SEG	3		0	8		1	0		0	0		0	1		0	
TER	3		2	7	1		2		1	0		0	0		0	
QUA	0		0	5	1		0		0	1	1	0	0		2	2
QUI	0		0	2		1	0	1	0	1		0	0		1	1
SEX	1		0	8			0		0	2		0	0		3	3
SÁB	0	1	1	6			0	1	0	0		0	0	1	3	3
DOM	4		1	0		2	0		0	0		1	1		0	
Total	11	1	4	36	2	4	2	2	1	4	1	1	2	1	9	9

Fonte: As autoras.

Tabela 2: Relação entre quantidades de instâncias detectadas e reais com os falsos positivos e falsos negativos para o conjunto Grupo 2

SEMANA – 12:00 horas		
TIPO	**QNT**	**%**
DETECÇÃO (D)	70	-
FALSOS POSITIVOS (FP)	12	17%
FALSOS NEGATIVOS (FN)	8	12%
REAL = D - FP + FN	**66**	-

Fonte: As autoras.

No grupo 2 (Figura 3 e Tabela 2), foram detectadas pelo código 70 instâncias, quando na realidade deveriam ter sido detectadas 66. Os falsos positivos representam 17% dos detectados (D) dentro desta amostra, e os falsos negativos representam 12% da contagem real.

Figura 4: Comparação Grupo 3

SEMANA 22:00											
dia	person	car			dog		suitcase	potted plant		motor cyde	truck
	D	D	FP	FN	D	FP	D	D	FP	FN	FN
SEG	0	1			0		0	0			
TER	0	4	1		0		0	2	2		
QUA	4	4		1	1	1	0	0			1
QUI	2	2		2	0		1	0			
SEX	4	7	1		0		0	0			
SÁB	1	5		2	0		0	0		1	
DOM	0	5	1		0		0	0		2	
Total	11	28	3	5	1	1	1	2	2	3	1

Fonte: As autoras.

Tabela 3: Relação entre quantidades de instâncias detectadas e reais com os falsos positivos e falsos negativos para o conjunto do Grupo 3

SEMANA – 22:00 horas		
TIPO	**QNT**	**%**
DETECÇÃO (D)	43	-
FALSOS POSITIVOS (FP)	6	14%
FALSOS NEGATIVOS (FN)	9	20%
REAL = D - FP + FN	**46**	-

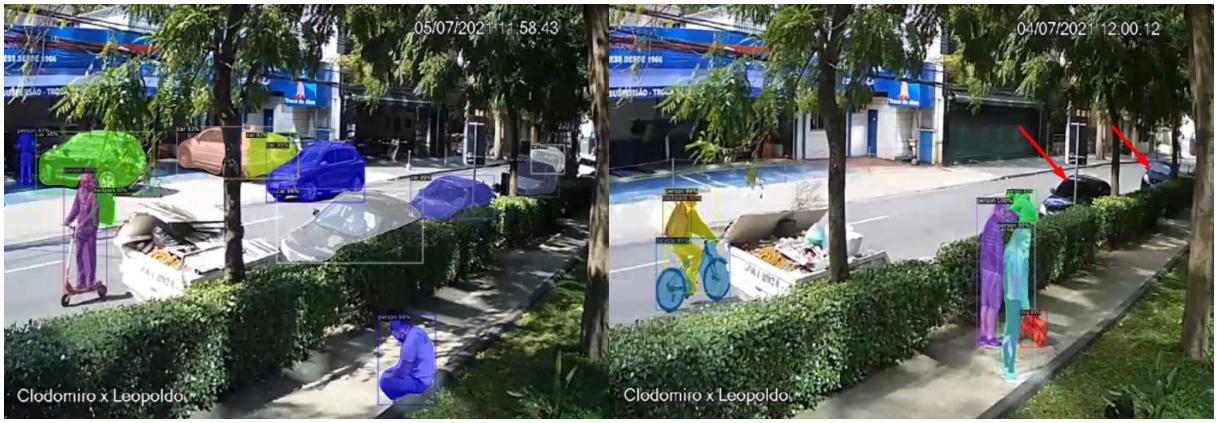
Fonte: As autoras.

No grupo 3 (Figura 4 e tabela 3), foram detectadas 43 instâncias pelo código, quando na realidade deveriam ter sido detectadas 46. Os falsos positivos representam 14% dos detectados (D) dentro desta amostra, e os falsos negativos representam 20% da contagem real. Para falsos positivos, a amostra da noite teve percentual menor entre os três conjuntos. E para falsos negativos a amostra do dia teve percentual menor entre os três conjuntos.

Na figura 5, à esquerda, verificam-se diversos objetos corretamente classificados durante o dia (12:00). Por outro lado, à direita, no mesmo horário, verifica-se dois falsos negativos apontados manualmente por setas vermelhas. Na figura 6 da esquerda, verifica-se diversos objetos corretamente classificados durante a noite

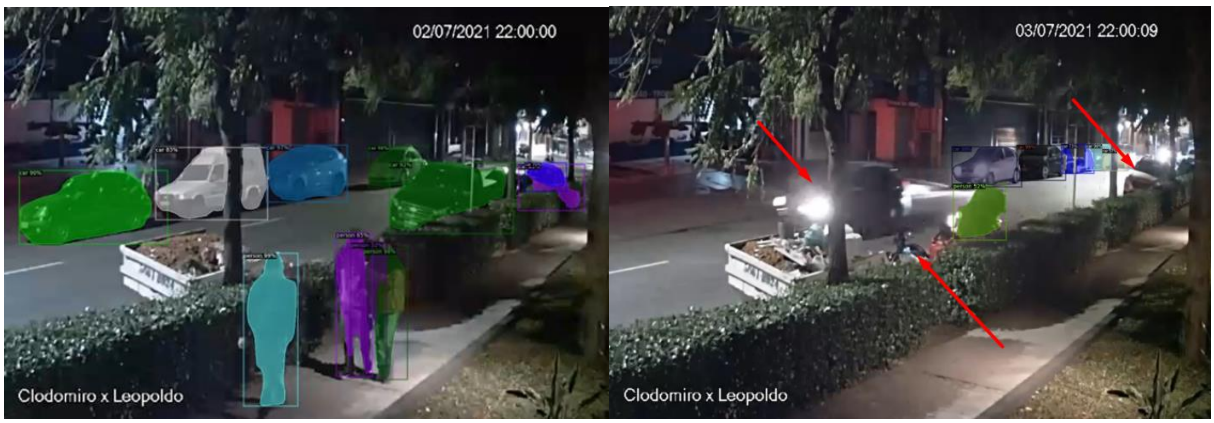
(22:00). Na figura 6 da direita, verificam-se dois carros e uma moto não detectados apontados por setas vermelhas.

Figura 5: Da esquerda para a direita, exemplo às 12 horas (DIA) segunda-feira e exemplo às 12 horas (DIA) domingo



Fonte: As autoras.

Figura 6: Da esquerda para a direita, exemplo às 22 horas (NOITE) sexta-feira e exemplo às 22 horas (NOITE) sábado



Fonte: As autoras.

Na figura 7 da esquerda, verificou-se um carro identificado erroneamente como mala (*suitcase*) indicado por seta vermelha. Na figura 7 da direita, a detecção classificou a caçamba como barco (*boat*)

Figura 7: Da esquerda para a direita, exemplo às 22 horas (NOITE) quinta-feira e exemplo às 7 horas (DIA) em 01/07/2021



Fonte: As autoras.

Na figura 7 da esquerda, o carro que se encontrava no mesmo local em diversos horários, no escuro, foi identificado apenas nas imagens das 20:00 e 24:00 horas. Houve situações, como a apresentada na figura 7 da direita, na qual uma instância é identificada e contada como se fossem duas instâncias, indicadas por setas vermelhas (cor azul e cor vermelha).

Figura 8: Da esquerda para a direita, exemplo às 24 horas e exemplo às 13 horas, ambas de 01/07/2021



Fonte: As autoras.

5 CONSIDERAÇÕES FINAIS

O estudo de caso foi realizado em uma rua de uso misto, comercial e residencial com a presença de bares e restaurantes. Considerando a detecção dos modais de trânsito urbanos pela perspectiva das horas do dia inteiro (Grupo 1), observa-se a presença de carros, pessoas, caminhões, bicicleta e ônibus. Os modais predominantes foram carros e pessoas, nos horários das 6:00 às 24:00 horas. Pode-se observar também um maior número de detecções dos modais pessoas e carros nos horários das 13:00 às 14:00 horas. Tem-se a impressão que o transporte público é quase inexistente no local, devido a uma única identificação manual, entretanto esta caracterização pode ter sido influenciada pela frequência da coleta de imagens (em horas cheias). Com relação à comparação entre dias da semana no horário diurno (Grupo 2) há predominância de carros de segunda a sábado, invertendo-se no domingo com presença maior de pessoas e bicicletas. No horário noturno (Grupo 3), o modal de transporte predominante por dia da semana é o carro seguido de pessoas de quarta a sexta. Observa-se também a presença de bicicletas e caminhões. O código Detectron/Mask R-CNN de detecção de instâncias em imagens associado ao *dataset* COCO foi capaz de identificar padrões de ocupação do espaço urbano coerentes com o contexto existente. Deve ser ponderado que as imagens capturadas foram tomadas no período da pandemia SARS-Cov-2, quando grande parte das atividades estavam suspensas ou com horários reduzidos.

A partir da aplicação exploratória descrita neste artigo, podemos listar algumas considerações, que podem auxiliar futuros trabalhos acerca da detecção de instâncias em imagens de câmeras urbanas através da aplicação do modelo Mask R-CNN pré-treinado com COCO *dataset.* É requerido conhecimento na linguagem de programação Python, a falta de competência neste quesito pode ser uma barreira sendo requerido ter apoio específico. Sugere-se complementar o código de forma a restringir as categorias a instâncias urbanas, como por exemplo remover vasos de plantas (*potted plants*) malas (*suitcases*), barcos (*boats*) entre outros. Não utilizar imagens noturnas, pois identificou-se que neste horário o código encontrou menos classes de objetos. Observou-se que a resolução das imagens não se apresentou como uma barreira para detecção pelo código, o que também foi observado em Kang *et al*. (2018).

Os próximos passos sugeridos para realizar um experimento completo seriam os seguintes: análises estatísticas (clusterização), medição da acurácia, métricas de classificação e uso de vídeos em vez de imagens estáticas. Sugere-se que o *dataset* utilizado seja revisado e complementado com foco em cenas urbanas complexas, corroborando com a observação de Lateef e Ruichek (2019) sobre *dataset* disponíveis.

A aplicação exploratória realizada cumpriu com a expectativa de fazer os autores experimentarem o ambiente de programação. O maior limitador para o desenvolvimento de uma aplicação de DL é o planejamento do tempo para cada atividade, uma vez que não haviam parâmetros para cada atividade a ser executada. A fim

de desenvolver pesquisas aplicadas robustas sobre diagnósticos de áreas urbanas, recomenda-se a aproximação de pesquisadores da ciência da computação aos da Arquitetura e Urbanismo.

AGRADECIMENTOS

O presente trabalho está sendo realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).

[1] Disciplina AQ 101: Tópicos Especiais II – Aprendizado de Máquina aplicado à Arquitetura, oferecida em junho de 2021, tem como no Programa de Pós-graduação em Arquitetura, Tecnologia e Cidade da Universidade Estadual de Campinas.

[2] https://github.com/facebookresearch/Detectron

[3] https://arxiv.org/abs/1703.06870

[4] https://arxiv.org/abs/1405.0312

[5] https://www.citycameras.prefeitura.sp.gov.br/ (acesso em: 08/07/2021). Desde junho/2022 o projeto está suspenso e fora de ar, para atualização de tecnologia da plataforma.

REFERÊNCIAS

ALPAYDIN, Ethem. **Machine Learning**. 3. ed. Massachusetts: The MIT Press, 2016.

ANDRADE, Fábio Felippe de. O método de melhorias PDCA. 2003. **Dissertação** (Mestrado em Engenharia de Construção Civil e Urbana) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2003. doi:https://doi.org/10.11606/D.3.2003.tde-04092003-150859

ARNOLD, Ludovic *et al*. An Introduction to Deep Learning. In: 18th European Symposium On Artificial Neural Networks, Computational Intelligence and Machine Learning - ESANN 2011 (Bruges, Belgium). **Anais…** ESANN, 2011, Bruges, Belgium. Disponível em: https://hal.science/hal-01352061/. Acesso em: 09 de junho de 2023.

CARRARA, Fabio *et al*. Adversarial image detection in deep neural networks. **Multimedia Tools and Applications**, v. 78, n. 3, p. 2815-2835, 2019. doi:https://doi.org/10.1007/s11042-018-5853-4

GAO, Lianli *et al*. Unsupervised urban scene segmentation via domain adaptation. **Neurocomputing**, v. 406, p. 295-301, 2020. doi:https://doi.org/10.1016/j.neucom.2020.01.117

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Massachusetts: The MIT press, 2016.

HALDAR, Malay *et al*. Applying deep learning to Airbnb search. In: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. **Proceedings…** SIGKDD, 2019. p. 1927-1935. doi:https://doi.org/10.1145/3292500.3330658

HE, Kaiming. *et al*. Mask R-CNN. In: IEEE international conference on computer vision. **Proceedings…** ICCV, 2017. p. 2961-2969. Disponível em: openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf. Acesso em: 09 de junho de 2023.

JI, Shunping *et al.* Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples. **Remote Sensing**, v. 11, n. 11, p. 1343, 2019. doi:https://doi.org/10.3390/rs11111343

KANG, Jian *et al.* Building instance classification using street view images. **ISPRS journal of photogrammetry and remote sensing**, v. 145, p. 44-59, 2018. doi:https://doi.org/10.1016/j.isprsjprs.2018.02.006

LATEEF, F.; RUICHEK, Y. Survey on semantic segmentation using deep learning techniques. **Neurocomputing**, v. 338, p. 321-348, 2019. doi:https://doi.org/10.1016/j.neucom.2019.02.003

LIN, Tsung-Yi *et al*. Microsoft coco: Common objects in context. In: European conference on computer vision. **Proceedings…** Springer, Cham, 2014. p. 740-755. doi:/https://doi.org/10.1007/978-3-319-10602-1_48

LOU, Lu *et al.* Detecting and counting the moving vehicles using mask R-CNN. In: 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS). **Proceedings…** IEEE, 2019. p. 987-992. doi:https://doi.org/10.1109/DDCLS.2019.8908877

RUDIO, Franz Victor. **Introdução ao projeto de pesquisa científica**. Rio de Janeiro: Vozes, 2003.

SHAWAHNA, A.; SAIT, S. M.; EL-MALEH, A. FPGA-based accelerators of deep learning networks for learning and classification: A review. **IEEE Access**, v. 7, p. 7823-7859, 2018. doi:https://doi.org/10.1109/ACCESS.2018.2890150

ZHAO, Wenzhi *et al*. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM). **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 151, p. 237-250, 2019. doi:https://doi.org/10.1016/j.isprsjprs.2019.03.019