

UNIVERSIDADE FEDERAL DE ALAGOAS
CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA CIVIL

ARTHUR RODRIGUES CARDOSO DE FARIAS

Aplicação de algoritmos computacionais para estudo de tráfego utilizando técnicas de Visão Computacional e *Deep Learning*

Maceió

2022

ARTHUR RODRIGUES CARDOSO DE FARIAS

Aplicação de algoritmos computacionais para estudo de tráfego utilizando técnicas de Visão Computacional e *Deep Learning*

Trabalho de Conclusão de Curso apresentado ao colegiado do curso de Engenharia Civil da Universidade Federal de Alagoas, como requisito parcial para obtenção do título de Engenheiro Civil.

Orientador: Prof^a. MSc. Aline Calheiros Espíndola

Maceió
2022

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

F224a Farias, Arthur Rodrigues Cardoso de.
Aplicação de algoritmos computacionais para estudo de tráfego utilizando técnicas de visão computacional e *deep learning* / Arthur Rodrigues Cardoso de Farias. – Maceió, 2022.
40 f. : il., grafs. e tabs. color.

Orientadora: Aline Calheiros Espíndola.
Monografia (Trabalho de conclusão de curso em Engenharia Civil) –
Universidade Federal de Alagoas. Centro de Tecnologia. Maceió, 2022.

Bibliografia: f. 38-40.

1. Contagem de veículos. 2. Estudo de tráfego. 3. Aprendizado profundo. 4. Processamento de imagens. 5. Visão computacional. 6. *You Only Look Once* (Algoritmo). I. Título.

CDU: 629.3:004.81:159.953.5

FOLHA DE APROVAÇÃO

ARTHUR RODRIGUES CARDOSO DE FARIAS

Aplicação de algoritmos computacionais para estudo de tráfego utilizando técnicas de Visão Computacional e *Deep Learning*.

Trabalho de Conclusão de Curso apresentado ao colegiado do curso de Engenharia Civil da Universidade Federal de Alagoas, como requisito parcial para obtenção do título de Engenheiro Civil.

Aprovado em 14 de dezembro de 2022.

Documento assinado digitalmente
 ALINE CALHEIROS ESPINDOLA
Data: 23/12/2022 14:39:01-0300
Verifique em <https://verificador.iti.br>

Prof^a. MSc. Aline Calheiros Espíndola (Orientador)
Universidade Federal de Alagoas

Banca Examinadora:

Documento assinado digitalmente
 EWERTON AMORIM DE OLIVEIRA
Data: 26/12/2022 15:54:19-0300
Verifique em <https://verificador.iti.br>

Eng. Ewerton Amorim de Oliveira (Examinador Externo)

Documento assinado digitalmente
 LUCIANA DA SILVA NASCIMENTO
Data: 26/12/2022 13:14:24-0300
Verifique em <https://verificador.iti.br>

Eng. Luciana da Silva Nascimento (Examinador Externo)

AGRADECIMENTOS

Agradeço primeiramente a Deus pelo discernimento e força que me concedeu durante todo o processo de elaboração deste trabalho. Gostaria de agradecer à minha família pelo constante apoio e carinho, especialmente à minha mãe Arlete e a minha irmã Raissíinha, que foram a minha principal rede de apoio durante toda a minha graduação.

Agradeço também a minha orientadora, Aline Espíndola, por todo incentivo e pelo tempo dedicado à revisão deste trabalho. Aproveitando o momento, gostaria de agradecer também a SMTT pela disponibilidade em compartilhar as filmagens que foram utilizadas nesse trabalho.

Sou muito grato também à professora Karoline Melo, que sempre me incentivou a arriscar em novos projetos e utilizar a universidade como um laboratório de experiência e criatividade. Posso dizer com tranquilidade que foi a minha mainha dentro da UFAL. Serei sempre grato pelo privilégio de ganhar uma amiga muito especial na minha vida.

Aos meus amigos do PET Engenharia Civil, vocês foram a minha família dentro da UFAL. Guardarei no fundo do meu coração para sempre todos os momentos que compartilhamos, todo o apoio mútuo que trocamos, risadas, brigas, caronas, lanches, viagens e todas as reuniões naquela salinha do PET. Obrigado por me ensinarem tanto.

Por fim, não poderia jamais de deixar de agradecer todos os meus amigos que de alguma forma contribuíram para a concretização desse sonho, seja com palavras de incentivo ou por estarem comigo lado a lado nos momentos de muita luta.

RESUMO

FARIAS, A.R.C. **Aplicação de algoritmos computacionais para estudo de tráfego utilizando técnicas de Visão Computacional e *Deep Learning***. Trabalho de Conclusão de Curso (Graduação – Engenharia Civil) – Universidade Federal de Alagoas, Maceió.

A combinação entre o processamento de imagens e a mineração de dados tem se mostrado cada vez mais eficaz na resolução de problemas do mundo real, sendo estes sistemas utilizados em larga escala em diversos campos de estudo. É importante frisar que um dos principais fatores que influenciam no dimensionamento dos pavimentos é o tráfego que solicitará determinada via durante sua vida útil de serviço. Dessa forma, a detecção, o rastreamento, a classificação e a contagem são parâmetros de grande relevância para monitoramento de tráfego e dimensionamento de rodovias. Diante dessa perspectiva, o presente trabalho, tem como objetivo desenvolver um algoritmo computacional capaz de efetuar a detecção e identificação dos veículos de projeto. Os algoritmos computacionais a serem implementados terão como base o algoritmo *You Only Look Once* (YOLO), através da linguagem computacional *Python*, e os resultados experimentais serão demonstrados a partir de vídeos do tráfego urbano de Maceió-AL fornecidos pela Superintendência Municipal de Transportes e Trânsito (SMTT).

Palavras-Chave: Contagem de veículos, Estudo de Tráfego, *Deep Learning*, Processamento de imagens, Visão computacional, YOLO.

ABSTRACT

The combination of image processing and data mining has proven to be increasingly effective in solving real-world problems, with these systems being used on a large scale in several fields of study. Given this scenario, one of the main factors that influence the dimensioning of pavements is the traffic it requests during its service life. Thus, the detection and the classification are parameters of great relevance for traffic monitoring and road sizing. Given this perspective, the present work aims to develop a prediction model using neural networks, capable of controlling the detection and identification of vehicles defined by the DNIT project. In this work, a model was trained with five classes and had the following evaluation metrics as the final result: "Precision = 86%", "Recall = 84.37%", "mAP = 51%" and "F1-Score = 75.37%". The computational algorithms were based on the You Only Look Once (YOLO) method, and the experimental results were generated from images of urban traffic in Maceió-AL provided by the Municipal Superintendence of Transport and Traffic. Although the final model is not yet ready to be applied in traffic studies, the present work shows the potential that Computer Vision and Deep Learning techniques have for the transport area.

Keywords: Vehicle counting, Traffic Study, Deep Learning, Image processing, Computer vision, YOLO.

LISTA DE FIGURAS

Figura 1 - Classificação do estado geral do pavimento das rodovias.	8
Figura 2 - Principais tipos de eixos rodoviários	12
Figura 3 - Valores de giro do veículo de projeto	13
Figura 4 - Possíveis configurações disponíveis para o tubo automático no dispositivo TimeMark III	15
Figura 5 - Disposição de lâminas de sensores de laço indutivo.	15
Figura 6 - Modelo de um neurônio artificial.	17
Figura 7 - Método de multicamadas para soluções não lineares.	17
Figura 8 - Exemplo da aplicação de bounding boxes.	18
Figura 9 - Exemplo de aplicação e funcionamento do YOLO.	19
Figura 10 – detecção de objetos de acordo com as métricas do dataset MS COCO.	20
Figura 11 -Detecção de objetos de acordo com as métricas do dataset.	21
Figura 12 - Localização da câmera de controle de tráfego, representada pelo ponto em vermelho.....	22
Figura 13 - Câmera de controle de tráfego em frente à entrada do bairro Benedito Bentes e próxima ao Shopping Pátio Maceió.	22
Figura 14 - Filmagens das câmeras de monitoramento para composição do banco de dados	23
Figura 15 - Localização do cruzamento em estudo e dos sentidos de tráfego para cada câmera de monitoramento.	24
Figura 16 - Exemplo de rotulação desenvolvida no programa labeling	25
Figura 17 - Ferramenta de catalogação tem como saída um arquivo no formato de texto	25
Figura 18 – Representação do cálculo da métrica Intersection over Union.....	28
Figura 19 - Diagrama de fluxo para construção do banco de dados e treinamento da rede neural.....	30
Figura 20 - Predição do primeiro modelo treinado.....	33
Figura 21 – Novas rotulações para a construção do segundo dataset.	34
Figura 22- Predição do segundo modelo treinado.	36

LISTA DE TABELAS

Tabela 1- Frota total de veículos por regiões brasileiras.....	7
Tabela 2- Principais dimensões básicas dos veículos de projeto.....	11
Tabela 3 - Limites legais para cada tipo de eixo de veículo.....	12
Tabela 4 - diferença entre TP, TN, FP e FN.....	27
Tabela 5 – Banco de dados customizado.	31
Tabela 6 - Matriz de confusão do primeiro treinamento.....	32
Tabela 7 - Cálculos das métricas de avaliação.	33
Tabela 8 - Segundo banco de dados customizado.	34
Tabela 9 - Matriz de confusão do segundo treinamento.	35
Tabela 10 - Resultado das métricas de avaliação do segundo treinamento	35

SUMÁRIO

1 INTRODUÇÃO	7
1.1 COMENTÁRIOS INICIAIS.....	7
1.2 JUSTIFICATIVA.....	8
1.3 OBJETIVOS	9
1.3.1 Objetivo geral.....	9
1.3.2 Objetivos específicos.....	9
1.4 DELIMITAÇÃO DO TRABALHO	10
2 REFERENCIAL TEÓRICO	10
2.1 ESTUDO DE TRÁFEGO.....	10
2.2 VEÍCULOS DE PROJETO	10
2.2.1 Tipos de Veículos de Projeto.....	11
2.3 CONTAGEM E CLASSIFICAÇÃO DE VEÍCULOS	13
2.3.1 Contagens Globais.....	13
2.3.2 Contagens Direcionais	14
2.3.3 Contagens Classificadoras	14
2.4 MÉTODOS DE CONTAGEM.....	14
2.4.1 Contagens manuais.....	14
2.4.2 Contagens automáticas por meio de sensores	14
2.4.3 Automatização através do aprendizado de máquina.....	15
3 MÉTODO	21
3.1 CONSTRUÇÃO DO BANCO DE DADOS/IMAGENS - <i>DATASET</i>	21
3.2 PRÉ-PROCESSAMENTO DE IMAGENS	24
3.2.1 Redimensionamento das imagens	24
3.3 ROTULAÇÃO DE IMAGENS	24
3.4 TREINAMENTO E VALIDAÇÃO DA REDE NEURAL - YOLO	25

3.4.1 Métricas de avaliação da detecção e classificação	26
3.4.2 Procedimento para criação do banco de dados e treinamento da rede neural.....	30
4 RESULTADOS E DISCUSSÕES.....	30
4.1 BANCO DE DADOS	31
4.2 DETECÇÃO DE VEÍCULOS	31
4.3 BANCO DE DADOS PARA O SEGUNDO TREINAMENTO.....	34
4.4 DETECÇÃO DE VEÍCULOS DO SEGUNDO MODELO TREINADO	34
5 CONCLUSÃO	36
REFERÊNCIAS	38

1 INTRODUÇÃO

1.1 COMENTÁRIOS INICIAIS

A maior parte do transporte de cargas produzidas e consumidas no Brasil (61,1%) é realizada pelo modo rodoviário, seguido dos modos ferroviário (20,7%), aquaviário (13,6%), dutoviário (4,2%) e aéreo (0,4%) (CNT, 2021).

Dessa forma, o modal de transporte rodoviário tem grande destaque na matriz de transporte brasileira, com importante contribuição para o crescimento do país e desenvolvimento econômico (CRISTINE, 2005). No entanto, também possui alguns problemas que prejudicam a sua operação e que conseqüentemente atrasam o desenvolvimento do país, sendo então necessários investimentos estratégicos na melhoria dos métodos construtivos e da manutenção das rodovias.

Conforme apresentado na Tabela 1, o Brasil tem cerca de 108 mil veículos registrados, e possui um crescimento de 66,5% entre 2010 e 2020. A região que tem maior número de veículos é o sudeste, mas as regiões que tiveram maiores crescimentos de frota foram as regiões norte e nordeste.

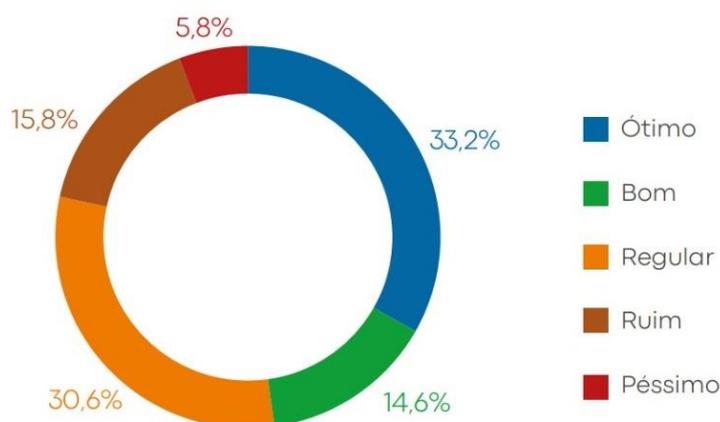
Tabela 1- Frota total de veículos por regiões brasileiras

Regiões	2010	2020	Crescimento 2010/2020 (%)
Norte	2.849.014	5.785.095	103,1%
Nordeste	9.469.880	18.841.281	99%
Sudeste	33.296.148	51.998.434	56,2%
Sul	13.383.052	21.156.154	58,1%
Centro-Oeste	5.819.880	10.167.407	74,7%
Brasil	64.817.974	107.948.371	66,5%

Fonte: CNT (2021).

Conforme a Figura 1, a Pesquisa CNT de Rodovias constatou que 52,2% das rodovias apresentaram algum tipo de problema no pavimento, evidenciando a baixa qualidade das rodovias brasileiras que apresentam um péssimo estado de conservação e carecem de restaurações, devido à presença de não-conformidades (deformações, buracos, trincas e remendos) que interferem na qualidade da pista de rolamento (CNT, 2021).

Figura 1 - Classificação do estado geral do pavimento das rodovias.



Fonte: Adaptado de CNT (2021).

Dada a situação das rodovias brasileiras, estudos na área de dimensionamento de pavimentos asfálticos e dos materiais que compõem sua estrutura tem se intensificado ao longo dos anos com o objetivo de desenvolver pavimentos que atendam as requisições de projeto, como tempo, qualidade e durabilidade (FILHO, *et al.*, 2018).

Devido a fatores como diferença de carga, composição de eixos, rodas e velocidades variadas, veículos distintos degradam de forma diferente os pavimentos. O entendimento e a quantificação dessas características que influenciam na degradação do pavimento são de fundamental importância para o desenvolvimento de métodos de dimensionamentos que levem em consideração de forma assertiva a influência do tráfego.

1.2 JUSTIFICATIVA

A qualidade das rodovias brasileiras é um fator que impacta a segurança, satisfação e economia dos usuários, sendo assim fundamental garantir que os pavimentos sejam corretamente dimensionados e conservados (FILHO, *et al.*, 2018).

Para a escolha do veículo de projeto, é necessário considerar a composição do tráfego que utilizará a infraestrutura. Essa composição é obtida por meio de contagens e de projeções futuras considerando um possível desenvolvimento do local de estudo (DNIT, 2006).

Devido a necessidade da determinação do valor 'N' no dimensionamento, é preciso o estabelecimento de procedimentos e/ou tecnologias que auxiliem nesse processo (DNIT, 2006). Normalmente, há duas formas de medir e verificar o comportamento do tráfego: através da contagem manual de veículos realizada por um operador humano ou por meio da instalação de sensores na pista de rolamento.

O procedimento de contagem manual de veículos, apesar de ser um procedimento simples, não consegue obter um resultado que represente o fluxo de veículos durante longos períodos. Com isso, acontecem situações de ausência total de informações sobre o tráfego ou a utilização de dados não condizentes com a realidade para estudos de projetos viários e de mobilidade.

Já a utilização de sensores requer um investimento financeiro considerável e possibilita a medição somente em pontos específicos. Apesar de serem extremamente precisos, esses sensores raramente apresentam a possibilidade de analisar diferentes parâmetros de tráfego de forma paralela.

Em contrapartida, os sistemas de monitoramento de vídeo baseados em processamento de imagens podem oferecer uma série de vantagens além das contagens de veículos, como uma possibilidade maior na criação de parâmetros de tráfego, como por exemplo, a quantidade de pedestres que atravessam determinada pista de rolamento (SAID *et al.*, 2011), qualidade das sinalizações verticais, classificações de veículos, mudanças de posição de veículos entre as faixas da pista de rolamento, identificação da presença de capacetes em motociclistas para prevenção de acidentes viários e etc (SILVA *et al.*, 2016).

Portanto, através da combinação entre o processamento de imagem e as técnicas de aprendizado de máquina, é possível obter o conhecimento do real comportamento do tráfego em uma rodovia, para que seja realizado um dimensionamento mais assertivo da infraestrutura.

1.3 OBJETIVOS

1.3.1 Objetivo geral

O objetivo deste estudo é desenvolver um algoritmo computacional capaz de efetuar a detecção e identificação dos veículos de projeto.

1.3.2 Objetivos específicos

Para alcançar o objetivo geral, tem-se como objetivos específicos:

- Construir um banco de dados personalizado onde as rotulações estejam de acordo com os veículos de projeto definidos pelo DNIT;
- Analisar a qualidade da rede neural desenvolvida em trecho teste na cidade de Maceió-AL;
- Analisar a viabilidade técnica para aplicação do modelo.

1.4 DELIMITAÇÃO DO TRABALHO

Toda as análises de processamento digital de imagens e aprendizagem profunda foram delimitadas pela *YOLO: Real-Time Object Detection*, através da linguagem de programação *Python*.

É importante destacar que os resultados experimentais foram demonstrados apenas a partir de vídeos do tráfego urbano de Maceió-AL, fornecidos pela Superintendência Municipal de Transportes e Trânsito (SMTT). Devido à limitação de espaço e capacidade de processamento dos computadores disponíveis para as análises, as gravações foram cedidas apenas nos horários de pico de um ponto crítico da cidade, durante o intervalo de uma semana.

Em relação à infraestrutura computacional, o treinamento dos protótipos criados para esse projeto foram executados no ambiente *Google Colaboratory*, sendo que as configurações utilizadas no ambiente para a execução dos modelos foram as padrões gratuitas disponíveis do Colab. Entretanto, para a rotulação do sistema de detecção e classificação, o processamento foi realizado por meio de um *notebook* alimentado por um processador *CPU Intel Core Duo i5-8265U(1,6 GHZ)*, *8 GB de RAM*, armazenamento *256GB SSD PCIe* e placa de vídeo dedicada *NVIDIA Geforce MX110 2GB*.

2 REFERENCIAL TEÓRICO

Para uma melhor compreensão do trabalho proposto, será apresentada informações que auxiliarão no embasamento teórico da temática desenvolvida.

2.1 ESTUDO DE TRÁFEGO

O estudo de tráfego permite a determinação quantitativa da capacidade das vias e, em consequência, o estabelecimento dos meios construtivos necessários à melhoria da circulação ou das características de seu projeto. Através dessas pesquisas, que fornecem os dados sobre o tráfego atual, e através do conhecimento da forma de geração e distribuição desse tráfego, obtém-se o prognóstico das necessidades de circulação no futuro, dado essencial para o planejamento de transporte (DNIT, 2006).

2.2 VEÍCULOS DE PROJETO

Pode-se definir Veículo de Projeto como o veículo teórico de uma certa categoria, cujas características físicas e operacionais influenciam diversos aspectos do dimensionamento geométrico da via. As características físicas e operacionais do veículo de projeto influenciam no dimensionamento da via de várias formas, que são apresentadas na sequência a seguir.

2.2.1 Tipos de Veículos de Projeto

O tráfego é responsável por grande parte dos danos causados ao pavimento ao longo de sua vida útil, devido a fatores como o volume de veículos, características de distribuição de eixos e rodas e das cargas aplicadas. De acordo com o DNIT (2006), há cinco grupos básicos de veículos de projeto:

- Veículo VP: Nessa categoria, se enquadram os veículos leves incluindo automóveis, minivans, vans, utilitários e similares, com no máximo dois eixos, destinados ao transporte de passageiros e pequenas quantidades de carga.
- Veículo CO: É o veículo que representa os veículos comerciais rígidos (não articulados) compostos de unidade tratora simples. Os veículos comerciais rígidos abrangem caminhões e ônibus convencionais, normalmente de 2 eixos e 6 rodas.
- Veículos O: É o veículo que representa os veículos comerciais rígidos de grandes dimensões, como os ônibus de longo percurso e de turismo, e caminhões longos (Ex. Ônibus e caminhões com eixo tandem).
- Veículo SR: É o veículo que representa os veículos comerciais articulados, compostos de unidade tratora simples e semi-reboque.
- Veículo RE: Veículos comerciais com reboque. É composto de uma unidade tratora simples, um semi-reboque e um reboque, frequentemente conhecido como bitrem. Seu comprimento é o máximo permitido pela legislação brasileira.

A Tabela 2 resume as principais dimensões básicas dos veículos de projeto, que são recomendados para utilização nos projetos geométricos de rodovias no Brasil.

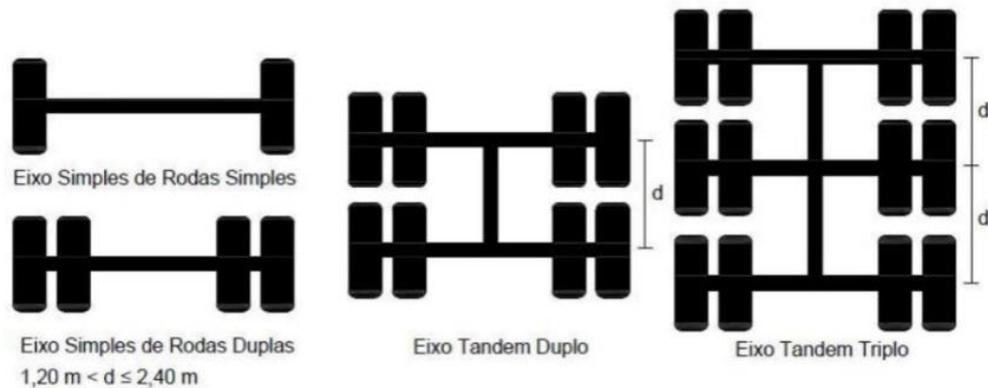
Tabela 2- Principais dimensões básicas dos veículos de projeto.

Características	Veículos leves (VP)	Caminhões e ônibus convencionais (CO)	Caminhões e ônibus longos (O)	Semirreboques (SR)	Reboques (RE)
Largura total (m)	2,1	2,6	2,6	2,6	2,6
Comprimento total (m)	5,8	9,1	12,2	16,8	19,8
Raio min. da roda externa dianteira (m)	7,3	12,8	12,8	13,7	13,7
Raio min. da roda interna traseira (m)	4,7	8,7	7,1	6,0	6,9

Fonte: Adaptado de DNIT (2020).

Conforme a Figura 2 a seguir, cada um desses tipos de veículos possui diferentes distribuições e conjuntos de eixos e rodas, sendo classificados isoladamente em quatro eixos: Eixo Simples de Rodas Simples (ESRS), Eixo Simples de Rodas Duplas (ESRD), Eixo tandem duplo (ETD) e Eixo Tandem Triplo (ETT). E para cada eixo existe um limite legal de carga máxima (Tabela 3).

Figura 2 - Principais tipos de eixos rodoviários.



Fonte: BOSSO (2018).

Tabela 3 - Limites legais para cada tipo de eixo de veículo.

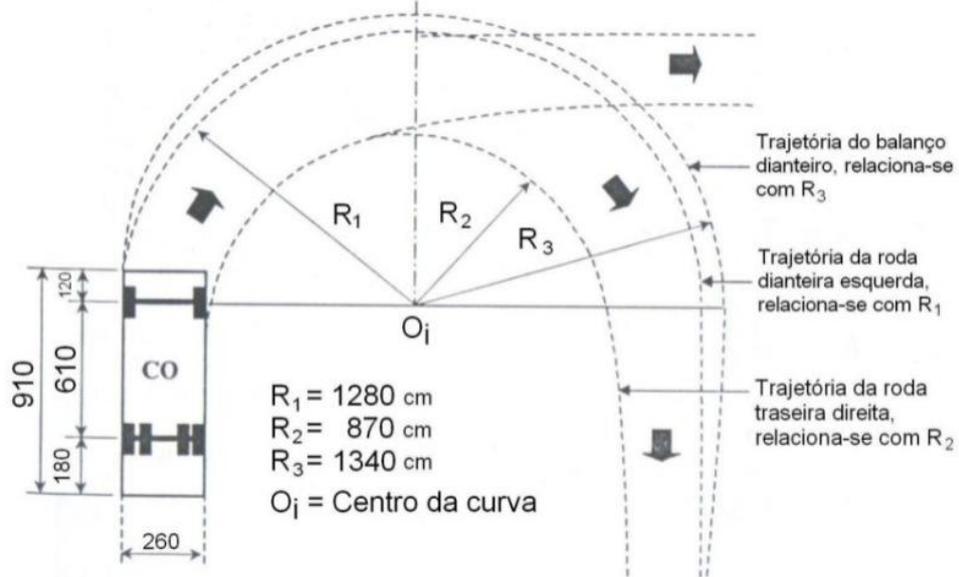
Limites Legais Resolução CONTRAN 12/98	Cargas Máximas (t)
Eixo Simples de Rodas Simples (ESRS)	6
Eixo Simples de Rodas Duplas (ESRD)	10
Eixo Tandem Duplo (ETD)	17
Eixo Tandem Triplo (ETT)	25,5
Peso Bruto Total (PBT)	45

Fonte: CONTRAN (1998).

Dada a importância do transporte rodoviário no Brasil (CNT, 2018), o veículo típico utilizado em projetos de engenharia de tráfego, de geometria de vias e de pavimentos é do tipo CO, englobando os ônibus e os caminhões convencionais e/ou veículos de carga, sendo que os caminhões leves e médios representam 64% da frota de caminhões estimada pelo DENATRAN.

A Figura 3 ilustra os menores valores de giro do veículo de projeto CO, os quais são importantes para dimensionamento de retorno para pista de tráfego contrário em vias expressas (ou de duas pistas).

Figura 3 - Valores de giro do veículo de projeto



Fonte: DNIT (2006).

Em relação ao peso, segundo a legislação de trânsito brasileira, os veículos registrados e licenciados são classificados em: leves (ciclomotor, motoneta, motocicleta, triciclo, quadriciclo, automóvel, utilitário, caminhonete e camioneta) e pesados (ônibus, micro-ônibus, caminhão, caminhão-trator, trator de rodas, trator misto, chassi-plataforma, motor-casa, reboque ou semirreboque e suas combinações) (DENATRAN, 2010).

Os veículos leves causam muito menos impacto ao pavimento e menos problemas no tráfego quando comparados com os pesados. Os veículos pesados/comerciais desgastam o pavimento com maior facilidade e em menor intervalo de tempo, além de possuírem um desempenho pior, com perda de velocidade em aclives acentuados, causando lentidão no tráfego. Dessa forma, pode-se dizer que o caminhão (tipo CO) é um fator determinante em qualquer projeto relacionado a dimensionamento de rodovias (CNT, 2018).

2.3 CONTAGEM E CLASSIFICAÇÃO DE VEÍCULOS

Os dados de tráfego são geralmente obtidos nas fases de Planos Diretores e Estudos de Viabilidade e as contagens volumétricas para estudo podem ser categorizadas em Contagens Globais, Contagens Direcionais e Contagens Classificatórias (DNIT, 2006).

2.3.1 Contagens Globais

São aquelas em que é registrado o número de veículos que circulam por um trecho de via, independentemente de seu sentido, grupando-os geralmente pelas suas diversas classes. São empregadas para o cálculo de volumes diários, preparação de mapas de fluxo e determinação de tendências do tráfego (DNIT, 2006).

2.3.2 Contagens Direcionais

São aquelas em que é registrado o número de veículos por sentido do fluxo e são empregadas para cálculos de capacidade, determinação de intervalos de sinais, justificação de controles de trânsito, estudos de acidentes, previsão de faixas adicionais em rampas ascendentes e etc. (DNIT, 2006).

2.3.3 Contagens Classificadoras

Nessas contagens são registrados os volumes para os vários tipos ou classes de veículos. São empregadas para o dimensionamento estrutural e projeto geométrico de rodovias e interseções, cálculo de capacidade, cálculo de benefícios aos usuários e determinação dos fatores de correção para as contagens mecânicas (DNIT, 2006).

2.4 MÉTODOS DE CONTAGEM

2.4.1 Contagens manuais

O método de contagem manual é realizado por mão de obra humana, com o auxílio de fichas e contadores manuais, que detectam os veículos passantes dentro de um determinado intervalo de tempo (DNIT, 2006). Este método tem a vantagem de possuir baixos custos, flexibilidade para a mudança de locais durante períodos curtos e facilidade na operação.

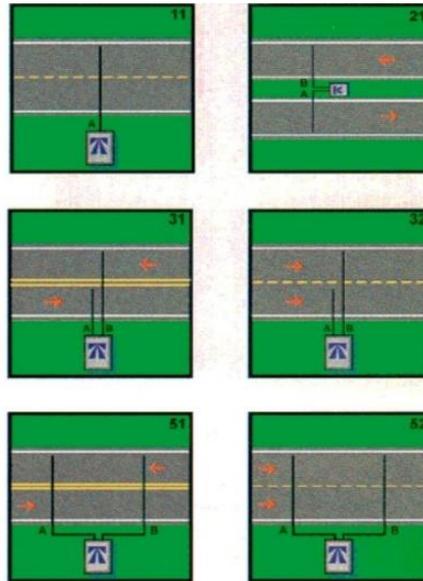
Entretanto, para esse método é preciso que pessoas sejam alocadas durante dias para a execução de um trabalho bastante repetitivo, além de dificultar um levantamento por longos períodos. Sendo assim, torna-se um impasse o conhecimento do comportamento do tráfego em períodos com variação devido à sazonalidade.

2.4.2 Contagens automáticas por meio de sensores

Diferentes tipos de contadores podem realizar contagens automáticas, por exemplo, tubos pneumáticos ou dispositivos magnéticos, sonoros, radar ou células fotoelétricas. Esses contadores geralmente são formados por dois componentes básicos: um detector de passagem de veículos e uma unidade acumuladora de dados (DNIT, 2006).

Os contadores automáticos possuem a desvantagem do alto custo e a exposição a roubos e vandalismos e a vantagem de possuírem uma alta precisão e possibilitarem a contagem tanto por tempos limitados quanto ilimitados. Alguns exemplos de contadores automáticos podem ser visualizados na Figura 4 e na Figura 5.

Figura 4 - Possíveis configurações disponíveis para o tubo automático no dispositivo *TimeMark III*.



Fonte: MIMBELA, *et al* (2000).

Figura 5 - Disposição de lâminas de sensores de laço indutivo.



Fonte: RITCHIE, *et al*. (2007).

2.4.3 Automatização através do aprendizado de máquina

Aprendizado de máquina pode ser definida como a capacidade de melhorar o desempenho de uma tarefa por meio das experiências prévias, ou seja, os programas que fazem uso desses princípios aprendem conforme as realizações de tarefas. Dessa forma, há diferentes aplicações de *Machine Learning*, como por exemplo: reconhecimento de fala, produção de programas com desempenho superior a grandes campeões em jogos, problemas logísticos, detecção de compras fraudulentas de cartões de crédito e predição de taxas de cura de pacientes.

Apesar da possibilidade de aplicar *Machine Learning* em diversas áreas, um dos fatores fundamentais para viabilizar sua utilização é o desenvolvimento da capacidade de processamento dos computadores, devido à necessidade de execução de milhares de tarefas durante períodos muito curtos. Por esse motivo, conforme hardwares de processamento mais

sofisticados foram lançados no mercado, a inteligência artificial e aprendizado de máquina se tornaram assuntos muito mais recorrentes. Dentre os principais métodos de ML estão: aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem por reforço e redes neurais profundas. Cada um desses métodos é utilizado nas mais diversas áreas, dependendo de fatores como complexidade do problema a ser solucionado e disponibilidade de dados. Neste trabalho será apresentado apenas o conceito e funcionamento das redes neurais profundas devido à sua fundamental aplicação para a realização do projeto (MITCHEL,1997).

2.4.3.1 Visão computação e *Deep Learning*

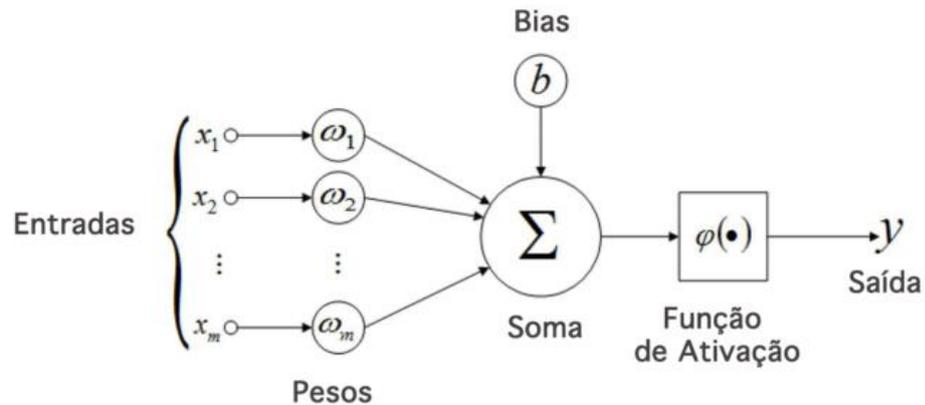
Visão computacional é a ciência responsável pela visão de uma máquina, pela forma como um computador enxerga o meio à sua volta, extraindo informações significativas a partir de imagens capturadas por câmeras de vídeo, sensores, scanners, entre outros dispositivos. Estas informações permitem reconhecer, manipular e pensar sobre os objetos que compõem uma imagem (BALLARD, 1982).

Deep Learning é o tipo de aprendizagem de máquina que se faz com rede neural profunda, em essência, uma percepção apurada de inteligência artificial, que se parece com a do ser humano e é capaz de gerar conteúdo baseados no aprendizado a partir dessa assimilação. Os algoritmos de aprendizagem profunda são capazes de analisar dados não-estruturados sem que haja algum tipo de pré-processamento ou supervisão (GOODFELLOW, *et al.*, 2016).

A palavra “*deep*” refere-se ao fato de estas serem construídas com o emprego de um maior número de camadas em relação às redes neurais tradicionais. Estas camadas adicionais são responsáveis por extrair ou representar a informação a partir de diferentes graus de abstração (VOINIGESCU,2017). É importante ressaltar que as tecnologias de aprendizagem profunda obtiveram sucesso em várias tarefas de visão computacional, como classificação e segmentação de imagens e detecção de objetos (IKRAM, 2019).

Existem três elementos básicos que compõem um modelo neural: Um conjunto de valores de entrada (x) e pesos (w), um somador para definir o valor da adição de todos os conjuntos e uma função de ativação, a qual tem o objetivo de restringir o valor de saída de um neurônio (Figura 6).

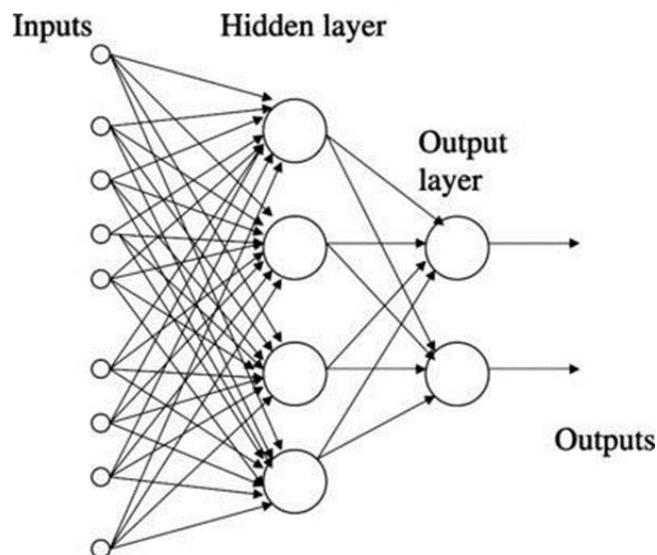
Figura 6 - Modelo de um neurônio artificial.



Fonte: HAYKIN (2010).

A partir desse modelo básico, é possível a solução de problemas lineares, mas para a solução de problemas não lineares, necessita-se que mais camadas de neurônios sejam acrescentadas, sendo chamadas assim de camadas escondidas, formando a *Rede Neural Multilayer Perceptron* (CHAUVIN, *et al.*, 1995). Dessa forma, a sua estrutura é composta por dados de entrada (*Inputs*), camadas escondidas (*Hidden Layers*), compostas nesse caso por quatro neurônios artificiais, e por fim, a camada final (*Output Layer*) formada por dois neurônios, sendo ela responsável por obter o resultado do problema o qual quer solucionar (Figura 7).

Figura 7 - Método de multicamadas para soluções não lineares.



Fonte: CHAUVIN, *et al.* (1995).

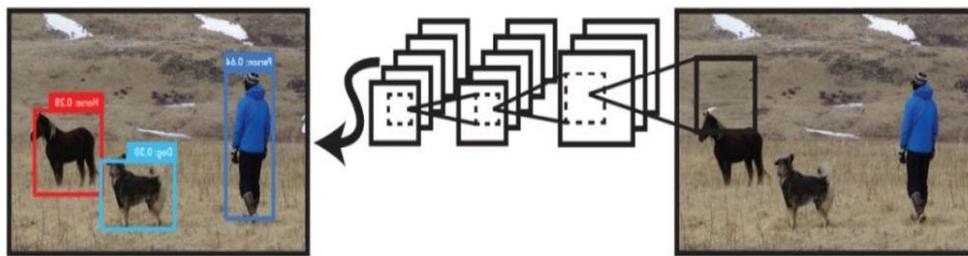
2.4.3.2 *You only look once* (YOLO)

O algoritmo YOLO (*You Only Look Once*) é um sistema de reconhecimento em tempo real, desenvolvido com o objetivo de detectar vários objetos em um único frame, além de gerar

caixas delimitadoras (*Bounding Box Predictions*) que indicam em que região da imagem os objetos estão localizados e a qual classe eles pertencem (BOCHKOVSKIY, *et al.*, 2020).

Para o seu funcionamento, a YOLO utiliza uma rede neural profunda, cuja arquitetura é chamada de *Darknet*, que é o mesmo nome do framework utilizado para implementar o detector. Vale ressaltar que seu desempenho difere-se das diversas arquiteturas de redes neurais pois, ele não somente realiza o treinamento da rede, como também a divisão da imagem em áreas de interesse, permitindo que o algoritmo também seja capaz de delimitar área em que os objetos serão encontrados (BOCHKOVSKIY, *et al.*, 2020). Na Figura 8, pode ser visto um exemplo de aplicação de *bounding boxes*.

Figura 8 - Exemplo da aplicação de *bounding boxes*.



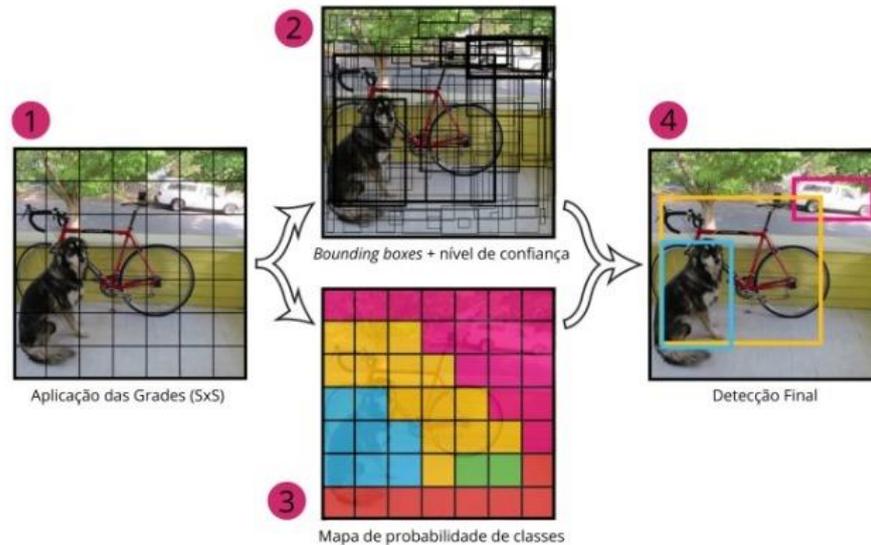
Fonte: REDMON, *et al.*, (2016).

O procedimento realizado pelo YOLO para detecção de objetos segue os seguintes passos:

- I. O primeiro processo que o YOLO realiza é selecionar a imagem na qual será feita a predição, e dividi-la em vários grids. Por padrão, a imagem é dividida em “7 x 7”, mas a rede neural pode ser adaptada para que ela possa dividir a imagem em “S x S”. Portanto, qualquer dimensão pode ser aplicada. Porém quanto mais divisões, mais complexa se tornará a rede.
- II. A partir daí, o algoritmo começa a aplicar o método de *bounding boxes* e tenta encontrar o objeto na imagem. Então, para cada grid, ele tentará fazer várias predições utilizando inúmeras *bounding boxes* ao mesmo tempo. Finalmente, ele entregará apenas *a bounding box* que tiver a maior acurácia, ou seja, a maior taxa de acerto.
- III. Entretanto, essa etapa tentará calcular a probabilidade do objeto ser realmente o que ele está esperando através do parâmetro acurácia.
- IV. Por fim, é feita a detecção final dos objetos, onde o YOLO mantém apenas as *bounding boxe*.

De uma forma resumida, esses passos podem ser visualizados na Figura 9 . Note-se que existe um enumeração para representar cada etapa.

Figura 9 - Exemplo de aplicação e funcionamento do YOLO.



Fonte: REDMON,*et al.*, (2016).

2.4.3.3 Versões da YOLO

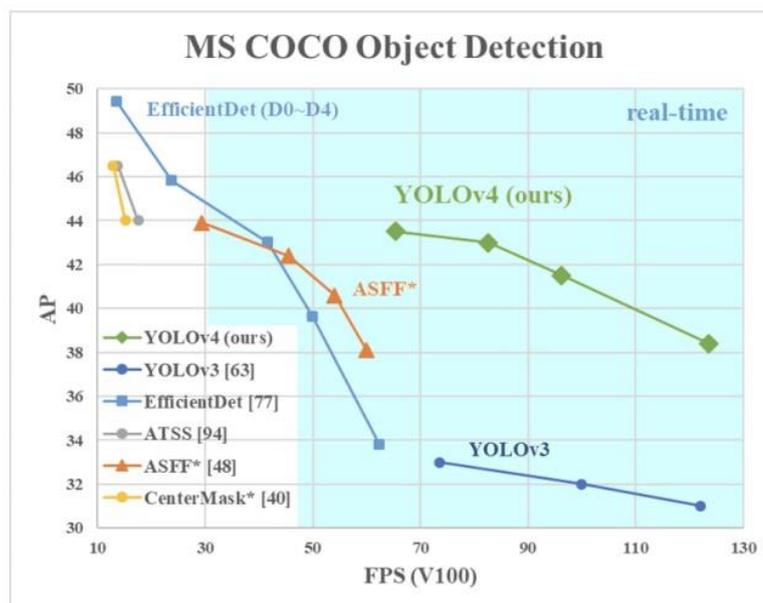
Em 2015, Joseph Redmon publicou sobre a primeira versão da YOLO em um trabalho intitulado como *You Only Look Once: Unified, Real-Time ObjectDetection*. Em dezembro de 2016, Redmon e Ali Farhadi introduziram a segunda versão YOLOv2, dessa vez intitulado como *YOLO9000: Better, Faster, Stronger*.

Em abril de 2018, a YOLOv3: *An Incremental Improvement* demonstrou uma grande melhora na eficiência da predição. No entanto, no geral ele não é mais rápido que a versão anterior. A principal novidade é a predição da imagem em três diferentes escalas, o que resolveu o principal problema da versão anterior, que era a dificuldade para reconhecer objetos muito pequenos na imagem. Essa novidade também é o principal motivo de não ser mais rápido que o seu antecessor, já que tal função exigiu mudanças na arquitetura e funcionamento, e por isso tornaram o processo mais lento (REDMON,*et al.*, 2018).

A quarta versão do YOLO foi lançada em abril de 2020, e as principais características que podem ser destacadas nessa versão são a melhoria na velocidade de inferência e acurácia. Outra característica importante é o fato de ser mais eficiente para rodar em GPU's, pois foi otimizada para utilizar menos memória. Essa versão demonstrou ser um ótimo detector de objetos para testes em tempo real de acordo com as métricas do MS COCO, um famoso *dataset*

utilizado para avaliar os sistemas de detecção de objetos. A Figura 10 demonstra a comparação entre a YOLOv4 e as outras tecnologias de visão computacional (BOCHKOVSKIY, *et al.*,2020).

Figura 10 – detecção de objetos de acordo com as métricas do *dataset* MS COCO.

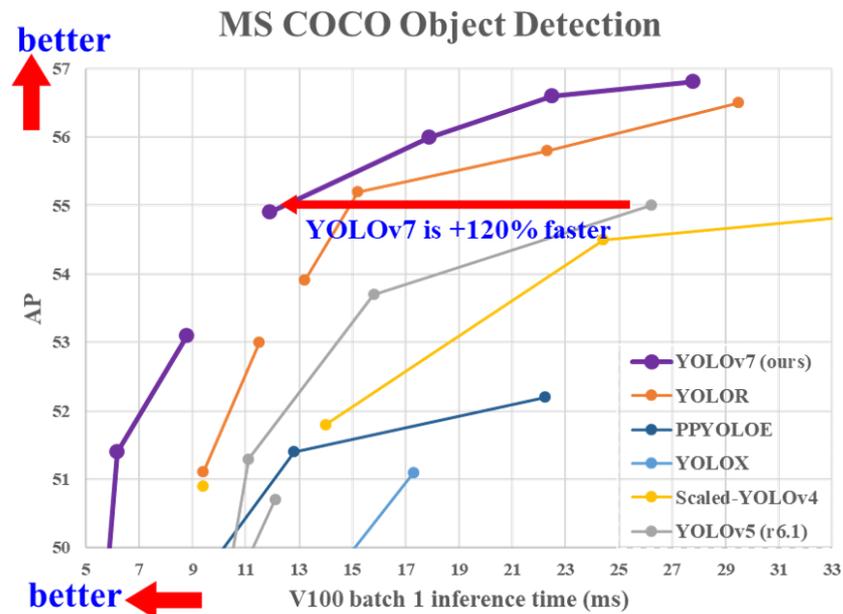


Fonte: BOCHKOVSKIY, *et al.*,(2020).

A quinta versão do algoritmo YOLOv5, lançado em junho de 2020 por Glenn Jocher, é o primeiro modelo da "família YOLO" a não ser lançado com um papel de acompanhamento e, da mesma forma, passando por "desenvolvimento contínuo" em seu repositório. Glenn estava mantendo uma versão do YOLOv3 implementada no *PyTorch*, mas como ele continuou a fazer melhorias na própria arquitetura, ele finalmente decidiu lançar um novo repositório com a marca YOLOv5.

Na YOLOv6, empiricamente esta versão demonstrou um aumento de desempenho. Além das alterações de arquitetura, o repositório YOLOv6 também implementa alguns aprimoramentos no pipeline de treinamento. Por fim, para o desenvolvimento da versão YOLOv7, os autores baseiam-se em pesquisas anteriores do projeto, com foco na quantidade de memória necessária para manter as camadas na memória junto com a distância que um gradiente leva para se propagar de volta pelas camadas, ou a seja, quanto menor o gradiente, maior é a eficiência da rede neural. A agregação de camada final que eles escolheram é o E-ELAN, uma versão estendida do bloco computacional ELAN (BOCHKOVSKIY, *et al.*,2022). A Figura 11 demonstra a comparação entre a YOLOv7 e as outras tecnologias de visão computacional.

Figura 11 -Detecção de objetos de acordo com as métricas do *dataset*.



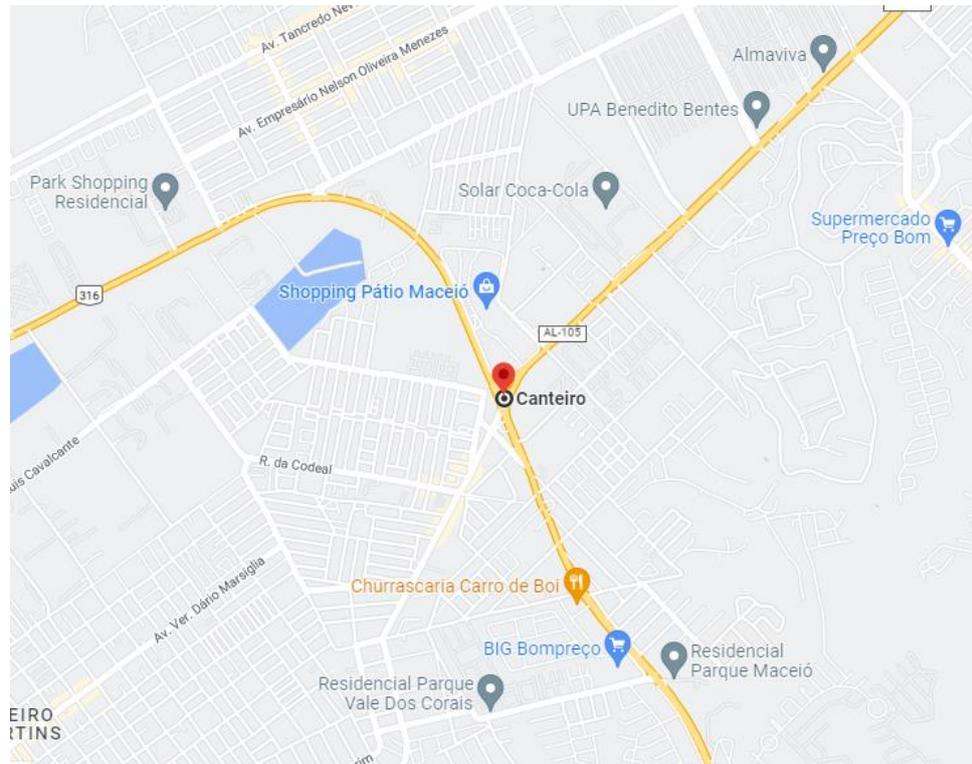
Fonte: BOCHKOVSKIY, *et al.*,(2022).

3 MÉTODO

3.1 CONSTRUÇÃO DO BANCO DE DADOS/IMAGENS - *DATASET*

O banco de dados ou *dataset* foi composto por imagens extraídas dos vídeos de câmeras de controle de tráfego que foram solicitadas à Superintendência Municipal de Transportes e Trânsito (SMTT) da cidade de Maceió-AL, contendo os trechos urbanos com alto volume de tráfego e diferentes tipos de veículos, dentro de um período de uma semana e nos horários de pico. Vale ressaltar que o local definido da coleta de dados da câmera de monitoramento, em frente à entrada do bairro Benedito Bentes (9°33'43.01"S 35°44'44.164"W), foi de acordo com a disponibilidade da SMTT. A Figura 12 apresenta a localização do trecho em estudo e a câmera de controle de tráfego está apresentada na Figura 13.

Figura 12 - Localização da câmera de controle de tráfego, representada pelo ponto em vermelho.



Fonte: Google Maps (2022).

Figura 13 - Câmera de controle de tráfego em frente à entrada do bairro Benedito Bentes e próxima ao Shopping Pátio Maceió.



Fonte: Google Street (2022).

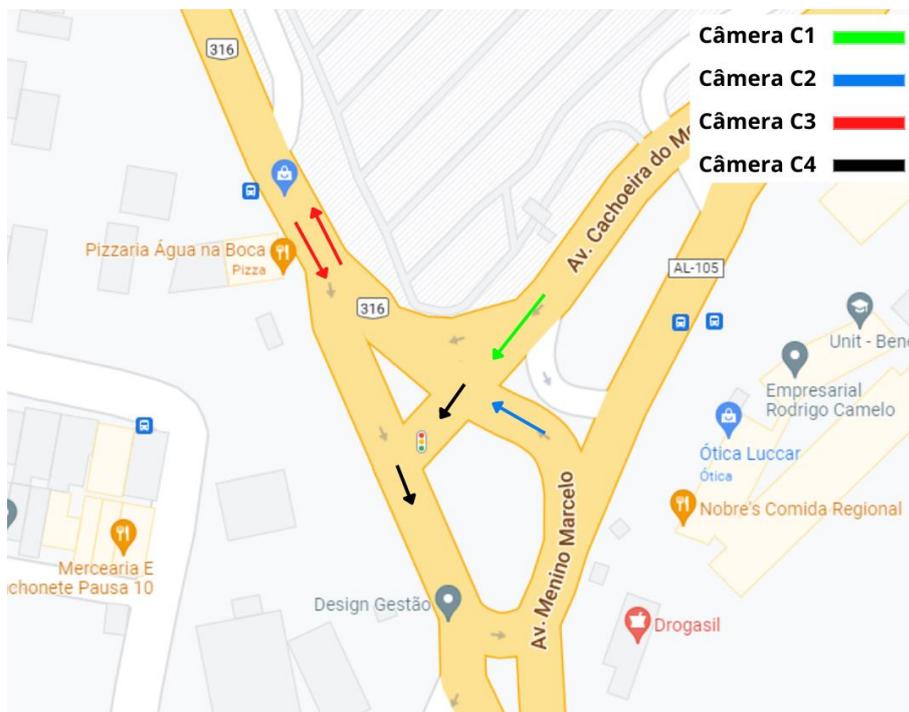
No dia 31 de agosto de 2022, foi aberto o processo de requerimento à Superintendência Municipal de Transportes e Trânsito (SMTT) da disponibilização das filmagens das câmeras de monitoramento. As filmagens cedidas pela SMTT do tráfego constam entre os dias 29 de agosto de 2022 e 2 de agosto de 2022, durante os horários de 6h30 e 8h30. Porém, vale destacar que cada dia de filmagem contém uma angulação diferente. Dessa forma, a construção do banco de imagens para treinamento e validação, foi composta por diferentes ângulos de filmagem entre os dias 29 e 2 de agosto (Figura 14). A Figura 15 ilustra a localização do cruzamento em estudo e dos sentidos de tráfego para cada câmera de monitoramento.

Figura 14 - Filmagens das câmeras de monitoramento para composição do banco de dados.



Fonte: Autor (2022).

Figura 15 - Localização do cruzamento em estudo e dos sentidos de tráfego para cada câmera de monitoramento.



Fonte: Autor (2022).

3.2 PRÉ-PROCESSAMENTO DE IMAGENS

Existem técnicas de organização e pré-processamento que podem ser utilizadas para melhorar o conjunto de dados antes de realizar o treinamento das redes. Tais técnicas são importantes para aumentar a eficácia de uma aplicação de aprendizado de máquina. A seguir, as técnicas consideradas no presente trabalho foram brevemente descritas.

3.2.1 Redimensionamento das imagens

As configurações de resolução das imagens fornecidas pela SMTT apresentam imagens com resolução 704x480 pixels, sendo que as redes neurais estudadas tomam apenas imagens quadradas. Dessa forma, para facilitar/viabilizar o treinamento da rede neural da YOLOv7, todas as imagens do banco de dados foram redimensionadas para 640x640 pixels.

3.3 ROTULAÇÃO DE IMAGENS

Para o treinamento da rede neural, são fornecidos como dados de entrada, as imagens que compõe o banco de dados e seus respectivos rótulos. Os rótulos inicialmente propostos foram os veículos de projetos (VP, CO, O, SR, RE e Outros). A categoria “Outros” abrange os objetos que aparecem nas filmagens e que não são classificados pelo DNIT como veículos de projeto, como por exemplo, motociclistas, ciclistas, pedestres e etc. A etapa de rotulação será desenvolvida no programa *labelimg*, conforme apresentado da Figura 16.

Figura 16 - Exemplo de rotulação desenvolvida no programa *labelimg*.



Fonte: Autor (2022).

A ferramenta de catalogação tem como saída um arquivo no formato de texto contendo as coordenadas da localização do objeto na imagem e sua respectiva classe, como pode ser observado na Figura 17.

Figura 17 - Ferramenta de catalogação tem como saída um arquivo no formato de texto.



Fonte: Autor (2022).

3.4 TREINAMENTO E VALIDAÇÃO DA REDE NEURAL - YOLO

O treinamento de redes neurais pode ser definido como um processo que envolve otimização matemática, envolvendo uma função de custo, que permite o ajuste automático dos pesos e vieses da rede neural. Durante este processo, os dados de treinamento com rótulos conhecidos serão implementados em uma função de otimização desenvolvida pelo algoritmo YOLOv7.

Para isso, foram utilizados alguns parâmetros de qualidade para designar a qualidade de aplicação da rede neural. Sendo as métricas mais comuns: *intersect over union* (interseção

sobre a união), *precision* (precisão), *recall* (sensibilidade), *F1-score*, *average precision* (precisão média) de cada tipo de rótulo e *mean average precision* (médias das precisões de todas as métricas) (NASCIMENTO, 2019). A explicação dessas métricas serão posteriormente explicadas na seção 3.4.1.

O conjunto de dados (*dataset*) foi dividido em duas partes, o primeiro conjunto de imagens foi utilizado para treinamento da rede neural e o segundo conjunto de imagens foi utilizado para a validação do modelo, essas duas partes foram divididas em uma proporção de 75% e 25%.

É importante frisar que as imagens de treinamento e validação serão selecionadas de forma aleatória, para que o modelo de detecção e classificação não seja enviesado durante o processo de treinamento.

3.4.1 Métricas de avaliação da detecção e classificação

A fim de avaliar o desempenho do treinamento das arquiteturas utilizando *datasets*, empregou-se métricas de desempenho difundidas em análises estatística e que são aplicadas a algoritmos que utilizam *deep learning*. A seguir, serão explicadas cada uma delas, indicando o cálculo realizado para sua determinação.

Para a determinação das métricas, é necessário identificar a quantidade de Verdadeiro-Positivo, Verdadeiro- Negativo, Falso-Positivo e Falso-Negativo.

Verdadeiro Positivo, ou, do inglês, *True Positive* (TP), indica a quantidade de objetos identificados que realmente estavam presentes na imagem, ou seja, é a quantidade de objetos que o algoritmo previu corretamente. Da mesma forma, um verdadeiro negativo ou *True Negative* (TN), é um resultado em que o modelo corretamente prevê a classe negativo.

O parâmetro Falso Positivo, ou *False Positive* (FP), indica a quantidade de previsões erradas do algoritmo, em que ele prevê a existência de um objeto, mas que não existe realmente na imagem. O Falso Negativo, ou *False Negative* (FN), é um parâmetro referente à quando o algoritmo não consegue identificar um objeto presente na imagem, ou seja, ele não define uma caixa delimitadora para esse objeto.

A matriz de confusão, ou do inglês *confusion matrix*, busca entender a relação entre acertos e erros que o modelo apresenta. A Tabela 4 ilustra a diferença entre *True Positive*, *True Negative*, *False Positive* e *False Negative*.

Tabela 4 - diferença entre TP, TN, FP e FN.

		<i>Predicted Label</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual label</i>	<i>Positive</i>	<i>True Positive</i> <i>TP</i>	<i>False Negative</i> <i>FN</i>
	<i>Negative</i>	<i>False Positive</i> <i>FP</i>	<i>True Negative</i> <i>TN</i>

Fonte: Autor (2022).

É importante ressaltar que a matriz de confusão é uma ferramenta gerada para avaliar o desempenho de um modelo de classificação, como a rede neural YOLO v7. Ela é composta por duas dimensões: as classes preditas pelo modelo (linhas) e as classes reais (colunas).

Na matriz de confusão gerada pela YOLO v7, cada célula representa o número de exemplos de uma determinada classe real que foram classificados como uma determinada classe predita pelo modelo. Por exemplo, se a célula (i, j) representa o número de exemplos da classe real i que foram classificados como a classe j pelo modelo, então o valor na célula (i, j) será alto se o modelo tem alta precisão em classificar os exemplos da classe i como a classe j.

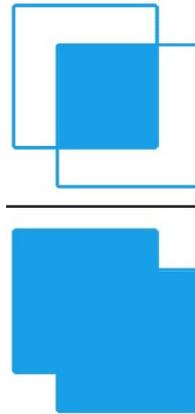
A matriz de confusão permite avaliar o desempenho do modelo de várias maneiras. Por exemplo, é possível verificar a precisão geral do modelo, que é dada pela soma dos valores na diagonal principal da matriz (isto é, os valores onde a classe real é igual à classe predita) dividida pelo número total de exemplos. Também é possível verificar a precisão de cada classe individualmente.

- *Intersection over Union (IoU)*

Para determinar o que é um verdadeiro positivo ou um falso positivo, é necessário delimitar um parâmetro que seja capaz de identificar o quanto uma caixa delimitadora satisfaz ao objeto real. Para isso, é definido a métrica Intersection over Union (Equação 1), chamada de IoU. Para seu cálculo, faz-se a divisão entre a área de interseção das caixas delimitadoras da predição e da real pela união das duas áreas (Figura 18).

$$IoU = \frac{\text{Área de interseção}}{\text{Área de união}} \quad (\text{Equação 1})$$

Figura 18 – Representação do cálculo da métrica *Intersection over Union*.



Fonte: Pyimagesearch (2016).

Ao avaliar os modelos de detecção de objetos, é preciso estabelecer um limite mínimo e máximo para este parâmetro. Definir um limite muito alto ,por exemplo “0.9” , pode resultar em previsões relevantes sendo descartadas. Por outro lado, definir um limite muito baixo pode gerar *bounding boxes* sem informações úteis. Em geral, na literatura, é considerado que uma detecção precisa é sempre superior a “0.5”. No entanto, este valor pode variar de acordo com o nível de precisão exigido em uma determinada aplicação (REZATOFIGHI et al., 2019).

- *Precision*

Precision ou precisão, é a razão entre o número de predições positivas corretas (Equação 2), e o número total de predições positivas . É um valor que varia entre 0 e 1, e quanto mais próximo de 1, maior é a precisão.

$$Precision = \frac{t_p}{t_p + f_p} \quad (\text{Equação 2})$$

- *Recall*

De modo complementar à precisão, o *recall* ou sensibilidade, analisa a quantidade de identificação de objetos que o algoritmo obteve. Para isso, calcula-se a razão dos verdadeiros positivos pelo total de objetos na imagem. O cálculo é dado pela Equação 3, a seguir:

$$Recall = \frac{t_p}{t_p + f_n} \quad (\text{Equação 3})$$

- *Mean Average Precision (mAP)*

Outra métrica calculada para comparação de desempenho é a precisão média, ou *Average Precision (AP)*, em que determina-se a acurácia das predições do algoritmo. Em termos gerais, avalia-se a quantidade de predições feitas e quantidade dessas que estavam corretas de acordo com a quantidade real de objetos existentes. Seu cálculo é dado pela razão entre a precisão e o *recall* expresso pela Equação 4, a seguir:

$$AP = \frac{Precision}{Recall} = \frac{TP + FP}{TP + FN} \quad (\text{Equação 4})$$

Geralmente, tem-se o resultado de AP para cada classe e, com isso, pode-se determinar a média desses valores para o algoritmo como um todo. Desta forma, calcula-se a métrica *Mean Average Precision (mAP)*. Em termos de expressão, para um algoritmo de identificação contendo n classes, tem-se (Equação 5):

$$mAP = \frac{\sum_{n=1}^n AP_i}{n} \quad (\text{Equação 5})$$

- *F1-Score:*

Outra métrica de desempenho é o *F1-score*, uma medida estatística voltada para determinar a acurácia e que envolve o cálculo da média harmônica da precisão e do recall. A expressão é dada pela Equação 6, abaixo:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (\text{Equação 6})$$

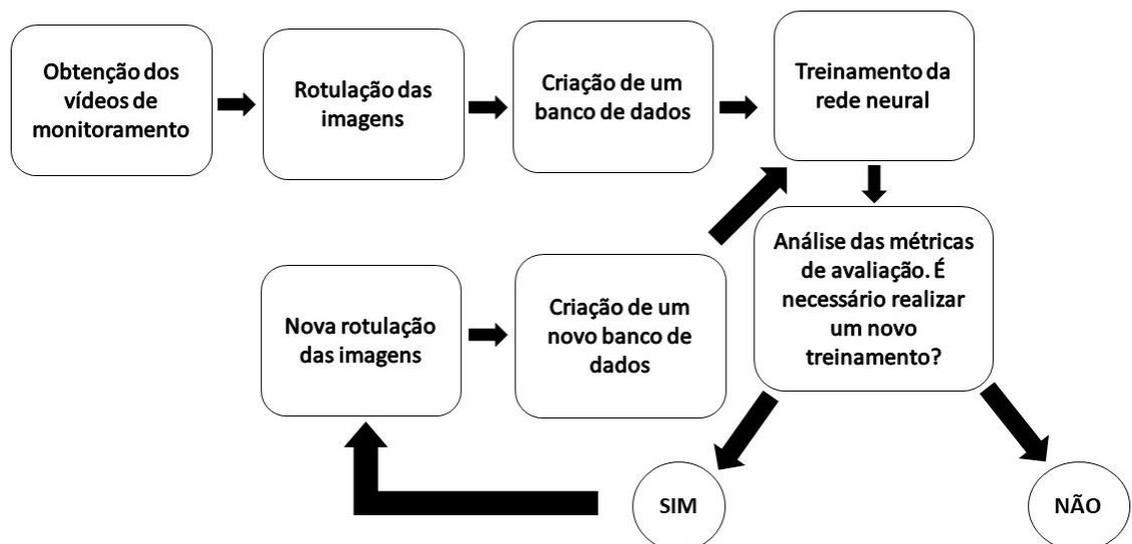
A escolha dos valores considerados satisfatórios para as métricas de avaliação geradas pela YOLO v7 depende de vários fatores, incluindo o contexto de aplicação e o nível de precisão exigido (REDMON,2017). Os valores comumente utilizados como referência são:

- *Accuracy* (precisão): Valores superiores a 80% são considerados bons.
- *Recall* (sensibilidade): Valores superiores a 70% são considerados bons.
- *F1-score*: Valores superiores a 70% são considerados bons.
- *Intersection over Union* (IoU): Valores superiores a 50% são considerados bons. Esses valores são apenas uma referência e podem variar dependendo do contexto de aplicação. É importante lembrar que as métricas de avaliação devem ser utilizadas em conjunto para avaliar o desempenho do modelo de forma mais completa

3.4.1 Procedimento para criação do banco de dados e treinamento da rede neural

A Figura 19 mostra os passos seguidos para rotular e treinar a rede neural empregada nesta pesquisa.

Figura 19 - Diagrama de fluxo para construção do banco de dados e treinamento da rede neural.



Fonte: Autor (2022).

4 RESULTADOS E DISCUSSÕES

Neste capítulo, apresenta-se os principais resultados do trabalho, sendo realizadas discussões acerca das explicações para possíveis desempenhos das redes neurais. A princípio, analisa-se o desempenho na detecção e contagem de veículos utilizando o *dataset* proposto.

4.1 BANCO DE DADOS

O banco de dados (*dataset*) criado e posteriormente utilizado na primeira rodada de treinamento, possuiu 2300 imagens para treinamento e 576 imagens para validação. Os objetos que não se enquadraram na definição de veículos de projeto, foram rotulados como “Outros”, como por exemplo, os motociclistas, ciclistas, pedestres e etc.

Uma vez que a quantidade de imagens e fonte de dados não é diversificada, algumas classes obtiveram poucas ocorrências. E para tentar manter o balanceamento entre as classes, optou-se por não rotular todos os objetos da classe “VP” e “Outros” que apareciam nas imagens. Entretanto, ainda assim as duas principais classes “VP” e “CO” apresentaram uma quantidade considerável de rotulações (Tabela 5).

Tabela 5 – Banco de dados customizado.

Classes	VP	CO	O	SR	RE	Outros
Quantidade de rótulos	2142	832	217	74	26	645
Frequência (%)	54,42	21,14	5,51	1,88	0,66	16,39

Fonte: Autor (2022).

4.2 DETECÇÃO DE VEÍCULOS

Na Tabela 6, a diagonal principal da matriz de confusão representa os verdadeiros positivos (*true positives*), que são aqueles em que as predições correspondem ao valor real. Todos os que estão fora da diagonal principal são os verdadeiros negativos, que o modelo detectou um veículo, mas errou a sua categoria. A linha “FN” representa os falsos negativos, ou seja, são aqueles nos quais havia um veículo na imagem, mas o modelo não detectou. A coluna “FP” representa os falsos positivos, que por sua vez, ocorrem quando não há objeto na imagem, mas o modelo detecta alguma categoria de veículo.

Tabela 6 - Matriz de confusão do primeiro treinamento.

ACURÁCIA DA PREVISÃO	VP	19%	2%	0%	0%	0%	0%	31%
	CO	0%	82%	21%	29%	0%	0%	44%
	O	0%	2%	36%	14%	50%	0%	3%
	SR	0%	1%	0%	43%	50%	0%	2%
	RE	0%	0%	7%	0%	0%	0%	0%
	Outros	0%	0%	0%	0%	0%	10%	20%
	FN	81%	13%	36%	14%	0%	90%	0%
	VP	CO	O	SR	RE	Outros	FP	
	VALOR REAL							

Fonte: Autor (2022).

Ao analisar matriz de confusão e o resultado das imagens e suas respectivas anotações, nota-se que há diversos problemas recorrentes na detecção de objetos. Dentre eles, pode-se citar em alguns casos, a observação da demarcação de caixas delimitadoras maiores que os objetos, o que resulta em extração de características que não são correspondentes àquela classe, mas sim, do cenário.

Para explicar o alto valor da métrica de Falso Negativo para as categorias “VP” e “Outros”, tem-se como correlação a presença de objetos sem anotações durante o processo de rotulação, o que faz com que a rede neural, durante o treinamento, interprete isto como um não objeto e o que acaba dificultando a identificação correta do veículo (Figura 20).

Figura 20 - Predição do primeiro modelo treinado.



Fonte: Autor (2022).

Com as variáveis disponíveis, foi possível realizar os cálculos das métricas de *Precision*, *Recall*, *Mean Average Precision* (*mAP* para IoU entre 0,5 e 0,95) e *F1-score*. Dessa forma, o resultado de cada uma das métricas pode ser observado na Tabela 7.

Tabela 7 - Cálculos das métricas de avaliação.

<i>Precision</i>	52 %
<i>Recall</i>	33,2 %
<i>mAP</i>	26 %
<i>F1-score</i>	37 %

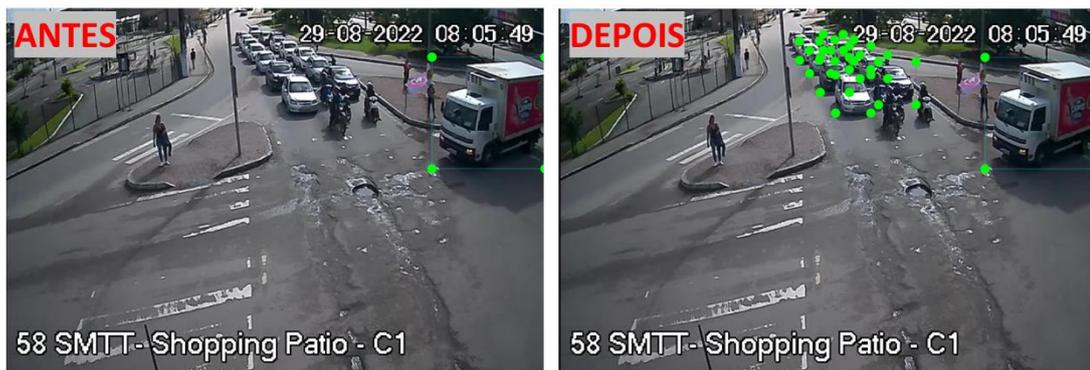
Fonte: Autor (2022).

Ao analisar o baixo desempenho das métricas de avaliação e a influência do alto número de falso negativos para as categorias “VP” e “Outros”, pode-se dizer que os baixos valores desses parâmetros podem explicar a falta de detecções durante os momentos de maiores densidades de veículos durante a filmagem, e por isso optou-se por realizar uma segunda rodada de treinamento a fim de aumentar as métricas de desempenho.

4.3 BANCO DE DADOS PARA O SEGUNDO TREINAMENTO

O novo banco de dados (*dataset*) utilizado na segunda rodada de treinamento, possuiu as mesmas 2300 imagens para treinamento e 576 imagens para validação. Entretanto, para diminuir a porcentagem do parâmetro “FN”, optou-se por rotular todos os objetos das imagens do novo *dataset*. A Figura 20 mostra a diferença entre as rotulações das imagens entre o primeiro e o segundo banco de imagens.

Figura 21 – Novas rotulações para a construção do segundo *dataset*.



Fonte: Autor (2022).

As classes continuaram as mesmas (VP, CO, SR e SE), mas é importante destacar que houve a exclusão da classe “Outros”, a fim de diminuir as intercessões das rotulações entre diferentes classes (Tabela 8).

Tabela 8 - Segundo banco de dados customizado.

Classes	VP	CO	O	SR	RE
Quantidade de rótulos	8568	834	217	74	26
Frequência (%)	88,16	8,58	2,23	0,76	0,27

Fonte: Autor (2022).

4.4 DETECÇÃO DE VEÍCULOS DO SEGUNDO MODELO TREINADO

Na Tabela 9, a diagonal principal da matriz de confusão ainda representa os verdadeiros positivos (TP). Assim como a Tabela 6, todos os que estão fora da diagonal principal são os verdadeiros negativos, que o modelo detectou um veículo, mas errou a sua categoria. A linha “FN” representa os falsos negativos, ou seja, são aqueles nos quais havia um veículo na imagem, mas o modelo não detectou. A coluna “FP” representa os falsos positivos, que por sua vez, ocorrem quando não há objeto na imagem, mas o modelo detecta alguma categoria de veículo.

Tabela 9 - Matriz de confusão do segundo treinamento.

ACURÁCIA DA PREVISÃO	VP	84%	1%	0%	0%	0%	39%
	CO	1%	83%	21%	29%	0%	46%
	O	0%	2%	36%	14%	50%	7%
	SR	0%	1%	0%	43%	50%	6%
	RE	0%	0%	7%	0%	0%	2%
	FN	15%	13%	36%	14%	0%	0%
		VP	CO	O	SR	RE	FP
		VALOR REAL					

Fonte: Autor (2022).

Com as variáveis disponíveis, foi possível calcular as métricas de *Precision*, *Recall*, *Mean Average Precision* (*mAP* para IoU entre 0,5 e 0,95) e *F1-score*. Dessa forma, o resultado de cada uma das métricas do segundo treinamento pode ser observado na Tabela 10.

Tabela 10 - Resultado das métricas de avaliação do segundo treinamento.

<i>Precision</i>	68,11 %
<i>Recall</i>	84,37 %
<i>mAP</i>	51 %
<i>F1-score</i>	75,37 %

Fonte: Autor (2022).

A partir da análise das métricas de avaliação para o segundo modelo, pode-se afirmar que houve um aumento significativo de desempenho em relação ao primeiro modelo. Isso se deve a uma maior eficácia no processo de rotulação de objetos presentes nas imagens, principalmente em momentos em pontos de retenção de veículos por conta do semáforo fechado, onde a densidade de veículos aumenta consideravelmente.

A diminuição de 81% para 15% de “FN” para a classe “VP” representa na prática a detecção de veículos mesmo em uma densidade considerável (Figura 21), e o aumento do parâmetro “TP” para essa mesma classe representa uma maior eficiência na detecção e classificação nesse segundo modelo.

Figura 22- Predição do segundo modelo treinado.



Fonte: Autor (2022).

Em relação as classes “CO”, “O”, “SR” e “RE”, como não houveram mudanças no processo de rotulação desses veículos de projeto durante a construção do segundo *dataset*, o desempenho dos parâmetros TP, FN e FP no segundo modelo continuaram os mesmos em relação ao primeiro modelo. Isso se deve a permanência do desbalanceamento significativo entre as quantidades de rótulos, sendo que as técnicas para promover esse equilíbrio entre quantidade de categorias não foram vistas nesse trabalho.

Os resultados da avaliação do modelo de rede neural mostraram valores satisfatórios para as métricas *de recall*, *F1-score* e *IoU*. No entanto, o valor obtido para a precisão foi de 68%, dessa forma, pode-se dizer que a métrica *Precision* ainda não atende ao valor considerado como satisfatório pela literatura. Isso indica que o segundo modelo teve uma boa performance na detecção e classificação dos objetos presentes nas imagens de teste, exceto no quesito precisão.

5 CONCLUSÃO

O presente trabalho teve como escopo a abordagem do uso de algoritmos de deep

learning para detecção e classificação e contagem de veículos em imagens de tráfego obtidas pela SMTT.

Dentro desse contexto, por meio dos resultados, verificou-se que foi possível criar um *dataset* através das filmagens de um ponto crítico do tráfego da cidade de Maceió, possuindo imagens e anotações de veículos de projeto definidos pela norma do DNIT. Sendo permitido a utilização desse banco de dados em projetos futuros, com possibilidade de ampliação, devido à identificação de soluções que podem melhorar os parâmetros de qualidade do treinamento da rede neural, dando assim a possibilidade de continuidade do trabalho.

Entretanto, mesmo com a melhora das métricas de avaliação no segundo treinamento, por conta da criação de uma nova base dados com rotulações mais precisas, há muitos passos que ainda precisam ser dados em trabalhos futuros para melhorar o detecção de todas as categorias de veículos de projeto, como por exemplo, a aplicação de técnicas que possam balancear a quantidade de rotulações entre diferentes classes. Dessa forma, pode-se dizer que o modelo ainda não está suficientemente validado para ser aplicado para estudos tráfego da cidade de Maceió-AL.

Por fim, como possíveis aprimoramentos para pesquisas e trabalhos futuros, sugere-se também o emprego de diferentes algoritmos de *deep learning* para comparação e verificação dos melhores modelos, uma vez que, neste trabalho, somente foi utilizado a configuração padrão da YOLOv7.

REFERÊNCIAS

- ALPAYDIN, E. **Introduction to machine learning**. 4. ed. Cambridge, Massachusetts: The MIT Press, 2020.
- BALLARD, D., BROWN, C. **Computer Vision**. PrenticeHall, Englewood Cliffs, 1982.
- BOCHKOVSKIY, A., WANG, C.-Y., LIAO, H.-Y. M. **YOLOv4: Optimal Speed and Accuracy of Object Detection**. 2020.
- BOCHKOVSKIY, A., WANG, C.-Y., LIAO, H.-Y. M. **YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors**. 2022.
- BOSSO, M. **Uso da tecnologia Weight-in-Motion para a caracterização do tráfego rodoviário e do excesso de carga em veículos comerciais**. São Paulo. 2018.
- CHEN, Z., ELLIS, T. AND VELASTIN, S. (2012). **Vehicle detection, tracking and classification in urban traffic**, 15th ITSC, pp. 951–956.
- CHAUVIN, Y., RUMELHART, D. **Backpropagation: Theory, Perceptron Learning Architectures, and Applications**. Erlbaum, Mahwah, NJ, 1995.
- CONFEDERAÇÃO NACIONAL DO TRANSPORTE (CNT). **Boletim Estatístico Janeiro 2021**. Disponível em: <<http://www.cnt.org.br>>. Acesso em: jun. 2022.
- CONFEDERAÇÃO NACIONAL DO TRANSPORTE (CNT). **Boletim Estatístico Janeiro 2020**. Disponível em: <<http://www.cnt.org.br>>. Acesso em: jun. 2022.
- DEPARTAMENTO NACIONAL DE TRÂNSITO (DENATRAN). **Resolução CONTRAN nº 12, de 06 de fevereiro de 1998. Estabelece os limites de peso e dimensões para veículos que transitem por vias terrestres**. 1998
- DEPARTAMENTO NACIONAL DE INFRAESTRUTURA DE TRANSPORTES (DNIT). **Manual de Estudos de Tráfego**. Rio de Janeiro, 2006. 384 p. (IPR. Publ., 723).
- FERRAZ, A. C. P.; TORRES, I. G. E. **Transporte Público Urbano**. São Carlos. 2004.

FILHO, J.; ROCHA, E. Estudo **Comparativo entre Pavimentos Flexível e Rígido na Pavimentação Rodoviária**. Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 03, Ed. 06, Vol. 02, pp. 146-163, junho de 2018. ISSN: 2448-0959.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.. **Deep Learning**. Cambridge: MIT Press, 2016.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3.ed. Estados Unidos: Prentice Wiley-Interscience, 2007.

IKRAM, C. **A benchmark for evaluating Deep Learning based Image Analytics**,. UNIVERSITY OF OSLO, 2019.

MIMBELA, L.; KLEIN, L. **A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems**. United States: Federal Highway Administration's Intelligent, 2000.

MITCHELL, T. M.: **Machine Learning**. 1ª Edição. McGraw-Hill, 1997.

NASCIMENTO, L. **Aplicação de aprendizagem profunda para detecção automática de trincas em pavimentos**. Trabalho de Conclusão de Curso – Faculdade de Engenharia Civil, Universidade Federal do Ceará, Fortaleza, 2019.

OPENCV. 2022 . **OpenCV modules**. Disponível em: <<https://docs.opencv.org/4.6.0/>> . Acesso em: jun. 2022.

PIETIKÄINEN, M; HADID, A.; ZHAO, G.; AHONEN, T. **Computer Vision Using Local Binary Patterns**. London: Springer Science & Business Media, 2011. v. 40, 207 p. ISBN: 978-0-85729-747-1.

PYIMAGESEARCH. **Intersection over Union (IoU) for object detection**. Disponível em: <pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection>. Acesso em: nov.2022.

PYTHON. 2022. **The Python Standard Library**. Disponível em: <docs.python.org/3/library/index.html> . Acesso em: jun. 2022.

REDMON, J. *et al.* **You Only Look Once: Unified, Real-Time Object Detection**. Las Vegas, 2016.

REDMON, J. *et al.* **YOLO9000: Better, Faster, Stronger**. Washington, 2016.

REDMON, J. *et al.* **YOLOv3: An Incremental Improvement**. Washington, 2018.

REZATOFIGHI, Hamid; TSOI, Nathan; GWAK, JunYoung; SADEGHIAN, Amir; REID, Ian; SAVARESE, Silvio. **Generalized intersection over union: A metric and loss for bounding box regression**. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. P. 658–666.

S. MESSELODI; C. M. MODENA; M. ZANIN; **A computer vision system for the detection and classification of vehicles at urban road intersections**; 2004; pp. 38.

SAID, Y., ATRI, M., TOURKI, R. **Human detection based on integral histograms of oriented gradients and svm**. In **Communications, Computing and Control Applications (CCCA)**. International Conference, páginas 1 –5. (2011).

SILVA, R.; AIRES, K.; SANTOS, T.; ABDALLA, K.; VERAS, R. **Segmentação, Classificação e Detecção de Motociclistas sem Capacete**. In: XI Simpósio Brasileiro de Automação Inteligente (SBAI), 11, 2013, Fortaleza. Anais... 2013.

VOINIGESCU, E. **Les Neurones ont une Forme Propice à L'apprentissage Profond**. 2017. Disponível em: <<https://www.cifar.ca/fr/nouvelles/2017/12/05/les-neurones-ont-une-forme-propice-%c3%a0-l-apprentissage-profond>> . Acesso em: jun. 2022.